Dissertations, Master's Theses and Master's Reports

2016

A SELF-NORMALIZING APPROACH TO THE SPECIFICATION TEST OF MIXED FREQUENCY MODELS

Henriette Groenvik
Michigan Technological University, hgronvik@mtu.edu

Recommended Citation

A SELF-NORMALIZING APPROACH TO THE SPECIFICATION TEST OF

MIXED FREQUENCY MODELS

By

Henriette Grønvik

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2016

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Mathematical Sciences.

Department of Mathematical Sciences

Thesis Advisor: *Dr. Yeonwoo Rho*

Committee Member: *Dr. Seokwoo Choi*

Committee Member: *Dr. Latika Lagalo*

Committee Member: *Dr. Qiuying Sha*

Department Chair: *Dr. Mark S. Gockenbach*

Contents

List of Figures

List of Tables

Abstract

In econometrics and finance, variables are collected at different frequencies. If a higher frequency variable can help predict a lower frequency variable, it would be of interest to construct such regression models. One straightforward solution is to flat aggregate the higher frequency variable to match the lower frequency. However, flat aggregation may overlook useful information in the higher frequency variable. On the other hand, keeping all higher frequencies may result in overly complicated models. In literature, mixed data sampling (MIDAS) regression models have been proposed to balance between the two. In this thesis the mixed frequency models are addressed, and we propose a new model specification test that can help decide between the simple aggregation and the MIDAS model.

Chapter 1

Introduction

The mixed frequency models are regression models with variables sampled at different time frequencies. The models can be used for forecasting. Often in financial and econometric studies, variables are collected at different time frequencies. For example, macroeconomic variables are often published at a lower frequency compared to the financial variables' frequency. It can be of interest to forecast a lower frequency variable with a higher frequency variable. The mixed frequency can be pairs of, for example, quarterly and monthly, monthly and daily, or monthly and intra-daily. One way to incorporate variables of different frequencies in a model is to aggregate the high frequency (HF) variable to match the low frequency (LF). A simple solution of aggregation is to average the HF observations. This method is called flat aggregation because the weights given to the HF observations are the same. A flat aggregation

model is simple and relatively easy to compute. However, one concern could be a potential loss of information in the HF variable when it is flat aggregated (Andreou et al., 2010). There might be some important information stored in the HF, which could be vital for forecasting.

Instead of assigning equal weights, the HF observations can be treated as individual predictors and their weights can be uniquely estimated. This might create a forecasting model with high accuracy. However, if the frequency difference between the dependent and the independent variables is large, for example yearly and daily, then the number of parameters would also be very large. This model may not be an efficient one, possibly introducing difficulties in estimations.

Ghysels et al. (2004), Ghysels et al. (2005), Ghysels et al. (2006), and Ghysels et al. (2007) introduced a mixed data sampling (MIDAS) model. This model is able to preserve information in the HF variable that could otherwise be lost in a flat aggregation model. The key factor is a lag polynomial that distributes exclusive weights to the HF observations while still keeping the model from having a large number of parameters. However, a MIDAS model may not always forecast better than a flat aggregation model. The forecasting efficiency depends upon the relationship between the regressors and the regressand. If the MIDAS model is not doing better than the flat aggregation model, which is the simpler model, then the flat aggregation scheme is preferred to the MIDAS model. Thus, a specification test for determining the

optimal model could be useful.

Andreou et al. (2010) addressed the model specification problem by introducing two hypothesis tests which test for whether a flat aggregation model or a MIDAS model should be chosen. Miller (2014b) also developed the idea further by using a variable addition test similar to the second test given in Andreou et al. (2010). The test in Miller (2014b) uses an ancillary regression that explores the possibility of dependence between the fitted residuals from the flat aggregation model and linearly transformed HF variable that is similar to the aggregation in a MIDAS model. Miller (2014b) mentioned a possible test statistic with a heteroskedasticity autocorrelation consistent (HAC) estimator. The HAC estimator, as described in Newey and West (1987) and Andrews (1991), is dependent upon a user-chosen bandwidth parameter. Miller (2014b) modified the test eliminating the need for a HAC estimator, but the test still needs the model builder to chose a parameter that has a considerable effect on the test's size and power. In this thesis, we propose to extend the self-normalizing approach as described in Shao (2010), to the mixed frequency model specification test. The self-normalizing approach is free of any user-chosen values, providing a test statistic with a pivotal limiting null distribution.

Chapter 2 overviews the mixed frequency models focusing on the MIDAS model and its extensions. Chapter 3 reviews the specification tests for mixed frequency models given in Andreou et al. (2010) and Miller (2014b). Chapter 4 presents our extension

of the self-normalizing method to the mixed frequency specification test. In Chapter

5 multiple power analyses are shown for the different specification tests. In Chapter

6 a conclusion is provided.

Chapter 2

Mixed Frequency Models

In order to illustrate the mixed frequency models, some notations have to be established. Let the dependent series be denoted by y_t which is measured at the lowest frequency where $t = 1, \ldots, T$. The independent predictor series $\boldsymbol{x}_t^{(m)} = (x_t, x_{t-1/m}, \ldots, x_{t-(m-1)/m})'$ is measured m times more frequently than y_t, i.e the HF variable. For example, y_t is observed at the end of quarter t and $\boldsymbol{x}_t^{(3)}$ is the vector of monthly observations in quarter t. The entries in $\boldsymbol{x}_t^{(3)}$ are then $x_{t-0/3} = x_t$ the third month, $x_{t-1/3}$ the second month, and $x_{t-2/3}$ the first month of quarter t. Note that for the rest of the thesis the forecasting of the dependent variable will be done with the use of only one regressor and for one period ahead, $t + 1$, unless something else is specified.

A classical solution for fitting two frequencies in one regression model is to use a simple aggregation scheme on the HF variable to match the lower frequency. One way is to distribute equal weights to the HF observations within one LF time unit. The flat aggregation model distributes the weight $\frac{1}{m}$ to each HF observations. This means that the observations $x_{t-i/m}$ for all $i = 0, \ldots, m-1$ are averaged over the LF time unit t. The averaged HF observations can then be used as the explanatory variable in a distributed lag model, which is given by

$$y_{t+1} = \beta_0 + \beta_1 \boldsymbol{\pi}' \boldsymbol{x}_t^{(m)} + \varepsilon_{t+1}, \tag{2.1}$$

where β_1 is the slope coefficient and $\boldsymbol{\pi} = (\frac{1}{m}, \ldots, \frac{1}{m})'$ is the $m \times 1$ flat aggregation vector. The model in (2.1) is only using one LF lag of the HF variable. However, it is possible to add more low frequency lags. Let L be the LF lag operator. Then $L^j \boldsymbol{x}_t^{(m)} = \boldsymbol{x}_{t-j}^{(m)}$ is a vector of HF observations from $t-j$. There might also be the case where y_t is serially correlated. It would, therefore, be of interest to extend the model in (2.1) to an autoregressive distributed lag (ADL) model (Andreou et al., 2011). The ADL(p, q) model with p autoregressive terms in the y_t and q LF lags of $\boldsymbol{x}_t^{(m)}$ is as follows

$$y_{t+1} = \beta_0 + \sum_{s=0}^{p} \gamma_s L^s y_t + \sum_{j=0}^{q} \beta_j L^j \boldsymbol{\pi}' \boldsymbol{x}_t^{(m)} + \varepsilon_{t+1}, \tag{2.2}$$

where γ_s are the autoregressive coefficients for the lagged y_{t-s}, and β_j are the slope coefficients for the LF lagged $\boldsymbol{\pi}' \boldsymbol{x}_t^{(m)}$. Clearly, the number of parameters to estimate

6

increases with the number of LF lags used.

There are also other types of simple aggregation that can be applied to the HF variable. One that is mentioned in Miller (2014b) is the end-of-period sampling. This implies that only the last HF observation in the time unit t will be given weight. In other words, in (2.1) the vector $\boldsymbol{\pi} = (1, 0, \ldots, 0)'$. A distributed lag or an ADL(p, q) model can also be used for this type of aggregation.

Flat aggregation models are easy to interpret. However, it is not always given that the relationship between y_t and $\boldsymbol{x}_t^{(m)}$ is a flat aggregation model. Hence, applying unique weights to the HF observations is another possibility. A step-weighting model is a model that distribute weights as slope coefficients for each of the HF observations. This model was considered in Armesto et al. (2010), and with one LF lag of $\boldsymbol{x}_t^{(m)}$ and lagged values of the LF it can be derived in this way

$$y_{t+1} = \beta_0 + \sum_{s=0}^{p} \gamma_s L^s y_t + \boldsymbol{\varpi}' \boldsymbol{x}_t^{(m)} + \varepsilon_{t+1}, \tag{2.3}$$

where $\boldsymbol{\varpi}$ is a $m \times 1$ vector with unique slope coefficients for each HF observations. Therefore, the number of parameters is dependent upon the value of m and the number of lags for the LF variable. Thus, if m is large, then the parameters in (2.3) will increase drastically. The parameters will also multiply when LF lags are added to the HF variable in (2.3) (Armesto et al., 2010). This makes the estimation of the

model difficult and complex.

The introduction of the MIDAS model in Ghysels et al. (2004) gave a new solution for how to estimate unique coefficients for the HF observations. The MIDAS model is similar to the step-weighting model, but has less parameters to estimate. The weights are given by a lag polynomial that is only dependent upon the placement of the HF observation, i, and a parameter $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ can be both a scalar or a vector. The weights from the lag polynomial are stored in the $m \times 1$ vector denoted $\boldsymbol{\pi}(\boldsymbol{\theta}) = (\pi(0, \boldsymbol{\theta}), \ldots, \pi(m-1, \boldsymbol{\theta}))'$. The basic MIDAS model can be derived as follows

$$y_{t+1} = \beta_0 + \beta_1 \boldsymbol{\pi}(\boldsymbol{\theta})' \boldsymbol{x}_t^{(m)} + \varepsilon_{t+1}. \tag{2.4}$$

In order for the slope coefficient β_1 to be uniquely estimated, we assume that $\sum_{i=0}^{m-1} \pi(i, \boldsymbol{\theta}) = 1$. The model is parsimonious because β_0, β_1, and $\boldsymbol{\theta}$ are the only parameters that need to be estimated. There are especially two lag polynomials that have been discussed in the literature. One is the two-parameter exponential Almon lag polynomial. The individual weights from this lag polynomial are given as follows

$$\pi(i, \boldsymbol{\theta}) = \frac{exp(\theta_1(i+1) + \theta_2(i+1)^2)}{\sum_{j=1}^{m} exp(\theta_1(j+1) + \theta_2(j+1)^2)}, \tag{2.5}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ are the only parameters, regardless of the value of m. Note that when $\theta_1 = \theta_2 = 0$, then the function in (2.5) equals $\frac{1}{m}$, which in turn transforms

8

(2.4) to the flat aggregation model in (2.1). It is $\boldsymbol{\theta}$ that decides the shape of the lag function. The weights will decrease with the increase of the lags when $\theta_2 \leq 0$. As seen in Figure 2.1, the shape of the lag polynomials in (2.5) when $\theta_1 = 0.08$ and $\theta_2 = -0.008$ and $\theta_1 = 0.1$ and $\theta_2 = -0.03$ are functions with a hump. This means that lag 1 is given smaller weights than some of the larger lags. In the two other examples in Figure 2.1, lag 1 has the largest weight.
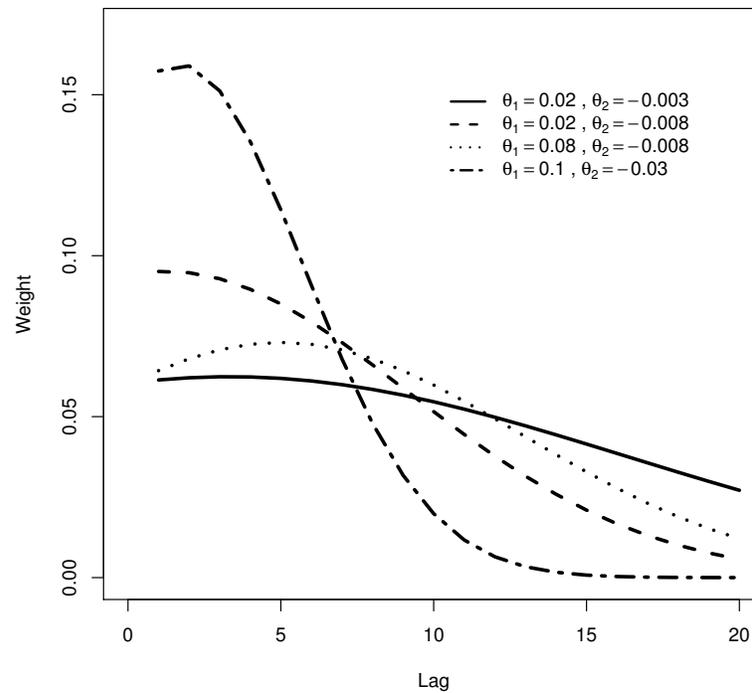
Figure 2.1: Examples of the exponential Almon lag polynomial for lags up to 20 based on different options for $\boldsymbol{\theta}$

The other popular option for the lag polynomial is described as the beta lag specification (Ghysels et al., 2007). This polynomial has two parameters, and resembles

the beta distribution. The weights for the HF observations are given as follows

$$\pi(i, \boldsymbol{\theta}) = \frac{f\left(\frac{i+1}{m}, \theta_1, \theta_2\right)}{\sum_{j=1}^{m} f\left(\frac{j}{m}, \theta_1, \theta_2\right)}, \tag{2.6}$$

where

$$f\left(\frac{i+1}{m}, \theta_1, \theta_2\right) = \frac{(i+1)^{\theta_1-1}(1-(i+1))^{\theta_2-1}\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}, \tag{2.7}$$

and $\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx$ is the gamma function. Note that (2.7) is the probability density function of the beta distribution. When $\theta_1 = \theta_2 = 1$, $\pi(i, \boldsymbol{\theta}) = \frac{1}{m}$ in (2.6), which is the flat aggregation weight. Examples of the beta lag specification are presented in Figure 2.2. It shows that a hump shaped function occurs at $\theta_1 = 3$ and $\theta_2 = 8$, and decreasing weights are observed for the other options of $\boldsymbol{\theta}$.

For the basic MIDAS model in (2.4) the lag polynomial is chosen by the user and the parameters, β_0, β_1, and $\boldsymbol{\theta}$, are estimated by nonlinear least squares.

Since the introduction of the MIDAS model, several additions and extensions to the model have been introduced. Ghysels et al. (2007) suggested the use of multiple regressors. This option has been discussed in later articles as well (see Clements and Galvão (2008), Clements and Galvão (2009), Andreou et al. (2010)). The model is called a M-MIDAS model where M stands for multiple regressors (Clements and Galvão (2008) and Clements and Galvão (2009)). The M-MIDAS model with n
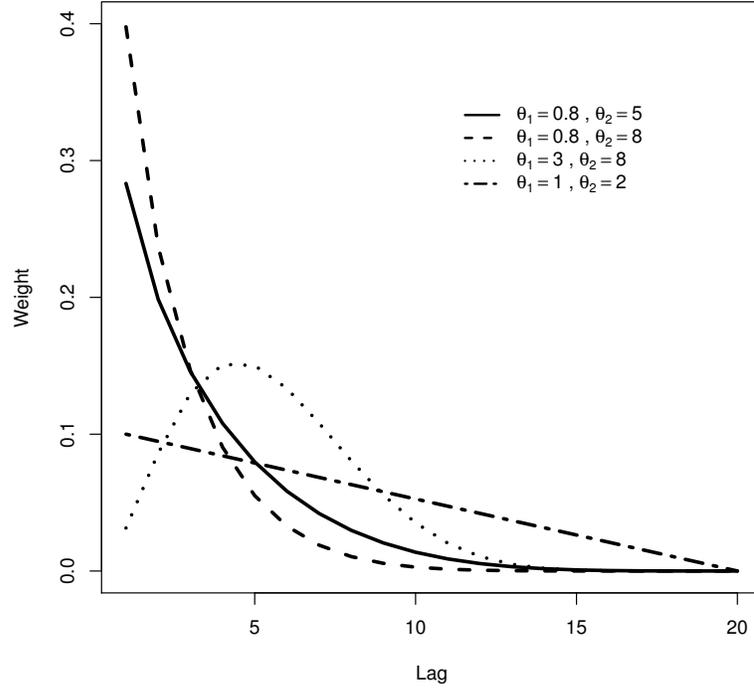
Figure 2.2: Examples of the beta lag polynomial for lags up to 20 based on different options for $\boldsymbol{\theta}$

regressors is derived in this way

$$y_{t+1} = \beta_0 + \sum_{j=1}^{n} \beta_{j,1} \boldsymbol{\pi}_j(\boldsymbol{\theta}_j)' \boldsymbol{x}_{j,t}^{(m)} + \varepsilon_{t+1}, \tag{2.8}$$

where $\boldsymbol{\pi}_j(\boldsymbol{\theta}_j)$ is the lag polynomial for the j predictor for $j = 1, \ldots, n$. Notice that the frequencies for the regressors are the same in (2.8), but it is possible to construct such model where the regressors have different frequencies.

It might be of interest to include an autoregressive term of y_t in a MIDAS model.

This type of model is called AR-MIDAS. This option was introduced in Ghysels et al. (2007). In Clements and Galvão (2008), it was shown that adding the AR-term straight into the model will create a spurious LF seasonal pattern in the $\boldsymbol{x}_t^{(m)}$. This event can be avoided if the autoregressive characteristic in y_t is employed as a common factor (Clements and Galvão, 2008). The following AR-MIDAS consists of one autoregressive LF lag

$$y_{t+1} = \beta_0 + \lambda y_t + \beta_1 \boldsymbol{\pi}(\boldsymbol{\theta})'(1 - \lambda L)\boldsymbol{x}_t^{(m)} + \varepsilon_{t+1}, \tag{2.9}$$

where λ is the AR coefficient and L is the LF lag operator for $t - 1$. If additional autoregressive lags are necessary, then they can be included without difficulty (see, Clements and Galvão (2008)).

In some situations observations of the HF variable might be available in the LF time unit that is being forecast. Armesto et al. (2010) labeled this intra-period forecasting. There are several different ways of incorporating the current available HF observations. The following equation is partially based on the intra-period MIDAS model given in Clements and Galvão (2009). Let s be the number of HF observations accessible in the LF time unit $t + 1$, which is being forecasted. Then $\boldsymbol{x}_{t+s/m}^{(m)}$ indicates that the s HF observations from $t + 1$ shall also be included. For example, if $s = 2$ and $m = 3$ (months in a quarter), then $t + 1/3$ is the first month of quarter $t + 1$, and $t + 2/3$ is the second month of $t + 1$. These are the intra-period observations that are

available for forecasting y_{t+1}. The MIDAS model with intra-period forecasting is as follows

$$y_{t+1} = \beta_0 + \beta_1 \boldsymbol{\pi}(\boldsymbol{\theta})' \boldsymbol{x}_{t+s/m}^{(m)} + \varepsilon_{t+1}, \tag{2.10}$$

where $\boldsymbol{\pi}(\boldsymbol{\theta})$ is a $(m+s) \times 1$ vector with weights from a chosen lag polynomial.

With a larger set of predictor variables, estimating a factor model can be beneficial. Marcellino and Schumacher (2010) introduced a factor MIDAS model. They discussed which factor estimation method is better, and three MIDAS variations are also compared. We will only give a brief summary of the basic factor MIDAS model given in Marcellino and Schumacher (2010). Let $\boldsymbol{X}_{t-i/m}$ be an $n \times 1$ vector of n multiple predictors that we wish to use in the forecasting. To reduce the dimensions of the model, the predictors are used to estimate $r \leq N$ factors, which are stored in $\boldsymbol{F}_{t-i/m}$. For the sake of simplicity assume $r = 1$ which means only one factor is necessary to explain the variability among the predictors. Denote the vector of estimated factors for time period t as $\hat{\boldsymbol{f}}_t^{(m)}$, then the estimated factors can replace the predictors in the MIDAS as follows

$$y_{t+1} = \beta_0 + \beta_1 \boldsymbol{\pi}(\boldsymbol{\theta})' \hat{\boldsymbol{f}}_t^{(m)} + \varepsilon_{t+1}. \tag{2.11}$$

Similar to the M-MIDAS model in (2.8), the factor MIDAS model can include multiple factors if that is applicable. A factor MIDAS with AR terms is also possible to implement (Marcellino and Schumacher, 2010).

Breitung and Roling (2015) introduced a nonparametric MIDAS model. They argue that the use of, for example, exponential Almon lag or beta lag polynomial can create a misconception of the true lag polynomial of the HF observations. The suggested nonparametric MIDAS model is similar to the step-weighting model in (2.3). For simplicity, the lagged values of the LF variable is not considered here. The difference between the models is that the nonparametric MIDAS model uses a penalized least-squares method to estimate ϖ, the vector of the weights for the HF observations from (2.3). The penalized least squares method utilizes the second differences of ϖ_i which is defined as

$$\bigtriangledown^2 \varpi_i = \varpi_i - 2\varpi_{i-1} + \varpi_{i-2} \qquad \text{for} \quad i = 2, \dots, m-1. \qquad (2.12)$$

The coefficients in ϖ are assumed to be a smooth function of i where the absolute value of $\bigtriangledown^2 \varpi_i$ are small (Breitung and Roling, 2015). Then the optimal estimate of ϖ is the one that minimizes the penalized least square function which is derived as

$$S(\beta_0, \varpi) = \sum_{t=1}^{T} \left(y_{t+1} - \beta_0 - \varpi' x_t^{(m)} \right)^2 + \lambda \sum_{i=2}^{m-1} \left(\bigtriangledown^2 \varpi_i \right)^2, \qquad (2.13)$$

where λ is the given smoothing parameter. Breitung and Roling (2015) suggested using Akaike information criterion (AIC) to find the optimal λ.

It is also important to mention that alternatives to MIDAS model also exists. One of

them is a mixed-frequency vector autoregressive (MF-VAR) model, which in Kuzin et al. (2011) was compared to the MIDAS model. The MF-VAR model assumes that the LF variable has missing observations. For example, if the LF variable is published every quarter, then the potential observations in the months between $t - 1$ and t will be considered missing in the LF variable. Hence, the LF variable can match the higher frequency by disaggregation. Kuzin et al. (2011) chose to describe the disaggregation of the LF variable as shown in Mariano and Murasawa (2003) and Mariano and Murasawa (2010). This equation links y_t and its unobserved monthly observations y_t^* as described below

$$y_t = \frac{1}{3}y_t^* + \frac{2}{3}y_{t-1/3}^* + y_{t-2/3}^* + \frac{2}{3}y_{t-3/3}^* + \frac{1}{3}y_{t-1}^*, \qquad (2.14)$$

where $x_{t-i/3}$ and $y_{t-i/m}^*$ are from a vector autoregressive model with k autoregressive lags (Kuzin et al., 2011). To forecast Kuzin et al. (2011) suggested to use a state space model and the Kalman filter. The disaggregation method in (2.14) was also applied in the factor analysis of the German GDP in Schumacher and Breitung (2008).

This review shows that the MIDAS model is diverse, and that there are many applications that can extend the basic MIDAS model in (2.4). However, there is a need to investigate whether or not a forecasting regression model will benefit from a MIDAS specification. The following chapter will discuss different hypotheses tests that will try to answer this question.

Chapter 3

Specification Tests

This chapter will give the details about existing model specification tests for mixed frequency models. Andreou et al. (2010) gave two test methods for testing if the flat aggregation model in (2.1) is a good fit. One of the tests is a variable addition test. For simplicity, the variable addition test will be denoted AGK for Andreou, Ghysels and Kourtellos. The variable addition approach was also used in Miller (2014b), which will be denoted Miller's test.

Some notations need to be introduced. Let $B_q(\cdot)$ be a (q)-vector of independent Brownian motions. Let $\rightarrow_D$ be convergence in distribution, and $O_p(1)$ denotes bounded in probability. Also note that $\lfloor rT \rfloor$ is the integer part of rT.

3.1   AGK's Specification Tests

The flat aggregation model as given in (2.1), is one simple solution to regress LF variables on HF frequency variable. However, as discussed earlier, the flat aggregation model might not be a good fit for the data. Therefore, it can be interesting to perform a test that can specify whether or not the flat aggregation model should be employed. In Andreou et al. (2010) two hypotheses tests for flat aggregation versus a MIDAS model were given. They developed a testing model which is an extension of the flat aggregation model in (2.1). The model is divided into two parts. One part consists of the flat aggregated HF observations, and the second part contains the MIDAS specified HF observations. The model can can be written as follows

$$y_{t+1} = \boldsymbol{\beta}_1^{*\prime} \boldsymbol{x}_t^{(m),FA} + \boldsymbol{\beta}_1^{*\prime} \boldsymbol{x}_t^{(m),NL}(\boldsymbol{\theta}) + w_{t+1}, \tag{3.1}$$

where $\boldsymbol{x}_t^{(m),FA} = (1, \boldsymbol{x}_t^{(m)})\boldsymbol{\pi}$ for $\boldsymbol{\pi} = (\frac{1}{m}, \ldots, \frac{1}{m})'$, and the second term is $\boldsymbol{x}_t^{(m),NL}(\boldsymbol{\theta}) = (1, \boldsymbol{x}_t^{(m)})\boldsymbol{\pi}(\boldsymbol{\theta}) - \boldsymbol{x}_t^{(m),FA}$. Here FA stands for flat aggregation and NL stands for nonlinear. The vector $\boldsymbol{\pi}(\boldsymbol{\theta})$ contains the weights from the exponential Almon lag polynomial given in (2.5).

The alternative model representation in (3.1) will be equal to the flat aggregation model in (2.1) when $\boldsymbol{\theta} = (0,0)'$. In this case $\boldsymbol{x}_t^{(m),NL}(\boldsymbol{\theta}) = 0$, and (3.1) reduces to the

18

flat aggregation term $\boldsymbol{\beta}_1^{*\prime} x_t^{(m),FA}$ and the error terms w_{t+1}. Hence, the hypotheses are

$$H_0 : \boldsymbol{\theta} = (0,0)', \text{flat aggregation}$$

$$H_1 : \boldsymbol{\theta} \neq (0,0)', \text{MIDAS}.$$

Note that $\boldsymbol{\beta}_1^*$ in (3.1) is still identified under the null, because it is restricted to be the slope coefficient for both $\boldsymbol{x}_t^{(m),FA}$ and $\boldsymbol{x}_t^{(m),NL}(\boldsymbol{\theta})$ (Andreou et al., 2010). It was also mentioned in Andreou et al. (2010) that the least square estimate of the slope coefficient $\hat{\boldsymbol{\beta}}_1^*$ will give an omitted variable bias when the flat aggregation is wrongly fitted. The size of the bias is dependent upon how different the MIDAS specification is from the flat weight scheme (Andreou et al., 2010). This bias will not exist when $E\left(\boldsymbol{x}_t^{(m),FA}\boldsymbol{x}_t^{(m),NL}(\boldsymbol{\theta})'\right) = 0$ or if the flat aggregation model is the better fit (Andreou et al., 2010).

The first test that Andreou et al. (2010) employed was a simple Lagrange multiplier (LM) test. The test looks at the difference between $S_0 = \hat{\varepsilon}_{t+1}'\hat{\varepsilon}_{t+1}$ where $\hat{\varepsilon}_{t+1}$ are the least squares residuals from (2.1), and $S_1 = \hat{w}_{t+1}'\hat{w}_{t+1}$ where $\hat{w}_{t+1}$ are the least squares residuals from (3.1). The LM test is given by

$$LM = \frac{S_0 - S_1}{S_0}, \tag{3.2}$$

where $LM \to_D \chi^2(2)$. Note that the degrees of freedom are dependent upon the

restrictions in (3.1). Since the model in (3.1) has two unknown parameters, namely the parameters in $\boldsymbol{\theta}$, the LM test will have two degrees of freedom.

The second test Andreou et al. (2010) proposed is a Wu-Hausman (WH) test. This test will be referred to as the AGK test. It tests for the omitted variable bias in (2.1), and the null hypothesis is $E(\varepsilon_{t+1} \mid \mathcal{F}_{t-1}^{(m)}) = 0$ where ε_{t+1} is the errors form (2.1) and $\mathcal{F}_{t-1}^{(m)}$ is the information set for the HF variable. The test requires an instrumental set of the HF observations. Andreou et al. (2010) suggested using all the HF observation in the instrumental vector $\boldsymbol{z}_t = \left(x_t, x_{t-1/m}, \ldots, x_{t-(m-1)/m}\right)'$. This vector will be helpful because if the flat aggregation model is false, then w_{t+1} might be correlated to $\boldsymbol{x}_t^{(m),FA}$. The instrumental vector $\boldsymbol{z}_t$, on the other hand, will be correlated to $\boldsymbol{x}_t^{(m),FA}$, but uncorrelated to w_{t+1} in (3.1). The test needs two regression models to be estimated. First, let $\boldsymbol{x}_t^{(m),FA} = \boldsymbol{P}\boldsymbol{z}_t + \boldsymbol{v}_t$, and $E(\boldsymbol{v}_t \mid \boldsymbol{x}_t^{(m),FA}) = 0$. The second model is $y_{t+1} = \beta_1 \boldsymbol{x}_t^{(m),FA} + \varepsilon_{t+1}$. Then the fitted residuals $\hat{\boldsymbol{v}}_t$ and $\hat{\varepsilon}_{t+1}$ are retained, and the test model will be

$$\hat{\varepsilon}_{t+1} = \boldsymbol{\gamma}'\boldsymbol{x}_t^{(m),FA} + \boldsymbol{\delta}'\boldsymbol{v}_t + \xi_{t+1}, \tag{3.3}$$

where $E(\xi_{t+1} \mid \boldsymbol{x}_t^{(m),FA}, \hat{\boldsymbol{v}}_t) = 0$. The null hypothesis is then $\boldsymbol{\delta} = \boldsymbol{0}$, and the test statistic is as follows

$$WH = TR^2, \tag{3.4}$$

where R^2 is the coefficient of determination for (3.3), and $WH \to_D \chi^2(2)$. Andreou

et al. (2010) reported that both the LM and the WH test had good sizes and power, but that the LM test showed better power. The problem with the AGK variable addition test is the size of the instrument set, $\boldsymbol{z}_t$. The use of all the HF observations might be preferred, but when m is large the dimension of the test model in (3.3) will also be large and the test might be difficult to interpret. Andreou et al. (2010) suggested to use a smaller set of the recent HF observations. The issue of large number of instruments has been discussed further in Miller (2014b), and an alternative solution to the issue was also given in Miller (2014b).

3.2 Miller's Specification Test

A variable addition test similar to the AGK test was introduced in Miller (2014b). However, a significant difference between the two methods is the choice of instrument variables. While all the HF observations are used in the AGK method, Miller (2014b) suggested using linear combinations of the HF variable where the number of combinations should be less than m. This practice let the user keep the degrees of freedom constant and uncorrelated to m.

For comparison to Miller's test we will represent the test model similar to the one

used in Miller (2014b). The model is as follows

$$y_{t+1} = \beta \boldsymbol{x}_t^{(m)\prime} \boldsymbol{\pi} + \boldsymbol{q}_t' \boldsymbol{\eta} + \varepsilon_{t+1}, \tag{3.5}$$

where $\boldsymbol{\pi}$ is a vector that contains the weight values for the HF observations, $\boldsymbol{q}_t$ is the nuisance terms, which can also be observed at a higher frequency. The model can be a flat aggregation specification if $\boldsymbol{\pi} = (\frac{1}{m}, \ldots, \frac{1}{m})'$, then (3.5) will be similar to (2.1). Miller (2014b) tested three different hypotheses sets. Test 1 and test 2, denoted τ_1 and τ_2, have the same null hypothesis

$$H_0 : \boldsymbol{\pi} \text{ is known and (3.5) is a flat aggregation model,}$$

while the alternative hypothesis are different as shown below

$$H_{1,\tau_1} : \boldsymbol{\pi} \text{ is a vector with unknown weights. The model in (3.5) is unrestricted}$$

and

$$H_{1,\tau_2} : \boldsymbol{\pi} \text{ is a vector with weights from a lag polynomial that is partially restricted,}$$
$$\text{as in a MIDAS model.}$$

The third test, τ_3, has hypotheses that are combinations of the hypotheses in the other two tests. The null hypothesis is equal to the alternative hypothesis in τ_2, while

the alternative hypothesis is the same as the alternative in τ_1. For the purpose of this thesis, we will focus on τ_2 because it proposes the alternative that a MIDAS model should be employed. Also note that τ_2 has the hypotheses that are similar to the ones in Andreou et al. (2010) which are found in Section 3.1. The implementation for the variable addition test for the three tests is based upon the specification test given in Miller (2014a). However, τ_3 is the only test investigated in that article.

The test consists of two steps. First, the model in (3.5) is estimated under the null hypothesis by the least square method. This means that $\boldsymbol{\pi} = \boldsymbol{\pi}_0 = (\frac{1}{m}, \ldots, \frac{1}{m})'$, which creates the flat aggregation model. Second, the fitted least square residuals $\hat{\varepsilon}_{t+1}$ from (3.5) are saved. These are vital for the variable addition test because they contain the information of whether or not the flat aggregation model is a good fit. The fitted residuals $\hat{\varepsilon}_{t+1}$ from (3.5) should be regressed upon $q + 1$ different linear transformations of the HF variable in this manner

$$\hat{\varepsilon}_{t+1} = \boldsymbol{x}_t^{(m)\prime} \boldsymbol{\Upsilon} \boldsymbol{\varphi} + e_t, \tag{3.6}$$

where $\boldsymbol{\Upsilon}$ is a $m \times (q+1)$ matrix where the first column is $\boldsymbol{\pi}_0$. The q remaining columns are vectors that will linearly transform $\boldsymbol{x}_t^{(m)}$. There are different opportunities for the set up of $\boldsymbol{\Upsilon}$. Miller (2014b) suggested having columns which are related to the weights given by lag polynomials as in (2.4). Another idea is to let $\boldsymbol{\Upsilon}$ be an $m \times m$ identity matrix. This will be similar to the AGK test where all the HF observations are used

as instruments. The vector $\boldsymbol{\varphi}$ contains the coefficients of $\boldsymbol{x}_t^{(m)\prime}\Upsilon$.

The reason for this ancillary regression is to test whether or not the null model is the right fit. If the flat aggregation model (the null model) is a good fit, then there should be no correlation between $\hat{\varepsilon}_{t+1}$ and the q last elements of $\boldsymbol{x}_t^{(m)\prime}\Upsilon$ (the first column is disregarded since it relates to the null model). However, if the null hypothesis is false, then the linear combinations of $\boldsymbol{x}_t^{(m)}$, other than the flat weights, might have an effect on the fitted residuals (Miller, 2014b). In that case, the ancillary regression model could indicate that a MIDAS model for y_{t+1} and $\boldsymbol{x}_t^{(m)}$ might be a better choice. The information for potential correlation between the HF variable and the fitted residuals will be found in the q last slope coefficients in $\boldsymbol{\varphi}$, corresponding to the q last columns of $\boldsymbol{x}_t^{(m)\prime}\Upsilon$. In order to find the coefficient vector that only consists of these q elements, let $\boldsymbol{\varphi}^{(q)} = A\boldsymbol{\varphi}$ where

$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0\ldots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 1 \end{bmatrix}. \tag{3.7}$$

Thus, the test will evaluate the significance of the coefficients in $\boldsymbol{\varphi}^{(q)}$. The new null hypothesis with $\boldsymbol{\varphi}^{(q)}$ is as follows

$$H_0 : \boldsymbol{\varphi}^{(q)} = \boldsymbol{0}, \tag{3.8}$$

with the alternative hypothesis being

$$H_1 : \boldsymbol{\varphi}^{(q)} \neq \mathbf{0}. \tag{3.9}$$

The estimation of $\boldsymbol{\varphi}$ is done by the least-squares method as follows

$$\hat{\boldsymbol{\varphi}}_T = (\Upsilon' M_{xx}^{(T)} \Upsilon)^{-1} \Upsilon' M_{x\varepsilon}^{(T)}, \tag{3.10}$$

where $M_{xx}^{(T)} = T^{-1} \sum_{j=1}^{T} \boldsymbol{x}_j^{(m)} \boldsymbol{x}_j^{(m)\prime}$ and $M_{x\varepsilon}^{(T)} = T^{-1} \sum_{j=1}^{T} \boldsymbol{x}_j^{(m)} \hat{\varepsilon}_{j+1}$. It follows that $\hat{\boldsymbol{\varphi}}_T^{(q)} = A\hat{\boldsymbol{\varphi}}_T$. Miller (2014b) mentioned the possibility of using a Wald test statistic. It is given by

$$V_T = T\hat{\boldsymbol{\varphi}}_T^{(q)\prime} (\hat{\Omega}_T^{(q)})^{-1} \hat{\boldsymbol{\varphi}}_T^{(q)}, \tag{3.11}$$

where $\hat{\Omega}_T^{(q)}$ is a consistent estimator of $\Omega^{(q)}$, the covariance matrix for $\hat{\boldsymbol{\varphi}}_T^{(q)}$. A consistent estimation of $\Omega^{(q)}$ requires a heteroskedasticity and autocorrelation consistent (HAC) estimator (Andrews, 1991). In the context of the the ancillary regression in (3.6), $\hat{\Omega}$ has the following form

$$\hat{\Omega} =$$

$$\left(\frac{1}{T} \sum_{j=1}^{T} \boldsymbol{x}_t^{(m)\prime} \Upsilon \Upsilon' \boldsymbol{x}_t^{(m)} \right)^{-1} \left\{ \frac{T}{T-q} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T} \right) \hat{\Lambda}(j) \right\} \left(\frac{1}{T} \sum_{j=1}^{T} \boldsymbol{x}_t^{(m)\prime} \Upsilon \Upsilon' \boldsymbol{x}_t^{(m)} \right)^{-1}, \tag{3.12}$$

where $k(\cdot)$ is a kernel function, S_T is the bandwidth parameter which has to be chosen by the user, and $\hat{\Lambda}(j)$ is the estimate of the autcovariance of $e_t \Upsilon \boldsymbol{x}_t^{(m)}$ where e_t are the error terms from (3.6). The HAC estimator is very sensitive to the choice of S_T. In Andrews (1991) optimal choices for $k(\cdot)$ and S_T were derived for a limited set of data generating process, but the choice of S_T is known to be difficult in general. Hence, finding a test statistic that is not dependent upon the HAC estimator would be interesting.

Miller (2014b) tested the statistic V_T in (3.11) on $\boldsymbol{x}_t^{(m)}$ as a series integrated by order 0, $I(0)$, and $\boldsymbol{x}_t^{(m)}$ as a series integrated by order 1, $I(1)$. The results indicated that $V_T \to_D \chi_q^q$ when $T \to \infty$ under the null hypothesis when $\boldsymbol{x}_t^{(m)}$ is $I(0)$. However, the test statistic has a complicated limiting distribution under the null when $\boldsymbol{x}_t^{(m)}$ is $I(1)$. Therefore, Miller (2014b) modified the test statistic V_T so that it would give a standard limiting distribution under the null hypothesis for both $I(0)$ and $I(1)$ series of $\boldsymbol{x}_t^{(m)}$.

The modification of the test will be called Miller's test. The main idea is to add more white noise to e_t in (3.6). The fitted residuals, $\hat{\varepsilon}_{t+1}$, from the null model in (3.5) is linearly transformed in this way

$$\hat{\varepsilon}_{t+1}^* = \hat{\varepsilon}_{t+1} + T^\epsilon u_t, \tag{3.13}$$

where $u_t \sim N(0, \sigma_u^2)$ and are independent and identically distributed, and $0 < \epsilon \leq 1/2$. The new series $\hat{\varepsilon}_{t+1}^*$ is regressed on $\boldsymbol{x}_t^{(m)'}\Upsilon$ as in (3.6) and $\hat{\boldsymbol{\varphi}}_T$ is estimated by least-squares. The new test statistic is denoted V_T^*. The reason for adding the error term u_t is to decrease the serial correlation and heteroskedasticity in $\hat{\varepsilon}_{t+1}$. The error term will dominate because it is multiplied by T^ϵ which increases with the increase of ϵ. The test still concerns the q last elements of $\boldsymbol{\varphi}$. However, the modified test statistic V_T^* will transform to a q times an F-statistic for the null hypothesis $H_0 : \boldsymbol{\varphi}^{(q)} = \boldsymbol{0}$ (Miller, 2014b). Hence, a linear regression model under the null hypothesis and a linear regression model under the alternative hypothesis need to be built. Then V_T^* will test for the difference between the sum of squares due to regression for the two models. Miller (2014b) proved that for the tests τ_1 and τ_2, the V_T^* test will have the limiting distribution χ_q^2 for both $\boldsymbol{x}_t^{(m)}$ as $I(0)$ and $I(1)$ under the null hypothesis. The choice of ϵ determines the degree of modification. In Miller (2014b) $\hat{\varepsilon}_{t+1}^*$ was labeled as unmodified when $\epsilon = 0$ and over-modified when $\epsilon = 1/2$. The modification has a positive effect on the size of the test, but causes a loss in power.

Both Andreou et al. (2010) and Miller (2014b) showed that using a variable addition test for specification testing is relatively simple. However, in the unmodified V_T in (3.11) there is a need to calculate a HAC estimator which will not work when $\boldsymbol{x}_t^{(m)}$ is $I(1)$. Miller (2014b) introduced a solution that can work when the HF variable is $I(1)$ without a direct estimation of $\Omega^{(q)}$. However, this method includes a parameter ϵ that needs to be chosen beforehand. As can be seen in the simulations in Miller

(2014b), the choice of ϵ has a considerable impact on the size and power of the test. Currently, there is no guide for finding the optimal value for ϵ. Therefore, finding an alternative way for testing the hypotheses without user-chosen values is of interest.

Chapter 4

Self-Normalizing Specification Test

In this chapter we will introduce our self-normalizing specification test. The set up of test will follow the specification test in Miller (2014b). This includes the same hypotheses and the use of an ancillary regression model.

In recent years a self-normalizing approach for conducting tests and constructing confidence intervals in stationary time series, has been explored. Lobato (2001) first introduced using a normalizing method to test whether or not a dependent process is uncorrelated up to a certain lag. The test statistic was shown to be pivotal under the null hypothesis. Shao (2010) took the idea of self-normalizing further to approximately linear statistics. The advantage of the self-normalizing approach is the

absence of user-chosen values. The core of the method is the use of a normalizing matrix using recursive estimates of the parameter being tested. The nuisance parameter Ω is canceled out by the normalizing matrix, which in turn will serve as an inconsistent estimator for Ω in the test statistic. The self-normalizing approach is similar to the fixed-b approach which was introduced in Kiefer and Vogelsang (2005) based on the work done in Kiefer et al. (2000), Bunzel et al. (2001), Kiefer and Vogelsang (2002a), and Kiefer and Vogelsang (2002b). The fixed-b scheme is a method where the bandwidth parameter in the HAC estimator is set to be bT, where b is a fixed constant. This creates an inconsistent HAC estimator of the variance matrix. Shao (2010) pointed out that the self-normalizing method is similar to the fixed-b method when $b = 1$ and the Bartlett kernel is used.

Since the generalization of the self-normalizing method in Shao (2010), developments of the method to other statistical inference problems for time series have been studied. The details of using self-normalizing method for change point detection was given in Shao and Zhang (2010). In Zhou and Shao (2013) the self-normalizing method was applied to linear regressions with fixed regressors and weakly dependent and stationary time series errors. The extension of the self-normalizing method to nonparametric regression models was shown in Kim et al. (2015). The need for a trimming and a bandwidth parameter is present, but the consistent estimation of the variance is still absent in this self-normalizing approach (Kim et al., 2015). The self-normalizing method has also been extended to functional time series (Shao, 2015). Using the

method for change point detection in the mean function for functional time series was developed in Zhang et al. (2011) as an extension of the univariate series case in Shao and Zhang (2010). These examples reflect the flexibility of the self-normalizing method in time series. For a full review of the self-normalizing method in other statistical inferences, see Shao (2015).

With the encouragement of the success of implementing the self-normalizing method to different types of inference problems in time series, we have examined the extension of the approach to the specification test of mixed frequency models. The main structure of the test is similar to the variable addition test in Miller (2014b). The hypotheses are the same as for the test τ_2, which is mentioned in the beginning of Section 3.2. The model in (3.5) is fitted under the null hypothesis, and the fitted residuals are regressed upon q transformations of $\boldsymbol{x}_t^{(m)}$ stored in the columns of the $m \times q$ matrix Υ. For the self-normalizing method, we need recursive estimates of $\boldsymbol{\varphi}$ from (3.6). The least-square recursive estimates are given below

$$\hat{\boldsymbol{\varphi}}_t = (\Upsilon' M_{xx}^{(t)} \Upsilon)^{-1} \Upsilon' M_{x\varepsilon}^{(t)} \qquad \text{for} \quad t = q, \dots, T, \tag{4.1}$$

where $M_{xx}^{(t)} = t^{-1} \sum_{j=1}^{t} \boldsymbol{x}_j^{(m)} \boldsymbol{x}_j^{(m)\prime}$ and $M_{x\varepsilon}^{(t)} = t^{-1} \sum_{j=1}^{t} \boldsymbol{x}_j^{(m)} \hat{\varepsilon}_{j+1}$. At time T, the estimate of $\hat{\boldsymbol{\varphi}}_T$ is the same as the estimate given in (3.10). The subsampling will start at time q, which is the number of columns in Υ, because that is the lowest value that allows for $\Upsilon' M_{xx}^{(t)} \Upsilon$ in (4.1) to be invertible.

The recursive estimates of $\hat{\varphi}_t$ are the main building blocks of the normalizing matrix. According to Shao (2010), the normalizing matrix can be derived as follows

$$W_T = T^{-2} \sum_{t=q+1}^{T} t^2 (\hat{\varphi}_t - \hat{\varphi}_T)(\hat{\varphi}_t - \hat{\varphi}_T)'. \tag{4.2}$$

This statistic is not dependent upon any unknown parameter like a bandwidth parameter or a trimming value. W_T is an inconsistent estimator of Ω and can replace the HAC estimator in a Wald test statistic. One assumption has to be made before the test statistic can be derived.

Let $\mathcal{A}[0,1]$ be the space of functions on $[0,1]$ which are right continuous, have left limits, and is defined on the Skorokhod metric (Billingsley, 1968). Also note that $\Rightarrow$ denotes weak functional convergence in $\mathcal{A}[0,1]$.

Assumption 4.0.1 *Let Δ be a $q \times q$ lower triangular matrix with nonnegative diagonal values and $r \in (0,1]$, assume*

$$T^{-1/2}\lfloor rT \rfloor (\hat{\varphi}_{\lfloor rT \rfloor} - \varphi) \Rightarrow \Delta B_q(r). \tag{4.3}$$

According to Assumption 4.0.1, $T^{1/2}(\hat{\varphi}_T - \varphi) \to_D \Delta B_q(1)$ when $r = 1$. Further, if we let

$$V_q = \int_0^1 (B_q(r) - rB_q(1))(B_q(r) - rB_q(1))'dr, \tag{4.4}$$

then we can show by Assumption 4.0.1 and the continuous mapping theorem that

$W_T \to_D \Delta V_q \Delta'$.

The self-normalized test statistic can now be found, and be used to test the null hypothesis where $\boldsymbol{\phi} = \mathbf{0}$. The test statistic is similar to the Wald statistic V_T in (3.11), except $\hat{\Omega}_T$ is replaced by W_T

$$K_T = T\hat{\boldsymbol{\varphi}}'_T W_T^{-1} \hat{\boldsymbol{\varphi}}_T. \tag{4.5}$$

This test statistic can be used to test $H_0 : \boldsymbol{\varphi}_T = \mathbf{0}$. The knowledge of the convergence of the parts in K_T can be used to show that under the null hypothesis

$$K_T \to_D U_q = B_q(1)' V_q^{-1} B_q(1). \tag{4.6}$$

The distribution U_q is not standard and was first introduced in Lobato (2001), where a table with the upper critical values for up to 20 degrees of freedom can be found. This amount puts a restriction to the dimension of Υ. Thus, using all the HF observations as instruments, as in the AGK test, is not possible if $m > 20$ without extending Lobato's table. The distribution U_q does not include any unknown parameters, so K_T is asymptotically pivotal.

As discussed in Section 3.2, Υ can have $q + 1$ columns where the first one is $\boldsymbol{\pi}_0$ and the remaining q are other linear combinations. The null hypothesis will then be to

test if $\boldsymbol{\varphi}_T^{(q)} = \boldsymbol{0}$ where $\boldsymbol{\varphi}^{(q)} = A\boldsymbol{\varphi}$ and A was defined in (3.7). This scenario is still functional under the self-normalizing method. The test statistic will be as follows

$$
\begin{aligned}
K_T &= T\hat{\boldsymbol{\varphi}}_T' A'W_T^{-1}A\hat{\boldsymbol{\varphi}}_T \\
&= T\hat{\boldsymbol{\varphi}}_T^{(q)'}W_T^{-1}\hat{\boldsymbol{\varphi}}_T^{(q)}.
\end{aligned}
\tag{4.7}
$$

The limiting distribution is still U_q because

$$
\begin{aligned}
U_q &= B_{q+1}(1)'A'(AV_{q+1}A')^{-1}AB_{q+1}(1) \\
&= B_q(1)'V_q^{-1}B_q(1).
\end{aligned}
\tag{4.8}
$$

This equality is not trivial, but it can be proved using similar argument in the proof of Theorem 1 in Kiefer et al. (2000).

The following theorem proves the limiting distributions for K_T.

Theorem 1 *Let K_T be the test statistic for τ_2, and under Assumption 4.0.1 we have that*

1. *Under H_0, $K_T \rightarrow_D U_q$*

2. *Under H_1, $K_T = O_p(T)$*

when $T \rightarrow \infty$

Proof: Under H_1, the test statistic will be

$$T\hat{\varphi}'_T W_T^{-1}\hat{\varphi}_T = T(\hat{\varphi}_T - \varphi)'W_T^{-1}(\hat{\varphi}_T - \varphi)$$

$$+ 2T\varphi'W_T^{-1}(\hat{\varphi}_T - \varphi) \qquad (4.9)$$

$$+ T\varphi'W_T^{-1}\varphi$$

Next step is to find how the three elements converges when $T \to \infty$. First,

$$T(\hat{\varphi}_T - \varphi)'W_T^{-1}(\hat{\varphi}_T - \varphi) \to_D U_q = O_p(1)$$

Second, let us divide the second term in (4.9) by $\sqrt{T}$ and derive the convergence as follows

$$\left(\frac{1}{\sqrt{T}}\right) 2T\varphi'W_T^{-1}(\hat{\varphi}_T - \varphi) = 2\varphi'\sqrt{T}W_T^{-1}(\hat{\varphi}_T - \varphi) \to_D 2\varphi'(V_q\Delta')^{-1}B_q(1) = O_p(1)$$

Thus, $2T\varphi'W_T^{-1}(\hat{\varphi}_T - \varphi) = O_p(\sqrt{T})$. Third, let us divide the third term in (4.9) by T, and the derive the convergence as follows

$$\left(\frac{1}{T}\right) T\varphi'W_T^{-1}\varphi \to_D \varphi'(\Delta V_q\Delta')^{-1}\varphi = O_p(1).$$

Thus, $T\varphi'W_T^{-1}\varphi = O_p(T)$. Therefore, is $K_T = O_p(T)$ under H_1. $\square$

The theorem gives some encouraging results in terms of asymptotic size and power of our method. The limiting distribution of our test statistic under the null hypothesis

35

is the same as the one in Lobato (2001). Thus, the critical values are available for testing, and Theorem 1(1) ensures that our test has asymptotically correct size. On the other hand, if a MIDAS model fits the data better than the flat aggregation, it would be easier to reject the null hypothesis because our test statistic, K_T, would have the same order as the sample size T. This implies our test has some power to reject the null hypothesis under the alternative hypothesis.

Chapter 5

Size and Power Analyses

In this chapter the results from size and power analyses are given to compare the self-normalizing test K_T in (4.5) to the AGK test, versions of Miller's test V_T^*, and the specification test V_T in (3.11) with a HAC estimator. The set up of the analyses will closely follow the small-sample analysis in Miller (2014b).

The data generating process (DGP) for the power analyses is inspired by the one given in Miller (2014b). The data is generated at the higher frequency, and is as follows

$$y_{t+1-i/m} = \beta x_{t-i/m} + \varepsilon_{t+1-i/m}, \tag{5.1}$$

where $\beta = 10$,

$$
\begin{pmatrix} \varepsilon_{t+1-i/m} \\ x_{t-i/m} \end{pmatrix} = B \begin{pmatrix} \varepsilon_{t+1-(i+1)/m} \\ x_{t-(i+1)/m} \end{pmatrix} + u_{t+1-i/m},
$$

for which $u_{t+1-i/m} \sim N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$ and independent and identically distributed, and $B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}$ where $(b_1, b_2) = \{(0,0), (0,0.3), (0.3,0.3), (0.3,-0.3), (0,0.5), (0.5,0.5), (0,0.8), (0.8,0.8)\}$ will make the vector with $\varepsilon_{t+1-i/m}$ and $x_{t-i/m}$, a vector autoregressive variable. Since $y_{t+1-i/m}$ is generated at a higher frequency, it needs to be aggregated to match the lower frequency. This can be done by $\sum_{i=0}^{m-1} y_{t-i/m}/m$ for all $t = 1, \ldots, T$, and for these analyses $T = 200$. The high frequencies considered in these analyses are $m = \{4, 20, 150\}$. Here $m = 4$ can represent the four quarters in a year, $m = 20$ can be the trading days in a month, while $m = 150$ is used to test when m is relatively large (Miller, 2014b).

The size and power of the tests are determined by lag functions in $\boldsymbol{\pi}$ in (3.5). The lag function is given by

$$
\pi(\theta) = \frac{(2 - s/m)^{4\theta}}{\sum_{s=1}^{m}(2 - s/m)^{4\theta}}, \tag{5.2}
$$

for $s = 1, \ldots, m$ and where $\theta = \{0, 0.1\}$ which is the only parameter that is changing. The size of the tests will be found when $\theta = 0$, then $\pi(\theta) = \frac{1}{m}$ in 5.2 and $\boldsymbol{\pi}_0 =$

38

$(\frac{1}{m}, \ldots, \frac{1}{m})$, and (3.5) is equal to the model under the null. When $\theta = 0.1$, the data will be generated as under the alternative and the power functions of the tests will be found. The cases where $\theta = 0.5, 1, 1.5, 2$ have also been simulated. However, the tests generate power close to 1.0 for these values of θ, and we choose not to display them here. In Figure 5.1 the distribution of the weights from the lag function $\pi(\theta)$ for select values of θ can be seen. The difference between the lag functions at $\theta = 0$ and $\theta = 0.1$ is not large. Therefore, discovering the power of the different tests at $\theta = 0.1$ will be interesting because it will tell us which test that can distinguish the minor change from a flat aggregation model to a model with decreasing weights.

Miller (2014b) suggested to set $q = 2$. Thus, the dimensions of Υ are $m \times (2 + 1)$. The first column is the vector $\boldsymbol{\pi}_0$, the flat aggregation vector. The two last columns of Υ carry vectors of linear transformations for $\boldsymbol{x}_t^{(m)}$, and following Miller (2014b) the columns are $0.9^{s-1}/\sum_{s=1}^{m} 0.9^{s-1}$ and $2(m+1-s)/(m(m+1))$, respectively.

The number of Monte Carlo simulations is set to 1000 and 5% significance level is used. The Υ matrix given in Miller (2014b) and Υ as an $m \times m$ identity matrix are used in the self-normalizing specification test. They are denoted Υ_1 and Υ_2 respectively. Note that Υ_2 cannot be used for $m > 20$ as there is no available critical values for U_q where $q > 20$. For comparison, the simulations are also done on three modifications of Miller's tests where $\epsilon = \{0, 1/5, 1/2\}$ and $\sigma_u^2 = 1$. The AGK test is also tested and is set up as in Miller (2014b). All of the HF observations are used as instruments,
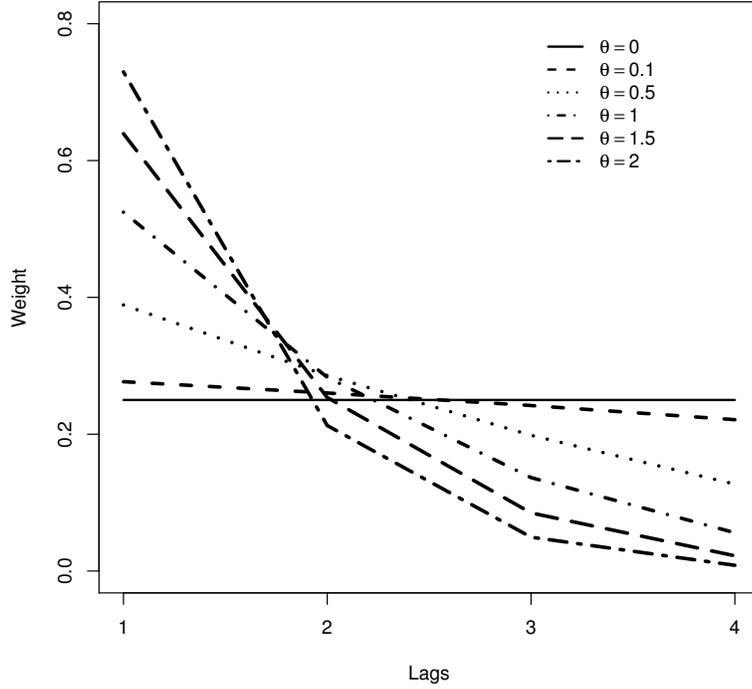
Figure 5.1: The weights given to each lag by $\pi(\theta) = \frac{(2-s/m)^{4\theta}}{\sum_{s=1}^{m}(2-s/m)^{4\theta}}$ for $m = 4$ and different options for θ

when an Υ with dimension $m \times m$, where the first column is $\boldsymbol{\pi}_0$ and the remaining columns are equal to the $m - 1$ last columns of an $m \times m$ identity matrix, is used. Further, the AGK test is implemented as a V_T^* test with $\epsilon = 0$ and the degrees of freedom $q = m - 1$. Miller (2014b) described V_T^* as an unmodified test. In addition, the specification test V_T in (3.11) using a HAC estimator for $\Omega^{(2)}$ is also computed. Both Miller's tests and the test with the HAC estimator employ Υ_1.

In the following tables "Miller" stands for Miller's test, "HAC" stands for the V_T test in (3.11), "AGK" stands for AGK test using the V_T^* with $\epsilon = 0$, and "SN" stands for

self-normalizing test K_T in (4.5).

In Table 5.1 the DGP gives no autoregressive lags for either $\varepsilon_{t+1-i/m}$ or $x_{t-i/m}$. The results show that for $m = 4$, all the tests regardless of the Υ are showing good size. When $m = 20$, the self-normalizing test with Υ_2 has an unacceptable size of 0.260. This means that already at $m = 20$, there is an overparamterization. The AGK test is doing better. However, AGK's test has large size when $m = 150$, while the self-normalizing test is doing well along side Miller's tests. The power of the tests are good, but for Miller's test with $\epsilon = 0.5$ the power is small especially for $m = 20$ and $m = 150$.

In Table 5.2, Table 5.5 and Table 5.7, the results for when only $x_{t-i/m}$ is correlated with its first lag are shown. All the tests display good size except for the self-normalizing method with Υ_2 when $m = 20$ and the AGK test when $m = 150$. The tests also have good power except Miller's test with $\epsilon = 0.5$ for $m = 20$ and $m = 150$

It is interesting to see in Table 5.3 when $m = 4$, the overmodification of Miller's test is needed for having an acceptable size. The self-normalizing method is only showing good size when Υ_2 is used. While in Table 5.4, the results show that when $m = 4$, both self-normalizing tests have good sizes, but Miller's test still needs overmodification to find a size close to 0.05. For this data generating process the self-normalizing process has an advantage since it is not dependent upon a user-chosen parameter compared to

Miller's test. This modification also has a negative effect on the power. At $m = 150$, Miller's test for $\epsilon = 0.5$ has a very small power compared to the other tests.

When the DGP uses the autoregressive coefficients $(0.5, 0.5)$, the results change based on m (Table 5.6). For $m = 4$, all the tests have unacceptable high sizes. However, when $m = 20$ and $m = 150$, Miller's test and the self-normalizing test with Υ_1 have good sizes.

A similar pattern can be seen for when the autoregressive coefficients are $(0.8, 0.8)$. These results can be found in Table 5.8. When $m = 4$ and $m = 20$, all the tests have sizes larger than 0.05, but when $m = 150$ all the tests preform well except the AGK test.

Overall, the results indicate that the performance of the different tests vary based on which version of the DGP is used. Clearly, there are multiple ways of generating data, which in turn has its effect on the result of a test. The main conclusion that we wish to draw is that the self-normalizing approach display similar sizes as Miller's test. The power of the self-normalizing test is overall good except for the case when $m = 4$ and $(b_1, b_2) = (0.8, 0.8)$. Therefore, may the self-normalizing test serve as a competitor for the specification testing of mixed frequency models.

Table 5.1

The size and power for the three options of higher frequency where the
diagonal entries of B are $(b_1, b_2) = (0, 0)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.052	1.000
	Miller, $\epsilon = 0.2$	0.067	1.000
	Miller, $\epsilon = 0.5$	0.049	0.943
	HAC	0.068	1.000
	AGK	0.051	1.000
	SN, Υ_1	0.041	0.996
	SN, Υ_2	0.031	1.000
$m = 20$	Miller, $\epsilon = 0$	0.044	1.000
	Miller, $\epsilon = 0.2$	0.050	1.000
	Miller, $\epsilon = 0.5$	0.049	0.350
	HAC	0.062	1.000
	AGK	0.053	1.000
	SN, Υ_1	0.058	0.996
	SN, Υ_2	0.260	1.000
$m = 150$	Miller, $\epsilon = 0$	0.061	1.000
	Miller, $\epsilon = 0.2$	0.045	0.760
	Miller, $\epsilon = 0.5$	0.042	0.101
	HAC	0.070	1.000
	AGK	0.242	0.693
	SN, Υ_1	0.043	0.999

Table 5.2

The size and power for the three options of higher frequency where the
diagonal entries of B are $(b_1, b_2) = (0, 0.3)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.056	1.000
	Miller, $\epsilon = 0.2$	0.063	1.000
	Miller, $\epsilon = 0.5$	0.047	0.979
	HAC	0.068	1.000
	AGK	0.057	1.000
	SN, Υ_1	0.050	1.000
	SN, Υ_2	0.031	0.998
$m = 20$	Miller, $\epsilon = 0$	0.035	1.000
	Miller, $\epsilon = 0.2$	0.058	1.000
	Miller, $\epsilon = 0.5$	0.050	0.584
	HAC	0.054	1.000
	AGK	0.056	1.000
	SN, Υ_1	0.051	1.000
	SN, Υ_2	0.260	1.000
$m = 150$	Miller, $\epsilon = 0$	0.060	1.000
	Miller, $\epsilon = 0.2$	0.049	0.967
	Miller, $\epsilon = 0.5$	0.040	0.137
	HAC	0.066	1.000
	AGK	0.243	0.940
	SN, Υ_1	0.042	0.999

Table 5.3

The size and power for the three options of higher frequency where the
diagonal entries of B are $(b_1, b_2) = (0.3, 0.3)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.125	1.000
	Miller, $\epsilon = 0.2$	0.118	1.000
	Miller, $\epsilon = 0.5$	0.059	0.918
	HAC	0.141	1.000
	AGK	0.121	1.000
	SN, Υ_1	0.103	0.994
	SN, Υ_2	0.073	0.990
$m = 20$	Miller, $\epsilon = 0$	0.053	1.000
	Miller, $\epsilon = 0.2$	0.057	1.000
	Miller, $\epsilon = 0.5$	0.052	0.540
	HAC	0.064	1.000
	AGK	0.064	1.000
	SN, Υ_1	0.055	0.997
	SN, Υ_2	0.264	1.000
$m = 150$	Miller, $\epsilon = 0$	0.057	1.000
	Miller, $\epsilon = 0.2$	0.046	0.961
	Miller, $\epsilon = 0.5$	0.039	0.133
	HAC	0.068	1.000
	AGK	0.247	0.832
	SN, Υ_1	0.040	0.998

Table 5.4

The size and power for the three options of higher frequency where the
diagonal entries of B are $(b_1, b_2) = (0.3, -0.3)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.092	1.000
	Miller, $\epsilon = 0.2$	0.092	1.000
	Miller, $\epsilon = 0.5$	0.064	0.825
	HAC	0.120	1.000
	AGK	0.101	1.000
	SN, Υ_1	0.071	0.971
	SN, Υ_2	0.064	0.965
$m = 20$	Miller, $\epsilon = 0$	0.059	1.000
	Miller, $\epsilon = 0.2$	0.055	0.989
	Miller, $\epsilon = 0.5$	0.056	0.214
	HAC	0.079	1.000
	AGK	0.059	0.970
	SN, Υ_1	0.059	0.949
	SN, Υ_2	0.257	0.960
$m = 150$	Miller, $\epsilon = 0$	0.055	0.981
	Miller, $\epsilon = 0.2$	0.047	0.502
	Miller, $\epsilon = 0.5$	0.048	0.083
	HAC	0.072	1.000
	AGK	0.247	0.427
	SN, Υ_1	0.046	0.936

Table 5.5

The size and power for the three options of higher frequency where the
diagonal entries of B are $(b_1, b_2) = (0, 0.5)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.057	1.000
	Miller, $\epsilon = 0.2$	0.056	1.000
	Miller, $\epsilon = 0.5$	0.046	0.987
	HAC	0.063	1.000
	AGK	0.053	1.000
	SN, Υ_1	0.049	1.000
	SN, Υ_2	0.036	0.999
$m = 20$	Miller, $\epsilon = 0$	0.041	1.000
	Miller, $\epsilon = 0.2$	0.057	1.000
	Miller, $\epsilon = 0.5$	0.047	0.825
	HAC	0.056	1.000
	AGK	0.052	1.000
	SN, Υ_1	0.048	1.000
	SN, Υ_2	0.258	1.000
$m = 150$	Miller, $\epsilon = 0$	0.056	1.000
	Miller, $\epsilon = 0.2$	0.046	0.999
	Miller, $\epsilon = 0.5$	0.040	0.195
	HAC	0.060	1.000
	AGK	0.246	1.000
	SN, Υ_1	0.034	0.999

Table 5.6

The size and power for the three options of higher frequency where the diagonal entries of B are $(b_1, b_2) = (0.5, 0.5)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.285	1.000
	Miller, $\epsilon = 0.2$	0.276	1.000
	Miller, $\epsilon = 0.5$	0.097	0.829
	HAC	0.323	1.000
	AGK	0.266	1.000
	SN, Υ_1	0.195	0.959
	SN, Υ_2	0.148	0.941
$m = 20$	Miller, $\epsilon = 0$	0.079	1.000
	Miller, $\epsilon = 0.2$	0.072	1.000
	Miller, $\epsilon = 0.5$	0.056	0.765
	HAC	0.086	1.000
	AGK	0.069	1.000
	SN, Υ_1	0.058	0.999
	SN, Υ_2	0.282	1.000
$m = 150$	Miller, $\epsilon = 0$	0.058	1.000
	Miller, $\epsilon = 0.2$	0.042	0.998
	Miller, $\epsilon = 0.5$	0.040	0.193
	HAC	0.065	1.000
	AGK	0.240	0.902
	SN, Υ_1	0.033	0.997

Table 5.7

The size and power for the three options of higher frequency where the
diagonal entries of B are $(b_1, b_2) = (0, 0.8)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.050	1.000
	Miller, $\epsilon = 0.2$	0.053	1.000
	Miller, $\epsilon = 0.5$	0.045	0.995
	HAC	0.073	1.000
	AGK	0.060	1.000
	SN, Υ_1	0.045	1.000
	SN, Υ_2	0.032	0.998
$m = 20$	Miller, $\epsilon = 0$	0.044	1.000
	Miller, $\epsilon = 0.2$	0.052	1.000
	Miller, $\epsilon = 0.5$	0.045	1.000
	HAC	0.057	1.000
	AGK	0.049	1.000
	SN, Υ_1	0.052	1.000
	SN, Υ_2	0.255	1.000
$m = 150$	Miller, $\epsilon = 0$	0.050	1.000
	Miller, $\epsilon = 0.2$	0.043	1.000
	Miller, $\epsilon = 0.5$	0.045	0.759
	HAC	0.063	1.000
	AGK	0.239	1.000
	SN, Υ_1	0.046	1.000

Table 5.8

The size and power for the three options of higher frequency where the diagonal entries of B are $(b_1, b_2) = (0.8, 0.8)$

	Test	$\theta = 0$	$\theta = 0.1$
$m = 4$	Miller, $\epsilon = 0$	0.651	0.763
	Miller, $\epsilon = 0.2$	0.645	0.753
	Miller, $\epsilon = 0.5$	0.360	0.442
	HAC	0.737	0.841
	AGK	0.611	0.711
	SN, Υ_1	0.486	0.594
	SN, Υ_2	0.385	0.474
$m = 20$	Miller, $\epsilon = 0$	0.301	1.000
	Miller, $\epsilon = 0.2$	0.286	1.000
	Miller, $\epsilon = 0.5$	0.123	0.970
	HAC	0.323	1.000
	AGK	0.197	0.998
	SN, Υ_1	0.205	0.979
	SN, Υ_2	0.407	0.983
$m = 150$	Miller, $\epsilon = 0$	0.056	1.000
	Miller, $\epsilon = 0.2$	0.043	1.000
	Miller, $\epsilon = 0.5$	0.038	0.722
	HAC	0.063	1.000
	AGK	0.243	0.954
	SN, Υ_1	0.050	0.999

Chapter 6

Conclusion

The variable addition test as a specification method for determining the best type of model for mixed frequency data, has been shown to be a good approach in Andreou et al. (2010) and Miller (2014b). The modification of the test introduced in Miller (2014b) is easy to compute, but the test still requests the user to choose a value of the parameter that determines the size of the modification. Therefore, we wish to find a method that is without any user-chosen values, but is still a powerful test. The self-normalizing approach is a simple and parameter free method. In this thesis, we have successfully extended the self-normalizing approach to the specification test of mixed frequency models. By following the set up of the test model in Miller (2014b), we have developed a test statistic that utilizes a normalizing matrix. There is no longer a necessity to estimate the variance matrix, since the normalizing matrix will

be an inconsistent estimator. The need for user-chosen values is not present. The test statistic is shown to follow a distribution that was introduced in Lobato (2001). This paper provides a table with critical values that can be used for the hypotheses testing. The nature of the self-normalizing test statistic makes it asymptotically pivotal under the null. The power analysis gave the promising results that in the majority of the DGPs used in this thesis, the self-normalizing specification test has good size and is powerful when a reduced set of linear combinations of the HF variable is tested. This is similar to the results of Miller's test. Thus, we propose the self-normalizing approach to the specification testing as a potential competitor to Miller's test.

This thesis focused on stationary regressors. Therefore, we suggest that further research can be done on DGPs with nonstationary regressors. In addition, conducting the self-normalizing specification test on an empirical data set could also be of interest.

References

Andreou, E., E. Ghysels, and A. Kourtellos (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics 158*(2), 246–261.

Andreou, E., E. Ghysels, and A. Kourtellos (2011). Forecasting with mixed-frequency data. In M. Clements and D. Hendry (Eds.), *Oxford handbook of economic forecasting*, pp. 225–245. Oxford: Oxford University Press.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society 59*(3), 817–858.

Armesto, M. T., K. M. Engemann, and M. T. Owyang (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review 92*(6), 521–536.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley.

Breitung, J. and C. Roling (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach. *Journal of Forecasting 34*(7), 588–603.

Bunzel, H., N. M. Kiefer, and T. J. Vogelsang (2001). Simple robust testing of hypotheses in nonlinear models. *Journal of the American Statistical Association 96*(455), 1088–1096.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics 26*(4), 546–554.

Clements, M. P. and A. B. Galvão (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics 24*(7), 1187–1206.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS touch: Mixed data sampling regression models. *Finance*.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. *Journal of Financial Economics 76*(3), 509 – 548.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131*(1), 59–95.

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

Kiefer, N. M. and T. J. Vogelsang (2002a). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica 70*(5), 2093–2095.

Kiefer, N. M. and T. J. Vogelsang (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory 18*(06), 1350–1366.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Kim, S., Z. Zhao, and X. Shao (2015). Nonparametric functional central limit theorem for time series regression with application to self-normalized confidence interval. *Journal of multivariate analysis 133*, 277–290.

Kuzin, V., M. Marcellino, and C. Schumacher (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting 27*(2), 529–542.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*(455), 1066–1076.

Marcellino, M. and C. Schumacher (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP*. *Oxford Bulletin of Economics and Statistics 72*(4), 518–550.

Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*(4), 427–443.

Mariano, R. S. and Y. Murasawa (2010). A coincident index, common factors, and monthly real GDP*. *Oxford Bulletin of Economics and Statistics 72*(1), 27–46.

Miller, J. I. (2014a). Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics 12*(3), 584–614.

Miller, J. I. (2014b). Simple robust tests for the specification of high-frequency predictors of a low-frequency series. Working paper, University of Missouri.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Schumacher, C. and J. Breitung (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*(3), 386–398.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 343–366.

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association 110*(512), 1797–1817.

Shao, X. and X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association 105*(491), 1228–1240.

Zhang, X., X. Shao, K. Hayhoe, and D. J. Wuebbles (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Statist. 5*, 1765–1796.

Zhou, Z. and X. Shao (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(2), 323–343.