A Ridge Approach for MIDAS Models

Corné Wiggers (409814)

Supervisor: Prof. dr. Ph.H.B.F. Franses
Second assessor: Prof. dr. R. Paap

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

June 2017

Abstract

In this paper, I introduce the Ridge approach (see Hoerl and Kennard (1970)) as a parametrization method for the MIDAS model, introduced by Ghysels et al. (2004). The low frequency and high frequency variables are sampled quarterly and monthly respectively. The Ridge approach assumes that the MIDAS parameters are equal to some function plus an error term. Furthermore, I investigate the performance of this method by comparing it with four benchmark methods that are in the common practice often used to estimate the MIDAS model. These are U-MIDAS, the Almon lag, the Exponential Almon lag and the Beta lag. Simulation results show that the Ridge approach is a very strong method. Its performance is always better or equal than the benchmark methods. In the empirical study, the GDP of the Netherlands is tried to be explained by some high frequency variable. The empirical results are less convincing. Although, the Ridge approach does not perform worse, it is not as dominant as it was in the simulations. This can be explained by the issue that it is still a big challenge to find the optimal shrinkage factor for the Ridge approach in the empirical study.

1 Introduction

It is always a big struggle if one has to deal with variables that are sampled at different frequencies. This is often the case for macroeconomic variables. For example the data about the GDP is only available quarterly, while the industrial production is sampled monthly. Industrial production is a good explanatory variable for GDP and therefore it is preferred to include this variable in a model to explain the GDP. However, because the variables are sampled at different frequencies, it is at fist glance not clear how to do this. To deal with this issue some methods were used to aggregate the high frequency variable to the low frequency variable (see Foroni and Marcellino (2013)). So in the example of quarterly and monthly observations, the monthly observations were aggregated to quarterly observations. However, aggregating the data always results in a loss of useful information. Ghysels et al. (2004) solved this problem by introducing the MIDAS model.

The MIDAS model has much in common with the distributed lag model. The MIDAS model uses all the information of the high frequency variable by regressing the lower frequency variable on a large number of lags of the higher frequency variable. However, new problems arises by using a large number of lags, because the large number of lags are likely to be highly correlated. This can result to some multicollinearity issues and therefore ordinary least squares cannot give accurate results. So it is not clear how to estimate the MIDAS models in a correct way and therefore the biggest challenge for MIDAS models is to find a good parametrization method.

In recent years many parametrization methods have been proposed. Many of them deal with the multicollinearity by imposing some structure on the parameters. Therefore, it is often assumed that the parameters lie on some kind of a polynomial. Examples of polynomials that are in the common practice often used, are the exponential Almon lag polynomial (see Almon (1965)) and the Beta lag polynomial (see Ghysels et al. (2007)). However, imposing such restrictions is not always a good idea. Franses (2016) shows that systematic patterns might be rare and hence imposing such restrictions results to misspecification and thus to biased estimates. To deal with this problem, I introduce in this paper a new approach in which I relax on these restrictions and therefore reduce the biased issue.

The new parametrization method uses a Ridge approach to estimate the MIDAS model and therefore I call this method the Ridge approach in this paper. Just like the parametrization methods mentioned above, it assumes some structure for the parameters. Only now the parameters does not have to fit exactly on the polynomial, so they can deviate from it. So by using the Ridge approach I impose some polynomial plus an error term on the parameters. In this way, the polynomial structure can deal with the multicollinearity and the error term can deal with the biased estimators. The challenge for this new parametrization method is to find a good balance between the multicollinearity issue and the biased estimators. To examine the performance of the Ridge approach in this paper, it is compared to the performance of several other commonly used parametrization methods for MIDAS models.

The paper is organized as follows. In section 2, first the MIDAS model is presented, then the Ridge approach and the other commonly used parametrization methods are introduced. The section finishes by explaining the used evaluation criterion that is used to evaluate the forecast performance of the different parametrization methods. In section 3, I use simulations to evaluate the performance of the Ridge approach. For these simulations two different DGPs are used. In section 4, I investigate the performance of the Ridge approach in two empirical studies. Section 5 concludes.

2 Methodology

It is always a big challenge to find a good parametrization method for the MIDAS model and therefore I introduce in this section a new parametrization method, which I call the Ridge approach. Also four commonly used parametrization methods are discussed. They are used as benchmark methods for the Ridge approach.

In this paper I use a similar notation as is used in Franses (2016). This means that for the variables sampled at the high frequency the time indicator is given by t, whereas the time indicator for the low frequency is denoted by T. The letter S is equal to the number of high frequency units that fits in a single low frequency time unit. In this paper, T is sampled quarterly and t monthly and hence S equals 3. Further the letters y and x are used for the high frequency and Y and X for the low frequency. Moreover I divide the sample in an estimation sample and a forecasts sample. In this way, I can use the estimation sample to estimate the MIDAS model and the forecast sample to evaluate the forecast performance of the different methods. In this paper, the estimation sample is given by $1 \dots T_{est}$ and the forecast sample by $T_{est} + 1 \dots T_{est} + h$.

This section starts with an introduction of the autoregressive MIDAS model. Then five different parametrization methods are discussed. First, the unrestricted parametrization method is explained, which uses ordinary least squares (OLS) to estimate the MIDAS model. Secondly, the Almon lag (AL) is introduced. Thirdly, I introduce the Ridge approach, which is actually a combination of the above two parametrization methods. Finally, the last two benchmark methods are introduced. These are the Exponential Almon lag (EAL) and the Beta lag (BL). The section ends with a short discussion of the used evaluation criterion. This is the root mean square error (RMSE) and it is used to compare the forecast performance of the Ridge approach with the benchmark methods.

2.1 The AR-MIDAS model

To deal with the different frequencies in the dependent and independent variables, Ghysels et al. (2004) introduced the Mixed Data Sampling (MIDAS) model. This model is closely related to the distributed lag model (DL), but this time the dependent variable is sampled on a low frequency, whereas the independent variables are sampled on a high frequency. The basic MIDAS model is given by

$$Y_T = \mu + \beta C(L^{\frac{1}{S}};\theta)x_t + \epsilon_T \tag{1}$$

Where $C(L^{\frac{1}{S}};\theta) = \sum_{i=0}^{n} c(i;\theta)L^{\frac{i}{S}}$. Here the weights are given by $c(i;\theta)$ and so $\sum_{j=0}^{n} c(j;\theta) = 1$. Further n is the number of included lags. I also include an autoregressive term (AR) in the model, as this is recommended by Franses (2016). Then the AR-MIDAS model becomes

$$Y_T = \mu + \rho Y_{T-1} + \beta C(L^{\frac{1}{S}};\theta)x_t + \epsilon_T \tag{2}$$

In this paper $\beta c(i;\theta)$ is often denoted by β_i, so that (2) becomes

$$Y_T = \mu + \rho Y_{T-1} + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_n x_{t-n} + \epsilon_T \tag{3}$$

Information criteria, such as BIC, can be used to determine the right number of lags, n. Finding a good parametrization method for the MIDAS regressions is one of biggest challenges. This is the result of the large lag order, n, as the included lags of the independent variable are likely to be highly correlated with each other. This leads to a multicollinearity issue and therefore OLS is not always feasible or accurate. To solve this problem, it is often assumed that there is some structure in the β parameters in (3). The structure can then be displayed by some function and so to derive the parameters we only have to estimate this function. In the common practice this function has often just two or three parameters. So instead of estimating all the β coefficients in (3), one only has to estimate the two or three parameters of the implied function. The smaller number of parameters results to more accurate estimators (smaller variance). However, if the assumed parameter structure is wrong, then imposing such a structure results to biased estimators. So imposing such a structure leads to a bias-variance tradeoff. In the following subsections, five different parametrization methods are discussed.

2.2 The Unrestricted MIDAS model

As mentioned above it is not always a good idea to impose a structure on the β parameters in (3), because if the assumed structure is wrong this results to biased estimators. Moreover, Franses (2016) concludes that it is unlikely that the parameters of the explanatory variables obey some certain convenient pattern. Foroni et al. (2015) investigated the performance of the Unrestricted MIDAS (U-MIDAS) model. Here the parameters in 3 are estimated just by using ordinary least squares. They showed that in the case of small differences in the sample frequencies, U-MIDAS works in most of the cases better than the restricted parametrization methods. However, this approach does not deal with multicollinearity and therefore as the differences in sample frequency increases (and so the number of included lags) it is likely that multicollinearity becomes a bigger issue. In this case the performance of U-MIDAS deteriorates, since U-MIDAS is then not able to estimate the parameters accurately.

2.3 Almon lag structure

The Almon lag (AL) structure was first introduced by Almon (1965). It is often used to reduce the multicollinearity in the distributed lag literature. The AL assumes that the parameters β_i are an unknown function of i. This function is a polynomial of order p and given by

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_p i^p \tag{4}$$

The coefficients, a_i, are unknown. Now I can substitute (4) in (3) and by making some rearrangements (3) can be converted to

$$Y_T = \mu + \rho Y_{T-1} + A\Pi X_t \tag{5}$$

Where $A = [a_0, a_1, \cdots, a_p]$, $X_t = [x_t, x_{t-1}, \cdots, x_{t-n}]$ and Π is given by

The restriction matrix, Π

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2^2 & \cdots & 2^p \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & n & n^2 & \cdots & n^p \end{bmatrix}$$

So by using this notation I just have to estimate p+1 parameters, a_i, instead of n+1 parameters, β_i. So this is a reduction of n-p parameters and therefore the parameters can be estimated more precisely. The model in (5) is estimated by using OLS. The β_i parameters in (3) can now be calculated by simply multiplying the estimated parameters $\hat{A}$ with the restriction matrix Π, that is

$$\hat{B}_{AL} = \hat{A}\Pi \tag{6}$$

Where $B = [\beta_0, \beta_1, \cdots, \beta_n]$. By using a higher order of p, the assumed function in (4) becomes a higher order polynomial. A higher order polynomial is of course more flexible and therefore the parameters are less restricted and this reduces the biased issue in the estimators. However, a higher p also leads to a higher number of parameters that has to be estimated and this results to less accurate estimators. In the common practice, a p of two or three is mostly used. In this paper I set p equal to two. The Almon polynomial is then not so flexible, but on the other hand it reduces the amount of parameters substantially.

2.4 The Ridge approach

Here, I introduce a new parametrization method for the AR-MIDAS regression. It combines the U-MIDAS model and the Almon lag polynomial using a Ridge approach. Ridge was first mentioned by Hoerl and Kennard (1970). For this new approach I assume that the parameters β_i in (3) are equal to a function plus an error term. In this way it is less restricted than other functional parametrization methods and so it can deal with the biased issue, but at the same time it is also able to deal with multicollinearity. For this method I use a Bayesian approach, which results to the Ridge coefficients.

The ridge parameters can be deduced by using a 2 step Bayesian approach. First, I impose a structure on the parameters β_i in (3). This can be any polynomial, but in this paper I have chosen to use the Almon polynomial. So the implementation is similar to that of the previous subsection. Also for the Ridge approach I use p = 2 in (4). So I again use the model that is given by (5). Only by this new approach the restrictions are relaxed and hence gives some room for an error term.

This error term is implemented in the next step. In this step I assume that the parameters of the explanatory variable, B, have the following prior function: $B_{Ridge} \sim N(\hat{B}_{AL}, \lambda\sigma^2 I)$. Here λ is the shrinkage factor. A large shrinkage factor corresponds to a large variance of the Ridge parameters and thus gives room for a large error term. On the other hand, a small shrinkage factor corresponds to a small error term. Now by using Bayes Rule it can be derived that

$$\hat{B}_{ridge} = ((X'X) + \frac{1}{\lambda}I)^{-1}(X'y + \frac{1}{\lambda}I\Pi\hat{B}_{Al}) \tag{7}$$

Here the optimal λ can be found by examining the performance of the ridge regression for different values of λ. Here I use the 5-fold cross validation method, so I split the sample data into five groups. Then for a particular lambda I use four groups to forecast the fifth group and do this for all the different group combinations. Then I estimate the average RMSE of all these different forecasts. The shrinkage factor which belongs to the lowest average RMSE is selected.

By using the Ridge parameters as shown in (7) it is clear that for a large shrinkage factor B_{Ridge} converges to B_{OLS} and that for a small shrinkage factor B_{Ridge} converges to the parameters of the Almon lag structure. So the Ridge parameters are related to the parameters of U-MIDAS and to that of the Almon lag.

2.5 Two other benchmark methods

Here two other parametrization methods for the MIDAS regression in (2) are mentioned. The first method imposes an Exponential Almon Lag (EAL) polynomial on the parameters, whereas the second method uses the Beta Lag (BL) polynomial. Both methods are often used in the common practice and have a reputation of giving good results.

The EAL was inspired by the Almon lag polynomial and introduced by Ghysels et al. (2007). By using the EAL polynomial the weights, c_i, in (2) are described by

$$c_i = c(i; \theta = [\theta_1, \theta_2]) = \frac{e^{\theta_1 i + \theta_2 i^2}}{\sum_{i=1}^{n} e^{\theta_1 i + \theta_2 i^2}} \tag{8}$$

For the second method a Beta lag structure is imposed. The Beta lag was first introduced by Ghysels et al. (2007). This means that the weights, c_i, in (2) are given by

$$c_i = c(i; \theta = [\theta_1, \theta_2]) = \frac{x_i^{\theta_1 - 1(1-x_i)^{\theta_2 - 1}}}{\sum_{i=1}^{n} x_i^{\theta_1 - 1}(1 - x_i)^{\theta_2 - 1}} \tag{9}$$

where $x_i = \frac{i}{N+1}$. Both polynomials are known to be very flexible and therefore they can take various different shapes. This includes increasing, decreasing and hump-shaped patterns. Imposing restrictions on the model results to biased estimators. So by using a more flexible polynomial this misspecification can be reduced, because the parameters are then less restricted. In this paper, I estimate the parameters by using nonlinear least squares (NLS). Further the MIDAS Matlab Toolbox is used for the implementation of these two methods. This toolbox is written by Eric Ghysels and his collaborators.

2.6 Evaluation criterion

Now the parametrization methods are known, I can use the estimation sample to estimate the parameters in (3). With the estimated MIDAS model, it is possible to make 1-step ahead forecasts. These forecasts can be calculated by

$$\hat{Y}_T = \hat{\mu} + \hat{\rho} Y_{T-1} + \hat{\beta}_0 x_t + \hat{\beta}_1 x_{t-1} + \cdots + \hat{\beta}_n x_{t-n} + \epsilon_T \qquad T = T_{est}+1, \ldots, T_{est}+h \tag{10}$$

By using this formula the forecast performance of the different parametrization methods can be evaluated. This can be done by comparing the forecasted values in (10) with the real values of the low frequency variable over this time period. As evaluation method, I use the same evaluation criterion as is used in Franses (2016). This is the Root Mean Square Error (RMSE). The RMSE squares the errors before they are averaged and therefore it gives a relatively high weight to large errors. So it is particularly useful when large errors are undesirable. The RMSE can be calculated in the following way

$$RMSE = \sqrt{\frac{1}{h} \sum_{T=T_{est}+1}^{T_{est}+h} (Y_T - \hat{Y}_T)^2} \tag{11}$$

3 Simulation

In this section I conduct some simulations to examine the performance of the Ridge regression and to compare the performance of this method with the other mentioned parametrization methods in section 2. The simulations allow us to have full control over the data generating process. So in this way I can examine the performance of the Ridge approach under self-induced circumstances.

In the simulations the high frequency variable, x, corresponds to monthly observations, whereas the low frequency variable, Y, corresponds to quarterly observations. For the simulations, I use two different DGPs. Furthermore, I assume that both DGPs at the high frequency are an autoregressive distributed lag model with lags (3,3) (ADL(3,3)). In the first simulation, I use a DGP in which there is not a clear pattern in the parameters of the explanatory variables and hence it favors U-MIDAS. On the other hand I use in the second simulation a DGP in which there is a much clearer structure. This benefits the restricted parametrization methods, which impose a structure on the parameters. It is expected that the Ridge approach performs well for both DGPs, as this method has both the characteristics of the unrestricted method (U-MIDAS) as that of the restricted methods (AL, EAL and BL).

In this section I first introduce the used MIDAS model and how it is derived. Secondly, I show how the data is generated and thirdly, the results are given and discussed.

3.1 MIDAS model

In this paper, the low frequent variable is sampled quarterly and the high frequent variable monthly, therefore I assume that the DGP at the high frequency is an ADL(3,3) model, so that

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \epsilon_t \tag{12}$$

Franses (2016) shows, by using the HILO transformation, that in this case a proper MIDAS model has the following form

$$\begin{aligned} Y_T = \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} \\ + \beta_5^* X_{1,T-1} + \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + \epsilon_T + \theta \epsilon_{T-1} \end{aligned} \tag{13}$$

Removing the moving average term, ϵ_{T-1}, will result to misspecification, however Franses (2016) shows that excluding the MA term is not very harmful. Therefore, I determine to exclude the moving average term out of the model. Furthermore a constant term is added to (13) and therefore the MIDAS model that is used in this paper has the following form

$$\begin{aligned} Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \\ \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + \epsilon_T \end{aligned} \tag{14}$$

This MIDAS model is equal to (3) when 8 lags are included in the model, so this corresponds to n = 8.

3.2 Data

The data is generated similar as in Franses (2016). This means that $x_t \sim N(1,1)$, $\epsilon_t \sim N(0,1)$ and $y_0/y_1/y_2 \sim N(0,1)$, then both DGPs at the high frequency are used to generate the missing y_t observations. In the first simulation I use the following high frequency DGP

$$y_t = \frac{\alpha}{2}y_{t-1} + \frac{\alpha}{4}y_{t-2} + \frac{\alpha}{8}y_{t-3} + x_t + 2x_{t-1} + 0.5x_{t-2} + 0.25x_{t-3} + \epsilon_t \quad (15)$$

And in the second simulation I use

$$y_t = \frac{\alpha}{2}y_{t-1} - \frac{\alpha}{4}y_{t-2} + \frac{\alpha}{8}y_{t-3} + x_t - 2x_{t-1} + 0.5x_{t-2} - 0.25x_{t-3} + \epsilon_t \quad (16)$$

For both the DGPs I consider 12 different cases. These are distinguished by the value of α and by the number of observations. I consider four different values of α. These are: $\alpha = 0.5$, $\alpha = 0.8$, $\alpha = 0.9$ and $\alpha = 0.95$ and for each different α I examine the performance of the different methods for N = 40, N = 120 and N = 260, where N is the number of simulated quarterly observations. So N = 40 corresponds to a period of 10 years, N = 120 to a period of 30 years and N = 260 to a period of 65 years. Because (14) is using 8 lags, the estimation starts at the third quarter (N = 3). Moreover to examine the forecast performances of the different methods I split the simulated data in an estimation sample and a forecast sample. This forecast sample will be equal to the last 5 quarterly observations. Therefore Y_3 to Y_{N-5} are used to estimate the MIDAS model and Y_{N-4} to Y_{40} are used to examine the predictions using the RMSE evaluation criterion. To get reliable results, I run every simulation a 1000 times and average the outputs.

3.3 Results

In table 1 the RMSEs of the different methods are given. We consider three different cases for N. It seems that for a small number of observations, N = 40, a lower shrinkage factor is chosen for the Ridge approach. The reason for this is that in this case the OLS parameters cannot be estimated accurately, due to the low number of observations and therefore the Ridge approach prefers to bind itself closer to the Almon polynomial. This is because the number of coefficients in the Almon structure is smaller and therefore they can be estimated more accurately. Of course this will come at the expense of some bias. Then if the bias-variance trade-off is favorable this will lead to better forecasts. However, by using the DGP in (15) a clear pattern in the parameters cannot be found and therefore the restricted parametrization methods produce biased estimators. Restricting the parameters too close to the Almon parameters is therefore

Table 1: Simulation results for the first DGP. This table reports the RMSE for five different parametrization methods, these are: Ridge, U-MIDAS (OLS), the Almon Lag (AL), the Exponential Almon Lag (EAL) and the Beta Lag (BL). The number between the parenthesis denotes the shrinkage factor, λ. The simulation results are based on a sample of N quarterly observations. Always the last five observations are used to evaluate the forecast performance and so the first N - 5 observations are used to estimate the model. For each simulation, the best result(s) is(are) given in bold. Furthermore, the monthly DGP is given by $y_t = \frac{\alpha}{2}y_{t-1} + \frac{\alpha}{4}y_{t-2} + \frac{\alpha}{8}y_{t-3} + x_t + 2x_{t-1} + 0.5x_{t-2} + 0.25x_{t-3} + \epsilon_t$, where $x_t \sim N(1,1)$, $\epsilon_t \sim N(0,1)$ and $y_0/y_1/y_2 \sim N(0,1)$. The MIDAS regression is

$Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + \epsilon_T$

alpha	N	Ridge	OLS	AL	EAL	BL
0.5	40	**2.5809** (10)	2.5814	3.3201	6.2899	6.2856
	120	**2.2137** (100000)	**2.2137**	3.1442	5.8222	5.8203
	260	**2.1135** (100000)	**2.1135**	3.0722	5.7851	5.7857
0.8	40	**2.9553** (10)	2.9606	4.3878	7.0246	7.0236
	120	**2.5240** (100000)	**2.5240**	4.0252	6.4244	6.4228
	260	**2.5003** (100000)	**2.5003**	4.0504	6.2970	6.2937
0.9	40	**3.0451** (100)	3.0455	4.6941	7.1125	7.0833
	120	**2.6900** (100000)	**2.6900**	4.4583	7.0602	7.0620
	260	**2.5730** (100000)	**2.5730**	4.3065	6.6245	6.6224
0.95	40	**3.2082** (100)	3.2092	4.9159	7.2257	7.2287
	120	**2.7552** (100000)	**2.7552**	4.5987	6.8383	6.8313
	260	**2.7097** (100000)	**2.7097**	4.5136	6.6772	6.6781

not desirable and this is also confirmed by the shrinkage factor in table 1. The shrinkage factor is always equal or larger than 10 and therefore the parameters still have much freedom.

For a larger N, the parameters can be estimated more accurately. This improves the prediction power of all the parametrization methods. However, it improves the prediction power of OLS more than that of the other methods. This is because multicollinearity was a bigger issue for U-MIDAS, due to the large number of parameters. With the large number of observation, there is more information available to estimate all the unbiased parameters accurately. For the Almon polynomial, multicollinearity was a smaller issue for N = 40, and therefore not much improvements are made for a larger number of observations. However, the estimated Almon parameters are still biased and so it still produces a weak forecast performance. So for a large N, OLS is able to estimate the parameters with almost the same standard error as the Almon polynomial, however the parameters of OLS are unbiased, whereas the parameters of the Almon method are biased. So U-MIDAS is in this case dominant to the Almon lag or the other benchmark methods and hence Ridge prefers to stay as close to U-MIDAS as possible. For this reason Ridge selects a very large shrinkage factor, so that the Ridge parameters converges to the OLS parameters.

The table shows that for the high frequency DGP the parametrization methods, corresponding to the EAL and BL structure, are not powerful methods. As I said before a clear pattern cannot be found and thus to bound the parameters to such a polynomial is not a very good idea and leads to biased coefficients. As a result the performance of these parametrization methods are weak in comparison with the methods which are not or less restricted to some polynomial, like OLS and Ridge.

Further, table 1 shows that the performance of OLS relative to that of the Almon lag improves for a larger value of α. This follows from the observation that for a larger value of α a higher shrinkage factor is selected.

Table 2: Simulation results for the second DGP. This table reports the RMSE results for five different parametrization methods, these are: Ridge, U-MIDAS (OLS), the Almon Lag (AL), the Exponential Almon Lag (EAL) and the Beta Lag (BL). The number between the parenthesis denotes the shrinkage factor, λ. The simulation results are based on a sample of N quarterly observations. Always the last five observations are used to evaluate the forecast performance and the first N - 5 observations are used to estimate the model. For each simulation, the best result is given in bold. Furthermore, the monthly DGP is given by $y_t = \frac{\alpha}{2}y_{t-1} - \frac{\alpha}{4}y_{t-2} + \frac{\alpha}{8}y_{t-3} + x_t - 2x_{t-1} + 0.5x_{t-2} - 0.25x_{t-3} + \epsilon_t$, where $x_t \sim N(1,1)$, $\epsilon_t \sim N(0,1)$ and $y_0/y_1/y_2 \sim N(0,1)$. The MIDAS regression is

$$Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + \epsilon_T$$

alpha	N	Ridge	OLS	AL	EAL	BL
0.5	40	**2.2689** (0.1)	2.2917	2.6133	2.5256	2.5014
	120	**1.9809** (1)	1.9818	2.4346	2.3642	2.3574
	260	1.9133 (1)	**1.9132**	2.4483	2.3747	2.3725
0.8	40	**2.5456** (0.1)	2.5875	2.8464	2.7746	2.7665
	120	**2.2228** (1)	2.2235	2.6682	2.6249	2.6247
	260	**2.1618** (1)	2.1619	2.6228	2.5803	2.5823
0.9	40	**2.6129** (0.1)	2.6899	2.8888	2.8025	2.7948
	120	**2.3228** (1)	2.3233	2.7455	2.6808	2.6839
	260	**2.2698** (1)	2.2700	2,6924	2.6672	2.6717
0.95	40	**2.6105** (0.1)	2.6887	2.8871	2.7992	2.7896
	120	**2.3223** (1)	2.3238	2.7323	2.6706	2.6715
	260	**2.2719** (1)	2.2722	2.7141	2.6721	2.6719

In table 2 the results are given of the second DGP, which is given by (16). For this DGP it is easier to find some kind of a structure in the parameters and therefore it favors the methods that use a polynomial. By considering the case of N = 40, the table shows that a low shrinkage factor is selected. This again follows from the multicollinearity issue due to the small number of observations. In this way the OLS coefficients cannot be estimated accurately and therefore the Ridge parameters prefer to stay close to the Almon polynomial, which can be estimated much more precisely. Moreover, for the used DGP in table 2, it is easier to find some structure and therefore the Almon parameters are not as biased as they were in table 1. Hence the Almon parametrization method is now much stronger than it was in the previous table. For this reason, Ridge wants to bind itself close to the Almon polynomial. This is confirmed by the small shrinkage factors, $\lambda = 0.1$, in table 2. These shrinkage values are smaller than the shrinkage values in table 1. Further it follows from table 2 that the Ridge approach outperforms all the benchmark methods for N = 40.

For higher N, the OLS performance increases. Therefore Ridge tends to get closer to the OLS parameters and picks a higher shrinkage factor. Nevertheless, also with a large number of observations the shrinkage factor does not become larger than 1. So it still prefers to keep close to the Almon polynomial. In this way the Ridge approach is even able in all of the cases, except one, to outperform all the other parametrization methods. So although the restriction parametrization methods are not able to beat U-MIDAS, a parametrization method which uses a restriction plus an error term actually can outperform U-MIDAS.

By looking to the performances of the EAL and BL parametrization methods, it can be concluded that the performances are worse than that of Ridge and OLS. However, they are relatively not as bad as they were in table 1. So this again shows that the parameters of the used DGP in table 2 can be better explained by some polynomial than the one used in table 1.

Finally, table 2 shows that for a larger value of α the difference between the RMSE of Ridge and OLS increases. However, it is difficult to draw some conclusions out of this, because this is for some part explained by the fact that a DGP with a larger α creates higher values of the dependent variable. So this means that the size of the observations are larger and therefore it is also likely that the size of the forecast errors are larger.

To summarize the two tables, it seems that especially for a small number of observations Ridge produces better forecasts than OLS, because in this case the OLS parameters cannot be estimated accurately due to multicollinearity. In the case of a large number of observations Ridge gives better or approximately equal results than OLS. So Ridge is never worse than OLS, but can produce better results which is most likely the case for a small number of observations or when there is some clear structure in the parameters. Moreover, Ridge always produces better results than the AL. Finally, for the used simulation DGPs in this paper, the EAL and BL parametrization methods are not good parametrization methods.

4 Empirical study

In the previous section, I have shown by using two different DGPs that the Ridge approach gives always better or approximately equal results than the benchmark methods. Ridge especially performs well if the datasets are small. However, by using the simulations I was able to have full control over the data generating process so the question remains if these results also apply for empirical data. Therefore, in this section I examine the performance of the Ridge approach for two small empirical datasets. In both datasets the low frequency variable is sampled quarterly and the high frequency variable monthly.

This section starts by introducing the empirical data. Secondly, some adjustments in the methodology are mentioned. Finally, the results are given and discussed.

4.1 Data

As mentioned above, I examine the Ridge approach for two different small datasets. All the data is taken from CBS StatLine. In the first dataset the seasonal adjusted GDP of the Netherlands is used for the low frequency variable and the seasonal adjusted average industrial day production for the high frequency variable. Both variables are on the time period from January 2005 until March 2017. This corresponds to 49 quarterly observations and 196 monthly observations.

In the second dataset the seasonal adjusted GDP of the Netherlands is again used for the low frequency variable and the total stock market value on EURONEXT Amsterdam for the high frequency variable. These variables are taken from January 2005 until March 2017. This corresponds to 46 quarters and 184 months.

By plotting the data it becomes clear that there is a trend in all the variables that are used in both datasets. To remove this trend I calculate the growth rate for all the variables. The growth rate is calculated by using the following formula

$$x_{growth,t} = \frac{x_t - x_{t-1}}{x_{t-1}} * 100 \tag{17}$$

So one observation is lost by using the growth rate. This reduces the total number of quarterly observations to 48 and 45 for the first and second dataset respectively. Figure 1 represents the first dataset and figure 2 displays the second dataset. I once again split the data in an estimation sample and a forecast sample. The first 37 observations are used to estimate the initial model for both datasets. So this means that for the first dataset the forecast sample contains 11 observations and for the second dataset the forecast sample contain 8 observations. I use a recursive forecast method. This method makes use of an increasing window to re-estimate the MIDAS model over the out-of-sample period.

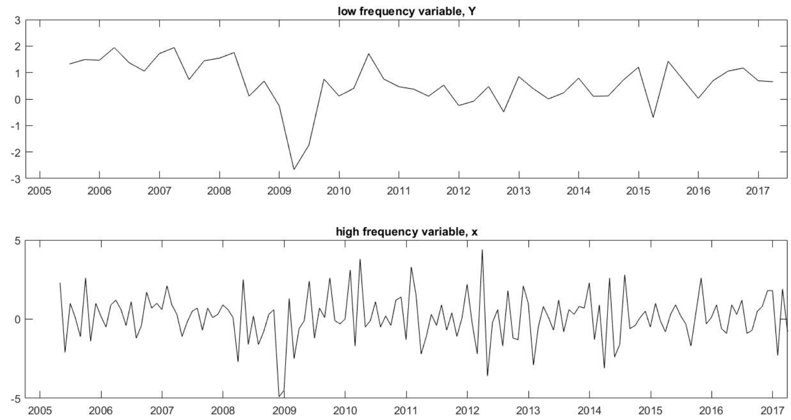

Figure 1: The first dataset. For the low frequency variable, Y, the growth rate of the seasonal adjusted GDP of the Netherlands is displayed and for the high frequency variable, x, the growth rate of the seasonal adjusted average industrial day production of the Netherlands is shown

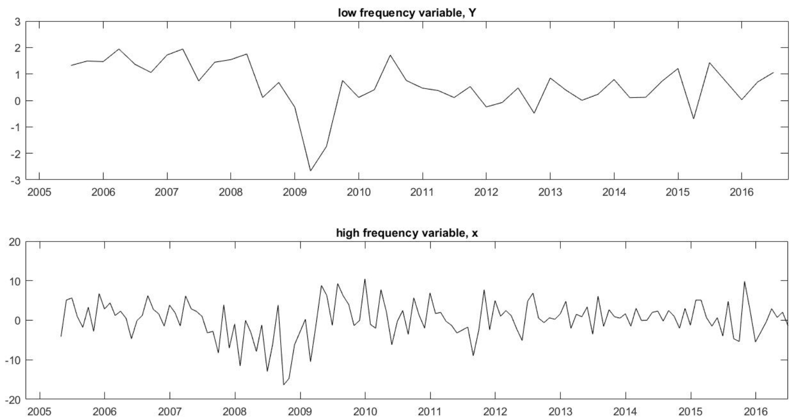

Figure 2: The second dataset. For the low frequency variable, Y, the growth rate of the seasonal adjusted GDP of the Netherlands is displayed and for the high frequency variable, x, the growth rate of the total stock market value on EURONEXT Amsterdam is shown

4.2 Data specific methodology

I use for the empirical studies the same MIDAS model that is used in the simulation section. So this model is once again described by (14).

In the empirical study I use another method to determine the shrinkage factor. The reason for this is that the 5-fold cross validation method does not give good results. This is because unlike in the simulation study, where the low frequent variable was generated by some DGP, the shrinkage factor in the empirical studies is not constant and changes over time. So by using the entire sample to determine the shrinkage factor, one gets some average shrinkage factor, which in most cases does not give good results for the one-step ahead forecasts. For this reason I do not use the 5-fold cross validation method. Alternatively, I consider two different cases. In the first case, no selection method for the optimal shrinkage factor is used. Instead, I examine the performance of the Ridge approach for 10 different shrinkage values. Here the shrinkage factor remains constant for all the recursive forecasts. In this way I can investigate how much the Ridge approach depends on the choice of the shrinkage factor. In the second case, I examine the performance of a method which tries to find the best shrinkage factor for every one-step ahead forecast. So the shrinkage factor can change over the out-of-sample period. Therefore, I adjust the selection method that is described in section 2. Now the shrinkage factor is determined by examining which shrinkage factor produces the lowest value of the RMSE for just the last q observations of the estimation sample. The RMSE can be calculated using the following formula

$$RMSE = \sqrt{\frac{1}{q} \sum_{T=T_{est}+1-q}^{T_{est}} (Y_T - \hat{Y}_T)^2} \tag{18}$$

I examine q for five different values, these are: q=1, q=2, q=3, q=5 and q=10.

4.3 Results

Table 3 gives the results of the two empirical studies. The RMSEs are reported for 10 different shrinkage factors. For the first dataset, the table shows that for an increasing shrinkage factor the Ridge performance converges to the U-MIDAS performance. Moreover, when the shrinkage factor decreases Ridge converges to the AL. Furthermore, the performance of Ridge are always between that of OLS and the AL. By considering the other benchmark methods, the table shows that the EAL and the BL are producing slightly better results than OLS. So for a large shrinkage factor, $\lambda \geq 1$, they also outperform Ridge, because Ridge is then close to OLS. But in the other cases Ridge is the stronger method.

The second empirical study gives some different results. Nonetheless, it can again be concluded that Ridge converges to OLS and the AL for a large and small shrinkage factor respectively. However, this time the Ridge approach outperforms both OLS and the AL for the relatively small shrinkage values of 0.00001, 0.0001 and 0.001. So improvements in the prediction performance of the AL can be made by relaxing the restrictions of the Almon polynomial a little bit. The other two benchmark methods are this time outperforming all the other parametrization methods. So although for some small shrinkage factor Ridge produces better results than OLS and the AL, it is not able to beat the EAL and the BL.

So by viewing the results of table 3 and by considering the results of the simulations in section 3, it seems that the Ridge approach has some upper bound. This can be reasoned by the observation that it has yet never produced worse than the worst method of OLS and the AL. So it seems that $RMSE_{Ridge} \leq max(RMSE_{OLS}, RMSE_{AL})$.

Table 3: Results of the first and second empirical study. For these studies, the growth rate of the seasonal adjusted GDP of the Netherlands is used for the low frequency variable (quarterly observations). In the first study the growth rate of the seasonal adjusted average industrial day production is used for the high frequency variable (monthly observations) and in the second study the growth rate of the total stock market value on EURONEXT Amsterdam is used for the high frequency variable. This table reports for 10 different shrinkage factors, λ, the RMSE for five different parametrization methods. These are: Ridge, U-MIDAS (OLS), the Almon lag (AL), the Exponential Almon lag (EAL) and the Beta lag (BL). For every shrinkage factor the best result(s) of Ridge, U-MIDAS and the Almon Lag is(are) given in bold. The used MIDAS regression is

$$Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + \epsilon_T$$

First dataset

λ	Ridge	OLS	AL	EAL	BL
0.00001	**0.7388**	0.8676	**0.7388**	0.8411	0.8414
0.0001	0.7392	0.8676	**0.7388**	0.8411	0.8414
0.001	0.7434	0.8676	**0.7388**	0.8411	0.8414
0.01	0.7732	0.8676	**0.7388**	0.8411	0.8414
0.1	0.8317	0.8676	**0.7388**	0.8411	0.8414
1	0.8616	0.8676	**0.7388**	0.8411	0.8414
10	0.8670	0.8676	**0.7388**	0.8411	0.8414
100	0.8676	0.8676	**0.7388**	0.8411	0.8414
1000	0.8676	0.8676	**0.7388**	0.8411	0.8414
10000	0.8676	0.8676	**0.7388**	0.8411	0.8414

Second dataset

λ	Ridge	OLS	AL	EAL	BL
0.00001	**0.8416**	0.9156	0.8422	0.8039	0.7992
0.0001	**0.8372**	0.9156	0.8422	0.8039	0.7992
0.001	**0.8270**	0.9156	0.8422	0.8039	0.7992
0.01	0.8762	0.9156	**0.8422**	0.8039	0.7992
0.1	0.9102	0.9156	**0.8422**	0.8039	0.7992
1	0.9150	0.9156	**0.8422**	0.8039	0.7992
10	0.9155	0.9156	**0.8422**	0.8039	0.7992
100	0.9156	0.9156	**0.8422**	0.8039	0.7992
1000	0.9156	0.9156	**0.8422**	0.8039	0.7992
10000	0.9156	0.9156	**0.8422**	0.8039	0.7992

Although Ridge is not always producing the best results in the empirical studies, it is maybe a good idea to use Ridge if you are doubting between the choice of using OLS or the AL. By always taking the optimal shrinkage factor for every one-step ahead forecast, it can be even argued that $RMSE_{Ridge} \leq min(RMSE_{OLS}, RMSE_{AL})$, which is also confirmed by the simulation results. So using a good shrinkage factor selection method can improve the performance of Ridge substantially.

Table 4: Results of the first and second empirical study. For these studies, the growth rate of the seasonal adjusted GDP of the Netherlands is used for the low frequency variable (quarterly observations). In the first study the growth rate of the seasonal adjusted average industrial day production is used for the high frequency variable (monthly observations) and in the second study the growth rate of the total stock market value on EURONEXT Amsterdam is used for the high frequency variable. This table reports for 5 different estimation samples, q, (that are used to determine the optimal shrinkage factor) the RMSE for five different parametrization methods. These are: Ridge, U-MIDAS (OLS), the Almon lag (AL), the Exponential Almon lag (EAL) and the Beta lag (BL). For every shrinkage factor the best result(s) of Ridge, U-MIDAS and the Almon Lag is(are) given in bold. The used MIDAS regression is

$Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + \epsilon_T$

Dataset1

q	Ridge	OLS	AL	EAL	BL
1	0.7970	0.8676	**0.7388**	0.8411	0.8414
2	0.7446	0.8676	**0.7388**	0.8411	0.8414
3	**0.7387**	0.8676	0.7388	0.8411	0.8414
5	0.7461	0.8676	**0.7388**	0.8411	0.8414
10	**0.7388**	0.8676	**0.7388**	0.8411	0.8414

Dataset2

q	Ridge	OLS	AL	EAL	BL
1	0.9217	0.9156	**0.8422**	0.8039	0.7992
2	0.9438	0.9156	**0.8422**	0.8039	0.7992
3	0.9320	0.9156	**0.8422**	0.8039	0,7992
5	0.9083	0.9156	**0.8422**	0.8039	0.7992
10	0.8972	0.9156	**0.8422**	0.8039	0.7992

Although table 3 gives the results of Ridge for different constant shrinkage factors, it is still not clear how to chose the correct shrinkage factor for every one-step ahead forecast. So in table 4 the results are given of the different parametrization methods when the shrinkage factor is selected, using the method that is explained in subsection 4.2. The λs that are used for every one-step ahead forecast are given in table A.1 and table A.2 of the appendix. By looking to the first dataset in table 4 it seems that this method works pretty well and gives even better results than both OLS and the AL for q = 3. Such good results could not be achieved in table 3.

However, by looking at the second dataset, the results are not so satisfactory. This time Ridge produces even worse results than both OLS and the AL for q $\leq$ 3. The best results are generated for q = 10. In this case the Ridge approach produces better results than OLS, but it is still much worse than that of the AL. So this shrinkage factor selection method is in the second empirical study never able to beat the AL, while this was the case for a constant shrinkage factor in table 3. Furthermore, it is still not clear what value to use for q, as the optimal q in the first dataset equals 3 but in the second dataset it equals 10. So it seems that q differs for each different dataset.

Table 4 shows that the EAL and the BL give similar results in both datasets. In the first dataset these results are worse than the results of the AL and the Ridge approach, but better than the performance of OLS. For the second dataset, the EAL and the BL outperforms all the other methods. So it seems that the parameters in the second dataset can be well explained by a flexible function like the Beta polynomial and the Exponential Almon polynomial.

5 Conclusion

In this paper I introduce a Ridge approach to estimate the MIDAS model. I assume that the dependent variable is hypothetically observed at the high frequency and from this assumption a correctly specified MIDAS model is derived. The goal of this paper is to examine the performance of the Ridge approach. Therefore, I compare it to four other parametrization methods, which are often used in the common practice and have a reputation of giving good results. The Ridge performance is investigated for four different situations. In the first part two different simulations are conducted. The simulations allow me to have full control over the data generating process. This allows me to investigate the Ridge approach under self-induced circumstances. In the second part the Ridge performance is evaluated in two empirical studies, where, as low frequency variable, the GDP of the Netherlands (quarterly observations) is tried to be explained by some high frequency variable (monthly observations).

In both the simulations the Ridge approach gives better or approximately equal results than the other parametrization methods. So it is never worse. It is especially a strong method if one has to deal with multicollinearity and this is often the case when there is just a limited number of observations available.

In the empirical studies, the Ridge results were not as good as in the simulations. This is because it is still not very clear how to determine the correct shrinkage factor for the Ridge method. However, by trying a lot of different shrinkage factors the results of the Ridge approach were in most cases competitive with that of the other benchmark methods. In some cases and for some specific shrinkage factors the Ridge approach was even able to outperform all the other benchmark methods. So if one finds a method that selects the optimal shrinkage factor, the Ridge approach will become a very strong parametrization method for the MIDAS model.

In this paper I only examined the Ridge approach for the MIDAS model where the low frequency variable is sampled quarterly and the high frequency variable is sampled monthly. The relative performance of Ridge can maybe improve for a larger sample difference, for example yearly and monthly observations. This is because then more variables are added and so it would become harder to estimate the model using OLS. Further research to this is recommended. Moreover, it is still not clear how to determine the right shrinkage factor. In this paper it becomes clear that Ridge is a strong parametrization method if the optimal shrinkage factor is selected. So I would also recommend further research to the selection of the optimal shrinkage factor. Moreover, in this paper I have used the Almon polynomial for the Ridge approach, but this can also be another polynomial. For example the Ridge approach can be applied on the Exponential Almon polynomial or the Beta polynomial. These polynomials are in general more flexible than the Almon polynomial. So maybe that applying the Ridge approach on these polynomials can improve the prediction power.

References

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 33(1):178–196.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (midas): Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1):57–82.

Foroni, C. and Marcellino, M. G. (2013). A survey of econometric methods for mixed-frequency data. Economics Working Papers ECO2013/02, European University Institute.

Franses, P. H. (2016). Yet another look at midas regression. Technical report.

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

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

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Appendix

Table A.1: Shrinkage factors of the first empirical study. This table reports the shrinkage factor for every one-step ahead forecast, h, for different values of q. Here the last q observations of the estimation sample are used to determine the shrinkage factor.

h	q = 1	q = 2	q = 3	q = 5	q = 10
1	1.0E-10	1.0E-10	1.0E-10	1.0E-10	1.0E-10
2	1.0E-10	1.0E-10	1.0E-10	1.0E-10	1.0E-10
3	1.0E+05	1.0E-10	1.0E-10	1.0E-10	1.0E-10
4	1.0E-10	1.0E-10	1.0E-10	1.0E-10	1.0E-10
5	1.0E-10	1.0E-10	1.0E-10	1.0E-10	1.0E-10
6	1.0E-01	1.0E-10	1.0E-10	1.0E-10	1.0E-10
7	1.0E+05	1.0E+05	1.0E-03	1.0E-10	1.0E-10
8	1.0E-03	1.0E+05	1.0E+05	1.0E-10	1.0E-10
9	1.0E-02	1.0E-02	1.0E-01	1.0E-03	1.0E-10
10	1.0E+05	1.0E+00	1.0E+00	1.0E+05	1.0E-10
11	1.0E+05	1.0E+05	1.0E+05	1.0E+05	1.0E-10

Table A.2: Shrinkage factors of the second empirical study. This table reports the shrinkage factor for every one-step ahead forecast, h, for different values of q. Here the last q observations of the estimation sample are used to determine the shrinkage factor.

h	q = 1	q = 2	q = 3	q = 5	q = 10
1	1.0E-10	1.0E-10	1.0E-10	1.0E-10	1.0E-01
2	1.0E-10	1.0E-10	1.0E-10	1.0E-10	1.0E-01
3	1.0E+05	1.0E-10	1.0E-10	1.0E-10	1.0E-01
4	1.0E-10	1.0E+05	1.0E+05	1.0E-03	1.0E-03
5	1.0E-10	1.0E-03	1.0E-03	1.0E-03	1.0E-03
6	1.0E-01	1.0E-03	1.0E-02	1.0E-02	1.0E-03
7	1.0E+05	1.0E+05	1.0E-02	1.0E+05	1.0E-03
8	1.0E-03	1.0E+05	1.0E+05	1.0E+05	1.0E-02