2019

# STATISTICAL METHODS FOR MIXED FREQUENCY DATA SAMPLING MODELS

Yun Liu
*Michigan Technological University*, yliu26@mtu.edu

# STATISTICAL METHODS FOR MIXED FREQUENCY DATA SAMPLING MODELS

By

Yun Liu

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Statistics

MICHIGAN TECHNOLOGICAL UNIVERSITY

2019

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Statistics.

Department of Mathematical Sciences

Dissertation Advisor: *Dr. Yeonwoo Rho*

Committee Member: *Dr. Hie Joo Ahn*

Committee Member: *Dr. Chee-Wooi Ten*

Committee Member: *Dr. Kui Zhang*

Department Chair: *Dr. Mark S. Gockenbach*

# Contents

# List of Figures

# List of Tables

# Preface

This dissertation is a result of my research in pursuing a Ph.D. degree in Statistics at Michigan Tech University. It includes previously published or ongoing papers in Chapters 2-3.

Chapter 2 contains one paper published on Communications in Statistics-Theory and Methods with Dr. Yeonwoo Rho. To test if the flat averaging is acceptable and enough, we constructed a specification test for the flat averaging against the mixed data sampling model. We illustrated a Durbin-Wu-Hausman type specification test constructed upon a two stage least squares estimation and focused on the choice of instrumental variables in the Durbin-Wu-Hausman type test. Details of the choice of instruments are presented to demonstrate its theoretical consistency when the frequency ratio is large enough.

Chapter 3 contains an ongoing paper with Dr. Yeonwoo Rho and Dr. Hie Joo Ahn. I made contributions on the theoretical proofs and simulations with Dr. Yeonwoo Rho, and incorporate with Dr. Hie Joo Ahn and Dr. Yeonwoo Rho on the application analysis. To reduce the complexity of the nonparametric mixed data sampling models, we introduced a nonparametric model with Fourier series expansion in the first part. Monte Carlo simulations are included to show the performance of the nonparametric models. Encouraged by the excellent performance with a single subject, we extended the nonparametric model with Fourier approximation by introducing a clustering approach with panel mixed sampling data. Simulation results, as well as a practical application, are provided to show clustering performance of estimated weights using Fourier approximation in practice.

# Acknowledgments

I would like to thank many people. This thesis would not be done without their encouragement and support.

First and foremost, my sincere acknowledgment should be given to my advisor, Prof. Dr. Yeonwoo Rho, for her excellent guidance and advice during the past five years. I am grateful for her efforts to encourage me for independent and critical thinking and focus on publicizing my work. She offered much help to make connection with other researchers. Apart from the academic guidance, she has been accommodating in improving my professional written skills more comprehensively and gave me valuable advice on my working experience and my future career.

I would like to give my sincere gratitude to Dr. Hie Joo Ahn for her kind support. She offered a lot of help and advice on empirical research. I am grateful to Dr. J. Isaac Miller and other researchers for constructive comments and suggestions.

I wish to acknowledge my colleagues and faculties in my department. With their warmhearted help, I could concentrate and spare no effort into my research.

Last but not least, I thank my parents, Mr. Xiaodong Liu and Mrs. Anmei Shao, for their everlasting support and love. Also, I wish to thank my boyfriend, Li Liu, for his encouragement and care. Without their trust, I could not make such significant progress. I would acknowledge my best friend, Ya Luo, for the memorable experience that we have.

Part of Chapter 2 is reprinted with permission from "On the choice of instruments in mixed frequency specification tests" by Yun Liu and Yeonwoo Rho, 2018. Communications in Statistics-Theory and Methods, page 1-16. 2018 by Taylor  Francis [53].

My research was supported in part by a research grant from the NSF grant CPS-1739422.

**Superior**, a high-performance computing infrastructure at Michigan Technological University, was used in obtaining the Monte Carlo simulation results presented in this publication.

# List of Abbreviations

| | |
|---|---|
| MIDAS | mixed data sampling |
| ADL | autoregressive distributed lag |
| DWH | Durbin-Wu-Hausman |
| VAT | variable addition test |
| (N)LS | (nonlinear) least squares |
| OLS | ordinary least squares |
| U-MIDAS | unrestricted-MIDAS |
| HF | high-frequency |
| LF | low-frequency |
| IV | instrumental variable |
| HAC | heteroscedasticity and autocorrelation consistent |
| AR | autoregressive |
| (R)MSE | root mean squared error |
| KPSS | Kwiatkowski-Phillips-Schmidt-Shin [49] |
| GLS | generalized least squares |
| MCP | minimax concave penalty [22] |
| SCAD | smoothly clipped absolute deviation [73] |
| GMM | generalized method of moments |

| | |
|---|---|
| GDP | gross domestic product |
| C-Lasso | classifier-Lasso [67] |
| IC | information criteria |
| AIC | Akaike information criterion |
| BIC | Bayesian information criterion |
| EM | expectation-maximization |
| i.i.d. | identically and independently distributed |
| SLS | smoothed least squares |
| ADMM | alternating direction method of multipliers |
| PPL | penalized profile likelihood |
| QLME | quasi-maximum likelihood estimation |
| TP | true positive |
| FP | false positive |
| TN | true negative |
| FN | false negative |
| ARI | adjusted Rand index |

# Abstract

The MIDAS models are developed to handle different sampling frequencies in one regression model, preserving information in the higher sampling frequency. Time averaging has been the traditional parametric approach to handle mixed sampling frequencies. However, it ignores information potentially embedded in high frequency. MIDAS regression models provide a concise way to utilize additional information in HF variables. While a parametric MIDAS model provides a parsimonious way to summarize information in HF data, nonparametric models would maintain more flexibility at the expense of the computational complexity. Moreover, one parametric form may not necessarily be appropriate for all cross-sectional subjects. This thesis proposes two new methods designed for mixed frequency data.

First part of this thesis proposes a specification test to choose between time averaging and MIDAS models. If time averaging is enough for given mixed frequency data, there is no need to use complicated nonlinear mixed frequency models. In such case, a specification test that justifies the use of the the simplest model, time averaging, is useful. We propose a specification test revising from a DWH type test. In particular, a set of instrumental variables is proposed and theoretically validated when the frequency ratio is large. As a result, our method tends to be more powerful than existing methods, as reconfirmed through the simulations.

The second part of the thesis provides a new way to identify groups in a panel data setting involving mixed frequencies. A flexible MIDAS model is proposed using a nonparametric approach. This nonparametric MIDAS model is further extended to a panel setting using a penalized regression idea. The estimated parameters can then be clustered using traditional clustering methods. The proposed clustering algorithm delivers reasonable clustering results both in theory and in simulations, without requiring prior knowledge about the true group membership information. An empirical application is presented to examine the panel MIDAS model.

# Chapter 1

# Introduction

Historically, time series data have been studied in numerous fields. Starting from the 1920s, the theoretical development of time series analysis started with stochastic processes. Later on, time series analysis has been developed rapidly. Various types of times series models have been developed and researched for different purposes theoretically and empirically. One of the primary purposes of time series analysis is to forecast future values of the interested response. Even though time series models can mimic the shape of series for forecasting, among different time-dependent variables, it is also important to measure their relationships, such as regression models. Nevertheless, in classic regression models, it is common that all records are supposed to to get a consistent sample size for all variables. While considering time-dependent variables, the observing periods may be inconsistent. Then, how about the regression models involving one or more time-dependent series as regressors comparing to the single-variable regression model? Even more, what if the response is a variable measured at a lower sampling frequency, but the regressors are measured at higher frequencies? This thesis aims to answer questions related to mixed frequencies.

## 1.1 Mixed Sampling Problem and Conventional Approaches

In recent years, datasets that involve different sampling frequencies have drawn substantial attention in various fields. In particular, sampling with different frequencies often arises in economic data. For instance, GDP is one of the most critical indicators of a country's economic status. However, due to the complexity involved in the measurement of GDP, it is only measured four times per year in the United States. On the other hand, many other variables, such as weekly initial claims, daily stock returns, etc., are available while one is waiting for the next release of GDP. In such case, potentially additional information in the more frequently observed (high frequency, HF) variables can be utilized in predicting the less frequently observed (low frequency, LF) variables such as GDP.

Several methods were introduced to handle mixed-frequency variables in the same regression model. These methods often transform the variables with higher observation frequency, matching the lowest frequency in the regression model and making the frequencies of all variables consistent. One conventional approach is time averaging of HF variables, where HF variables are aggregated using a predetermined fixed-weight function. This approach has different names given the weight vector in time averaging. For example, if the weight vector is chosen to be zero except for the end of the period, it is called the end-of-period data sampling. Another example is the flat aggregation, which uses flat weights to average HF records.

Although predetermined and fixed time averaging provides a simple solution to the mixed frequency issue, it may ignore some useful information in the HF variable if the predetermined weights are not properly chosen. On the contrary, the ADL model uses all HF variables as regressors. This approach requires estimating all the coefficients of regressors. Since the weights are determined by the data, allowing to retain most information in the HF variables. However, this model may not be optimal in forecasting, since the estimated weights may follow too closely to the data. Besides, when the frequency ratio between the HF and LF variables is large, the ADL model may require the estimation of too many parameters.

## 1.2 Innovative Approaches for Mixed-Frequency Data

The MIDAS regression model [32] was proposed to balance the complexity and flexibility of the time averaging and the ADL model. In MIDAS models, the weight function is written as a nonlinear parametric function with only a few parameters. The elements in its weight function do not move as freely as the ones in the ADL model due to the parametric restrictions. They are still more flexible than those in time averaging since data control parameters in the weight function. This idea of concise yet data-driven reduction of information embedded in high sampling frequency has driven a recent surge of interest in MIDAS models. Due to the robustness in no small frequency ratio, the MIDAS models have been drawing a large amount of attention recently. For example, Götz et al. [36] proposed a mixed frequency error correction model based on MIDAS models focusing on possibly co-integrated non-stationary processes with different sampling frequencies. Miller [58] introduced the co-integrated MIDAS time averaging regression models which generalized the nonlinear MIDAS regression models, proposed a test strategy for such models against the linear MIDAS regression models. Ghysels and Miller [29] showed the effects of the mixed-frequency data as well as temporal aggregation on the size of the common co-integration tests. Ghysels et al. [34] introduced a Granger causality test with mixed-frequency data. As noted earlier, MIDAS models involve nonlinear estimation. Andreou et al. [1] explored an NLS estimator for the MIDAS regression model and derived its asymptotic properties. The estimator is a so-called MIDAS-NLS estimator. They also showed that, in the presence of the mixed frequency effect, the LS estimator with a flat aggregation is asymptotically biased if the HF variable is serially correlated. It was demonstrated that the MIDAS-NLS estimator is relatively more efficient than the LS estimator as well.

Regression models that we have mentioned previously are parametric MIDAS models. Even though parametric MIDAS indeed make regression models more flexible than predetermined and fixed weights, it is still highly dependent on the weight function and the number of parameters chosen in the weight. In case that the inappropriate user-chosen components, for example, weight functions or the predetermined number of parameters, limit the flexibility in a way, a nonparametric MIDAS model was proposed in 2015 [12]. Instead of estimating parameters in weight functions by minimizing the MSE, Breitung and Roling [12] introduced a tuning parameter which penalizes the variability of aggregated weights. The objective function combines the MSE and the smooth spline term so that it provides a trade-off between the goodness-of-fit and the term which penalizes sharp changes of aggregated weights. It makes the nonparametric MIDAS approach more helpful that users do not need to make an appropriate decision of the number of parameters, or the weight function. However, this approach requires to calculate the optimal tuning parameter to estimate all aggregated weights. It means that the computation is more complicated compared to the methods mentioned above. Another approach, semi-parametric MIDAS model, which is proposed by Chen and Ghysels [15], provides volatility predicting combining kernel-based nonparametric and lag polynomial parametric methods. Kernel-based nonparametric approaches are applied to estimate news impact curves such as realized volatility or HF returns of HF regressors. Lag polynomials embody the parametric temporal dependence part. Semi-parametric models utilize lag polynomial to aggregate information of HF variables. Whereas, the computational complexity remains a fatal problem. As mentioned in Chen and Ghysels [15], it may take about 20 hours for estimating, while parametric models may only take a few minutes. Refer to Ghysels et al. [34] for more MIDAS approaches and their applications.

# Chapter 2

# Choice of IVs in Specification Test: MIDAS vs Time Averaging

## 2.1 Specification Test for MIDAS Models

As we mentioned in the Introduction, the flat averaging seems to be the most straightforward approach to deal with the mixed-frequency problem given the computational complexity. If the flat weight is enough to maintain adequate information of HF variables, it is not necessary to go over parametric MIDAS, let alone nonparametric MIDAS models. However, if the mixed frequency effect exists, the MIDAS model should be chosen over a time averaging model. This motivates a specification test that helps decide between the time averaging and the MIDAS models. There have yet been only a handful of such tests. Andreou et al. [1] presented a DWH type test, designed to see whether there is an omitted variable bias caused by overlooking the MIDAS effect. Miller [59] presented two VAT statistics. In particular, the second VAT statistic, called a modified VAT statistic, was designed for nonstationary HF variables. The modified VAT statistic is robust to the MIDAS models with covariates in deterministic and stochastic trends when the frequency ratio is large. Groenvik and Rho [38] further extended Miller's first VAT statistic using a self-normalized approach.

The methods mentioned above rely on the choice of IVs or other types of user-chosen parameters. In particular, Andreou et al. [1] briefly mentioned using all or part of HF variables as IVs for the DWH test. However, with such choice of IVs , it is possible that the 2SLS estimators may not be consistent when the chosen IVs are correlated with the error process. Furthermore, existing a large number of possibly weak IVs may also lead to the inconsistency of 2SLS estimators in the DWH test [14]. Therefore, the choice of IVs in the specification test context should be carefully examined. In Chapter 2, we shall further explore the DWH specification test introduced in Andreou et al. [1]. More precisely, there has not yet been practical guidance so far for choosing appropriate IVs in the DWH test. We shall propose a set of IVs that is suitable for this test.

Some notations of Chapter 2 will be defined here. Others will be clarified corresponding to the contents in this chapter. These notations are used in Appendix A.1 as well. Let $T$ be the sample size at low frequency, and $m$ be the frequency ratio between the two sampling frequencies. $\mathbf{j}_t$ is a $T \times 1$ vector with the $t$-th element being 1 and the rest 0. $\mathbf{j}$ is a $T \times 1$ vector of 1's. In Chapter 2, symbols $\mathbf{y} = (y_1, \cdots, y_T)'$, $\mathbf{x}_t = \left(x_t, x_{t-1/m}, \cdots, x_{t-(m-1)/m}\right)'$ and $\mathbf{z}_t = (z_{1,t}, \cdots, z_{p,t})'$ are reserved for the LF variable and the HF variable , and $p$ instrumental variables, respectively. We use $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_m)'$ to indicate an $m \times 1$ weight vector to aggregate the HF variable such that $\pi_i \geq 0$ and $\sum_{i=1}^{m} \pi_i = 1$. For matrix $A$, the matrix $P_A = A(A'A)^{-1}A'$ denotes the projection matrix onto the space spanned by the columns of $A$, and $M_A = I - P_A$ where $I$ or $I_.$ indicates the identity matrix. $\mathbf{u}_t = [u_{t,1}, \cdots, u_{t,q}]'$ is defined as $q$ additional LF covariates including the intercept at time $t$ in MIDAS models in Chapter 3 and Appendix A.2. For convenience, we define the following matrices: $X = [\mathbf{x}_1, \cdots, \mathbf{x}_T]'$, $Z = [\mathbf{z}_1, \cdots, \mathbf{z}_T]'$, $X^A = [\mathbf{j}, X\boldsymbol{\pi}_0] = [\mathbf{x}_1^A, \cdots, \mathbf{x}_T^A]'$ and $U = [\mathbf{u}_1, \cdots, \mathbf{u}_T]'$ where $\mathbf{x}_t^A = (1, x_t^A)'$ is the $t$-th row of $X^A$ and $\boldsymbol{\pi}_0$ is the predetermined weight vector.

## 2.2 Choice of IVs Based on the DWH Test

Consider a dataset with different sampling frequencies. Let $\{y_t\}_{t=1}^T$ and $\{\mathbf{x}_t\}_{t=1}^T$ be the variables observed at lower and higher sampling frequencies, respectively. The

MIDAS model is constructed, aiming to model the LF variable using HF variable:

$$y_{t+h} = \beta_0 + (\mathbf{j}_t' X \boldsymbol{\pi}(\boldsymbol{\theta}))\, \beta_1 + \varepsilon_t, \quad t = 1, \ldots, T. \tag{2.1}$$

The error process $\{\varepsilon_t\}$ is stationary and uncorrelated with $\{\mathbf{x}_t\}$. The vector $\boldsymbol{\pi}(\boldsymbol{\theta}) = (\pi_1(\boldsymbol{\theta}), \ldots, \pi_m(\boldsymbol{\theta}))'$ consists of a function of a finite dimensional unknown parameter $\boldsymbol{\theta}$ such that $\pi_i(\boldsymbol{\theta}) \geq 0$ and $\sum_{i=1}^m \pi_i(\boldsymbol{\theta}) = 1$. This vector dictates how much weight would be assigned when aggregating the HF variable, $\mathbf{x}_t$.

In a time averaging model, $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ is a predetermined fixed-weight vector that does not depend on any unknown parameter $\boldsymbol{\theta}$. Without loss of generality, let the number of aggregated lags be the same as the frequency ratio $m$. Then the regression model (2.1) becomes

$$y_t = \beta_0^A + (\mathbf{j}_t' X \boldsymbol{\pi}_0)\, \beta_1^A + \varepsilon_t^A = \beta_0^A + x_t^A \beta_1^A + \varepsilon_t^A. \tag{2.2}$$

Consider the test between time averaging (2.2) and MIDAS aggregation (2.1), i.e. $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$ versus $H_a : \boldsymbol{\pi} = \boldsymbol{\pi}(\theta)$. The two commonly used weights for time averaging are the flat aggregation $\boldsymbol{\pi}_0 = (1/m, \ldots, 1/m)'$ and the end-of-period sampling $\boldsymbol{\pi}_0 = (1, 0, \ldots, 0)'$. In this article, a more general scenario of the end-of-period sampling is considered: a fixed number, $n$, of elements in $\boldsymbol{\pi}_0$ are assigned with positive values, where $n$ is independent of $m$. For brevity, we assign the first $n$ elements and leave the rest as zero, i.e. $\boldsymbol{\pi}_0 = (\pi_{0,1}, \ldots, \pi_{0,n}, 0, \ldots, 0)'$ where $\pi_{0,i} > 0$ for $i = 1, \cdots, n$ and $\sum_{i=1}^n \pi_{0,i} = 1$. The LS principle can be applied to estimate the parameters $\beta_0^A$ and

$\beta_1^A$ in model (2.2) when the null hypothesis is true. We call this estimator, $\widehat{\boldsymbol{\beta}}^A = (\widehat{\beta}_0^A, \widehat{\beta}_1^A)' = (X^{A\prime} X^A)^{-1} X^{A\prime}\mathbf{y}$, the NULL-LS estimator. By comparing models (2.1) and (2.2), the error process (2.2) can be rewritten as $\varepsilon_t^A = \varepsilon_t + \mathbf{j}_t' X \left(\boldsymbol{\pi}(\boldsymbol{\theta}) - \boldsymbol{\pi}_0\right) \beta_1$. Under the null, $\varepsilon_t^A$ is uncorrelated with $x_t^A$ since $\varepsilon_t^A = \varepsilon_t$. However, under the alternative, $\varepsilon_t^A$ is correlated with $x_t^A$ due to the omitted variable. Therefore, testing whether $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ is equivalent to testing whether the NULL-LS estimator is consistent.

To test the consistency of the NULL-LS estimator using a DWH-type test, another estimator that is consistent under both the null and the alternative is required. This estimator may not be efficient under the null. See Lee [52], for example. The 2SLS estimator with proper IVs could be such an estimator. Assume that the IVs $\mathbf{z}_t$ are correlated with $x_t^A$, but uncorrelated with $\varepsilon_t^A$. Consider a two stage regression model: the time-averaging model (2.2) and an auxiliary regression of the flat aggregated term $x_t^A$ on the IVs $\mathbf{z}_t$ given as

$$y_t = \beta_0 + x_t^A \beta_1 + \varepsilon_t^A \quad \text{and} \quad x_t^A = \mathbf{z}_t' \Gamma + \tilde{\varepsilon}_t, \tag{2.3}$$

where $E\left(\tilde{\varepsilon}_t | x_t^A\right) = 0$. The 2SLS estimator is $\widehat{\boldsymbol{\beta}} = (X^{A\prime} P_Z X^A)^{-1}(X^{A\prime} P_Z \mathbf{y})$. The bias of the 2SLS estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be written as

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (X^{A\prime} P_Z X^A)^{-1}(X^{A\prime} P_Z)\boldsymbol{\varepsilon}^A, \tag{2.4}$$

where $\boldsymbol{\varepsilon}^A = (\varepsilon_1^A, \ldots, \varepsilon_T^A)'$. The following Assumption 2.1 is for the consistency of the NULL-LS under the null and for the consistency of the 2SLS estimator under both the null and the alternative.

**Assumption 2.1.** *Consider the time-averaging model and the auxiliary regression in (2.3).*

*(a)* $T^{-1}X^{A\prime}X^A \xrightarrow{p} E\left(\mathbf{x}_t^A \mathbf{x}_t^{A\prime}\right) = Q_{XX}$ *for some positive definite matrix* $Q_{XX}$*;*

*(b)* $T^{1/2}\left(T^{-1}X^{A\prime}\boldsymbol{\varepsilon}^A - E\left(\mathbf{x}_t^A \varepsilon_t^A\right)\right) \xrightarrow{d} N(\mathbf{0}, \Omega)$ *for some matrix* $\boldsymbol{\Omega}$*. Under the null,* $E\left(\mathbf{x}_t^A \varepsilon_t^A\right) = 0$*;*

*(c) Rank of* $Z$ *is no less than the column rank of* $X^A$*;*

*(d)* $T^{-1}Z'Z \xrightarrow{p} E\left(\mathbf{z}_t \mathbf{z}_t'\right) = Q_{ZZ}$ *for some positive definite matrix* $Q_{ZZ}$*;*

*(e)* $T^{-1}X^{A\prime}Z \xrightarrow{p} E\left(\mathbf{x}_t^A \mathbf{z}_t'\right) = Q_{XZ}$ *for some positive definite matrix* $Q_{XZ}$ *with rank as the column rank of* $X^A$*;*

*(f)* $T^{-1}Z'\boldsymbol{\varepsilon}^A \xrightarrow{p} E\left(\mathbf{z}_t \varepsilon_t^A\right) = \mathbf{0}$*;*

*(g)* $T^{-1/2}Z'\boldsymbol{\varepsilon}^A \xrightarrow{d} N(\mathbf{0}, \Sigma_{Z\varepsilon})$ *for some positive definite matrix* $\Sigma_{Z\varepsilon}$*.*

Assumptions 2.1(a) and 2.1(b) ensure the consistency of the NULL-LS estimator. Assumption 2.1(a) indicates that $X^A$ has full column rank. Assumption 2.1(b) implies the relation between the time-averaging term $X^A$ and the error process $\boldsymbol{\varepsilon}^A$, and their product should be asymptotically normal. Under the null, $X^A$ and $\boldsymbol{\varepsilon}^A$ should not be correlated, leading $E\left(\mathbf{x}_t^A \varepsilon_t^A\right) = 0$. Under the alternative, $X^A$ and $\boldsymbol{\varepsilon}^A$ are allowed to be correlated, i.e., $E\left(\mathbf{x}_t^A \varepsilon_t^A\right) \neq 0$. The variance-covariance matrix $\Omega$ in Assumption 2.1(b) can be consistently estimated. This can be done, for example, using heteroskedasticity and autocorrelation consistent (HAC) estimators [2, 61]. Assumptions 2.1(d)–(g) hold under both hypotheses. These ensure the consistency of the 2SLS estimator. In particular, Assumption 2.1(d) requires that $Z$ and $\boldsymbol{\varepsilon}^A$ should be uncorrelated. The number of IVs should be greater than or equal to the rank of $X^A$. Refer to Ruud [66] for more details and explanations.

Now we derive our test statistic. If Assumption 2.1 holds, the asymptotic distributions of $\widehat{\boldsymbol{\beta}}^A$ under the null and $\widehat{\boldsymbol{\beta}}$ under both hypotheses can be written as followings:

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}^A - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, V^A) \text{ under } H_0 \quad \text{and} \quad \sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, V) \text{ under } H_0 \text{ and } H_a, \tag{2.5}$$

where

$$V^A = Q_{XX}^{-1} \Omega Q_{XX}^{-1},$$

and

$$V = \left(Q_{XZ} Q_{ZZ}^{-1} Q'_{XZ}\right)^{-1} \left(Q_{XZ} Q_{ZZ}^{-1} \Sigma_{Z\varepsilon} Q_{ZZ}^{-1} Q'_{XZ}\right) \left(Q_{XZ} Q_{ZZ}^{-1} Q'_{XZ}\right)^{-1}.$$

Since both $\widehat{\boldsymbol{\beta}}^A$ and $\widehat{\boldsymbol{\beta}}$ are consistent under the null, the difference between the two estimators, $\widehat{\Delta} = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^A$ converges to zero in probability. The main idea of the

DWH test is to test whether $\widehat{\Delta}$ is significantly different from $\mathbf{0}$. This is equivalent to test whether $X^{A\prime}P_Z M_{X^A}\mathbf{y}$ is significantly different from $\mathbf{0}$, since $\widehat{\Delta}$ can be written as $\widehat{\Delta} = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^A = (X^{A\prime}P_Z X^A)^{-1}(X^{A\prime}P_Z M_{X^A}\mathbf{y})$ and $(X^{A\prime}P_Z X^A)^{-1}$ is positive definite.

We can easily see that

$$P_Z Z = Z, \quad M_Z Z = \mathbf{0}, \quad M_{X^A} X\boldsymbol{\pi}_0 = \mathbf{0}, \quad \mathbf{j}' M_{X^A}\mathbf{y} = 0, \quad \text{and} \tag{2.6}$$

$$X^{A\prime}P_Z \mathbf{M}_{X^A}\mathbf{y} = [\mathbf{j},\ X\boldsymbol{\pi}_0]'\, P_Z M_{X^A}\mathbf{y} = (0,\ (X\boldsymbol{\pi}_0)' P_Z M_{X^A}\mathbf{y})'. \tag{2.7}$$

Thus, $(X\boldsymbol{\pi}_0)'P_Z M_{X^A}\mathbf{y}$ should be approximately zero under the null. Let $\widehat{\boldsymbol{\varepsilon}} = M_Z X\boldsymbol{\pi}_0$ and $\widehat{\boldsymbol{\varepsilon}}^A = M_{X^A}\mathbf{y}$ indicate the fitted residuals from (2.3). Consider a regression model $\widehat{\boldsymbol{\varepsilon}}^A = X^A\boldsymbol{\alpha} + \widehat{\boldsymbol{\varepsilon}}\delta + \boldsymbol{v}$. Applying Frisch−Waugh−Lovell (FWL) theorem, the OLS estimator $\widehat{\delta}$ of $\delta$ is

$$\widehat{\delta} = \{(M_{X^A}M_Z X\boldsymbol{\pi}_0)'(M_{X^A}M_Z X\boldsymbol{\pi}_0)\}^{-1}\,(M_{X^A}M_Z X\boldsymbol{\pi}_0)' M_{X^A}\widehat{\boldsymbol{\varepsilon}}^A. \tag{2.8}$$

The latter part of $\widehat{\delta}$ can be derived as $(M_{X^A}M_Z X\boldsymbol{\pi}_0)'M_{X^A}\widehat{\boldsymbol{\varepsilon}}^A = (X\boldsymbol{\pi}_0)'M_{X^A}\mathbf{y} - (X\boldsymbol{\pi}_0)'P_Z M_{X^A}\mathbf{y}$. Since the third relation shown in (2.6) indicates that $(X\boldsymbol{\pi}_0)'M_{X^A}\mathbf{y} = 0$, $\widehat{\delta} = 0$ is equivalent to $(X\boldsymbol{\pi}_0)'P_Z M_{X^A}\mathbf{y} = 0$. Hence, testing whether $\widehat{\Delta}$ approaches to zero in probability can be viewed as testing if the coefficient $\widehat{\delta}$ is significantly different from zero. Consider the test statistic

$$\lambda_T = T\widehat{\delta}' \left(\mathbf{b}'(\widehat{V} - \widehat{V}^A)\mathbf{b}\right)^{-1} \widehat{\delta}, \tag{2.9}$$

where $\mathbf{b}' = -\left[(M_{X^A} M_Z X \boldsymbol{\pi}_0)'(M_{X^A} M_Z X \boldsymbol{\pi}_0)\right]^{-1} \left[(X\boldsymbol{\pi}_0)' P_Z X^A\right]$, and $\widehat{V}$ and $\widehat{V}^A$ are consistent estimators of $V$ and $V^A$, respectively.

**Theorem 2.1.** *Suppose Assumption 2.1 holds. Under the null hypothesis,* $\lambda_T \xrightarrow{d} \chi^2_1$.

The proof of Theorem 2.1 is presented in Appendix A.1.1. Assumption 1 holds only when the IVs $\mathbf{z}_t$ are chosen carefully. More specifically, $\mathbf{z}_t$ should be correlated with the time-averaging term, $x_t^A$, but uncorrelated with $\varepsilon_t^A$. This is to ensure Assumptions 2.1(e) and 2.1(f). Otherwise, the consistency of the 2SLS estimator may not be guaranteed. However, in practice, it is difficult to find such IVs. Andreou et al. [1] suggested using all or part of HF variables as IVs. However, they did not provide any practical guidance that is theoretically supported. In fact, with their suggested choice of IVs, it is possible that the chosen IVs are correlated with the error process. In this case, the 2SLS estimators would not be consistent, which may lower the power. In what follows, we shall propose a set of IVs that is theoretically valid for the DWH-type specification test. To derive theoretical properties, we assume the following conditions on the IVs and the data generating process.

**Assumption 2.2.** *Consider assumptions for* $k = 0, 1, \cdots, m-1, t = 1, \cdots, T$, *(a)* *The HF processes* $\{x_{t-k/m}\}$ *and* $\{\varepsilon_{t-k/m}\}$ *are independently, identically distributed (i.i.d.) or follow stationary AR(1) processes with finite second moment respectively;*

*(b)* $\{\varepsilon_{t-k/m}\}$ *is uncorrelated with* $\{x_{t-k/m}\}$*;*

*(c)* *Suppose* $\boldsymbol{\varepsilon}_{t,m} = (\varepsilon_t, \varepsilon_{t-1/m}, \cdots, \varepsilon_{t-(m-1)/m})'$ *with mean zero and positive definite covariance matrix, the error process* $\{\varepsilon_t\}$ *is an aggregated term of* $\boldsymbol{\varepsilon}_{t,m}$ *with the weight vector* $\boldsymbol{\pi}(\theta) = (\pi_1(\theta), \cdots, \pi_m(\theta))'$*, i.e.,* $\varepsilon_t = \boldsymbol{\varepsilon}_{t,m}{}'\boldsymbol{\pi}(\theta)$ *where* $\pi_j(\theta) = (2 - j/m)^{4\theta} / \sum_{i=1}^{m}(2 - i/m)^{4\theta}$.

Under Assumption 2.2, the LF response variable $\{y_t\}$ is viewed as an MIDAS aggregation of the underlying HF true process $\{y_{t-k/m}\}$, where $y_{t-k/m} = \beta_0 + x_{t-k/m}\beta_1 + \varepsilon_{t-k/m}$. Nevertheless $\{y_{t-k/m}\}$ is not observed in practice.

If we choose too many HF lags as IVs, it might lead to a problem of a large number of weak IVs. As a consequence, the 2SLS estimator may be biased towards the NULL-LS estimator. The bias tends to get worse when there is a more excessive number of IVs compared to the number of endogenous regressors. A brief explanation is presented by Greene [37]. Based on the number of the parameters in (2.3) and the consideration on possibly weak IVs, we shall construct the number of IVs as $p = 2$, $\mathbf{z}_t = (z_{1,t}, z_{2,t})'$, $t = 1 \cdots, T$, as linear combinations of the HF regressor. Inspired by Miller [59], we propose to choose weights of the IVs $\mathbf{z}_t$ as the following two decreasing sequences:

$$\begin{aligned}\Upsilon_1 &= (f_1(1), f_1(2), \cdots, f_1(m))', \text{ where } f_1(j) = \frac{0.9^{j-1}}{\sum_{i=1}^{m} 0.9^{i-1}}, \quad \text{and} \\ \Upsilon_2 &= (f_2(1), f_2(2), \cdots, f_2(m))', \text{ where } f_2(j) = \frac{m+1-j}{\sum_{i=1}^{m}(m+1-i)}.\end{aligned} \tag{2.10}$$

These weights are designed to decrease exponentially and linearly fast. This is to mimic the behaviors of the MIDAS weights with exponential Almon lag and beta polynomials. Then the two IVs can be written in a vector form as $\mathbf{z}_t' = \mathbf{x}_t{}'\Upsilon$, where $\Upsilon = [\Upsilon_1,\, \Upsilon_2]$. The following theorem demonstrates that the proposed IVs are approximately valid when the frequency ratio is large.

**Theorem 2.2.** *Let $Z_r = X\Upsilon_r = (z_{r,1}, \cdots, z_{r,T})'$ for $r = 1, 2$, where $\Upsilon_r$ be as presented in (2.10), be the two IVs. Assume that Assumption 2.2 holds. Write $Z = [Z_1, Z_2]$.*

*(a) Under the null hypothesis, $Z$ satisfies Assumption 2.1.*

*(b) Under the alternative hypothesis, $Z$ satisfies Assumptions 2.1(a)–(e). For any sample size $T$, Assumptions 2.1(f) and (g) are fulfilled approximately, as the frequency ratio $m$ approaches infinity. In fact, $E(z_{r,t}\varepsilon_t^A) = O(m^{-1})$ for $r = 1, 2$.*

The proof of Theorem 2.2 can be found in Appendix A.1.2. Under both the null and the alternative, it is easy to see that $z_{r,t}$ is correlated with $x_t^A$. The main result of Theorem 2.2 is that $z_{r,t}$ and $\varepsilon_t^A$ are asymptotically uncorrelated when the frequency ratio is large, with the rate $E(z_{r,t}\varepsilon_t^A) = O(m^{-1})$. Hence, the 2SLS estimator using our choice of the IVs is consistent when the frequency ratio $m$ is large. On the other hand, when $m$ is small, $T^{-1}Z'\boldsymbol{\varepsilon}^A$ converges, in probability, to a nonzero constant. Thus, the DWH specification test with our choice of IVs would only work when $m$ is large enough. This explains the low power of our test in finite samples when $m$ is small in the next section.

## 2.3 Monte Carlo Simulations

To compare the performance of estimation, we examine finite sample sizes and powers of our method and two other comparable methods in literature: the second test presented in Andreou et al. [1] (AGK, hereafter) and the unmodified VAT test in Miller [59]. The algorithms of the methods are briefly introduced as follows.

**Algorithm 1:** Our Method

**Data:** HF $\mathbf{x}_t$, LF $y_t$

1. $x_t^A = \mathbf{x}_t{}'\boldsymbol{\pi}_0$; IVs $\mathbf{z}_t = \mathbf{x}_t{}'\Upsilon$ with $\Upsilon$ in (2.10).
2. Obtain fitted error: regress $y_t$ on $x_t^A$ to get $\widehat{\varepsilon}_t^A$; regress $x_t^A$ on $\mathbf{z}_t$ to get $\widehat{\varepsilon}_t$.
3. Regress $\widehat{\varepsilon}_t^A$ on $x_t^A$ and $\widehat{\widehat{\varepsilon}}_t$ using $\widehat{\varepsilon}_t^A = \alpha_0 + x_t^A\alpha + \widehat{\widehat{\varepsilon}}_t\delta + \upsilon_t$.

**Result:** Test if $\widehat{\delta}$ is significantly different from 0 using a $t$ test and a HAC estimator [2, 61].

**Remark 2.1.** The AGK method can be implemented using Algorithm 1. To limit the number of IVs, the first two regressors of the HF variable are used in our simulations.

**Algorithm 2:** Miller's Method

**Data:** HF $\mathbf{x}_t$, LF $y_t$

1. $x_t^A = \mathbf{x}_t{}'\boldsymbol{\pi}_0$; $\mathbf{z}_t = \mathbf{x}_t{}'\Upsilon$ with $\Upsilon$ in (2.10).
2. Obtain fitted error: regress $y_t$ on $x_t^A$ to get $\widehat{\varepsilon}_t^A$.
3. Regress $\widehat{\varepsilon}_t^A$ on $x_t^A$ and $\mathbf{z}_t$ using $\widehat{\varepsilon}_t^A = \alpha_0 + x_t^A\alpha + \mathbf{z}_t'\phi + \upsilon_t$.

**Result:** Test if $\widehat{\phi}$ of $\phi$ is significantly different from 0 using a Wald statistic and a HAC estimator [2, 61].

**Remark 2.2.** Our method and Miller's unmodified VAT are similar. Both methods utilize the two MIDAS-type aggregations, $\mathbf{z}_t$, of the HF variable. While our method uses $\mathbf{z}_t$ as IVs under the classical framework with omitted variables, Miller's use of $\mathbf{z}_t$ is more direct. Miller's method searches whether the elements of $\mathbf{z}_t$ have any significant effect on residual of $y_t$ after taking time averaging into account.

To make the results comparable, we use a simulation setting similar to the one proposed by Miller [59]. At HF level, data are generated with $y_{t-k/m} = x_{t-k/m}\beta + \varepsilon_{t-k/m}$ for $t = 1, \ldots, T$, $k = 0, \ldots, m-1$. The HF processes $\{x_{t-k/m}\}$ and $\{\varepsilon_{t-k/m}\}$ are generated as stationary AR(1) processes given by $\varepsilon_{t-k/m} = c\varepsilon_{t-(k+1)/m} + \eta_{t-k/m}$ and $x_{t-k/m} = dx_{t-(k+1)/m} + \tilde{\eta}_{t-k/m}$, where $\{\eta_{t-k/m}\}$ and $\{\tilde{\eta}_{t-k/m}\}$ are i.i.d. $N(0, 1)$. Let $\beta = 10$. Denote $\mathbf{y}_t = (y_t, y_{t-1/m}, \cdots, y_{t-(m-1)/m})'$ and $\boldsymbol{\varepsilon}_{t,m}$ be the unobserved HF response and the error process between time $t-1$ and $t$. Let $\boldsymbol{\pi}_0 = \mathbf{j}/m$ and $\boldsymbol{\pi}(\theta) = (\pi_1(\theta), \cdots, \pi_m(\theta))$, where $\pi_j(\theta)$ is defined in Assumption 2.2(c). The LF processes are generated as $y_t = \mathbf{y}_t'\boldsymbol{\pi}(\theta)$ and $\varepsilon_t = \boldsymbol{\varepsilon}_{t,m}'\boldsymbol{\pi}(\theta)$. Here, $\theta = \theta_0 = 0$ indicates the flat aggregation, which corresponds to the null. If $\theta \neq 0$, the weights are no longer flat. Let $\theta = \theta_0 + k$ where $k \in \{0.1, 0.2, \cdots, 1.9, 2.0\}$ represent MIDAS-type alternatives. The nominal level is 0.05. $R = 2000$ Monte Carlo replications are generated. The sample sizes is $T \in \{125, 512\}$. The frequency ratio is $m \in \{4, 150, 365\}$.

**Table 2.1**

Empirical Sizes and Powers of our method (new), AGK, and Miller's method in the Representative Simulation Model

| T | m | c | $k$ | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 125 | 4 | 0.0 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | New | **5** | **0** | **0** | **0** | **10** | **66** | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | 0.8 | Miller | 7 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | 6.2 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | New | **5.3** | **4** | **1** | **4** | **14** | **46** | **81** | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 150 | 0.0 | Miller | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | 6 | **26** | **39** | **46** | **51** | **56** | **60** | **63** | **67** | **70** | **72** | **74** | **76** | **77** | **79** | **90** | **81** | **82** | **83** | **84** | **84** |
| | | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | 0.8 | Miller | **5.7** | **72** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | **5.7** | **8** | **13** | **20** | **28** | **35** | **40** | **45** | **51** | **56** | **60** | **64** | **66** | **69** | **71** | **73** | **75** | **76** | **77** | **79** | **80** |
| | | | New | **5.7** | **73** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 365 | 0.0 | Miller | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | 5.8 | **14** | **18** | **20** | **23** | **25** | **27** | **30** | **33** | **35** | **37** | **39** | **41** | **43** | **45** | **47** | **49** | **52** | **54** | **55** | **56** |
| | | | New | **4.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | 0.8 | Miller | 6.5 | **70** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | 5.8 | **7** | **9** | **12** | **15** | **16** | **18** | **21** | **23** | **27** | **28** | **30** | **32** | **35** | **37** | **39** | **41** | **43** | **45** | **47** | **48** |
| | | | New | **4.6** | **76** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 512 | 4 | 0.0 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | **5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | New | 5.4 | **6** | **5** | **3** | **1** | **0** | **0** | **0** | **0** | **1** | **3** | **14** | **42** | **79** | 96 | 99 | 100 | 100 | 100 | 100 | 100 |
| | | 0.8 | Miller | 5.6 | **62** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | **5.1** | **61** | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | New | 5.4 | **6** | **6** | **5** | **4** | **3** | **2** | **3** | **4** | **5** | **9** | **14** | **24** | **40** | **57** | **75** | **88** | 96 | 99 | 100 | 100 |
| | 150 | 0.0 | Miller | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | **5.5** | **22** | **53** | **75** | **85** | 90 | 93 | 94 | 95 | 96 | 96 | 97 | 97 | 97 | 98 | 98 | 98 | 98 | 98 | 98 | 98 |
| | | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | 0.8 | Miller | 6.1 | **23** | **71** | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | **5.1** | **6** | **8** | **12** | **17** | **24** | **31** | **38** | **46** | **53** | **60** | **67** | **72** | **77** | **80** | **84** | **86** | **89** | 90 | 91 | 92 |
| | | | New | 5.4 | **25** | **73** | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 365 | 0.0 | Miller | 5.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | 4.7 | **10** | **23** | **35** | **45** | **52** | **56** | **59** | **62** | **65** | **67** | **69** | **70** | **71** | **73** | **73** | **74** | **75** | **76** | **77** | **78** |
| | | | New | **5.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | 0.8 | Miller | 5.6 | **24** | **71** | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | AGK | **4.7** | **5** | **5** | **7** | **8** | **11** | **12** | **16** | **19** | **23** | **26** | **30** | **33** | **36** | **40** | **42** | **45** | **47** | **51** | **53** | **55** |
| | | | New | **5.3** | **29** | **78** | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

All values are shown as percentage. The nominal level is 0.05. Monte Carlo replication 2000. Bold numbers for $k = 0.0$ represent the rejection rates closest to 0.05 under the null. Bold cells for $k \neq 0.0$ indicate the rejection rates less than 0.90 under the local alternatives.

Table 1 presents the empirical sizes and the powers of our method, the AGK method and Miller's method when $c \in \{0, 0.8\}$ and $d = 0$. The results of more comprehensive settings are presented in Appendix B.1, which are consistent with what we observe in Table 1. When $k = 0$, sizes closest to 0.05 are presented in boldface. In all our simulation settings, all methods seem to have reasonable sizes. Our and the AGK method tend to have more cases in which sizes are closer to the nominal level, while Miller's unmodified VAT tends to slightly over-reject.

When $k \neq 0$, empirical rejection rates represent the powers of the tests. Powers less than 0.9 are shown in boldface. When $m$ is small, our method is not as powerful as the AGK method or Miller's unmodified VAT. These two methods have much better performance under all alternatives. For $T = 125$, when the HF error is AR(1), our method is less powerful when the effect size is small ($k \leq 0.6$), whether the HF error is i.i.d. or not. When $T = 512$, the power of our method is not very large when the effect size is not large enough. This observation is consistent with Theorem 2.2. When $m$ is small, the 2SLS estimator would not be consistent using the chosen instruments. If $m = 4$, the two weighted functions in constructing the instruments are almost identical. Therefore, when $m$ is small, $m = 4$, the AGK method seems to be good enough by choosing the most recent two HF variables (out of four). Miller's unmodified VAT is another attractive alternative when $m$ is small since it is as powerful as the AGK method.

However, when $m$ is large, the effect of a careful choice of instruments is more visible. When $m$ is 150, the power of the AGK method never exceeds 0.90 for all alternatives. In the meantime, our method tends to have higher power under almost all alternatives. Miller's unmodified VAT tends to be just a little less powerful than our method for small effective sizes. Additionally, as the sample size increases ($T = 512$), the AGK method becomes more powerful for large local alternatives, while all three methods reduce the power when the effective sizes are small. Except for a few small effect sizes with the AR(1) HF error process, our method has the highest power for most cases. Similar conclusions can be drawn for $m = 365$.

**Remark 2.3.** When our method works, i.e., when $1/m$ is small enough, our method and Miller's method have similar finite sample performance, though our test tends to have slightly better sizes and powers. Given their similar formulation, as mentioned in Remark 3.3, this similarity is somewhat expected. If one is interested in the comparison between the two methods, it would be interesting to consider more than one regressors. In this case, our method calls for more than two instruments, $\mathbf{z}_t$ would be different, making it easier to see the difference between the two methods. However, this is out of the scope of this chapter. We leave it as future work.

# Chapter 3

# Panel Nonparametric MIDAS

## 3.1 Nonparametric MIDAS and Its Extension with Panel Data

Except for focusing on the test for parametric models, we would concentrate on the nonparametric MIDAS as well. When the frequency ratio between HF and LF variables is relatively large, the nonparametric MIDAS proposed by Breitung and Roling [12] takes a long time to handle the complex computation. Inspired by the versatility of Fourier approximation, we shall control the frequency ratio to reduce the complexity of the nonparametric MIDAS by introducing Fourier series expansion.

Fourier approximation has been successfully applied in many aspects of macroeconomics and finance since Gallant [27]. In particular, many researchers have demonstrated a remarkable performance to capture a nonlinear trend. Becker et al. [6] introduced a test using a likelihood ratio approach to identify time variation in coefficients. These coefficients are parameterized using the Fourier expansion. Later, Becker et al. [7] modified the standard KPSS test proposed by Kwiatkowski et al. [49] for stationarity against a unit root. They used Fourier approximation for the deterministic trends in regression models to make the model more general than the one in the standard KPSS test. Moreover, Enders and Lee [20] proposed a Lagrange Multiplier unit root test relying on the availability of Fourier approximation on a series with several smooth structural breaks. Rodrigues and Taylor [65] generalized the procedure of the unit root test on local generalized least squares (GLS) de-trending and applied a Fourier approximation on the unknown deterministic trend. Later, Güriş [41] eliminated the tendency of nonstationary in structural breaks and nonlinearity in traditional unit root tests using Fourier expansion. These work demonstrated that Fourier approximation is capable of approximating most functions to any degree of accuracy as long as we use a sufficient number of parameters. In addition, HF regressors are linearly transformed using polynomials and trigonometric terms in a Fourier transformation. This makes our approach computationally efficient as it only requires OLS estimation rather than nonlinear estimation in the parametric MIDAS models.

It is also faster than Breitung and Roling [12]'s method, where a panelized optimization has to be conducted. In finite sample simulation, we compare the performance of our new nonparametric MIDAS model using Fourier transformation with the penalized nonparametric MIDAS introduced by Breitung and Roling [12] using the MSE of weight functions.

Tracing back to 1960s, Okun [62] analyzed the relationship between the deviation of the unemployment rate and the growth rate of GDP empirically. Later, Okun's law has been widely recognized in economics as a tool for short-run trend analysis. The data were measured quarterly for both the response and the regressor. In recent years, Economou and Psarianos [18], Guisinger et al. [40], Micallef [57], as well as some other literature, applied Okun's law to examine the labor market on the dataset from different countries. Moreover, Ball et al. [4] pointed out that the breakdowns in the law are exaggerated or flawed. However, even Okun's law is used extensively, the measuring frequencies are consistent for the unemployment rate and GDP. Apart from the consistent-frequency variables, it is reasonable to include more related regressors with higher frequencies, for instance, initial claims which could capture the size of layoffs in the labor market. Initial Claims have been known to be the most timely indicator of joblessness among professional forecasters. It is a report filed by individuals who are seeking to receive jobless benefits. As an informative indicator, it is usually measured weekly rather than quarterly. To maintain as much information in Okun's law using the weekly measured claims, introducing a MIDAS model in the law could be a wise and proper choice. Furthermore, according to Ball et al. [4], the significance of the variables of more flexible labor markets across states could be various depending on different aspects, such as geographical location, education, labor structure, etc. A recent study by Guisinger et al. [39] showed that the relationships in Okun's law are diverse across states. Guisinger et al. [39] argued that the difference in estimated coefficients is likely to represent heterogeneity in the functioning of the labor market. Such heterogeneity can be accounted for by the industry composition, union power, demographic characteristics, educational attainment, and so on. Since it requires more work to fit Okun's law state by state, clustering all state-level data as a whole panel data is worthy of study.

Encouraged by the performance of Fourier approximation in nonparametric MIDAS models, we extend such nonparametric MIDAS by introducing a clustering approach with panel MIDAS data for such empirical application. Clustering algorithms are widely used to visualize the impact of relations between subjects and modified not specifically only for mixed-frequency data. Su et al. [67] modified the traditional Lasso penalty in regression models into C-Lasso to penalize the difference between the estimated parameters in subjects and the estimated group-average parameters. C-Lasso requires a predetermined maximum of the group number and a choice of the tuning parameter. By minimizing the IC that they proposed in their paper, users can determine a value for the tuning parameter. However, it is crucial that it requires a user-chosen number of clusters. When the number of subjects is large, the possible number of clusters may vary a lot. Setting the maximum group number as the number of subjects would result in a relatively long-time computation. Ma and Huang [55] introduced a penalized method on the regression model for subjects and applied concave penalty functions to divide subjects into groups based on their intercepts. The penalty functions that they used are the MCP [73] and the SCAD [22], which not only share the sparsity properties like Lasso but also are asymptotically unbiased. Later on, Ma and Huang [54] extended their work on the intercepts to cluster subjects based on the part of the regressors instead of the intercepts only. Zhu and Qu [74] modified the regression model by introducing subject-wise B-spline smoothing functions to approximate covariates. Rather than exploring the patterns of parameters directly, they focused more on investigating the longitudinal trajectories over time. Casarin et al. [13] considers a more general situation where the parameters can change over time as the regime changes, using a Bayesian Markov switching model.

After introducing a new MIDAS model, we focus on a panel MIDAS regression model where the HF regressors are aggregated using nonparametric weight functions. Given the advantages of Fourier transformation and the concave penalties introduced in the clustering algorithm, we then extend our nonparametric model with Fourier transformation to include more subjects for clustering performance. In the first step, the MIDAS weights and other coefficients are chosen with incentives to have almost the same if two subjects would have similar estimated coefficients. This part is handled using the idea of the feature selection. In particular, we use the MCP penalty as it is well known to provide unbiased estimations. The only major assumption that we need is the sparsity assumption, which requires the number of groups to be much smaller than the number of subjects. We apply the nonparametric MIDAS weights with Fourier expansion as introduced above to handle the HF part, as different subjects may have different forms of weights. What more, with the help of Fourier approximation, the number of parameters in distinct periods can be unified. In the next step, the estimated coefficients are clustered using conventional clustering methods such as K-means clustering. This would work reasonably well as the coefficients that are already chosen to be very close to each other in the first step, presumably if they are in the same group. To the best of our knowledge, there has been only a couple of research articles that handle panel MIDAS models. An unpublished article [48] proposes a GMM approach for panel data with parametric MIDAS models. Coffey et al. [16] proposed a regression model for time-course gene expression data. They extend the linear mixed effects and P-spline smoothing model for clustering multiple gene expression profiles. Without knowing the true clusters of profiles, profiles are randomly assigned to a predetermined number of clusters. After estimating parameters multiple times using the EM algorithm and changing the starting points, the AIC or the BIC can be used to determine the number of clusters properly among user-chosen candidates. Similarly to Su et al. [67], this method also requires a predetermined number of clusters and to find a proper value of clusters, users have to go through many, even all possible candidates to compare their performance based on the AIC or BIC. All technical proofs and full simulation results can be found in the appendix.

The rest of this chapter is organized as follows. In Section 3.2, we focus on applying a Fourier series approximation to estimate the coefficients of HF regressors in the MIDAS model. We show that Fourier expansion provides an accurate estimation of aggregated weights theoretically and empirically. In particular, in Section 3.2.2, the median RMSEs of parameter estimation and the one-step-ahead forecast are chosen to present the performance of our method compared to the nonparametric approach proposed by Breitung and Roling [12]. In Section 3.3, we introduce a more general model with panel data and other LF covariates. Simulation results are provided to show clustering performance of estimated weights using Fourier approximation empirically. Besides, We provide conditions for the proposed estimator. Section 3.4 provides an empirical application of our clustering method by revisiting Okun's law. We analyze the importance of the HF initial claims on predicting the unemployment rate across states. Heterogeneity in the functioning of the labor market is examined, and states are clustered based on the predicted behavior of the unemployment rate as well as the initial claims.

Apart from the notations defined in Chapter 2, some notations are defined throughout this Chapter additionally. Others will be clarified with respect to contents. The following notations are used in Appendix A.2 as well. For an $m \times n$ matrix $A$ with its

$(i, j)$th element being $a_{ij}$, $||A||_p$ to indicate $p$-norm induced by corresponding vector norms, that is, $||A||_p = \frac{\sup_{x \neq 0} ||Ax||_p}{||x||_p}$. In particular, $||A||_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|$ and $||A||_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|$. For a symmetric and positive definite matrix $A$, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ indicate the smallest and largest eigenvalues of $A$, respectively. In this case, $||A||_2 = \lambda_{\max}(A)$. $I_p$ is an identity matrix of size $p$ and $\otimes$ denotes the Kronecker product. For any real number $x$, $\lfloor x \rfloor$ denotes the largest integer that is smaller than or equal to $x$.

## 3.2 Nonparametric MIDAS for Single Subject

### 3.2.1 Nonparametric MIDAS with Fourier Expansion

Consider the following one-HF-variable MIDAS model with the lead $h \geq 0$:

$$y_{t+h} = \sum_{i=1}^{q} \alpha_i u_{t,i} + \sum_{j=0}^{m-1} \beta_j^* x_{t,j} + \varepsilon_{t+h} = \mathbf{u}_t' \boldsymbol{\alpha} + \mathbf{x}_t{}' \boldsymbol{\beta}^* + \varepsilon_{t+h}, \tag{3.1}$$

for $t = 1, \cdots, T$. Here, $\boldsymbol{\alpha}_i$ be the corresponding coefficient. $\mathbf{x}_t = (x_{t,0}, \cdots, x_{t,m-1})'$ is the HF variable and $\boldsymbol{\beta}^* = (\beta_0^*, \cdots, \beta_{m-1}^*)'$ is the weight that aggregates $\mathbf{x}_t$ to the LF. $\varepsilon_{t+h}$ is the error process. To introduce Fourier approximation in (3.1), consider the following Dirichlet condition for a periodic function $f(\cdot)$.

1. The function is periodic on the whole real values $\mathbb{R}$, i.e. $f(x) = f(x+P)$ for $x \in \mathbb{R}$ where $P$ is the period of function $f(x)$;

2. $f(x)$ has a finite number of maxima and minima;

3. $f(x)$ has an at most finite number of discontinuous points in one period;

4. $f(x)$ is integrable over the period.

Any non-periodic function defined on a finite interval can be viewed to be extended on $\mathbb{R}$. Hence, there exists a Fourier series expansion for such non-periodic function. In the subsequent argument, we assume that there is an underlying weight function $\beta^*(\cdot)$ defined on $[0,1]$ which satisfies the Dirichlet condition. The weight function $\beta_j^*$ in MIDAS regression model (3.1) are viewed as a realization from this function $\beta^*(\cdot)$, $\beta_j^* = \beta^*(j/m)$ for $j = 0, \cdots, m-1$. Since $\beta^*(\cdot)$ satisfies the Dirichlet condition, it can be approximated with appropriately chosen orders in Fourier expansion.

Formally, we assume the following condition for the weights $\beta_j^*$:

**Assumption 3.1.** *For any $r \in [0,1]$, $\beta^*_{\lfloor rm \rfloor} \to \beta^*(r)$ as $m \to \infty$. Here, $\beta^*(\cdot)$ is defined over $[0,1]$, has a finite number of maxima and minima, has a finite number of discontinuous points, and is integrable over $[0,1]$.*

For large enough $L$ and $K$,

$$\beta_j^* \approx \beta^*(j/m) \approx \sum_{l=0}^{L} \beta_l (j/m)^l + \sum_{k=1}^{K} \left(\beta_{1,k} \sin(2\pi k \cdot j/m) + \beta_{2,k} \cos(2\pi k \cdot j/m)\right). \quad (3.2)$$

Consequently, the MIDAS model (3.1) with Fourier approximation of the parameters becomes

$$\begin{aligned} y_{t+h} \approx & \sum_{i=1}^{q} \alpha_i u_{t,i} + \varepsilon_{t+h} \\ & \sum_{j=0}^{m-1} \left( \sum_{l=0}^{L} \beta_l \left(\frac{j}{m}\right)^l x_{t,j} + \sum_{k=1}^{K} \left( \beta_{1,k} \frac{j \sin(2\pi k)}{m} x_{t,j} + \beta_{2,k} \frac{j \cos(2\pi k)}{m} x_{t,j} \right) \right), \end{aligned} \quad (3.3)$$

where $\widetilde{x}_{t,l}$, $\widetilde{x}_{t,k}^{(s)}$ and $\widetilde{x}_{t,k}^{(c)}$ are transformed HF data for $l = 0, \cdots, L$ and $k = 1, \cdots, K$,

$$\widetilde{x}_{t,l} = (j/m)^l x_{t,j}, \quad \widetilde{x}_{t,k}^{(s)} = \sin(2\pi k \cdot j/m) x_{t,j}, \; \widetilde{x}_{t,k}^{(c)} = \cos(2\pi k \cdot j/m) x_{t,j}.$$

Denote the transformation matrix $M$ as the following:

$$M = \begin{bmatrix} (0/m)^0 & (1/m)^0 & \cdots & ((m-1)/m)^0 \\ \vdots & \vdots & & \vdots \\ (0/m)^L & (1/m)^L & \cdots & ((m-1)/m)^L \\ \sin(2\pi \cdot 1 \cdot 0/m) & \sin(2\pi \cdot 1 \cdot 1/m) & \cdots & \sin(2\pi \cdot 1 \cdot (m-1)/m) \\ \cos(2\pi \cdot 1 \cdot 0/m) & \cos(2\pi \cdot 1 \cdot 1/m) & 0 \cdots & \cos(2\pi \cdot 1 \cdot (m-1)/m) \\ \vdots & \vdots & & \vdots \\ \sin(2\pi \cdot K \cdot 0/m) & \sin(2\pi \cdot K \cdot 1/m) & \cdots & \sin(2\pi \cdot K \cdot (m-1)/m) \\ \cos(2\pi \cdot K \cdot 0/m) & \cos(2\pi \cdot K \cdot 1/m) & \cdots & \cos(2\pi \cdot K \cdot (m-1)/m) \end{bmatrix} \tag{3.4}$$

then the transformed data becomes $\widetilde{X} = XM'$, where $\widetilde{X} = [\widetilde{\mathbf{x}}_1, \cdots, \widetilde{\mathbf{x}}_T]'$. The vector $\widetilde{\mathbf{x}}_t$ is defined as $(\widetilde{x}_{t,0}, \widetilde{x}_{t,1}, \cdots, \widetilde{x}_{t,L}, \widetilde{x}^{(s)}_{t,1}, \widetilde{x}^{(c)}_{t,1}, \cdots, \widetilde{x}^{(s)}_{t,K}, \widetilde{x}^{(c)}_{t,K})'$, . The MIDAS model with Fourier expansion in (3.3) can be written as

$$\mathbf{y} = U\boldsymbol{\alpha} + X\boldsymbol{\beta}^* + \boldsymbol{\varepsilon} \approx U\boldsymbol{\alpha} + \widetilde{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{3.5}$$

Denote $W = (U, \widetilde{X})$ to be the new dataset. The model becomes

$$\mathbf{y} \approx W\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

Using $\boldsymbol{\beta}^* \approx M'\boldsymbol{\beta}$, the OLS estimator, $\widehat{\boldsymbol{\beta}^*}$, of $\boldsymbol{\beta}^*$ can be derived as

$$\begin{aligned}\widehat{\boldsymbol{\beta}^*} &= M'C_s\widehat{\boldsymbol{\gamma}} = M'C_s(W'W)^{-1}W'\mathbf{y} = M'C_s(W'W)^{-1}W'(W\boldsymbol{\gamma}+\boldsymbol{\varepsilon}) \\ &= M'C_s(\boldsymbol{\gamma}+(W'W)^{-1}W'\boldsymbol{\varepsilon}) = \boldsymbol{\beta}^* + M'C_s(W'W)^{-1}W'\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta}^* + M'C_s\left(\frac{1}{T}W'W'\right)^{-1}\left(\frac{1}{T}W'\boldsymbol{\varepsilon}\right),\end{aligned} \tag{3.6}$$

where $C_s = \left[\mathbf{0}_{(l+1+2K)\times q}, I_{L+1+2K}\right]$ is a $(L+1+2K)\times(q+L+1+2K)$ matrix. To show that $\boldsymbol{\beta}$ can be estimated consistently via the OLS estimator $\widehat{\boldsymbol{\beta}^*}$, we assume some regular conditions.

**Assumption 3.2.** *Consider $T\times m$ regressors $X$, $T\times 1$ vector $\mathbf{y}$ and $T\times q$ covariates $U$. $M$ is the transformation matrix for $\mathbf{x}_t$.*

1. $\mathbf{y} = U\boldsymbol{\alpha} + X\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$ *where $\boldsymbol{\varepsilon}$ is the error process vector.*

2. *The regressors $\mathbf{u}_t$, $\mathbf{x}_t$ are orthogonal to $\varepsilon_t$, i.e. $E(\mathbf{u}_t\varepsilon_t) = \mathbf{0}$, $E(\mathbf{x}_t\varepsilon_t) = \mathbf{0}$.*

3. *$T^{-1}X'X$, $T^{-1}U'U$ is positive definite with finite samples. Moreover,*

$$T^{-1}X'X \xrightarrow[T\to\infty]{p} E(\mathbf{x}_t\mathbf{x}_t') =: \Sigma_{XX}, \qquad T^{-1}U'X \xrightarrow[T\to\infty]{p} E(\mathbf{u}_t\mathbf{x}_t') =: \Sigma_{UX},$$
$$T^{-1}X'U \xrightarrow[T\to\infty]{p} E(\mathbf{x}_t\mathbf{u}_t') =: \Sigma_{XU}, \qquad T^{-1}U'U \xrightarrow[T\to\infty]{p} E(\mathbf{u}_t\mathbf{u}_t') =: \Sigma_{UU}.$$

4. *The process $\{\mathbf{x}_t\varepsilon_t\}$ and $\{\mathbf{u}_t\varepsilon_t\}$ are martingale difference sequences with finite a second moment.*

In classical linear regression models, the error term is assumed to be identically and independently normally distributed. Whereas, in economic fields, it may be too narrow to assume i.i.d. error term. The assumptions shown above are better suited. In deriving the asymptotic distribution of the OLS estimator, the distributional assumption is not specified in the above assumptions. See Section 2.3 in Hayashi [44] for more details. Given the assumptions listed above, $\boldsymbol{\beta}^*$ can be consistently estimated by the OLS estimator $\widehat{\boldsymbol{\beta}}$. The asymptotic distribution of $\widehat{\boldsymbol{\beta}^*}$ is

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}^*} - \boldsymbol{\beta}^*\right) \xrightarrow[T\to\infty]{d} N\left(\mathbf{0}_m, M'C_s Q_W^{-1} Q_{W\boldsymbol{\varepsilon}} Q_W^{-1} C_s' M\right), \tag{3.7}$$

where

$$Q_W = \begin{bmatrix} \Sigma_{UU} & \Sigma_{UX}M' \\ M\Sigma_{XU} & M\Sigma_{XX}M' \end{bmatrix}, \quad Q_{W\boldsymbol{\varepsilon}} = \begin{bmatrix} E(\mathbf{u}_t\varepsilon_t\varepsilon_t\mathbf{u}_t') & E(\mathbf{u}_t\varepsilon_t\varepsilon_t\mathbf{x}_t')M' \\ ME(\mathbf{x}_t\varepsilon_t\varepsilon_t\mathbf{u}_t') & ME(\mathbf{x}_t\varepsilon_t\varepsilon_t\mathbf{x}_t')M' \end{bmatrix}.$$

For inference, a consistent estimator $\widehat{Q}_{W\boldsymbol{\varepsilon}}$ of $Q_{W\boldsymbol{\varepsilon}}$ can be used, for instance, using the HAC estimation [2].

### 3.2.2 Simulation: Nonparmetric MIDAS

Given the nonparametric MIDAS with one subject, our method is compared with the nonparametric MIDAS approach proposed by Breitung and Roling [12]. The data is generated as the following. For $j = 0, \cdots, m-1,\ t = 1, \cdots, T$,

$$y_{t+h} = \alpha_0 + \sum_{j=0}^{m-1} \beta_j^* x_{t,j} + \varepsilon_{t+h}, \quad x_{t,j} = c + dx_{t,j-1} + \tilde{\varepsilon}_{t,j}, \tag{3.8}$$

where $\varepsilon_{t+h} \sim i.i.d.N(0, 0.125)$, $\tilde{\varepsilon}_{t,j} \sim i.i.d.N(0,1)$, $\alpha_0 = 0.5$, $\beta_j^* = \alpha_1 \omega_j(\boldsymbol{\theta})$. $\alpha_1$ are chosen from $\{0.2, 0.3, 0.4\}$, $T \in \{100, 200, 400\}$ and the frequency ratio $m \in \{20, 40, 60, 150, 365\}$. For the AR(1) HF regressor, $c = 0.5$, $d = 0.9$.

**Remark 3.1.** *In the current framework, we are absorbed in showing the performance of Fourier approximation in MIDAS regression models. More complicated settings, such as including LF variables in the model, are acceptable to use our nonparametric MIDAS. Fourier approximation is only required for HF variables to reduce and unify the frequency ratio.*

To show that our method has a satisfactory performance on estimating various shapes of weights in the MIDAS regression model, we consider five different shapes for $\omega_j(\boldsymbol{\theta})$. The first four are generated discretely by four functions suggested in Breitung and Roling [12]. The last one is typically the weight of an end-of-period sampling. The following weight functions satisfy the Dirichlet conditions, then the Fourier series expansion exists for each functions. However, since the value of parameter $K$ and $L$ are predetermined, we choose $K = 3$, $L = 2$ in the following Monte Carlo simulation.

1. Exponential Decline: $\omega_j(\theta_1, \theta_2) = \dfrac{\exp\{\theta_1 j + \theta_2 j^2\}}{\sum_{i=1}^{m} \exp\{\theta_1 i + \theta_2 i^2\}}$, $\theta_1 = 7 \times 10^{-4}$, $\theta_2 =$

$-6 \times 10^{-3}$;

2. Hump-Shaped: $\omega_j(\theta_1, \theta_2) = \dfrac{\exp\{\theta_1 j - \theta_2 j^2\}}{\sum_{i=1}^{m} \exp\{\theta_1 i - \theta_2 i^2\}}$, $\theta_1 = 0.08$, $\theta_2 = \theta_1/10$, $\theta_1/20, \theta_1/30$;

3. Linear Decline: $\omega_j(\theta_0, \theta_1) = \dfrac{\theta_0 + \theta_1(j-1)}{\theta_0(m) + \theta_1(m)(m+1)/2}$, $\theta_1 = 1$, $\theta_2 = 0.05$;

4. : $\omega_j(\theta_1, \theta_2) = \dfrac{\theta_1}{m}\left(\sin\left(\theta_2 + 2\pi\dfrac{j}{m-1}\right)\right)$, $\theta_2 = 0.01$, $\theta_1 = 5, 5/2, 5/3$;

5. Discrete: $\omega_j = (0, 0, \cdots, 0, 5/m, \cdots, 5/m)$ where we assign value $5/m$ to the last one fifth elements and 0 to the rest.

The first weight function is also known as the Exponential Almon Lag proposed by Ghysels et al. [33], which is able to mimic various shapes with a few parameters. The formula that we concentrate on is designed by two parameters. The cyclical weight and the end-of-sampling weight illustrate the flexibility of our methods. All weights are positive and normalized, to sum up to one.

To compare our method with the nonparametric approach proposed by Breitung and Roling [12], we present the median RMSE of parameters $\beta^*_\cdot$'s among all replications in Table 3.1 and the one-step-ahead forecast of the response in Table 3.2, correspondingly. RMSE of estimated $\widehat{\boldsymbol{\beta}^*}$ is calculated as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|M'\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_2^2}.$$

In consideration of the computational complexity, the number of replications is set to be 1000 for RMSE of $\beta_\cdot^*$'s and 250 for RMSE of the one-step-ahead forecast. In particular, the one-step-ahead forecast is calculated via the following steps.

1. Obtain the estimated $\widehat{\boldsymbol{\beta}}^*$ in the regression model $y_{t+h} = \mathbf{x}_t{}'\boldsymbol{\beta}^* + \varepsilon_{t+h}$ for $t = 1, \cdots, T/2$. Denote the estimated parameter as $\widehat{\boldsymbol{\beta}^*}$.

2. Get the predicted response $\widehat{y}_{T/2+h+1}$ by using the estimated parameter $\widehat{\boldsymbol{\beta}}^*_{T/2}$ and one-step-ahead regressor $\mathbf{x}_{T/2+1}$, i.e. $\widehat{y}_{T/2+h+1} = \mathbf{x}_{T/2+1}{}'\widehat{\boldsymbol{\beta}}^*_{T/2}$.

3. Repeat step 1-2 by using one more step ahead of the regressor and the response to get the estimated response $\widehat{y}_{T/2+h+k}$ for $k = 2, \cdots, T/2$. Especially, in $k$-th repeated process, use the observations $\mathbf{x}_{t+k-1}$ and $y_{t+h+k-1}$ for $t = 1, \cdots, T/2$ to get the estimator $\widehat{\boldsymbol{\beta}}^*_{T/2+k-1}$. Obtain the estimated $y_{T/2+h+k}$ by using $\mathbf{x}_{T/2+k}$ and the estimated parameter $\widehat{\boldsymbol{\beta}}^*_{T/2+k-1}$.

4. After calculating the estimated response $\widehat{y}_{t+h}$ for $t = T/2+1, \cdots, T$, we compare the estimated responses with the observed responses and calculate the RMSE of the predicted variable.

$$RMSE = \sqrt{\frac{1}{T/2}\sum_{k=1}^{T/2}(\widehat{y}_{T/2+h+k} - y_{T/2+h+k})^2}.$$

We present the main idea of the nonparametric MIDAS proposed in Breitung and Roling [12] as well to have a more intuitive understanding of the advantage of Fourier transformation in MIDAS. The nonparametric MIDAS in Breitung and Roling [12] takes advantage of the cubic smooth spline. The least-squares objective function is penalized by the sum of the second difference of weights to balance the goodness of fit and the smoothness of weights. Suppose that the MIDAS model is shown in (3.1). The penalized least-squares objective function is

$$Q_{BR} = \sum_{t=1}^{T} \left( y_{t+h} - \alpha_0 - \sum_{i=0}^{m-1} x_{t,i}\beta_i^* \right)^2 + \lambda_{BR} \sum_{i=2}^{m} \left( \bigtriangledown^2 \beta_i^* \right)^2 ,$$

where $\bigtriangledown^2 \beta_i^* = (\beta_i^* - 2\beta_{i-1}^* + \beta_{i-2}^*)$ indicates the second difference of weights. The SLS estimator [12] becomes

$$\widehat{\boldsymbol{\beta}}_{BR}^* = \arg\min_{\beta^*} \left( \|\mathbf{y} - X\boldsymbol{\beta}^*\|_2^2 + \lambda_{BR} \|D\boldsymbol{\beta}^*\|_2^2 \right),$$

where

$$D_{(m-2)\times(m+1)} = \begin{pmatrix} 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}.$$

In such nonparametric MIDAS model, $\lambda_{BR}$ is a tuning parameter which has to be predetermined. Breitung and Roling [12] minimized the modified AIC to choose $\lambda$. They introduced a peudo-dimension $s_{\lambda_{BR}}$, which can be treated as the dimension of spanned space of estimated parameters.

$$AIC_{\lambda_{BR}} = \log\left(\|\mathbf{y} - \widehat{\mathbf{y}}_{BR}\|_2^2\right) + \frac{2(s_{\lambda_{BR}} + 1)}{T - s_{\lambda_{BR}} + 2},$$

where $\widehat{\mathbf{y}}_{BR} = X(X'X + \lambda_{BR}D'D)^{-1}X'\mathbf{y}$. Except for common sense that the pseudo-dimension $s_{\lambda_{BR}}$ is supposed to be an integer, it is also allowed to be any real value in $[2, m-1)$. It results in estimated parameters to be more smooth expectantly. In addition, they proposed a way to minimize AIC by solving the first-order condition. More details can be found in Breitung and Roling [12].

**Table 3.1**

Median RMSE of the Estimated Parameter $\boldsymbol{\beta}$

| | | | $m = 20$ | | | 40 | | | 60 | | | 150 | | | 365 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | Method | $\alpha_1 = 0.2$ | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 |
| Exp Decline | 100 | B&R | 0.9066 | 0.9195 | 0.9168 | 0.6490 | 0.6583 | 0.6603 | 0.5718 | 0.5716 | 0.5665 | 0.4814 | 0.4828 | 0.4862 | 0.4468 | 0.4395 | 0.4469 |
| | | Fourier | 0.5695 | 0.5829 | 0.5851 | 0.2480 | 0.2247 | 0.2278 | 0.1345 | 0.1338 | 0.1342 | 0.0592 | 0.0721 | 0.0877 | 0.0825 | 0.1208 | 0.1600 |
| | 200 | B&R | 0.8630 | 0.8790 | 0.8801 | 0.5914 | 0.5901 | 0.5890 | 0.4986 | 0.4958 | 0.4944 | 0.4016 | 0.3986 | 0.4002 | 0.3803 | 0.3778 | 0.3796 |
| | | Fourier | 0.4162 | 0.4157 | 0.4068 | 0.1520 | 0.1546 | 0.1591 | 0.0894 | 0.0922 | 0.0938 | 0.0490 | 0.0632 | 0.0806 | 0.0803 | 0.1183 | 0.1578 |
| | 400 | B&R | 0.8383 | 0.8441 | 0.8489 | 0.5426 | 0.5444 | 0.5417 | 0.4382 | 0.4435 | 0.4391 | 0.3403 | 0.3418 | 0.3401 | 0.3132 | 0.3127 | 0.3121 |
| | | Fourier | 0.2850 | 0.2818 | 0.2851 | 0.1109 | 0.1105 | 0.1104 | 0.0648 | 0.0663 | 0.0678 | 0.0436 | 0.0600 | 0.0775 | 0.0787 | 0.1175 | 0.15620 |
| Hump Shaped | 100 | B&R | 0.9052 | 0.9172 | 0.9172 | 0.6465 | 0.6580 | 0.6595 | 0.5710 | 0.5703 | 0.5662 | 0.4814 | 0.4829 | 0.4863 | 0.4467 | 0.4394 | 0.4469 |
| | | Fourier | 0.5695 | 0.5829 | 0.5851 | 0.2269 | 0.2247 | 0.2280 | 0.1338 | 0.1338 | 0.1330 | 0.0458 | 0.0458 | 0.0458 | 0.0224 | 0.0224 | 0.0245 |
| | 200 | B&R | 0.8639 | 0.8796 | 0.8776 | 0.5877 | 0.5894 | 0.5906 | 0.4995 | 0.4948 | 0.4942 | 0.4016 | 0.3982 | 0.4001 | 0.3800 | 0.3778 | 0.3795 |
| | | Fourier | 0.4162 | 0.4157 | 0.4069 | 0.1520 | 0.1546 | 0.1591 | 0.0894 | 0.0917 | 0.0933 | 0.0316 | 0.0316 | 0.0316 | 0.0173 | 0.0173 | 0.0173 |
| | 400 | B&R | 0.8390 | 0.8439 | 0.8490 | 0.5418 | 0.5426 | 0.5412 | 0.4383 | 0.4442 | 0.4391 | 0.3403 | 0.3416 | 0.3406 | 0.3132 | 0.3127 | 0.3121 |
| | | Fourier | 0.2850 | 0.2818 | 0.2851 | 0.1109 | 0.1105 | 0.1103 | 0.0640 | 0.0656 | 0.0663 | 0.0224 | 0.0224 | 0.0224 | 0.0100 | 0.0141 | 0.0141 |
| Linear Decline | 100 | B&R | 0.9064 | 0.9191 | 0.9147 | 0.6471 | 0.6575 | 0.6594 | 0.5711 | 0.5707 | 0.5660 | 0.4812 | 0.4828 | 0.4863 | 0.4467 | 0.4394 | 0.4469 |
| | | Fourier | 0.5694 | 0.5829 | 0.5851 | 0.2269 | 0.2247 | 0.2280 | 0.1338 | 0.1338 | 0.1330 | 0.0458 | 0.0458 | 0.0458 | 0.0224 | 0.0224 | 0.0224 |
| | 200 | B&R | 0.8635 | 0.8786 | 0.8787 | 0.5875 | 0.5888 | 0.5905 | 0.4995 | 0.4946 | 0.4942 | 0.4016 | 0.3982 | 0.4004 | 0.3800 | 0.3778 | 0.3795 |
| | | Fourier | 0.4162 | 0.4157 | 0.4068 | 0.1520 | 0.1546 | 0.1591 | 0.0894 | 0.0917 | 0.0933 | 0.0316 | 0.0316 | 0.0316 | 0.0141 | 0.0141 | 0.0141 |
| | 400 | B&R | 0.8379 | 0.8441 | 0.8483 | 0.5417 | 0.5425 | 0.5411 | 0.4464 | 0.4440 | 0.4393 | 0.3403 | 0.3416 | 0.3401 | 0.3134 | 0.3127 | 0.3121 |
| | | Fourier | 0.2850 | 0.2818 | 0.2851 | 0.1109 | 0.1058 | 0.1100 | 0.0640 | 0.0656 | 0.0663 | 0.0224 | 0.0224 | 0.0224 | 0.0100 | 0.0100 | 0.0100 |
| Cyclical | 100 | B&R | 0.9144 | 0.9256 | 0.9257 | 0.6500 | 0.6576 | 0.6598 | 0.5696 | 0.5695 | 0.5663 | 0.4812 | 0.4828 | 0.4863 | 0.4467 | 0.4394 | 0.4469 |
| | | Fourier | 0.5689 | 0.5825 | 0.5694 | 0.2269 | 0.2247 | 0.2280 | 0.1338 | 0.1338 | 0.1330 | 0.0458 | 0.0458 | 0.0458 | 0.0224 | 0.0224 | 0.0224 |
| | 200 | B&R | 0.8677 | 0.8774 | 0.8807 | 0.5877 | 0.5908 | 0.5880 | 0.4995 | 0.4943 | 0.4942 | 0.4016 | 0.3982 | 0.4001 | 0.3800 | 0.3778 | 0.3795 |
| | | Fourier | 0.4163 | 0.4159 | 0.4061 | 0.1520 | 0.1546 | 0.1591 | 0.0894 | 0.0917 | 0.0933 | 0.0316 | 0.0316 | 0.0316 | 0.0141 | 0.0141 | 0.0141 |
| | 400 | B&R | 0.8426 | 0.8456 | 0.8472 | 0.5418 | 0.5450 | 0.5401 | 0.4407 | 0.4432 | 0.4394 | 0.3403 | 0.3416 | 0.3401 | 0.3132 | 0.3127 | 0.3121 |
| | | Fourier | 0.2848 | 0.2818 | 0.2850 | 0.1109 | 0.1105 | 0.1100 | 0.0640 | 0.0656 | 0.0663 | 0.0224 | 0.0224 | 0.0224 | 0.0104 | 0.0104 | 0.0100 |
| Discrete | 100 | B&R | 1.0838 | 1.2615 | 1.4540 | 0.6703 | 0.7102 | 0.7537 | 0.5769 | 0.5933 | 0.6000 | 0.4830 | 0.4848 | 0.4894 | 0.4465 | 0.4401 | 0.4474 |
| | | Fourier | 0.7854 | 0.9965 | 1.2264 | 0.3655 | 0.4854 | 0.6180 | 0.2347 | 0.3186 | 0.4098 | 0.0911 | 0.1257 | 0.1628 | 0.0400 | 0.0529 | 0.0686 |
| | 200 | B&R | 1.0064 | 1.1517 | 1.3293 | 0.6107 | 0.6425 | 0.6789 | 0.5077 | 0.5130 | 0.5272 | 0.4022 | 0.4006 | 0.4030 | 0.3800 | 0.3780 | 0.3796 |
| | | Fourier | 0.6779 | 0.9027 | 1.1426 | 0.3237 | 0.4551 | 0.5925 | 0.2119 | 0.3036 | 0.3960 | 0.0843 | 0.1208 | 0.1587 | 0.1122 | 0.0500 | 0.0656 |
| | 400 | B&R | 0.9444 | 1.0658 | 1.2004 | 0.5613 | 0.5858 | 0.6167 | 0.4469 | 0.4591 | 0.4681 | 0.3410 | 0.3434 | 0.3428 | 0.3134 | 0.3130 | 0.3124 |
| | | Fourier | 0.6053 | 0.8492 | 1.1063 | 0.3055 | 0.4411 | 0.5813 | 0.2030 | 0.2960 | 0.3905 | 0.0806 | 0.1183 | 0.1568 | 0.0332 | 0.0490 | 0.0648 |

All RMSE's are the presented value times $10^{-2}$. The number of MC replications is 1000.

**Table 3.2**

Median RMSE of One-Step-Ahead Prediction

| | T | Method | $m = 20$ $\alpha_1 = 0.2$ | 0.3 | 0.4 | 40 0.2 | 0.3 | 0.4 | 60 0.2 | 0.3 | 0.4 | 150 0.2 | 0.3 | 0.4 | 365 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Exp Decline | 100 | B&R | 0.2106 | 0.2076 | 0.2107 | 0.1990 | 0.2001 | 0.2000 | 0.1967 | 0.1984 | 0.2021 | 0.2424 | 0.2320 | 0.2381 | 0.3187 | 0.3186 | 0.3220 |
| | | Fourier | 0.1358 | 0.1366 | 0.1367 | 0.1375 | 0.1380 | 0.1378 | 0.1380 | 0.1366 | 0.1409 | 0.1515 | 0.1630 | 0.1880 | 0.2598 | 0.3504 | 0.4526 |
| | 200 | B&R | 0.2044 | 0.2056 | 0.2062 | 0.1888 | 0.1887 | 0.1881 | 0.1848 | 0.1842 | 0.1830 | 0.1897 | 0.1890 | 0.1895 | 0.2602 | 0.2631 | 0.2603 |
| | | Fourier | 0.1310 | 0.1319 | 0.1303 | 0.1317 | 0.1295 | 0.1296 | 0.1316 | 0.1309 | 0.1310 | 0.1343 | 0.1370 | 0.1421 | 0.2457 | 0.3346 | 0.4289 |
| | 400 | B&R | 0.2015 | 0.2028 | 0.2032 | 0.1837 | 0.1857 | 0.1857 | 0.1760 | 0.1757 | 0.1743 | 0.1680 | 0.1684 | 0.1681 | 0.1926 | 0.1923 | 0.1924 |
| | | Fourier | 0.1280 | 0.1283 | 0.1277 | 0.1280 | 0.1289 | 0.1282 | 0.1276 | 0.1280 | 0.1283 | 0.1316 | 0.1349 | 0.1400 | 0.2376 | 0.3257 | 0.4220 |
| Hump Shaped | 100 | B&R | 0.2105 | 0.2077 | 0.2107 | 0.1990 | 0.2001 | 0.2001 | 0.1968 | 0.1984 | 0.2021 | 0.2423 | 0.2320 | 0.2381 | 0.3181 | 0.3189 | 0.3228 |
| | | Fourier | 0.1358 | 0.1365 | 0.1367 | 0.1374 | 0.1379 | 0.1374 | 0.1376 | 0.1363 | 0.1397 | 0.1380 | 0.1360 | 0.1371 | 0.1379 | 0.1385 | 0.1416 |
| | 200 | B&R | 0.2045 | 0.2057 | 0.2063 | 0.1887 | 0.1884 | 0.1880 | 0.1850 | 0.1841 | 0.1830 | 0.1897 | 0.1890 | 0.1895 | 0.2601 | 0.2631 | 0.2601 |
| | | Fourier | 0.1310 | 0.1320 | 0.1302 | 0.1317 | 0.1294 | 0.1296 | 0.1314 | 0.1309 | 0.1305 | 0.1311 | 0.1312 | 0.1298 | 0.1296 | 0.1324 | 0.1326 |
| | 400 | B&R | 0.2017 | 0.2028 | 0.2031 | 0.1837 | 0.1859 | 0.1856 | 0.1757 | 0.1756 | 0.1743 | 0.1680 | 0.1684 | 0.1681 | 0.1926 | 0.1923 | 0.1925 |
| | | Fourier | 0.1279 | 0.1283 | 0.1277 | 0.1279 | 0.1287 | 0.1278 | 0.1273 | 0.1274 | 0.1273 | 0.1281 | 0.1279 | 0.1278 | 0.1275 | 0.1285 | 0.1294 |
| Linear Decline | 100 | B&R | 0.2106 | 0.2076 | 0.2107 | 0.1991 | 0.2001 | 0.2002 | 0.1974 | 0.1986 | 0.2021 | 0.2423 | 0.2320 | 0.2381 | 0.3181 | 0.3189 | 0.3228 |
| | | Fourier | 0.1359 | 0.1366 | 0.1368 | 0.1375 | 0.1378 | 0.1377 | 0.1377 | 0.1360 | 0.1402 | 0.1380 | 0.1362 | 0.1376 | 0.1378 | 0.1377 | 0.1405 |
| | 200 | B&R | 0.2045 | 0.2056 | 0.2062 | 0.1887 | 0.1883 | 0.1881 | 0.1851 | 0.1844 | 0.1830 | 0.1897 | 0.1890 | 0.1895 | 0.2601 | 0.2631 | 0.2183 |
| | | Fourier | 0.1311 | 0.1319 | 0.1304 | 0.1317 | 0.1296 | 0.1296 | 0.1313 | 0.1309 | 0.1306 | 0.1314 | 0.1313 | 0.1300 | 0.1292 | 0.1307 | 0.1314 |
| | 400 | B&R | 0.2015 | 0.2028 | 0.2073 | 0.1838 | 0.1857 | 0.1856 | 0.1757 | 0.1757 | 0.1743 | 0.1680 | 0.1684 | 0.1681 | 0.1926 | 0.1923 | 0.1925 |
| | | Fourier | 0.1281 | 0.1284 | 0.1277 | 0.1280 | 0.1288 | 0.1280 | 0.1274 | 0.1276 | 0.1277 | 0.1281 | 0.1280 | 0.1281 | 0.1271 | 0.1275 | 0.1283 |
| Cyclical | 100 | B&R | 0.2102 | 0.2078 | 0.2119 | 0.1990 | 0.1999 | 0.2000 | 0.1971 | 0.1983 | 0.2021 | 0.2423 | 0.2320 | 0.2381 | 0.3181 | 0.3189 | 0.3228 |
| | | Fourier | 0.1353 | 0.1367 | 0.1368 | 0.1376 | 0.1381 | 0.1371 | 0.1375 | 0.1367 | 0.1398 | 0.1380 | 0.1362 | 0.1373 | 0.1380 | 0.1374 | 0.1395 |
| | 200 | B&R | 0.2035 | 0.2061 | 0.2063 | 0.1890 | 0.1890 | 0.1880 | 0.1849 | 0.1842 | 0.1832 | 0.1897 | 0.1890 | 0.1890 | 0.2601 | 0.2631 | 0.2601 |
| | | Fourier | 0.1310 | 0.1320 | 0.1301 | 0.1315 | 0.1295 | 0.1297 | 0.1314 | 0.1310 | 0.1305 | 0.1310 | 0.1309 | 0.1309 | 0.1293 | 0.1307 | 0.1308 |
| | 400 | B&R | 0.2017 | 0.2026 | 0.2029 | 0.1839 | 0.1859 | 0.1856 | 0.1758 | 0.1756 | 0.1745 | 0.1680 | 0.1684 | 0.1684 | 0.1926 | 0.1923 | 0.1925 |
| | | Fourier | 0.1280 | 0.1282 | 0.1278 | 0.1278 | 0.1286 | 0.1278 | 0.1273 | 0.1273 | 0.1274 | 0.1281 | 0.1278 | 0.1278 | 0.1270 | 0.1276 | 0.1276 |
| Discrete | 100 | B&R | 0.2113 | 0.2116 | 0.2140 | 0.1998 | 0.2011 | 0.2005 | 0.1967 | 0.1989 | 0.2029 | 0.2420 | 0.2325 | 0.2382 | 0.3177 | 0.3188 | 0.3236 |
| | | Fourier | 0.1365 | 0.1406 | 0.1427 | 0.1400 | 0.1432 | 0.1447 | 0.1409 | 0.1437 | 0.1507 | 0.1442 | 0.1479 | 0.1566 | 0.1433 | 0.1512 | 0.1626 |
| | 200 | B&R | 0.2039 | 0.2083 | 0.2090 | 0.1891 | 0.1898 | 0.1890 | 0.1853 | 0.1845 | 0.1836 | 0.1899 | 0.1888 | 0.1894 | 0.2603 | 0.2633 | 0.2604 |
| | | Fourier | 0.1320 | 0.1349 | 0.1354 | 0.1335 | 0.1338 | 0.1368 | 0.1336 | 0.1367 | 0.1411 | 0.1361 | 0.1409 | 0.1498 | 0.1360 | 0.1452 | 0.1532 |
| | 400 | B&R | 0.2023 | 0.2039 | 0.2049 | 0.1844 | 0.1861 | 0.1857 | 0.1758 | 0.1757 | 0.1750 | 0.1680 | 0.1682 | 0.1682 | 0.1927 | 0.1923 | 0.1923 |
| | | Fourier | 0.1290 | 0.1312 | 0.1317 | 0.1296 | 0.1331 | 0.1355 | 0.1300 | 0.1328 | 0.1371 | 0.1329 | 0.1387 | 0.1464 | 0.1320 | 0.1411 | 0.1511 |

The number of MC simulation is 250.

In general, Fourier approximation presents a better performance in both RMSE of the parameter and the one-step-ahead forecast. First, in Table 3.1, estimation accuracy increases as the frequency ratio become larger using either approach. In exponential decline, hump-shaped and linear decline cases, Fourier approximation improves accuracy compared with the nonparametric method substantially. The improvement is similar in exponential Almon lag and hump-shaped cases and becomes considerable when the sample size or the frequency ratio is enlarged. Fourier approximation captures the flexibility of two-parameter exponential Almon lag more precisely than the nonparametric approach. Expectantly, in the linear decline case, Fourier approximation provides a much more accurate estimation with enlarged frequency ratios. Fourier approximation contains a linear term so that it could estimate the linear pattern better. However, in the cyclical case, Fourier expansion keeps performing similar median RMSE with different sample sizes. The nonparametric approach outperforms Fourier approximation and provides more accurate estimations when the sample size increases. Even though Fourier approximation contains trigonometric terms, the number of parameters in Fourier expansion may have an impact on the estimation accuracy. An appropriate choice of the number of parameters is crucial. Second, in Table 3.2, we present the median RMSE of the one-step-ahead forecast. Even though both two methods perform more accurate forecast as the sample size or the frequency ratio increases in all five cases, Fourier approximation is still superior to the nonparametric estimation slightly. Among five cases, Fourier approximation provides a much more precise forecast in linear decline case with large frequency ratios.

## 3.3 Panel Nonparametric MIDAS

In Section 3.2.1, we have introduced Fourier approximation in MIDAS models. The overall performance of Fourier approximation is laudable in general. Given the complexity of panel data, Fourier approximation could be a wise choice, especially when frequency ratios of distinct subjects are not consistent, or the ratios are significantly large, for example, 365 for daily vs yearly data. Fourier approximation transforms inconsistent frequency ratios to a fixed, predetermined small number, which could reduce the computational complexity efficiently. Given the performance of introducing Fourier series expansion in MIDAS models presented in Section 3.2, we extend the nonparametric model to a cross-subsectional data.

### 3.3.1 Panel MIDAS with Fourier Transformation

Suppose there are $n$ subjects. For simplicity, we assume that all subjects have the same sample size $T$ and frequency ratio $m$. The arguments in this chapter should still be suitable with different sample sizes and frequency ratios for different subjects at the expense of more complicated notations and slight changes in the results. See Remark 3.2.

For the $i$-th subject, let $\mathbf{u}_{i,t}$ be the $q$-vector of covariates including the intercept at time $t$, $t = 1, \ldots, T$, and $\boldsymbol{\alpha}_i$ be the corresponding coefficient. Consider the following MIDAS model with the lead $h \geq 0$:

$$y_{i,t+h} = \mathbf{u}'_{i,t}\boldsymbol{\alpha}_i + \mathbf{x}'_{i,t}\boldsymbol{\beta}^*_i + \varepsilon_{i,t+h}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, n, \tag{3.9}$$

or

$$\mathbf{y}_i = U_i\boldsymbol{\alpha}_i + X_i\boldsymbol{\beta}^*_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, n, \tag{3.10}$$

where $\mathbf{y}_i = (y_{i,1+h}, \ldots, y_{i,T+h})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1+h}, \ldots, \varepsilon_{i,T+h})'$, $\boldsymbol{\beta}^*_i = (\beta^*_{i,0}, \ldots, \beta^*_{i,m-1})'$. $\varepsilon_{i,t+h}$ is the error process for the $i$-th subject. $X_i$ is a $T \times m$ matrix with the $t$-th row being $\mathbf{x}'_{i,t} = (x_{i,t,0}, x_{i,t,1}, \ldots, x_{i,t,m-1})$, and $U_i$ is a $T \times q$ matrix with the $t$-th row being $\mathbf{u}'_{i,t} = (u_{i,t,1}, \ldots, u_{i,t,q})$. With basically the same formulation of single-subject MIDAS model, the MIDAS model (3.10) is an extension of (3.1) with panel data.

Consider smoothing the MIDAS weight vector $\boldsymbol{\beta}^*_i$ using the Fourier approximation. For each subject $i = 1, \ldots, n$, define Fourier transformed HF variables $\widetilde{X}_i = X_i M'$, where $M$ is the same as the transformation matrix in Section 3.2.1. For all $i$, $X_i\boldsymbol{\beta}^*_i \approx \widetilde{X}_i\boldsymbol{\beta}_i$, as along as $L$ and $K$ are large enough and the underlying MIDAS weight functions $\boldsymbol{\beta}^*_i(\cdot)$ satisfy the Dirichlet conditions.

Let $W_i = (U_i, \widetilde{X}_i)$ and $\boldsymbol{\gamma}_i = (\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i')'$. The equation (3.10) can be rewritten as

$$\mathbf{y}_i = (U_i, X_i) \begin{pmatrix} \boldsymbol{\alpha}_i \\ \\ \boldsymbol{\beta}_i^* \end{pmatrix} + \boldsymbol{\varepsilon}_i \approx (U_i, \widetilde{X}_i) \begin{pmatrix} \boldsymbol{\alpha}_i \\ \\ \boldsymbol{\beta}_i \end{pmatrix} + \boldsymbol{\varepsilon}_i = W_i \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \tag{3.11}$$

Concatenating $\mathbf{y}_i$ in (3.11) into $\mathbf{y}$, a vector of length $nT$,

$$\mathbf{y} \approx W \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{3.12}$$

where $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')'$, $W = \text{diag}(W_1, \ldots, W_n)$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_n')'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_n')'$. Let $p = q + 2K + L + 1$. In our formulation, $\boldsymbol{\gamma}_i$ is a vector of length $p$ and $\boldsymbol{\gamma}$ is of length $np$.

**Remark 3.2.** Allowing different sample sizes and frequency ratios for different subjects can be done at the expense of complicity in notations. The major complication arises from the need to use different $M_i$ for each $i$ in $\widetilde{X}_i = X_i M_i'$, where $m_i$ replaces $m$ in (3.4) for the $i$-th state. $\mathbf{y}$ is a vector of length $\sum_{i=1}^{n} T_i$ rather than $nT$. As this makes the notations for the subsequent proofs more complicated without adding fundamental differences, we do not pursue this generalization at the current stage. On the other hand, we should use the same $L$ and $K$ for all subjects $i = 1, \ldots, n$, unlike the case for $T$ or $m$. This is because it is necessary to compare $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$ directly and their dimensions need to be matched. It can be interpreted that although different subjects may have different degrees of HF information, they need to be eventually matched after smoothing to compare different subjects.

Now, we introduce the estimation of parameters in (3.12) if the subjects can be separated into a small number of groups. Denote the number of groups as $G$. The advantage of the proposed procedure is that it does not require any prior knowledge of group information or the number of groups. The only information required is to identify group-specific parameters. Here, we focus on the case where all elements in $\boldsymbol{\gamma}_i$ are the same within a group. It is possible to relax this assumption by letting some of $\boldsymbol{\gamma}_i$ be individual-specific, rather than assuming all parameters are strongly tied with groups. It is also possible to extend the model further by allowing for more than one HF variables. See Remark 3.3 for a brief discussion on these two extensions.

The OLS solution of (3.12) is $\boldsymbol{\gamma}$ that minimizes

$$\frac{1}{2}||\mathbf{y} - W\boldsymbol{\gamma}||_2^2. \tag{3.13}$$

However, the OLS estimator of $\boldsymbol{\gamma}$ would not reflect the relevant group information. We propose a panelized regression method to force all elements in $\boldsymbol{\gamma}_i$ to have similar values within a group. Our method is based on the observation that if two subjects $i$ and $j$ belong to the same group, the difference of their group-specific parameter would be zero, i.e., $\boldsymbol{\eta}_{ij} = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j = \mathbf{0}$. Under the circumstances, the OLS estimator of $\boldsymbol{\eta}_{ij}$ would also be somewhat close to a zero vector, though it would not be exactly zero. Nevertheless, since subject $i$ and $j$ are in the same group, $\boldsymbol{\eta}_{ij}$ should better estimated to be exactly zero, rather than somewhat close to zero. This can be forced by imposing a penalty for small values of $\boldsymbol{\eta}_{ij}$. In particular, if the number of groups $N$ is much smaller than the number of subjects $n$, only a small number of $\boldsymbol{\eta}_{ij}$ would be nonzero. Therefore, we consider the following penalized objective function:

$$Q(\boldsymbol{\gamma}) = \frac{1}{2}||\mathbf{y} - W\boldsymbol{\gamma}||_2^2 + \sum_{1 \le i < j \le n} \rho(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j, \lambda_1), \tag{3.14}$$

where $\rho(\cdot, \cdot)$ is an appropriate penalty function and $\lambda_1$ is the tuning parameter. By introducing $\boldsymbol{\eta}_{ij} = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j$, minimizing (3.14) is equivalent to minimizing

$$Q(\boldsymbol{\gamma}, \boldsymbol{\eta}) = \frac{1}{2}||\mathbf{y} - W\boldsymbol{\gamma}||_2^2 + \sum_{1 \le i < j \le n} \rho(\boldsymbol{\eta}_{ij}, \lambda_1) \ \text{ subject to } \ \boldsymbol{\eta}_{ij} = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j, \tag{3.15}$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}'_{12}, \ldots, \boldsymbol{\eta}'_{n-1,n})'$. Following Boyd et al. [10], we solve this constrained optimization problem using a variant of the augmented Lagrangian

$$\begin{aligned} Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi}) =& \frac{1}{2}||\mathbf{y} - W\boldsymbol{\gamma}||_2^2 + \sum_{i<j} \rho(\boldsymbol{\eta}_{ij}, \lambda_1) \\ &+ \frac{\lambda_2}{2} \sum_{i<j} ||\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j - \boldsymbol{\eta}_{ij}||_2^2 + \sum_{i<j} \boldsymbol{\xi}'_{ij}(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j - \boldsymbol{\eta}_{ij}), \end{aligned} \tag{3.16}$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}'_{12}, \boldsymbol{\xi}'_{13}, \ldots, \boldsymbol{\xi}'_{n-1,n})'$ and $\boldsymbol{\xi}_{ij}$ are $p$-vectors of Lagrangian multipliers. As proposed in Boyd et al. [10], the optimazition problem in (3.16) can be solved using the ADMM algorithm.

Refer to the algorithm, at the $(s+1)$-th step, estimated parameters $\boldsymbol{\gamma}^{s+1}$, $\boldsymbol{\eta}^{s+1}$ and $\boldsymbol{\xi}^{s+1}$ should be updated as

$$\begin{cases} \boldsymbol{\gamma}^{s+1} = \arg\min_{\boldsymbol{\gamma}} Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}^s, \boldsymbol{\xi}^s), \\ \boldsymbol{\eta}^{s+1} = \arg\min_{\boldsymbol{\eta}} Q_{\lambda_2}(\boldsymbol{\gamma}^{s+1}, \boldsymbol{\eta}, \boldsymbol{\xi}^s), \\ \boldsymbol{\xi}_{ij}^{s+1} = \boldsymbol{\xi}_{ij}^s + \lambda_2(\boldsymbol{\eta}_{ij}^{s+1} - \boldsymbol{\gamma}_i^{s+1} + \boldsymbol{\gamma}_j^{s+1}), \end{cases} \tag{3.17}$$

where $\boldsymbol{\eta}^s$ and $\boldsymbol{\xi}^s$ are the estimates in the $s$-th iteration.

By gathering terms only related to $\boldsymbol{\gamma}$, the first function in (3.17) is equivalent to minimizing

$$Q_{\lambda_2}^{\gamma}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{y} - W\boldsymbol{\gamma}\|_2^2 + \frac{\lambda_2}{2}\|D\boldsymbol{\gamma} - (\boldsymbol{\eta} + \boldsymbol{\xi}/\lambda_2)\|_2^2, \tag{3.18}$$

where $D_{ij} = (\boldsymbol{e}_i - \boldsymbol{e}_j)' \otimes I_p$ and $D = (D'_{12}, D'_{13}, \cdots, D'_{n-1,n})'$. $\boldsymbol{e}_i$ is an $n$-dimension vector with the $i$-th element as one and the rest as zeros. $I_p$ is an identity matrix with rank $p$. Therefore, $\boldsymbol{\gamma}^{s+1} = (W'W + \lambda_2 D'D)^{-1} (W'\mathbf{y} + \lambda_2 D'(\boldsymbol{\eta}^s + \boldsymbol{\xi}^s/\lambda_2))$.

The MCP is shown to be nearly unbiased and is applicable here to update $\boldsymbol{\eta}^{s+1}$ [74]. The penalty function of the MCP is $\rho(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j, \lambda_1) = \rho_\theta(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2, \lambda_1)$ where $\rho_\theta(a, b) = b\int_0^a (1 - \frac{u}{\theta b})_+ du$. As a consequence, when the MCP is selected, $\boldsymbol{\eta}_{ij}^{s+1}$ can be updated by

$$\boldsymbol{\eta}_{ij}^{s+1} = \begin{cases} \tilde{\boldsymbol{\eta}}_{ij}^{s+1}, & \text{if } \|\tilde{\boldsymbol{\eta}}_{ij}^{s+1}\|_2 \geq \theta\lambda_1, \\ \dfrac{\theta\lambda_2}{\theta\lambda_2 - 1}\left(1 - \dfrac{\lambda_1/\lambda_2}{\|\tilde{\boldsymbol{\eta}}_{ij}^{s+1}\|_2}\right)_+ \tilde{\boldsymbol{\eta}}_{ij}^{s+1}, & \text{if } \|\tilde{\boldsymbol{\eta}}_{ij}^{s+1}\|_2 < \theta\lambda_1, \end{cases} \tag{3.19}$$

where $\tilde{\boldsymbol{\eta}}_{ij}^{s+1} = \boldsymbol{\gamma}_i^{s+1} - \boldsymbol{\gamma}_j^{s+1} - \boldsymbol{\xi}_{ij}^{s}/\lambda_2$ and $\theta > 1/\lambda_2$ for the global convexity of the second minimization function in (3.17) [69].

If the minimization function of $\boldsymbol{\eta}^{s+1}$ is non-convex, assigning appropriate initial values becomes essential. A proper start leads to an ideal solution. Inspired by Zhu and Qu [74], we would summarize the whole algorithm in Algorithm 3.

**Algorithm 3:** The Clustering Algorithm: Fourier Transformed Data

**Initialization:**

$\boldsymbol{\xi}^0 = \mathbf{0}$, $\boldsymbol{\gamma}^0 = (W'W)^{-1}(W'\mathbf{y})$, $\boldsymbol{\eta}^0 = \arg\min_{\boldsymbol{\eta}} Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi})$, where $\lambda_2$ and $\theta > 1/\lambda_2$ are fixed.

**for** $s = 0, 1, 2, \cdots$ **do**

  $\boldsymbol{\gamma}^{s+1} = (W'W + \lambda_2 D'D)^{-1}(W'\mathbf{y} + \lambda_2 D'\tilde{\boldsymbol{\eta}}^s)$.

  $\boldsymbol{\eta}^{s+1} = \arg\min_{\boldsymbol{\eta}} Q_{\lambda_2}(\boldsymbol{\gamma}^{s+1}, \boldsymbol{\eta}, \boldsymbol{\xi}^s)$,

  $\boldsymbol{\xi}_{ij}^{s+1} = \boldsymbol{\xi}_{ij}^{s} + \lambda_2(\boldsymbol{\eta}_{ij}^{s+1} - \boldsymbol{\gamma}_i^{s+1} + \boldsymbol{\gamma}_j^{s+1})$, for all $1 \leq i < j \leq n$.

  **if** *the stopping criteria are true* **then**

    Break

  **end**

**end**

The tuning parameter $\lambda_1$ is chosen by minimizing

$$BIC_{\lambda_1} = \log\left(\frac{\|\mathbf{y} - W\widehat{\boldsymbol{\gamma}}\|_2^2}{n}\right) + \frac{\log(n) \cdot \left(\widehat{G}p\right)}{n}. \tag{3.20}$$

The estimated number of groups, $\widehat{G}$, can be obtained by $\boldsymbol{\eta}$. We expected to have

$\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$ in the same cluster if $\widehat{\gamma}_i = \widehat{\gamma}_j$. However, as a penalty $\boldsymbol{\eta}_{ij}$ has been imposed in the clustering algorithm, the equality of two estimated parameters are not achievable. As a result, the MCP penalty is utilized on $\widehat{\boldsymbol{\eta}}_{ij}$. Two parameters $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$ are clustered in the same group if $\widehat{\boldsymbol{\eta}}_{ij} = \mathbf{0}$. Only if the tuning parameter $\lambda_1$ is given, $\widehat{G}$ and the estimated coefficients $\widehat{\gamma}$ can be evaluated. Hence, we assign different values to $\lambda_1$ and calculate the corresponding BIC's shown in (3.20). $\lambda_1$ is selected when BIC reaches the minimum.

In Algorithm 3, let $\boldsymbol{\kappa}_{ij}^{s+1} = \boldsymbol{\gamma}_i^{s+1} - \boldsymbol{\gamma}_j^{s+1} - \boldsymbol{\eta}_{ij}^{s+1}$, $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_{12}', \cdots, \boldsymbol{\kappa}_{n-1,n}')'$ and $\boldsymbol{\tau}_k^{s+1} = -\lambda_2 \left( \sum_{i=k} (\boldsymbol{\eta}_{ij}^{s+1} - \boldsymbol{\eta}_{ij}^{s}) - \sum_{j=k} (\boldsymbol{\eta}_{ij}^{s+1} - \boldsymbol{\eta}_{ij}^{s}) \right)$, $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_n)'$. At any step $s^*$, if for some small values $\epsilon^\kappa$ and $\epsilon^\tau$, $\|\boldsymbol{\kappa}^{s^*}\|_2 \leq \epsilon^\kappa$ and $\|\boldsymbol{\tau}^{s^*}\|_2 \leq \epsilon^\tau$, the algorithm stops. According to Zhu and Qu [74], $\epsilon^\kappa$ and $\epsilon^\tau$ are defined as

$$\epsilon^\kappa = \sqrt{np}\epsilon^{abs} + \epsilon^{rel} \|D' \boldsymbol{\xi}^{s^*}\|_2, \quad \epsilon^\tau = \sqrt{|\mathcal{I}|p}\epsilon^{abs} + \epsilon^{rel} \max\{\|D\boldsymbol{\eta}^{s^*}\|_2, \|\boldsymbol{\eta}^{s^*}\|_2\},$$

where $\mathcal{I} = \{(i,j) : 1 \leq i < j \leq n\}$, $|\mathcal{I}|$ indicates the cardinality of $\mathcal{I}$. $\epsilon^{abs}$ and $\epsilon^{rel}$ are predetermined small values.

**Theorem 3.1.** *The clustering algorithm ensures convergence, s.t.*

$$\|\boldsymbol{\kappa}^{s+1}\|_2^2 \to 0 \quad \textit{and} \quad \|\boldsymbol{\tau}^{s+1}\|_2^2 \to 0,$$

*as* $s \to \infty$.

The proof of Theorem 3.1 can be found in Appendix A.2.1. Theorem 3.1 demonstrates that the clustering algorithm is convergent as the number of iteration, $s$, approaches infinity. The stopping criteria can be satisfied at some step eventually.

**Remark 3.3.** It is possible to extend the setting to allow for more than one HF variables and subject-specific variables. All coefficients in (3.12) are group-specific. If there are subject-specific coefficients, a similar argument would still work, although some rates and conditions would change. In particular, the number of coefficients that are subject-specific should be added following a similar argument in Ma and Huang [54, 55]. If there are more than one group-specific HF variables, it is enough to stack all corresponding coefficients in $\boldsymbol{\gamma}$.

Next, we show some theoretical properties of the estimators solving the optimization problem in (3.14). Suppose the true group memberships are known. Let the number of groups be $G$. For $g = 1, \ldots, G$, let $\mathcal{G}_g$ be the set of subject indices that corresponds to the $g$-th group. Assume $\mathcal{G}_1, \ldots, \mathcal{G}_G$ are mutually exclusive and $\mathcal{G}_1 \cup \ldots \cup \mathcal{G}_G = \{1, \ldots, n\}$. This means that each subject belongs to exactly one group. Denote $|\mathcal{G}_g|$ to be the number of elements in $\mathcal{G}_g$ for $g = 1, \ldots, G$. Define $g_{\min} = \min_{g=1,\ldots,G} |\mathcal{G}_g|$ and $g_{\max} = \max_{g=1,\ldots,G} |\mathcal{G}_g|$.

Let the true parameter of the $i$-th subject as $\boldsymbol{\gamma}_i^0$, and $\boldsymbol{\varphi}_g^0$ is the true common vector for group $\mathcal{G}_g$. Take $\boldsymbol{\gamma}^0 = (\boldsymbol{\gamma}_1^{0\prime}, \cdots, \boldsymbol{\gamma}_n^{0\prime})'$ and $\boldsymbol{\varphi}^0 = (\boldsymbol{\varphi}_1^{0\prime}, \cdots, \boldsymbol{\varphi}_G^{0\prime})'$. Each $\boldsymbol{\gamma}_i$ is the individual-specific coefficient. Let $\boldsymbol{\varphi}_g$ be the common value for the $\boldsymbol{\gamma}_i$'s from group

$\mathcal{G}_g$, then $\boldsymbol{\gamma}_i = \boldsymbol{\varphi}_g$ for all $i \in \mathcal{G}_g$ and any $g = 1, \cdots, G$. In other words, $\boldsymbol{\varphi}_g$ indicates the $g$-th panel-specific parameter. $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \cdots, \boldsymbol{\varphi}_G')'$. $\widehat{\boldsymbol{\gamma}}$ is the estimated parameters of all subjects, and the estimated panel effects $\widehat{\boldsymbol{\varphi}}_1, \cdots, \widehat{\boldsymbol{\varphi}}_{\widehat{G}}$ are the distinct values of $\widehat{\boldsymbol{\gamma}}$ where $\widehat{G}$ is the estimated number of panels. So denote the estimated group as $\widehat{\mathcal{G}}_g := \{i : \widehat{\boldsymbol{\gamma}}_i = \widehat{\boldsymbol{\varphi}}_g, 1 \leq i \leq n\}$ for $1 \leq g \leq \widehat{G}$. According to Ma and Huang [55], the clustering algorithm allows to get $\widehat{\boldsymbol{\eta}}_{ij} = \mathbf{0}$. Then, $\widehat{\boldsymbol{\varphi}}_g$ would eventually be $\widehat{\boldsymbol{\varphi}}_g = |\widehat{\mathcal{G}}_g|^{-1} \sum_{i \in \widehat{\mathcal{G}}_g} \widehat{\boldsymbol{\gamma}}_i$ for the $g$-th group.

Let $\Pi$ be a $n \times G$ matrix with the $(i, g)$-th element being 1 if $i$-th subject belongs to $g$-th group, and 0 otherwise. Then

$$\boldsymbol{\gamma} = (\Pi \otimes I_p)\boldsymbol{\varphi} = \Gamma \boldsymbol{\varphi}, \tag{3.21}$$

where $\Gamma = (\Pi \otimes I_p)$. Consider an estimator $\widehat{\boldsymbol{\gamma}}^{or}$ of $\boldsymbol{\gamma}^0$. By (3.21), we define an oracle estimator $\widehat{\boldsymbol{\gamma}}^{or} = (\Pi \otimes I_p)\widehat{\boldsymbol{\varphi}}^{or}$. We call this an oracle estimator since it utilizes the knowledge of the true group memberships in $\Pi$, which is infeasible in practice.

For the oracle estimator, We use the OLS estimator $\widehat{\boldsymbol{\gamma}}^{or}$ of $\boldsymbol{\gamma}^0$

$$\widehat{\boldsymbol{\gamma}}^{or} = (W'W)^{-1}W'\mathbf{y}, \quad \widehat{\boldsymbol{\varphi}}^{or} = (\Gamma'W'W\Gamma)^{-1}\Gamma'W'\mathbf{y}, \tag{3.22}$$

assuming that $\Gamma'W'W\Gamma$ is invertible. This is the case as we assume $n \ll G$. Using this OLS estimator and (3.21), the oracle estimator of $\boldsymbol{\gamma}$ is

$$\widehat{\boldsymbol{\gamma}}^{or} = \Gamma\widehat{\boldsymbol{\varphi}}^{or} = \Gamma(\Gamma' W' W \Gamma)^{-1}\Gamma' W' \mathbf{y}. \tag{3.23}$$

Before introducing the theoretical properties, we formally organize the assumptions.

**Assumption 3.3.** *There are $G$ distinct functions $\beta_g^*(\cdot)$ that satisfy the conditions in Assumption 3.1. In particular, for any $r \in [0,1]$, $\beta_{i,\lfloor rm \rfloor}^* \to \beta_g^*(r)$ as $m \to \infty$ for all $i \in \mathcal{G}_g$.*

**Assumption 3.4.** *We also assume that the number of clusters is much smaller than the number of subjects, i.e., $G \ll n$.*

**Assumption 3.5.** *Assume $\lambda_{\min}(\sum_{i\in\mathcal{G}_g} W_i'W_i) \geq c|\mathcal{G}_g|T$, $\lambda_{\max}(\sum_{i\in\mathcal{G}_g} W_i'W_i) \leq c'nT$, $\max_{1\leq i\leq n} \lambda_{\max}(W_i'W_i) \leq c''T$ and $\lambda_{\max}(\Gamma'W'W\Gamma) \leq c^*|\mathcal{G}_g|T$ for some constant $c$, $c'$, $c''$ and $c^*$ that does not depend on $g = 1, \ldots, G$. In addition, We further assume that for any $\epsilon > 0$, there exist $0 < M_1, \ldots, M_4 < \infty$ such that*

$$P\left(\sup_{i=1,\ldots,n} \|U_i'U_i\|_\infty > \sqrt{qT}M_1\right) < \epsilon, \quad P\left(\sup_{i=1,\ldots,n} \|X_i'X_i\|_\infty > \sqrt{mT}M_2\right) < \epsilon,$$
$$P\left(\sup_{i=1,\ldots,n} \|U_i'X_i\|_\infty > \sqrt{mT}M_3\right) < \epsilon, \quad P\left(\sup_{i=1,\ldots,n} \|X_i'U_i\|_\infty > \sqrt{qT}M_4\right) < \epsilon.$$

**Assumption 3.6.** *The penalty function $\rho(t,\lambda)$ is a symmetric, nondecreasing, and concave in $t$ for $t \in [0,\infty)$. Let $\rho(t) = \lambda^{-1}\rho_\theta(t,\lambda)$. There exists a constant $0 < c_\rho < \infty$ such that $\rho(t)$ is a constant for all $t \geq a\lambda$. $\rho(t)$ is differentiable and $\rho'(t)$ is continuous* *except for a finite number of* $t$. $\rho(0) = 0$ *and* $\rho'(0_+) = 1$.

**Assumption 3.7.** *There exists* $\tilde{c} > 0$ *such that*

$$E\left\{\exp\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\nu_{i,t}\varepsilon_{i,t}\right)\right\} \leq \exp\left(\tilde{c}\sum_{i=1}^{n}\sum_{t=1}^{T}\nu_{i,t}^2\right)$$

*for any real numbers* $\nu_{i,t}$ *for* $i = 1, \ldots, n$ *and* $t = 1, \ldots, T$. *Furthermore, assume that* $Var(\varepsilon_{i,t}) = O(\tilde{c})$ *which is independent to* $n$, $G$ *and* $T$.

Assumption 3.3 is required for the feature selection technique that we use in (3.14), as the methods require sparsity. Assumption 3.5 is reasonable considering the usual assumption that the smallest eigenvalue of $W_i'W_i$ is bounded by $cT$ where $T$ is the sample size and $c$ is some constant. This condition can be relaxed allowing different $c_g$ for different groups. In such case, our results would not hold if the number of clusters $G$ grows to infinity. It would still work as long as $G$ is finite by choosing $c = \min_{g=1,\ldots,G} c_g$ in the statement of Theorem 3.2. Moreover, Assumption 3.5 is stated for heterogeneous case. For homogenous model, the only difference that we should assume is that $\lambda_{\min}(\sum_{i\in\mathcal{G}_g} W_i'W_i) \geq cnT$, and $\lambda_{\max}(\Gamma'W'W\Gamma) \leq c^*nT$ for some constant $c$ and $c^*$ that does not depend on $g = 1, \ldots, G$ since $\max|\mathcal{G}_g| = n$. Assumption 3.6 is are adapted from Ma and Huang [55] and is conventional in literature. Assumption 3.7 holds for independent subgaussian vector $\boldsymbol{\varepsilon}$, which is commonly assumed in high dimensional settings. The variance of the sub-Gaussian process is bounded by the parameter $\tilde{c}$. The following theorem provides conditions for the convergence of the oracle estimator $\widehat{\boldsymbol{\gamma}}^{or}$.

**Theorem 3.2.** *If Assumptions 3.3–3.7 hold, then*

$$P(||\widehat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0||_\infty \leq \phi_{n,T,G,\zeta}) \geq 1 - e^{-\iota},$$

*where* $\phi_{n,T,G,\zeta} = \frac{\sqrt{2\tilde{c}}}{c} B_{q,m}^{1/2} \frac{(m\tilde{M}g_{\max})^{1/2}(Gp)^{3/4}}{g_{\min}T^{3/4}}(Gp+2\sqrt{Gp}\sqrt{\zeta}+2\zeta)^{1/2}$. $B_{q,m} = [q^{1/2}+ m^{1/2}(L+1+2K)]^{1/2}$, $\tilde{M} = \max\{M1, M2, M3, M4\}$, $\iota = \min\{\zeta, -\log(\epsilon)\} - \log(2)$ *for* $\epsilon$ *defined in Assumption 3.5.*

*Furthermore, for any vector* $c_n \in \mathbb{R}^{Gp}$ *such that* $\|c_n\|_2 = 1$*, following the Lindeberg-Feller Central Limit Theorem, the asymptotic distribution of* $\hat{\boldsymbol{\gamma}}^{or}$ *is*

$$c_n'(\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0) \to N(0, \sigma_\gamma^2),$$

*where* $\sigma_\gamma^2 = Var(\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0)$.

With an appropriate choice of $\zeta_{n,T,G}$, we can show that the conditions of convergence of the oracle estimator. Note that we fix the frequency ratio $m$ and the number of transformed parameters $p$ for simplification. $g_{min} < n/G$ in all cases in Theorem 3.2, Corollary 3.1 and Theorem 3.3.

**Corollary 3.1.** *The oracle estimator* $\widehat{\boldsymbol{\gamma}}^{or}$ *converges to the true parameter* $\boldsymbol{\gamma}^0$ *in probability under one of the following conditions:*

1. *$n$ is fixed. Let $\zeta = o(T^{3/2})$ as $T \to \infty$;*

2. *$n \to \infty$. Whether $T$ is fixed or $T \to \infty$,*

   (a) *when $G$ is fixed, $g_{min} = O(n^{1/2+\tilde{\alpha}_0})$ for some constant $\tilde{\alpha}_0 < 1/2$, $\zeta = o(n^{2\tilde{\alpha}_0}T^{3/2})$ approaches to infinity;*

   (b) *when $G \to \infty$,*

      i. *suppose $g_{min} = O(n^{7/9+\tilde{\alpha}_1})$ for some constant $\tilde{\alpha}_1 < 2/9$, $\zeta = O(G)$ approaches to infinity;*

      ii. *suppose $g_{min} = O(n^{5/7+\tilde{\alpha}_2})$ for some constant $\tilde{\alpha}_2 < 2/7$, $\zeta = o(n^{7\tilde{\alpha}_2/2}T^{3/2}) \gg G$ approaches to infinity.*

Corollary 3.1 lists the convergent conditions of the oracle estimator for a large enough $\zeta_{n,T,G}$ with respect to different conditions of $n, T$ and $G$. The following theorem indicates that the our estimator $\widehat{\boldsymbol{\gamma}}$ of parameter $\boldsymbol{\gamma}$ converges to the oracle estimator in probability, which further demonstrates that our estimator converges to the true parameter.

**Assumption 3.8.** *The minimal difference of the common values between two panels is*

$$b_{n,T,G} = \min_{i \in \mathcal{G}_g, j \in \mathcal{G}_{g'}, g \neq g'} \|\boldsymbol{\gamma}_i^0 - \boldsymbol{\gamma}_j^0\|_2 = \min_{g \neq g'} \|\boldsymbol{\varphi}_g^0 - \boldsymbol{\varphi}_{g'}^0\|_2 > a\lambda_1 + 2p\phi_{n,T,G},$$

*for some constant $a > 0$.*

Assumption 3.8 limits the minimum difference between the averages of parameters of all groups. In other words, the clustering works appropriately when the difference of pairwise groups is large enough. The following theorem shows that our estimator enjoys oracle property without prior knowledge of true group memberships. $\zeta^*$ is a parameter introduced in the proof of Theorem 3.3.

**Theorem 3.3.** *Suppose Assumption 3.8 holds. Consider the following conditions:*

1. *As $n \to \infty$ with $T$ fixed, suppose that conditions in Theorem 2 are satisfied, $g_{min} \gg (p + 2\sqrt{p} + 2)^{1/2} \max(n, \zeta^*)^{1/2}$. Let $\zeta^* \to \infty$.*

2. *As $T, n \to \infty$. Consider $g_{min} \gg (p+2\sqrt{p}+2)^{1/2} \max(n, \zeta^*)^{1/2} T^{1/4}$. Let $\zeta^* \to \infty$.*

   *(a) Consider $G \to \infty$. Let $\frac{n^{7/13}}{T^{1/13}} \ll g_{min} < n/G$, $\zeta \le G$ and $\zeta \to \infty$.*

   *(b) When $G \ll \zeta \to \infty$*

      *i. When $G$ is fixed, let $g_{min} = O(n^{1/4+\tilde{\alpha}_3})$ for some positive constant $\tilde{\alpha}_3 < 3/4$ and $\zeta = o(n^{4\tilde{\alpha}_3} T^{1/2})$, $\zeta \to \infty$.*

      *ii. When $G \to \infty$, for some positive constant $\tilde{\alpha}_4 < 6/11$, let $g_{min} = O(n^{5/11+\tilde{\alpha}_4})$ and $G \le n/g_{min}$, $\zeta = o(n^{11\tilde{\alpha}_4/2} T^{1/2})$ and $\zeta \to \infty$.*

*Under one of these conditions, for $\lambda_1 \gg p\phi_{n,T,G}$ where $\phi_{n,T,G}$ is given in Theorem 3.2, the local minimizer $\widehat{\gamma}$ of (3.14) is almost surely the same as the oracle estimator $\widehat{\gamma}^{or}$, that is,*

$$P(\widehat{\gamma} = \widehat{\gamma}^{or}) \to 1$$

*as* $nT \to \infty$.

Theorem 3.3 focuses on the second level of convergence. Considering additional conditions except for those listed in Corollary 3.1, our estimator $\widehat{\boldsymbol{\gamma}}$ converges to the oracle estimator $\widehat{\boldsymbol{\gamma}}^{or}$ in probability one.

**Corollary 3.2.** *Suppose that Assumption 3.3–3.7 and Assumption 3.8 hold. Then* $\widehat{\boldsymbol{\gamma}}$ *converges to* $\boldsymbol{\gamma}$ *in distribution given any case of the following conditions:*

1. *as* $n \to \infty$ *with* $T$ *fixed, consider the conditions in Corollary 3.1 under the same circumstance, when* $(p+2\sqrt{p}+2)^{1/2}(\max(n, \zeta^*))^{1/2} \ll g_{min} = O(n^{7/9+\tilde{\alpha}_0}) \leq n/2$, *let* $\zeta^* \to \infty$.

2. *as* $n, T \to \infty$,

    (a) *when* $G$ *is fixed,* $g_{min} = O(n^{1/2+\tilde{\alpha}_4})$ *for some constant* $\tilde{\alpha}_4 < 1/2$, *let* $\zeta = o(\min(n^{1+4\tilde{\alpha}_4}T^{1/2}, n^{2\tilde{\alpha}_4}T^{3/2}))$ *approach to infinity,* $\zeta^* \to \infty$.

    (b) *when* $G \to \infty$,

        i. *suppose* $\max\left(\frac{n^{7/13}}{T^{1/13}}, (p+2\sqrt{p}+2)^{1/2} \max(n, \zeta^*)^{1/2}\right) \ll g_{min} = O(n^{7/9+\tilde{\alpha}_3})$ *for some constant* $\alpha_3 < 2/9$, *let* $\zeta = O(G)$ *approach to infinity,* $\zeta^* \to \infty$;

        ii. *suppose* $g_{min} = O(n^{5/7+\tilde{\alpha}_5})$ *for some constant* $\tilde{\alpha}_5 < 2/7$, *let* $\zeta = o(\min(n^{10/7+11/2\tilde{\alpha}_5}T^{1/2}, n^{7\tilde{\alpha}_5/2}T^{3/2}))$.

Throughout two steps of convergence, with adequately chosen values of parameters, our estimator is shown to be consistent under different circumstances. As $T \to \infty$ with other parameters fixed, the convergence of our estimator to the oracle estimator cannot be guaranteed. The proof of Theorem 3.3 with the heterogeneous case is shown in Appendix A.2.2.2. We present the proof as well as the required conditions of the homogeneous case in Appendix A.2.2.3. So far, we have shown that under some conditions, our estimator converges to the oracle estimator, and the oracle estimator converges to the true parameter as well theoretically. More simulation results will be presented in the following to show the performance empirically and illustrate the robustness of our method compared with other clustering methods.

### 3.3.2 Simulation: Panel MIDAS

Except for the nonparametric MIDAS that we introduced, we shall consider two more clustering approaches as a comparison. One is the cross-sectional extension of the nonparametric MIDAS proposed in Breitung and Roling [12]. The other is proposed in Su et al. [67]. We would present the models and algorithms of their approaches in advance.

#### 3.3.2.1 Comparable Clustering Method

**1. Breitung and Roling [12]: Nonparametric MIDAS**

In (3.10), the MIDAS regression model without Fourier transformation of each subject is

$$\mathbf{y}_i = U_i \boldsymbol{\alpha_i} + X_i \boldsymbol{\beta}_i^* + \boldsymbol{\varepsilon}_i, \quad i = 1, \cdots, n.$$

For more than one subject, we can write the penal MIDAS as

$$\mathbf{y}_i = (U_i, X_i) \begin{pmatrix} \boldsymbol{\alpha_i} \\ \\ \boldsymbol{\beta}_i^* \end{pmatrix} = \widetilde{W}_i \boldsymbol{\gamma}_i^*, \quad \text{or} \quad \mathbf{y} = \widetilde{W} \boldsymbol{\gamma}^* + \boldsymbol{\varepsilon},$$

where $\widetilde{W}_i = (U_i, X_i)$ is the raw observations, $\boldsymbol{\gamma}_i^* = (\boldsymbol{\alpha_i}', \boldsymbol{\beta_i^*}')'$, $\boldsymbol{\gamma}^* = (\gamma_1^{*\prime}, \cdots, \gamma_n^{*\prime})'$.

Refer to the main idea of Breitung and Roling [12], we consider to introduce the cubic smoothing spline penalty which rejects too sharped changes of parameters when estimating the parameter $\boldsymbol{\gamma}^*$ of the raw data. Then, the penalized objective function will be given as

$$Q(\boldsymbol{\gamma}^*) = \frac{1}{2} \|\mathbf{y} - \widetilde{W} \boldsymbol{\gamma}^*\|_2^2 + \frac{1}{2} \theta_{\gamma^*} \boldsymbol{\gamma}^{*\prime} \mathbf{A} \boldsymbol{\gamma}^*, \tag{3.24}$$

where $\theta_{\gamma^*}$ is the predetermined smoothing parameter, $\mathbf{A} = I_n \otimes (A'A)$. $A$ is defined as

$$A_{(m-2)\times m} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}.$$

According to Zhu and Qu [74], solve the constrained optimization function

$$Q_{\lambda_2}(\boldsymbol{\gamma}^*, \boldsymbol{\eta}, \boldsymbol{\xi}) = Q(\boldsymbol{\gamma}^*) + \sum_{i<j} \rho(\boldsymbol{\eta}_{ij}, \lambda_1) + \frac{\lambda_2}{2} \sum_{i<j} ||\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_j^* - \boldsymbol{\eta}_{ij}||_2^2 + \sum_{i<j} \boldsymbol{\xi}_{ij}'(\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_j^* - \boldsymbol{\eta}_{ij}). \tag{3.25}$$

The clustering algorithm of (3.25) is similar to Algorithm 3.

**Algorithm 4:** The Clustering Algorithm: Raw Data And Smooth Penalty

**Initialization:**

$\boldsymbol{\xi}^0 = \mathbf{0}$, $\boldsymbol{\gamma}^0 = \left(\widetilde{W}'\widetilde{W} + \theta_{\gamma^*}\mathbf{A}\right)^{-1}\left(\widetilde{W}'\mathbf{y}\right)$, $\boldsymbol{\eta}^0 = \arg\min_{\boldsymbol{\eta}} Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi})$, where $\lambda_2$ and $\theta > 1/\lambda_2$ are fixed.

**for** $s = 0, 1, 2, \cdots$ **do**

$\boldsymbol{\gamma}^{s+1} = (W'W + \lambda_2 D'D + \theta_{\gamma^*}\mathbf{A})^{-1}(W'\mathbf{y} + \lambda_2 D'\tilde{\boldsymbol{\eta}}^s)$.

$\boldsymbol{\eta}^{s+1} = \arg\min_{\boldsymbol{\eta}} Q_{\lambda_2}(\boldsymbol{\gamma}^{s+1}, \boldsymbol{\eta}, \boldsymbol{\xi}^s)$,

$\boldsymbol{\xi}_{ij}^{s+1} = \boldsymbol{\xi}_{ij}^s + \lambda_2(\boldsymbol{\eta}_{ij}^{s+1} - \boldsymbol{\gamma}_i^{s+1} + \boldsymbol{\gamma}_j^{s+1})$, for all $1 \leq i < j \leq n$.

**if** *the stopping criteria are true* **then**

Break

**end**

**end**

Algorithm 4 follows the same idea of Zhu and Qu [74]. However, in Zhu and Qu [74], the model introduces B-splines to approximate observations, while Algorithm 4 uses all HF regressors. Moreover, an additional tuning parameter, $\theta_{\gamma^*}$, is required to be predetermined. Refer to Breitung and Roling [12], Zhu and Qu [74], we select $\theta_{\gamma^*}$ by minimizing the AIC given by

$$AIC_{\theta_{\gamma^*}} = \sum_{i=1}^{n} \left( \log \left( \frac{\|\mathbf{y}_i - W_i \widehat{\boldsymbol{\gamma}}_i\|_2^2}{T} \right) + \frac{2df_i}{T} \right),$$

where $df_i = \mathrm{tr}\{W_i(W_i'W_i + \theta_{\gamma^*} A'A)^{-1} W_i'\}$. The selection of $\lambda_1$ here, is by minimizing

$$BIC_{\lambda_1} = \log \left( \frac{\|\mathbf{y} - W\widehat{\boldsymbol{\gamma}}\|_2^2}{n} \right) + \frac{\log(n) \left( \widehat{G} + \frac{1}{n} \sum_{i=1}^{n} df_i \right)}{n}.$$

With fixed $\lambda_1$, we can obtain $AIC_{\theta_{\gamma^*}}$ for different values of $\theta_{\gamma^*}$. Then, fix $\theta_{\gamma^*}$ with the minimum BIC, we can calculate $BIC_{\lambda_1}$ based on the determined $\theta_{\gamma^*}$.

**2. Su et al. [67]: PPL Estimation**

As mentioned in the introduction, Su et al. [67] introduced C-Lasso for clusters to identify relatively significant differences between parameters and group averages. The PPL function mentioned in Su et al. [67] is

$$Q(\boldsymbol{\gamma}^*) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \phi(w_{it}; \boldsymbol{\gamma}_i^*, \widehat{\mu}_i(\boldsymbol{\gamma}_i^*)). \tag{3.26}$$

By introducing the group Lasso penalty, the PPL criterion function becomes

$$Q_{G,\lambda_{PPL}} = Q(\boldsymbol{\gamma}^*) + \frac{\lambda_{PPL}}{G} \sum_{i=1}^{N} \prod_{g=1}^{G_0} \|\boldsymbol{\beta}_i - \boldsymbol{\alpha}_g\|_2, \tag{3.27}$$

where $\lambda_{PPL}$ is a tuning parameter. The C-Lasso estimation $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\alpha}}$, respectively. Without any prior knowledge of the true clusters, the PPL C-Lasso estimation requires a predetermination of reasonable maximum value, $G_0$, of groups. An appropriate choice of $(\lambda_{PPL}, G_0)$ can be found by minimizing IC based on all possible values of clusters less than $G_0$ as long as predetermined values of $\lambda_{PPL}$. To start the algorithm, Su et al. [67] suggested a natural initial value as $\widehat{\boldsymbol{\alpha}}_g^{(0)} = 0$ for all $g = 1, \cdots, G_0$ and $\widehat{\boldsymbol{\gamma}}^{*(0)}$ as the QMLE of $\boldsymbol{\gamma}_i^*$ in each subjects. More details can be found in Su et al. [67].

**Algorithm 5:** PPL Algorithm Given $G_0$ and $\lambda_{PPL}$

**Initialization:** $\widehat{\boldsymbol{\alpha}}^{(0)} = (\widehat{\boldsymbol{\alpha}}_1^{(0)}, \cdots, \widehat{\boldsymbol{\alpha}}_{G_0}^{(0)})'$, $\widehat{\boldsymbol{\gamma}}^{*(0)} = (\widehat{\boldsymbol{\gamma}}_1^{*(0)}, \cdots, \widehat{\boldsymbol{\gamma}}_n^{*(0)})'$ s.t. $\sum_{i=1}^n \|\widehat{\boldsymbol{\gamma}}_i^{*(0)} - \widehat{\boldsymbol{\alpha}}_g^{(0)}\| \neq 0$ for all $g = 2, \cdots, G_0$.

**for** $s = 1, 2, \cdots$ **do**

  **for** $g = 1, 2, \cdots G_0$ **do**

    Obtain the estimator $(\widehat{\boldsymbol{\gamma}}^{*(s,G)}, \widehat{\boldsymbol{\alpha}}_g^{(s)})$ of $(\boldsymbol{\gamma}^*, \boldsymbol{\alpha}_g)$ by minimizing the following objective function $Q_{G,\lambda_{PPL}}^{(s,g)}(\boldsymbol{\gamma}^*, \boldsymbol{\alpha}_g)$.

    **if** $g = 1$ **then**

      $Q_{G,\lambda_{PPL}}^{(s,g)}(\boldsymbol{\gamma}^*, \boldsymbol{\alpha}_g) = Q(\boldsymbol{\gamma}^*) + \frac{\lambda_{PPL}}{N} \sum_{i=1}^N \|\boldsymbol{\gamma}_i^* - \boldsymbol{\alpha}_g\| \prod_{k=2}^G \|{\boldsymbol{\gamma}_i}^{*(s-1,k)} - \boldsymbol{\alpha}_k^{(s-1)}\|$ ;

    **else if** $g \neq G$ **then**

      $Q_{G,\lambda_{PPL}}^{(s,g)}(\boldsymbol{\gamma}^*, \boldsymbol{\alpha}_g) = Q(\boldsymbol{\gamma}^*) + \frac{\lambda_{PPL}}{N} \sum_{i=1}^N \|\boldsymbol{\gamma}_i^* - \boldsymbol{\alpha}_g\| \prod_{j=1}^{g-1} \|\widehat{\boldsymbol{\gamma}}_i^{*(s,j)} - \boldsymbol{\alpha}_j^{(s)}\| \prod_{k=g+1}^G \|{\boldsymbol{\gamma}_i}^{*(s-1,k)} - \boldsymbol{\alpha}_k^{(s-1)}\|$;

    **else**

      $Q_{G,\lambda_{PPL}}^{(s,g)}(\boldsymbol{\gamma}^*, \boldsymbol{\alpha}_g) = Q(\boldsymbol{\gamma}^*) + \frac{\lambda_{PPL}}{N} \sum_{i=1}^N \|\boldsymbol{\gamma}_i^* - \boldsymbol{\alpha}_g\| \prod_{k=1}^{G-1} \|\widehat{\boldsymbol{\gamma}}_i^{*(s,k)} - \boldsymbol{\alpha}_k^{(s)}\|$ ;

    **end**

  **end**

  **if** *the stopping criteria are true* **then**

    Break

  **end**

**end**

Su et al. [67] provided a stopping criteria for the algorithm in the supplementary material.

$$\widehat{Q}^{(s-1)}_{G,\lambda_{PPL}} - \widehat{Q}^{(s)}_{G,\lambda_{PPL}} \le \epsilon_{tl} \text{ and } \frac{\sum_{g=1}^{G} \left\| \widehat{\boldsymbol{\alpha}}_g^{(s)} - \widehat{\boldsymbol{\alpha}}_g^{(s-1)} \right\|^2}{\sum_{g=1}^{G} \left\| \widehat{\boldsymbol{\alpha}}_g^{(s-1)} \right\|^2 + 10^{-4}} \le \epsilon_{tl}, \tag{3.28}$$

where $\epsilon_{tl}$ is a predetermined small value indicating the tolerance level.

#### 3.3.2.2 Comparing Criteria and Settings

To show the performance of clustering results, we present the estimated number of groups $\widehat{G}$ and the Rand index [64]. The Rand index is designed to check if two subjects from the same group are still assigned to the same group, while two from different groups are separated. Define true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) as in Table 3.3 for any subject indices $i$ and $j$, $1 \le i < j \le n$. For example, TP indicates the number of pairs of indices $(i, j)$

**Table 3.3**
Confusion Matrix for Clustering

| | | Actual | |
|---|---|---|---|
| | | $i, j \in \mathcal{G}_g$ | $i \in \mathcal{G}_g, j \notin \mathcal{G}_g$ |
| Predict | $i, j \in \widehat{\mathcal{G}}_g$ | TP | FP |
| | $i \in \widehat{\mathcal{G}}_g, j \notin \widehat{\mathcal{G}}_g$ | FN | TN |

that are in the same group and predicted to be in the same group. The Rand index is defined as

$$Rand = \frac{TP + TN}{TP + TN + FP + FN}.$$

However, even though we expect to have a good clustering performance when the Rand index is approaching 1, random labeling independently is still a problem. When the number of clusters in each group is large, it is quite possible to get a large Rand index. For example, if each group contains 100 samples, 99 different clusters are generated in one group and 98 different clusters in another group. TN would be large with different assigned clusters index so that the Rand index increases. In such case, when the number of clusters increases, the Rand index can get close to 1 regardless of the quality of clusters [26]. Nevertheless, such random label assignments would lead to an ARI close to zero or even negative. As a result, we consider the ARI [46] to eliminate the effect of the independent clustering. The adjusted Rand index is defined as

$$ARI = \frac{Rand - E(Rand)}{\max(Rand) - E(Rand)}.$$

Since the ARI is the normalized difference between the Rand index and its expectation, the ARI is expected to be zero for the independent clustering case. The Jaccard Index is also considered as a measure of the accuracy of clustering:

$$Jaccard = \frac{TP}{TP + FP + FN}.$$

As shown in Algorithm 3, $\lambda_2$ and $\theta$ are two parameters that controls the performance of clustering besides $\lambda_1$. In the code example provided by Zhu and Qu [74], $\lambda_2 = 1$ while $\theta = 2$ guarantee the update formula of $\boldsymbol{\eta}^{s+1}$ in (3.17) being a convex function with respect to $\boldsymbol{\eta}_{ij}$ for all combinations of $i$ and $j$. To compare the effect of parameters on the performance of clustering, we restrict two clusters constructed by the exponential decline and the cyclical function presented in subsection 3.2.2. In each cluster, 15 data processes are generated. So, 30 coefficient vectors are clustered, and two groups are expected after clustering. Each data process follows (3.8) shown in subsection 3.2.2. $\lambda_2 = 1$, $\theta \in \{2, 2.5\}$, $\lambda_1 \in \{1, \cdots, 4.5\}$, $\beta_0 = 0$, $T \in \{100, 200, 400\}$, $m = \{20, 40\}$ and $\alpha_1 \in \{0.2, 0.3, 0.4\}$ in this section. Since when $\theta$ exceeds 2.5, most of the clustering performances are almost the same based on the results that we calculated, then we only present results for $\theta = 2$ and 2.5. 200 samples are generated to evaluate the average performance due to the computational complexity of the non-parametric MIDAS in Breitung and Roling [12]. The clustering algorithm was forced to stop at the 5,000-th iterations if the stopping conditions cannot be satisfied before the final iteration. Median RMSE of estimated $\widehat{\boldsymbol{\gamma}}$ is chosen to present the estimation performance. RMSE of estimated $\widehat{\boldsymbol{\gamma}}$ is calculated as the following:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_2^2}.$$

#### 3.3.2.3 Clustering Performance

As we previously highlighted, the nonparametric MIDAS proposed by Breitung and Roling [12] (B&R's method) along with the clustering algorithm requires more tuning parameters. Due to the sensitivity of the clustering results with respect to the chosen values of tuning parameters, it is quite crucial to make appropriate choices. However, in this section, it does take a long time to find an appropriate $\theta_{\gamma^*}$ because of the complexity of calculation and more combinations of tuning parameters. Furthermore, refer to the simulation results presented by Breitung and Roling [12], the choice of $\theta_{\gamma^*}$ is sensitive to the sample size, it is hard to determine a proper range of $\theta_{\gamma^*}$. Based on the range of $\theta_{\gamma^*}$ set in Section 3.2.2, we choose the value within the range $[0, 100]$ using the AIC. Besides, all three indexes, as well as the number of clusters and RMSE's, are the average values of all 200 samples.

In Table 3.4, we set $T = 100$, $m = 20$, $\alpha_1 = 0.4$. When $\theta = 2$, even though B&R's method shows better clustering performance than our method, the RMSE indicates that the estimation of parameters is much worse using B&R nonparametric MIDAS generally. Focusing on our method, we can tell that even the performance becomes better as $\lambda_1$ increases, three indexes keep showing that the performance is not good enough. B&R's method shows better clustering performance as $\lambda_1$ keeps increasing, expecially when $\lambda_1$ exceeds 2.5. Moreover, with the BIC chosen $\lambda_1$, small values of the

**Table 3.4**
Clustering Performance with Different Settings of $\theta$ and $\lambda_1$: 200 MC Samples, $T = 100$, $\alpha_1 = 0.4$, $m = 20$

| $\theta$ | $\lambda_1$ | Method | Rand | ARI | Jaccard | Clusters | RMSE($*10^{-2}$) |
|---|---|---|---|---|---|---|---|
| 2 | 1 | Our | 0.531 | 0.030 | 0.026 | 26.45 | 0.5246 |
| | | B&R | 0.530 | 0.756 | 0.027 | 27.05 | 0.4270 |
| | 1.5 | Our | 0.545 | 0.057 | 0.059 | 23.62 | 0.5741 |
| | | B&R | 0.950 | 0.899 | 0.899 | 3.57 | 0.5984 |
| | 2 | Our | 0.526 | 0.020 | 0.021 | 26.32 | 0.6197 |
| | | B&R | 0.950 | 0.899 | 0.899 | 3.63 | 0.7630 |
| | 2.5 | Our | 0.483 | 0.000 | 0.483 | 1.00 | 0.6620 |
| | | B&R | 0.995 | 0.989 | 0.989 | 2.17 | 0.8954 |
| | 3 | Our | 0.517 | 0.007 | 0.517 | 1.07 | 0.6937 |
| | | B&R | 0.998 | 0.996 | 0.996 | 2.05 | 1.0139 |
| | 3.5 | Our | 0.483 | 0.000 | 0.483 | 1.00 | 0.7408 |
| | | B&R | 0.999 | 0.998 | 0.998 | 2.01 | 1.1296 |
| | 4 | Our | 0.483 | 0.000 | 0.480 | 1.20 | 0.7676 |
| | | B&R | 0.995 | 0.989 | 0.989 | 2.12 | 1.2880 |
| | 4.5 | Our | 0.483 | 0.000 | 0.483 | 1.00 | 0.8055 |
| | | B&R | 0.984 | 0.967 | 0.966 | 2.47 | 1.3130 |
| | $\lambda_{1,BIC}$ | Our=3.157 | 0.498 | 0.029 | 0.497 | 1.06 | 0.7094 |
| | | B&R=2.226 | 0.951 | 0.905 | 0.931 | 2.51 | 0.8385 |
| 2.5 | 1 | Our | 0.671 | 0.326 | 0.319 | 13.52 | 0.5308 |
| | | B&R | 0.962 | 0.924 | 0.922 | 3.19 | 0.4368 |
| | 1.5 | Our | 0.906 | 0.810 | 0.805 | 5.43 | 0.5789 |
| | | B&R | 0.985 | 0.983 | 0.966 | 2.13 | 0.6120 |
| | 2 | Our | 0.968 | 0.935 | 0.933 | 3.00 | 0.6321 |
| | | B&R | 0.998 | 0.996 | 0.995 | 2.06 | 0.7533 |
| | 2.5 | Our | 0.999 | 0.998 | 0.998 | 2.01 | 0.6618 |
| | | B&R | 1.000 | 1.000 | 1.000 | 2.00 | 0.8597 |
| | 3 | Our | 1.000 | 1.000 | 1.000 | 2.00 | 0.6897 |
| | | B&R | 1.000 | 1.000 | 1.000 | 2.00 | 0.9593 |
| | 3.5 | Our | 1.000 | 1.000 | 1.000 | 2.00 | 0.7325 |
| | | B&R | 1.000 | 1.000 | 1.000 | 2.00 | 1.0465 |
| | 4 | Our | 1.000 | 1.000 | 1.000 | 2.00 | 0.7736 |
| | | B&R | 1.000 | 1.000 | 1.000 | 2.00 | 1.1210 |
| | 4.5 | Our | 1.000 | 1.000 | 1.000 | 2.00 | 0.8146 |
| | | B&R | 1.000 | 1.000 | 1.000 | 2.00 | 1.1837 |
| | $\lambda_{1,BIC}$ | Our= 2.190 | 0.998 | 0.996 | 0.996 | 2.05 | 0.6534 |
| | | B&R= 1.107 | 0.994 | 0.987 | 0.987 | 2.18 | 0.4388 |

adjusted Rand index illustrate the lousy performance of clustering with our method. Nevertheless, since all indexes are close to 1, B&R's method outperforms our method when $\theta \geq 2$.

However, when $\theta = 2.5$, with a proper choice of tuning parameter $\lambda_1$, our and B&R's method seems to have similar clustering performance when $\lambda_1$ exceeds 2.5. First of all, the RMSE of estimated parameters with our method is much better than B&R's method, which coincides with the conclusion when $\lambda_1 = 2$. Then, compare the average of clusters and three indexes. With small values of $\lambda_1$ ($\lambda_1 < 2$), our clustering performance is not good based on the adjusted Rand index, even though the performance becomes better as $\lambda_1$ increases. However, as $\lambda_1$ exceeds 1.5, all three indexes show that the performance becomes remarkable, especially when $\lambda_1 > 2$, the number of clusters are really close to the true number of panels. In the meantime, B&R's method shows better performance when $\lambda_1$ is small because of the choice of $\theta_{\gamma^*}$. However, it is hard to set the range for appropriate $\theta_{\gamma^*}$ which is required additionally in B&R's method. Third, comparing to the clustering performance with B&R's method, even though B&R's method has better clustering results when $\lambda_1 \leq 2.5$, our method with the BIC chosen $\lambda_1$ has much better clustering performance. It indicates that the BIC chosen $\lambda_1$ is an appropriate way to choose tuning parameters and our method outperforms B&R's method. Finally, since our method reduces the frequency ratio between HF and LF variables, it is much faster than B&R's method, especially when $m$ becomes large, such as $m = 40$. According to our simulation experience, our method becomes significantly fast when $m$ exceeds 60. In addition, comparing to the estimating performance with single subject shown in Table 3.1, both two methods have similar performance on the RMSE of the estimated parameter. Therefore, with an appropriate choice of $\theta$, our method works much better than B&R's method, based on no matter whether the computational complexity or the number of required tuning parameters given the similar clustering performance.

Apart from the results of one setting in Table 3.4, Table 3.5 shows the clustering performance of all three approaches with respect to different settings. To compare the performance of the estimation, we further include the linear regression (lm) subject by subject. The median RMSE of $\widehat{\boldsymbol{\beta}}$ is calculated with the case $\theta = 2.5$ for our and B&R method, as it intends to result in better grouping results. The frequency ratios $m$ selected in Table 3.5 are 20 and 40 to save workload on B&R's method. Other than that, the sample size $T$ and the scale $\alpha_1$ of weights are the same as what is considered in Section 3.2.2. Su's method is included as an alternative for the comparison of the estimation accuracy. In Su's method, we fix the max number of groups as two for the grid search to save the calculation load since we have prior knowledge of the true number of clusters. However, in practice, it could be a problem with an improperly chosen number. In general, all three clustering methods have correct clustering results with the BIC chosen tuning parameters, so the grouping information is not presented in the table. The accuracy of estimation by two nonparametric methods outperforms Su's clustering approach or the subject-level linear regression,

**Table 3.5**
Median RMSE of Overall Performance for Different Settings: $\theta = 2.5$

| | | | $T$ | | |
|---|---|---|---|---|---|
| $m$ | $\alpha_1$ | method | 100 | 200 | 400 |
| 20 | 0.2 | our | 0.4031 | 0.3466 | 0.3059 |
| | | B&R | 0.3945 | 0.3279 | 0.2261 |
| | | Su | 9.6804 | 8.4929 | 7.9253 |
| | | lm* | 8.2571 | 5.5480 | 3.7683 |
| | 0.3 | our | 0.5163 | 0.4691 | 0.4315 |
| | | B&R | 0.4306 | 0.3531 | 0.2404 |
| | | Su | 7.4175 | 6.8505 | 6.2241 |
| | | lm | 8.2573 | 5.5478 | 3.7685 |
| | 0.4 | our | 0.6392 | 0.5966 | 0.5558 |
| | | B&R | 0.4496 | 0.3663 | 0.2482 |
| | | Su | 5.8207 | 5.4699 | 5.1157 |
| | | lm | 8.2573 | 5.5478 | 3.7685 |
| 40 | 0.2 | our | 0.1587 | 0.1442 | 0.1304 |
| | | B&R | 0.1487 | 0.1103 | 0.1005 |
| | | Su | 8.8707 | 7.5578 | 6.2652 |
| | | lm | 13.7938 | 8.4324 | 5.5765 |
| | 0.3 | our | 0.2152 | 0.2012 | 0.1828 |
| | | B&R | 0.1670 | 0.1221 | 0.0922 |
| | | Su | 7.2194 | 5.8779 | 4.3612 |
| | | lm | 13.7938 | 8.3948 | 5.5765 |
| | 0.4 | our | 0.2744 | 0.2603 | 0.2145 |
| | | B&R | 0.1789 | 0.1364 | 0.0959 |
| | | Su | 5.8455 | 4.8590 | 4.6615 |
| | | lm | 13.7938 | 8.3948 | 5.5765 |

All RMSE are the presented value times $10^{-2}$.
* method without clustering.

even though Su's method and the linear regression tend to become more accurate as the sample size increases. B&R's nonparametric MIDAS seems to have the best performance for all settings. It is reasonable that B&R's method outperforms our method since applying Fourier approximation results in a two-layer estimation of parameters. It may reduce the estimating accuracy in a way. Though our method is not better than B&R's approach comparing the values directly, the accuracy is still acceptable and remarkable.

As the scale $\alpha_1$ is enlarged, the accuracy of estimation by the panel MIDAS with Fourier approximation is decreased. However, as the sample size $T$ or the frequency ratio $m$ increases, our approach tends to have a more accurate estimation on the weights. Same circumstances occur on B&R's and Su's method. Overall, the improvement is more significant for our and B&R's method comparing to their original scale of RMSE. Even though Su's approach would not perform better on the accuracy aspect, it is notable that our nonparametric and Su's method have similar computing time, while B&R's method tends to run triple or even much longer. As a balance of the computing time and estimated accuracy, our approach seems to be the best choice among all three approaches.

#### 3.3.2.4 One-Step-Ahead Forecast with Clustering

In reality, grouping subjects with respect to their weight function may not be the only thing that we are interested in. Apart from the clustering performance and the estimation of parameters, it is also interesting to explore the performance of one-step-ahead forecast with the help of clustering. To see more general behavior of the forecast with clustering, we shall consider all possible pairs of weight functions introduced in subsection 3.2.2. Given the computational complexity of the forecast along with the clustering algorithm, we only generate 250 samples in MC simulation. Furthermore, all possible settings are included in this subsection, i.e. $p \in \{20, 40\}$, $T \in \{100, 200, 400\}$ and $\alpha_1 \in \{0.2, 0.3, 0.4\}$. The RMSE of the predicted variable is calculated by

$$RMSE = \sqrt{\frac{1}{nT/2}\sum_{k=1}^{T/2}\sum_{j=1}^{n}(\widehat{y}_{j,T/2+h+k} - y_{j,T/2+h+k})^2}.$$

Moreover, the linear regression (lm) is included in the forecast as well, calculating subject by subject. Similarly to what we have done in the previous part, the following table will present the median of all RMSE of 250 samples.

Table 3.6 shows the one-step-ahead forecast performance of all three clustering methods and linear regression calculated subject by subject. For our and B&R's method, $\theta = 2.5$ is set for comparison. Sample sizes, frequency ratios and the scale $\lambda_1$ are chosen in the same way as in Section 3.3.2.3. It is notable that the subject-level linear regression method outperforms all the clustering approaches. The estimated parameters may contain possible group information after grouping using these clustering approaches. When comparing the median RMSE of one-step-ahead forecast, the linear regression does not require the group information which may affect the estimated parameters. It may be the reason that the subject-level linear regression has more accurate prediction on the one-step-ahead forecast. Comparing the accuracy among clustering approaches generally, our method tends to have more accurate forecasting

**Table 3.6**
Overall One-Step-Ahead Forecast for Different Settings: $\theta = 2.5$

| $m$ | $\alpha_1$ | method | T 100 | 200 | 400 |
|---|---|---|---|---|---|
| 20 | 0.2 | our | 0.7700 | 0.7437 | 0.7164 |
| | | B&R | 0.9942 | 0.7925 | 0.7214 |
| | | Su | 2.5779 | 2.6228 | 2.7425 |
| | | lm* | 0.1619 | 0.1401 | 0.1319 |
| | 0.3 | our | 0.7911 | 0.7591 | 0.7192 |
| | | B&R | 0.9937 | 0.8214 | 0.7197 |
| | | Su | 2.4774 | 2.4952 | 2.5131 |
| | | lm | 0.1619 | 0.1401 | 0.1319 |
| | 0.4 | our | 0.8072 | 0.7722 | 0.7290 |
| | | B&R | 1.0281 | 0.8336 | 0.7289 |
| | | Su | 2.2377 | 2.2315 | 2.2493 |
| | | lm | 0.1619 | 0.1401 | 0.1319 |
| 40 | 0.2 | our | 0.7591 | 0.7173 | 0.7081 |
| | | B&R | 0.7781 | 0.7336 | 0.7139 |
| | | Su | 2.6582 | 2.5276 | 2.4786 |
| | | lm | 0.2916 | 0.1617 | 0.1398 |
| | 0.3 | our | 0.7836 | 0.7150 | 0.7051 |
| | | B&R | 0.8010 | 0.7243 | 0.7144 |
| | | Su | 2.5103 | 2.3803 | 2.3066 |
| | | lm | 0.2916 | 0.1621 | 0.1398 |
| | 0.4 | our | 0.8058 | 0.7252 | 0.7176 |
| | | B&R | 0.8166 | 0.7277 | 0.7257 |
| | | Su | 2.2844 | 2.1698 | 2.0982 |
| | | lm | 0.2916 | 0.1621 | 0.1398 |

* method without clustering.

performance.

As sample size $T$ or the frequency ratio $m$ is enlarged, the forecast accuracy of our method is improved, while the forecast tends to be worse as the scale $\alpha_1$ increases. The change of the forecasting accuracy is not significant when $m$ changes. B&R's method shares similar circumstances. Su's method has much worse accuracy on the one-step-ahead forecasting for all settings. In the meantime, Su's method also requires a prior knowledge of the true number of groups. It may not be a preferable method among all three methods in our framework.

Broadly speaking, our method has more accurate forecast than B&R's method. Both two alternatives present more accurate forecast as the sample size $T$ or the frequency ratio $m$ increases, while the forecasting accuracy becomes worse when the scale $\alpha_1$ increases. When the sample size $T$ is not large enough, for example, $T < 400$, our method is better than B&R's with all different values of other parameters. The difference is quite significant for small sample sizes and frequency ratios. As the sample size $T = 400$, both two methods have similar one-step-ahead forecasting accuracy. Balancing the accuracy of forecast, the estimation of parameters and the computing time, our method could be a wise choice in such panel MIDAS model.

### 3.3.3 Selection of Tuning Parameters

According to the clustering performance shown above, choosing the tuning parameters, $\lambda_2$, $\theta$ and $\lambda_1$ is an essential task which can affect the clustering performance significantly. The choice of $\lambda_1$ depends on the predetermination of $\theta$. In other words, it is crucial for users to make a wise choice of $\theta$. As tables are shown above, the Rand index, the ARI and the Jaccard index are three approaches to compare the clustering performance and to select the tuning parameters. It is required to know the true clusters at first. However, in practice, we may not know the true clusters of panels. The main method to determine $\theta$ and $\lambda_1$ is the BIC. However, selecting $\theta$ and $\lambda_1$ based on the BIC may not be appropriate if the initial values are not assigned properly. Here, we shall propose guidance to select the tuning parameters $\theta$ by calculating the globally convex interval introduced in [73]. When $\theta$ lies in such an interval, the convexity of the objective function (3.16) would be ensured to use (3.19).

In this section, we will give a brief introduction of the guidance choosing tuning parameters. More details can be found in [73]. Consider the convergence of the objective function $Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi})$. $Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi})$ is supposed to converge to a global coordinate-wise minimum shown in (3.17). However, if the second minimization function is not convex, it is hard to guarantee the convergence of $Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi})$. Then, let $c^*(\lambda_1)$ be the minimal eigenvalue of $W(\Pi \otimes I_p)/n = W\Gamma/n$ where $\Pi$ and $\Gamma$ are introduced in (3.21), the objective function $Q_{\lambda_2}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\xi})$ with MCP is convex if $\theta > 1/c^*(\lambda_1)$. Refer to [11], we define the globally convex interval of $\lambda_1$ to be $(\lambda_1^*, \infty)$ where

$$\lambda_1^* = \inf\{\lambda_1 : \theta > 1/c^*(\lambda_1)\}. \tag{3.29}$$

Then, given a value of $\theta$, the solution can be found by

1. Select $\lambda_1$ using BIC;

2. Determine the globally convex interval of $\lambda_1$;

3. Reduce $\theta$ if $\lambda_1$ lies inside the globally convex interval. The convexity can be guaranteed as well. Otherwise, enlarge $\theta$ to make the objective function more convex.

To illustrate the performance of choosing the tuning parameters $\theta$ and $\lambda_1$, we shall take two samples in MC simulation for example.

**Table 3.7**
Selection of $\lambda_1$ given $\theta$

| | Sample | $\theta = 2$ | $\theta > 2$ |
|---|---|---|---|
| $\lambda_1$ by BIC | 4 | 4.5 | 3.5 |
| | 5 | 5.0 | 4.0 |
| $c^*(\lambda_1)$ | 4 | 0.1452 | 0.0681 |
| | 5 | 0.1420 | 0.0694 |
| Globally Convex Interval | 4 | $(6.89, \infty)$ | $(14.69, \infty)$ |
| | 5 | $(7.04, \infty)$ | $(14.41, \infty)$ |

Take sample 4 as an example. We start to examine the globally convex interval from $\theta = 2$. BIC chosen $\lambda_1$ is 4.5, and following the guidance, the globally convex interval for $\theta$ is calculated to be $(6.89, \infty)$. To make the objective function more convex, we enlarge $\theta$. In the simulation, we set $\theta$ from 2.1 to 16. Since the clustering performance keeps the same with different $\theta$'s, according to the way that we construct the design matrix, the convex intervals are the same as well. Then, we present the results of all settings in one column. The interval $(14.68, \infty)$ implies that $\theta$ is expected to be around 15 to guarantee the convexity of the objective function. According to the clustering performance, the guidance does offer an appropriate value for $\theta$. However, since we examine more settings of $\theta$ and it results in the same globally convex interval, we would suggest that in our framework, choosing any value greater than 2 is proper as well. We can draw a similar conclusion on choosing $\theta$ when we focus on sample 5. In general, when knowing the true panels, we can compare the clustering performance with different values of tuning parameters, and such guidance can give us an idea of properly chosen values. However, in reality, we may not have prior knowledge of the true panels most of the time. In such a case, this approach provides a guide to find appropriate values of parameters efficiently.

## 3.4 Okun's Law: Countrywide Unemployment-Losses Relation

Except for the simulation performance, we shall present an empirical application to cluster states based on the behavior related to the labor markets. To have a general idea of how different the relations could be across states between the unemployment rate, the initial unemployment claims and the growth rate of GPD, we consider the mixed-frequency panel data in practice.

### 3.4.1 Labor Market Panel Data and Model Description

Nowadays, Okun's law is famous in the prediction of labor markets. It is a negative correlation between output growth and unemployment rate, which is named after the economist Arthur Okun. Okun [62] first documented that for each 1 percent-point in real GNP growth rate is accompanied by the 0.3 percentage-point decrease in the unemployment rate. Meanwhile, economists have observed that Okun's law model might have limitation to capture the sudden and abrupt rise in unemployment rate due to a spike in job loss during economic downturns, though the model's performance in predicting the unemployment rate, in the long run, is robust (e.g., Karg [47], Lee [51], Moazzami and Dadgostar [60]). To take the nonlinear trend in the unemployment-rate dynamics, we include the weekly initial unemployment claims in the Okun's law model except for the quarterly observed unemployment rate. The weekly initial claims capture the job loss in the economy and have the highest frequency among the variables measuring the labor market slack. Once we measure how much the weekly initial claims help to predict the quarterly unemployment rate, we can utilize the correlation coefficient to predict the unemployment rate on a weekly basis.

To analyze the recent U.S. labor markets, we focus on digging the relationship among three variables based on an extension of Okun's law and identify states that share similar characteristics from the lens of our extended Okun's law framework. Including the quarterly observed GDP, all three variables are collected from 2005 Q2 (the second quarter) to 2018 Q2. The response, the growth rate of GDP as well as the LF regressor, the unemployment rate, are measured quarterly, while the HF regressor is the initial claims observed weekly. There are 51 states in the panel data. Missing initial claims are imputed by the average of records collected in the same week. The data that we construct for $i$-th state is

$$\begin{aligned} u_{i,t} &= \frac{\mathrm{GDP}_{i,t} - \mathrm{GDP}_{i,t-1}}{\mathrm{GDP}_{i,t-1}}, \\ x_{i,t,j} &= \text{initial claims}_{i,t,j}, \\ y_{i,t} &= \text{unemploy}_{i,t} - \text{unemploy}_{i,t-1}, \end{aligned} \tag{3.30}$$

where $t$ is the index of the quarter and $j$ is the index of the week in the $t$-th quarter. The intercept is included in the LF regressor term $\mathbf{u}_{i,t}$. Therefore, Okun's law can be modified as

$$y_{i,t} = u_{i,t}\alpha_i + \mathbf{x}'_{i,t}\boldsymbol{\beta^*}_i + \varepsilon_{i,t},$$

where $\mathbf{x}_{i,t} = (x_{i,t,1}, \cdots, x_{i,t,m_t})'$. $\boldsymbol{\beta}^*$ is hard to define since the number of weeks $m_{i,t,m_t}$ could be different with respect to the quarter $t$. It ranges from 12 to 14, so the regression model cannot be simply constructed. Then, Fourier approximation outperforms other approaches mentioned previously at this point.

Given the transformed HF to be $\tilde{\mathbf{x}}_{i,t} = M_i\mathbf{x}_{i,t}$, the nonparametric MIDAS with Fourier approximation is

$$y_{i,t} = u_{i,t}\alpha_i + \tilde{\mathbf{x}}'_{i,t}\boldsymbol{\beta}_i + \varepsilon_{i,t}, \tag{3.31}$$

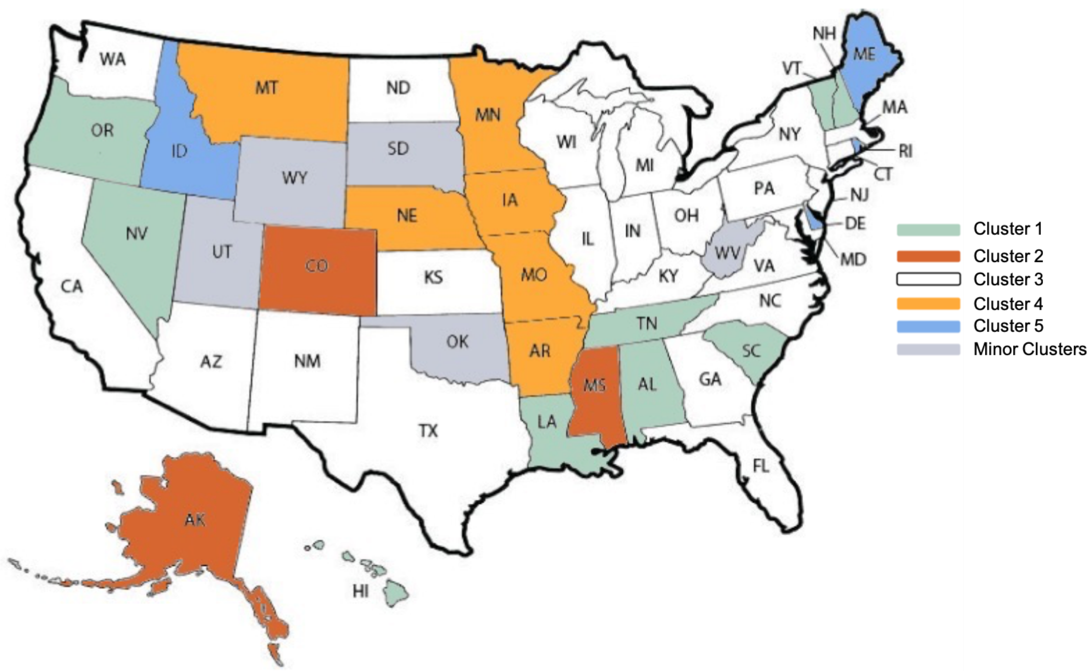
where $\boldsymbol{\beta}_i = (\beta_{i,1}, \cdots, \beta_{i,L+2K+1})'$ for $L$ and $K$ being the number of parameters in Fourier approximation. $(\alpha_i, \boldsymbol{\beta}'_i)$ are forced to have the similar values if they belong to the same group. The difference in estimated coefficients across states could represent heterogeneity in the functioning of the labor market.

### 3.4.2 Clustering Analysis on the US Labor Markets

According to the algorithm for panel nonparametric MIDAS with Fourier expansion, we set $\theta = 3$ and $\lambda_1 = 21$ chosen by the BIC. Except for $\theta = 3$, we tried different values such as $\theta = 2, 5, 8, 10$. With small values of $\theta$, for example, $\theta = 2, 3, 5$, with BIC chosen $\lambda_1$, the clustering results are quite similar. The only significant change is whether some individual states are clustered in the same group or separated. A few states would be isolated to become a single-subject group with different $\theta$. States are eventually divided into 10 groups in Figure 3.1. These groups contain one major group, four moderate groups as well as several single-subject groups, which are shaded in the same color. About half of the states (22 states) share similar relationships among the unemployment rate, initial claims and output growth. In other moderate clusters, there are 3-9 states. In either major or moderate groups, states are clustered regionally. With the chosen value, states which are close to each other tend to share similar correlation among labor market variables and the growth of GDP. Geographical proximity seems to be an essential factor in determining a cluster or in determining the intrinsic characteristics of a states labor market. For example, Minnesota, Iowa, Missouri, and Arkansas are vertically connected, and Nebraska is right next to Iowa. California, Arizona, New Mexico, Texas are connected and most of the northeastern states are collected in the major group. At the same time, it is notable that states that belong to the same cluster are scattered. States that belong to the major cluster, Cluster 3, are observed in the southwest as well as in the northeast, though these two subgroups are disconnected. Besides, Oregon and Nevada, Tennessee and Alabama are observed to belong in the same group. Having that said, states that are geographically close to each other might share the similar characteristics in their functioning of labor markets, but this result suggests that it is not the necessary condition to determine a cluster. Such finding is consistent with the conclusion in Guisinger et al. [39] that states within the same geographical region can have heterogeneous business cycle experiences. Nonetheless, the observation that states close to each other tend to share similar correlation among labor market variables and output growth suggests that some economic information of adjacent states can help to predict a states labor market outcome.

Although states in the same cluster share the similarity in the attributes of labor markets by state, there still exists difference within each group. For example, in

**Figure 3.1:** Labor Market Heterogeneity by State




Cluster 3 the states that are located in the west including California, Texas, and New Mexico, are mostly oil-producing states, while most of the states in the east including Michigan, Ohio, Kentucky and Indiana tend to have a high share of manufacturing employment. According to Hamilton and Owyang [42], oil-producing states and manufacturing-intensive states tend to have quite different business cycles. Though they might have different cyclical characteristics, our result suggests that oil-producing states in the southwest and manufacturing-intensive states in the northeast might share similar attributes in the labor market functioning. In fact, Guisinger et al. [39] claims that there are multiple factors that determine the coefficients of Okuns law such as industrial composition, labor-market regulation, the demographic composition of the population and so on. This finding implies that whether states belong to the same group or not is not determined by a single structural characteristic of a states economy.

In general, the clustering results imply that expansionary policies are likely to have quite different effects on the labor market outcomes at the state level. Given that estimating the effect of policy on a state's labor market in real time is challenging, understanding how much weekly initial claims help to predict the unemployment rate in the state level will be able to guide policymakers to adjust their policy implementation in a timely manner.

# Chapter 4

# Conclusion

## 4.1 Summary of Results

The last more than a decade has witnessed the dramatic developments of approaches to coping with the mixed-frequency sampling problem in the regression models. Conventionally, HF variables are aggregated by predetermined and fixed weights. MIDAS models were proposed to assign more flexibility on the weights to maintain more information in HF variables. Compared to parametric models, nonparametric models were introduced to gain more flexibility of the fitted weights at the expense of the computation complexity.

### 4.1.1 On the Choice of IVs

In Chapter 2, we considered a DWH test to choose between the time-averaging models and MIDAS models. For the DWH test, the instruments need to be carefully chosen to avoid the problems involved with weak instruments and correlation with the error terms. However, there had not yet been rigorous work regarding the proper choice of instruments.

The main contribution of Chapter 2 is that a set of instruments has been proposed with a theoretical validation. In particular, the proposed instruments would only work when the frequency ratio is large enough. The Monte Carlo simulations reconfirm our theoretical findings. The DWH test with our proposed instruments is more potent in finite samples compared to the one with a less careful choice of instruments. However, this is only the case when the frequency ratio is large enough. Therefore, our proposed specification test would be useful when handling two extremely different sampling frequencies, such as monthly versus hourly observations. On the other hand, if the frequency ratio is very small, taking a few most recent HF variables as the instruments or taking Miller's approach would be better.

The primary purpose of Chapter 2 is to provide an insight into the proper choice of instruments. To keep the exposition concise, we limited the scope of Chapter

2 using somewhat strong assumptions. Now that we understand the behavior of the instruments better, an extension of Chapter 2 to accommodate more than one regressors and general data generating process is underway.

### 4.1.2 Panel MIDAS with Nonparametric Approach

We introduced Fourier expansion approximation into MIDAS models to estimate the weight function in Chapter 3. With properly predetermined numbers of polynomial terms as well as trigonometric terms, we showed that Fourier expansion would be an appropriate approach theoretically and empirically. Comparing to the nonparametric approach in MIDAS models, Fourier expansion approximation could be more effective along with precise estimations. By using the MC simulation, empirical MSE, and one step ahead prediction indicate that Fourier expansion in MIDAS models outperforms nonparametric method in our framework in general. On the other hand, in some cases, the nonparametric approach would have slightly better performance. Considering the workload of the nonparametric approach, it remains a crucial problem to balance the complexity of calculation and the accuracy of estimation.

As considering a more general model with panel data, we proposed a clustering algorithm to stratify the estimated weights with Fourier expansion approximation. With an accurate estimation of weight functions, the algorithm provided a clear path to get convergent clustering. When the true clusters are known at first, the clustering performance is evaluated by three indexes as well as the number of clusters. We present how the tuning parameter $\lambda_1$ controls the performance of clustering of two panels. Moreover, we demonstrated the effect of another parameter $\theta$ in the clustering algorithm on clustering using MC simulations.

In practice, the true clusters may be unknown before clustering. The BIC would not choose the tuning parameters appropriately if the initial values are not correctly assigned. As a result, we propose an approach to select the tuning parameters $\theta$ and $\lambda_1$. Simulated examples indicate that the approach shows an optimal path to select parameters. Furthermore, the results of such guidance explain the reason of lousy clustering performance with some values of tuning parameters. US labor market is analyzed by digging the relationship between the quarterly unemployment rate, the weekly initial claims, and the quarterly GDP growth. Clustering the relationships by state, we observe that the groups are regionally related. Apart from the regional impact, other factors may also affect the clustering of the state-level behavior of the labor market.

## 4.2 Discussion and Future Research

While this thesis presents a few new approaches for MIDAS models, there still are more opportunities to extend the current work. The first possible future direction is exploring a wider variety of IVs in the specification test. As we mentioned in the DWH type test, the choice of IVs could be in other forms. The only two IVs examined in Chapter 2 are inspired by Miller [59]. However, the construction of IVs could be highly data-related. Except for the data, the user-determined shape of IVs could also be a crucial factor that affects the performance of the MIDAS model. It is worth exploring other kinds of IVs in the DWH type test. Leaving the choice of IVs alone, more general models could be taken into consideration, such as allowing the error to be a random walk or including more regressors in the MIDAS model, etc.

The second compelling direction for extending this thesis work is applying the specification tests to real data. In particular, the research on the MIDAS methods related to the US labor markets in Chapter 3 could be worth discussing using our choice of instruments. Our specification test may not have enough power for this data because the frequency ratio for our labor market example in our empirical analysis in Chapter 3 is around 13. Such frequency ratio is too small for our specification test to have enough power to judge the necessity of the MIDAS model. Nonetheless, finding out whether the flat aggregation would be enough for all states would still be an appealing topic. This direction may also involve other difficulties such as the consideration of the financial crisis of 2007 - 2008. We shall explore the details in future work.

The third future work is refining our empirical analysis in Chapter 3. For example, mentioned above, the data between 2007 and 2008 may require adjustment. Some variables may have better performance on the MIDAS models using different transformations, such as taking logarithm, scaling data state by state or of all states together and so on. Models including intercepts or other possibly related variables that we did not consider in the exploration of the US labor markets could be useful alternatives. All these modifications would be considered in our next step.

The fourth direction is refining the theory in Chapter 3. We have demonstrated the clustering method in the panel MIDAS model theoretically and empirically. In our method, all parameters are considered in the algorithm, including the HF variable as well as additional LF covariates. However, it is possible that some LF covariates may contribute significantly when we intend to measure the proximity of panels based on the similarity of the MIDAS coefficients, or more generally, some of the coefficients. The substantial contribution from these variables could be eliminated if we exclude them in the clustering algorithm. For example, in the investigation of the US labor markets, the difference of weekly initial claims and the quarterly GDP between subjects are aggregated to measure the distinction of panels. To avoid the potentially significant effect from GDP, clustering only the initial claims may be more reasonable to some extent. Such a topic would involve a more general discussion of our method. More in-depth exploration may help to gain a clearer and direct thought about such distinction.

Last but not least is applying our proposed methods outside of econometrics and finance. The MIDAS models are developed for better forecast mainly in finance and economics area and have demonstrated its potential in providing a more accurate forecast. However, these areas are not the only options that care about the quality of the forecast.

As one such application utilizing the forecasting ability of MIDAS models, we are considering an example related to power grids and electricity demand forecast. In the electricity demand forecasting, it is known that variables such as temperature, natural gas price, renewable fuels productivity, and time trend are considerable. It would be interesting to find out the form of how these variables affect electricity demand using a MIDAS model. These variables are related not only to the personal usage but also to the electricity demand from enterprises, such as the locally seasonal temperature, industrial structure, the use of reverse cycle air conditioning, etc. Since temperature can be collected at a high sampling frequency such as hourly or daily, whereas the electricity demand can be measured based on the monthly bill, it would be interesting to find whether a MIDAS approach would increase accuracy in electricity demand forecasting. Our choice of instruments could be helpful in forecasting the electricity demand, which could further affect the electricity price in marketing. Moreover, it points out some directions to investigate the significance of some parameters, for example, renewable fuels, in different seasons or in different regions.

The above mentioned application of our MIDAS model to the power grid may also extend to determining insurance premium for power grids. In particular, in recent decades, the rise of cyber threats has drawn substantial attention. Malicious cyber attacks could lead to massive economic loss. To avoid comparatively huge loss in each cyber attack, introducing the ecosystem with cyberinsurance on power grid would promote the technological development protecting the critical infrastructure. For instance, a proper forecast of the electricity demand can help to determine the potential loss more accurately, which is essential to formulate the premium of the cyber insurance on power grid reasonably. In the power system, the substations and utilities are closely connected. As a result, regional or enterprise-level clustering may offer a more integral and acceptable determination of the electricity demand as well as possibly the insurance premium. Based on the findings that we have for the labor market, our method is likely to be useful for clustering the effect of related parameters on the power-use to understand more thoroughly about the power system and plays an influential role on the prediction of electricity demands. We presented a brief discussion on how the panel MIDAS could be useful in the cyber insurance based on the power grid. However, this application still requires more detailed discussion and data collection, which is out of the scope of this dissertation. We leave it as future work.

# References

[1] E. Andreou, E. Ghysels, and A. Kourtellos. Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158:246–261, 2010.

[2] D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

[3] J. Bai, E. Ghysels, and J. H. Wright. State space models and midas regressions. *Econometric Reviews*, 32(7):779–813, 2013.

[4] L. Ball, D. Leigh, and P. Loungani. Okun's law: Fit at 50? *Journal of Money, Credit and Banking*, 49(7):1413–1441, 2017.

[5] R. J. Barro and J. W. Lee. A new data set of educational attainment in the world, 19502010. *Journal of Development Economics*, 104(C):184–198, 2013.

[6] R. Becker, W. Enders, and S. Hurn. A general test for time dependence in parameters. *Journal of Applied Econometrics*, 19(7):899–906, 2004. ISSN 1099-1255. doi: 10.1002/jae.751. URL `http://dx.doi.org/10.1002/jae.751`.

[7] R. Becker, W. Enders, and J. Lee. A stationarity test in the presence of an unknown number of smooth breaks. *Journal of Time Series Analysis*, 27(3): 381–409, March 2006.

[8] B. S. Bernanke and R. S. Grkaynak. *Is Growth Exogenous? Taking Mankiw, Romer, and Weil Seriously*, pages 11–72. MIT Press, January 2002. URL `http://www.nber.org/chapters/c11063`.

[9] B. S. Bernanke and R. S. Grkaynak. *Is Growth Exogenous? Taking Mankiw, Romer, and Weil Seriously*, pages 11–72. MIT Press, January 2002. URL `http://www.nber.org/chapters/c11063`.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[11] P. Breheny and J. Huang. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.*, 5 (1):232–253, 03 2011.

[12] J. Breitung and C. Roling. Forecasting inflation rates using daily data: A non-parametric midas approach. *Journal of Forecasting*, 34(7):588–603, 2015.

[13] R. Casarin, C. Foroni, M. G. Marcellino, and F. Ravazzolo. Uncertainty through the lenses of a mixed-frequency bayesian panel markov switching model. 2017.

[14] J. C. Chao and N. R. Swanson. Consistent estimation with a large number of weak instruments. *Econometrica*, 73(5):1673–1692, 2005.

[15] X. Chen and E. Ghysels. News–good or bad–and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, 24(1):46–81, 2011.

[16] N. Coffey, J. Hinde, and E. Holian. Clustering longitudinal profiles using p-splines and mixed effects models applied to time-course gene expression data. *Computational Statistics & Data Analysis*, 71:14 – 29, 2014. ISSN 0167-9473.

[17] J. Durbin. Errors in variables. *Review of the International Statistical Institute*, 22(1/3):23–32, 1954.

[18] A. Economou and I. N. Psarianos. Revisiting okuns law in european union countries. *Journal of Economic Studies*, 43(2):275–287, 2016.

[19] W. Enders and J. Lee. A unit root test using a fourier series to approximate smooth breaks*. *Oxford Bulletin of Economics and Statistics*, 74(4):574–599. doi: 10.1111/j.1468-0084.2011.00662.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.2011.00662.x`.

[20] W. Enders and J. Lee. A unit root test using a fourier series to approximate smooth breaks. *Oxford Bulletin of Economics and Statistics*, 74(4):574–599, 2012. URL `https://EconPapers.repec.org/RePEc:bla:obuest:v:74:y:2012:i:4:p:574-599`.

[21] C. L. Epstein. How well does the finite fourier transform approximate the fourier transform? *Communications on Pure and Applied Mathematics*, 58(10):1421–1435.

[22] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456): 1348–1360, 2001.

[23] M. Fidino and S. B. Magle. Using fourier series to estimate periodic patterns in dynamic occupancy models. *Ecosphere*, 8(9):e01944–n/a, 2017. ISSN 2150-8925.

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

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

[26] E. B. Fowlkes and C. L. Mallows. A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383):553–569, 1983.

[27] A. Gallant. On the bias in flexible functional forms and an essentially unbiased form: The fourier flexible form. *Journal of Econometrics*, 15(2):211 – 245,

1981. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(81)90115-9. URL http://www.sciencedirect.com/science/article/pii/0304407681901159.

[28] A. Gallant and G. Souza. On the asymptotic normality of fourier flexible form estimates. *Journal of Econometrics*, 50(3):329 – 353, 1991. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(91)90024-8. URL http://www.sciencedirect.com/science/article/pii/0304407691900248.

[29] E. Ghysels and I. Miller. Testing for cointegration with temporally aggregated and mixed-frequency time series. *Journal of time series analysis*, 36(6):797–816, 2015.

[30] E. Ghysels and H. Qian. Estimating midas regressions via ols with polynomial parameter profiling. *Econometrics and Statistics*, 2018.

[31] E. Ghysels and H. Qian. Estimating midas regressions via ols with polynomial parameter profiling. Technical report, 2018.

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

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

[34] E. Ghysels, J. B. Hill, and K. Motegi. Testing for granger causality with mixed frequency data. *Journal of Econometrics*, 129(1):207–230, 2016.

[35] G. H. Golub and C. F. Van Loan. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996. ISBN 0-8018-5414-8.

[36] T. Götz, A. Hecq, and J.-P. Urbain. Forecasting mixed frequency time series with ecm-midas models. Research Memorandum 012, Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR), 2012.

[37] W. H. Greene. *Econometric Analysis, 7th Edition*. Prentice Hall, Upper Saddle River, 2012.

[38] H. Groenvik and Y. Rho. A self-normalizing approach to the specification test of mixed-frequency models. *Communications in Statistics - Theory and Methods. Accepted*, 2017.

[39] A. Y. Guisinger, R. Hernandez-Murillo, M. T. Owyang, and T. M. Sinclair. A State-Level Analysis of Okuns Law. Working Papers (Old Series) 1523, Federal Reserve Bank of Cleveland, Oct. 2015. URL `https://ideas.repec.org/p/fip/fedcwp/1523.html`.

[40] A. Y. Guisinger, R. Hernandez-Murillo, M. T. Owyang, and T. M. Sinclair. A state-level analysis of okun's law. *Regional Science and Urban Economics*, 68: 239 – 248, 2018. ISSN 0166-0462. doi: https://doi.org/10.1016/j.regsciurbeco.2017.11.005. URL `http://www.sciencedirect.com/science/article/pii/S0166046217300327`.

[41] B. Güriş. A New Nonlinear Unit Root Test with Fourier Function. MPRA Paper 82260, University Library of Munich, Germany, Oct. 2017.

[42] J. D. Hamilton and M. T. Owyang. The Propagation of Regional Recessions. *The Review of Economics and Statistics*, 94(4):935–947, November 2012.

[43] J. A. Hausman. Specification tests in econometrics. *Econometrica*, 46(6):1251–1271, 1978.

[44] F. Hayashi. *Econometrics*. Princeton University Press, 2011.

[45] D. Hsu, S. Kakade, T. Zhang, et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17, 2012.

[46] L. Hubert and P. Arabie. Comparing partitions. 2:193–218, 02 1985.

[47] B. Karg. Okuns law and long term co-integration analysis for oecd countries (1987-2012). *EMAJ: Emerging Markets Journal*, 6(1), 2016.

[48] L. Khalaf, M. Kichian, C. Saunders, and M. Voia. Dynamic panels with midas covariates: Estimation and fit. Technical report, Technical report, 2013.

[49] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1):159 – 178, 1992.

[50] A. N. Langville and W. J. Stewart. The kronecker product and stochastic automata networks. *Journal of Computational and Applied Mathematics*, 167 (2):429 – 447, 2004. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2003.10.010. URL `http://www.sciencedirect.com/science/article/pii/S0377042703009312`.

[51] J. Lee. The robustness of okun's law: Evidence from oecd countries. *Journal of Macroeconomics*, 22(2):331 – 356, 2000. ISSN 0164-0704.

[52] M. Lee. *Micro-econometrics: Methods of moments and limited dependent variables (Second Edition)*. Springer New York, 12 2010.

[53] Y. Liu and Y. Rho. On the choice of instruments in mixed frequency specification tests. *Communications in Statistics - Theory and Methods*, pages 1–16, 11 2018. doi: 10.1080/03610926.2018.1528367.

[54] S. Ma and J. Huang. Estimating subgroup-specific treatment effects via concave fusion. 07 2016.

[55] S. Ma and J. Huang. A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association*, 112(517):410–423, 2017.

[56] N. Mankiw, D. Romer, and D. Weil. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(May):407–437, 1992.

[57] B. Micallef. Empirical estimates of okuns law in malta. *Applied Economics and Finance*, 4(1), 2016.

[58] I. J. Miller. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics*, 12:584–614, 2014.

[59] J. I. Miller. Simple robust tests for the specification of high-frequency predictors of a low-frequency series. *Econometrics and Statistics*, 5:45 – 66, 2018.

[60] B. Moazzami and B. Dadgostar. Okun's law revisited: evidence from oecd countries. *Int. Bus. Econ. Res. J.*, 8:21–24, 02 2011. doi: 10.19030/iber.v8i8.3156.

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

[62] A. M. Okun. Potential gnp: Its measurement and significance. *Reprinted as Cowles Foundation Paper 190*, 84, 1962.

[63] P. Perron, M. Shintani, and T. Yabu. Testing for flexible nonlinear trends with an integrated or stationary noise component. *Oxford Bulletin of Economics and Statistics*, 79(5):822–850.

[64] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[65] P. Rodrigues and R. Taylor. The flexible fourier form and local generalised least squares de-trended unit root tests*. *Oxford Bulletin of Economics and Statistics*, 74(5):736–759, 2012.

[66] P. Ruud. *An Introduction to Classical Econometric Theory*. Oxford University Press, 2000.

[67] L. Su, Z. Shi, and P. C. B. Phillips. Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264, 2016.

[68] Y. Uematsu and S. Tanaka. Macroeconomic forecasting and variable selection with a very large number of predictors: A penalized regression approach. Technical report, March 2017. URL `https://ssrn.com/abstract=2927876`.

[69] H. Wang, Z. Shi, and C.-S. Leung. Admm-mcp framework for sparse recovery with global convergence. *IEEE Transactions on Signal Processing*, Sep 2018, Forthcoming.

[70] D.-M. Wu. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, 41(4):733–750, 1973.

[71] T. Y. Xue Wang, Mike Mingcheng Wei. Minimax concave penalized multi-armed bandit model with high-dimensional convariates. In *35th International Conference on Machine Learning, ICML 2018*, pages 8249–8266. International Machine Learning Society (IMLS), Jan 2018.

[72] G. U. Yule. On the time-correlation problem, with especial reference to the variate- difference correlation method. *Journal of the Royal Statistical Society*, 84(4):497–537, 1921. ISSN 09528385. URL `http://www.jstor.org/stable/2341101`.

[73] C. Zhang. Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.*, 38(2):894–942, 04 2010.

[74] X. Zhu and A. Qu. Cluster analysis of longitudinal profiles with subgroups. 12: 171–193, 01 2018.

# Appendix A

# Theoretical Proofs

## A.1 Proof of Theorems in Chapter 2

### A.1.1 Test Statistic $\lambda_T$ and Asymptotic Distribution

*Proof of Theorem 2.1.* It is easy to see that under the null, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}^A$ is $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}^A - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, V^A)$. Under both the null and the alternative, the asymptotic distribution $\widehat{\boldsymbol{\beta}}$ is $\sqrt{T}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, V)$. Moreover, for some matrix $V^*$, we are able to derive $\sqrt{T}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^A) \xrightarrow{d} N(\mathbf{0}, V^*)$. Following the argument in

Section 5.1 of [52], the asymptotic distribution of $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^A$ can be derived as

$$T\widehat{\boldsymbol{\Delta}}' \left(\widehat{V} - \widehat{V}^A\right)^{-1} \widehat{\boldsymbol{\Delta}} \xrightarrow{d} \chi^2_{rank(V-V^A)}. \tag{A.1}$$

By noting that $(X\boldsymbol{\pi}_0)'P_Z M_{X^A}\mathbf{y} = (0,1)\left(X^{A\prime}P_Z X^A\right)\widehat{\boldsymbol{\Delta}}$ and $(X\boldsymbol{\pi}_0)'M_{X^A}\mathbf{y} = 0$, $\widehat{\delta}$ can be rewritten as

$$\begin{aligned}
\widehat{\delta} &= \left[(M_{X^A}M_Z X\boldsymbol{\pi}_0)'(M_{X^A}M_Z X\boldsymbol{\pi}_0)\right]^{-1}\left(-(X\boldsymbol{\pi}_0)'P_Z M_{X^A}\mathbf{y}\right) \\
&= -\left[(M_{X^A}M_Z X\boldsymbol{\pi}_0)'(M_{X^A}M_Z X\boldsymbol{\pi}_0)\right]^{-1}(0,1)\left(X^{A\prime}P_Z X^A\right)\widehat{\boldsymbol{\Delta}} \\
&= -\left[(M_{X^A}M_Z X\boldsymbol{\pi}_0)'(M_{X^A}M_Z X\boldsymbol{\pi}_0)\right]^{-1}\left((X\boldsymbol{\pi}_0)'P_Z X^A\right)\widehat{\boldsymbol{\Delta}} \\
&= \mathbf{b}'\widehat{\boldsymbol{\Delta}}.
\end{aligned} \tag{A.2}$$

where $\mathbf{b}' = -\left[(M_{X^A}M_Z X\boldsymbol{\pi}_0)'(M_{X^A}M_Z X\boldsymbol{\pi}_0)\right]^{-1}\left((X\boldsymbol{\pi}_0)'P_Z X^A\right)$. Thus,

$$\sqrt{T}\widehat{\delta} = \sqrt{T}\mathbf{b}'\widehat{\delta} = \sqrt{T}\left[\mathbf{b}'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - \mathbf{b}'\left(\widehat{\boldsymbol{\beta}}^A - \boldsymbol{\beta}\right)\right]. \tag{A.3}$$

The asymptotic distribution of $\mathbf{b}'\widehat{\boldsymbol{\beta}}^A$ is $\sqrt{T}\mathbf{b}'\left(\widehat{\boldsymbol{\beta}}^A - \boldsymbol{\beta}\right) \xrightarrow{d} N(0, \mathbf{b}'V^A\mathbf{b})$ under the null. The asymptotic distribution of $\mathbf{b}'\widehat{\boldsymbol{\beta}}$ is $\sqrt{T}\mathbf{b}'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(0, \mathbf{b}'V\mathbf{b})$ under both the null and the alternative. Since the estimator $\mathbf{b}'\widehat{\boldsymbol{\beta}}^A$ is still consistent and efficient under the null, while the estimator $\mathbf{b}'\widehat{\boldsymbol{\beta}}$ is consistent under the null and the alternative, then

$$T\left[\mathbf{b}'\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^A\right)\right]'\left(\mathbf{b}'\widehat{V}\mathbf{b} - \mathbf{b}'\widehat{V}^A\mathbf{b}\right)^{-1}\left[\mathbf{b}'\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^A\right)\right] \xrightarrow{d} \chi^2_{rank(\mathbf{b}'(V-V^A)\mathbf{b})}. \tag{A.4}$$

Therefore,

$$T\widehat{\delta}'\left(\mathbf{b}'(\widehat{V}-\widehat{V}^A)\mathbf{b}\right)^{-1}\widehat{\delta}\xrightarrow{d}\chi^2_{rank(\mathbf{b}'(V-V^A)\mathbf{b})}. \tag{A.5}$$

Note that under our settings, $\mathbf{b}$ is a column vector with two elements. The rank of $\mathbf{b}'(V-V^A)\mathbf{b}$ is one. Hence, the degree of freedom of $\chi^2$ distribution is one. □

### A.1.2 Theoretical Verification of the Chosen Set of Instruments

*Proof of Theorem 2.2.* It is obvious that our choice of IVs follows Assumption 2.1(c). Following Slutsky's theorem, it is straightforward to show that our choice of IVs satisfies Assumption 2.1(d) and 2.1(e). So, the main part is to show that our choice of IVs satisfies Assumption 2.1(f), i.e., $E(Z'\boldsymbol{\varepsilon}^A)$ is zero or approximates to zero as the frequency ratio $m$ approaches infinity. Assumption 2.1(g) follows.

Under the null hypothesis, $\widehat{\boldsymbol{\beta}}^A$ is consistent to estimate $\boldsymbol{\beta}$, then the error process $\{\varepsilon_t\}$ is exactly $\{\varepsilon_t^A\}$ in (2.2). Therefore, following Assumption 2.2(b), $\varepsilon_t^A=\varepsilon_t=\boldsymbol{\varepsilon}_{t,m}{}'\boldsymbol{\pi}(\theta)$, $\mathbf{z}_t{}'=\mathbf{x}_t{}'\Upsilon$,

$$T^{-1}Z'\boldsymbol{\varepsilon}^A=T^{-1}Z'\boldsymbol{\varepsilon}=T^{-1}\sum_{t=1}^{T}\mathbf{z}_t\varepsilon_t=T^{-1}\sum_{t=1}^{T}\Upsilon'\mathbf{x}_t\boldsymbol{\varepsilon}_{t,m}{}'\boldsymbol{\pi}(\theta)\xrightarrow{p}\mathbf{0}.$$

It follows that the asymptotic distribution is $T^{-1/2}Z'\boldsymbol{\varepsilon}^A \xrightarrow{d} N(0, \Sigma_{Z\varepsilon})$ for some matrix $\Sigma_{Z\varepsilon}$.

Under the alternative hypothesis, $\widehat{\boldsymbol{\beta}}^A$ is not consistent, the true model is the MIDAS model in (2.1), i.e. $\mathbf{y} = X(\theta)\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $X(\theta) = [\mathbf{j}, X\boldsymbol{\pi}(\theta)]$. Recall that $X^A = [\mathbf{j}, X\boldsymbol{\pi}_0]$. Let $\mathbf{x}_t^{A\prime}$ and $\mathbf{x}_t(\theta)'$ be $t$-th row of $X^A$ and $X(\theta)$, respectively. Comparing the MIDAS model with the regression model in (2.2), $\mathbf{y} = X^A\boldsymbol{\beta}^A + \boldsymbol{\varepsilon}^A$, it is easy to show that $\boldsymbol{\beta}^A$ can be written as $\boldsymbol{\beta}^A = \left\{E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)\right\}^{-1}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right)\right\}\beta$, then

$$
\begin{aligned}
\varepsilon_t^A &= y_t - \mathbf{x}_t^{A\prime}\boldsymbol{\beta}^A = y_t - \mathbf{x}_t^{A\prime}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)\right\}^{-1}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right)\right\}\boldsymbol{\beta} \\
&= \left(\mathbf{x}_t(\theta)' - \mathbf{x}_t^{A\prime}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)\right\}^{-1}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right)\right\}\right)\boldsymbol{\beta} + \varepsilon_t \qquad \text{(A.6)} \\
&= A\boldsymbol{\beta} + \varepsilon_t,
\end{aligned}
$$

where $A = \mathbf{x}_t(\theta)' - \mathbf{x}_t^{A\prime}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)\right\}^{-1}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right)\right\}$. Let $J_m = \mathbf{j}\mathbf{j}'$ be a all-ones matrix with dimension $m$. According to the property of $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}(\theta)$, we have $\boldsymbol{\pi}_0'\mathbf{j} = 1$ and $\boldsymbol{\pi}(\theta)'\mathbf{j} = 1$.

Since the HF processes $\{x_{t-k/m}\}$ and $\{\varepsilon_{t-k/m}\}$ are assumed to be i.i.d. or follow stationary AR(1) processes with finite second moment, respectively, for $k = 0, 1, \cdots, m-1, t = 1, \cdots, T$ and $\sum_{i=1}^m \pi_i = 1$, denote the variance-covariance matrix of $\mathbf{x}_t$ as

$\Phi = E(\mathbf{x}_t\mathbf{x}_t') - E(\mathbf{x}_t)E(\mathbf{x}_t)'$, then $E(\mathbf{x}_t) = \mu\mathbf{j}$, $E(\mathbf{x}_t\mathbf{x}_t') = \Phi + \mu^2 J_m$.

$$A = \begin{pmatrix} 1 & \mathbf{x}_t'\boldsymbol{\pi}(\theta) \end{pmatrix} - \begin{pmatrix} 1 & \mathbf{x}_t'\boldsymbol{\pi}_0 \end{pmatrix} \left\{E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)\right\}^{-1}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right)\right\}$$

where

$$E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right) = \begin{bmatrix} 1 & \boldsymbol{\pi}_0'E(\mathbf{x}_t) \\ \boldsymbol{\pi}_0'E(\mathbf{x}_t) & \boldsymbol{\pi}_0'E(\mathbf{x}_t\mathbf{x}_t')\boldsymbol{\pi}_0 \end{bmatrix} = \begin{bmatrix} 1 & \mu \\ \mu & \boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)\boldsymbol{\pi}_0 \end{bmatrix},$$

$$E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right) = \begin{bmatrix} 1 & \boldsymbol{\pi}(\theta)'E(\mathbf{x}_t) \\ \boldsymbol{\pi}_0'E(\mathbf{x}_t) & \boldsymbol{\pi}_0'E(\mathbf{x}_t\mathbf{x}_t')\boldsymbol{\pi}(\theta) \end{bmatrix} = \begin{bmatrix} 1 & \mu \\ \mu & \boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)\boldsymbol{\pi}(\theta) \end{bmatrix}.$$

Assuming that $E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)$ is invertible (if $E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)$ is not invertible, we can get the generalized inverse), then we can derive

$$\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t^{A\prime}\right)\right\}^{-1}\left\{E\left(\mathbf{x}_t^A\mathbf{x}_t(\theta)'\right)\right\} = \begin{bmatrix} 1 & (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\mu\boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)(\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\theta)) \\ 0 & (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) \end{bmatrix},$$

Therefore,

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & \mathbf{x}_t{}'\boldsymbol{\pi}(\theta) \end{pmatrix} - \begin{pmatrix} 1 & \mathbf{x}_t{}'\boldsymbol{\pi}_0 \end{pmatrix} \left\{ E\left(\mathbf{x}_t^A \mathbf{x}_t^{A\prime}\right) \right\}^{-1} \left\{ E\left(\mathbf{x}_t^A \mathbf{x}_t(\theta)'\right) \right\} \\
&= \begin{pmatrix} 1 & \mathbf{x}_t{}'\boldsymbol{\pi}(\theta) \end{pmatrix} \\
&\quad - \begin{pmatrix} 1 & (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1} \left\{ \mu\boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)(\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\theta)) + \mathbf{x}_t{}'\boldsymbol{\pi}_0\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) \right\} \end{pmatrix} \\
&= \begin{pmatrix} 0 & \mathbf{x}_t{}'\boldsymbol{\pi}(\theta) - (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1} \left\{ \mu\boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)(\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\theta)) + \mathbf{x}_t{}'\boldsymbol{\pi}_0\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) \right\} \end{pmatrix}.
\end{aligned}
\tag{A.7}
$$

Next, calculate $E\left(\mathbf{z}_t \varepsilon_t^A\right)$ where $\mathbf{z}_t' = \mathbf{x}_t{}'\Upsilon$,

$$
E\left(\mathbf{z}_t \varepsilon_t^A\right) = E\left(\mathbf{z}_t(A\boldsymbol{\beta} + \varepsilon_t)\right) = E\left(\mathbf{z}_t A\boldsymbol{\beta}\right) = E\left(\mathbf{z}_t A \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{bmatrix}\right). \tag{A.8}
$$

Combine (A.7) with (A.8), then

$$
\begin{aligned}
&E\left(\mathbf{z}_t \varepsilon_t^A\right) \\
&= \beta_1 E\left(\mathbf{z}_t \left(\mathbf{x}_t{}'\boldsymbol{\pi}(\theta) - (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1} \left\{ \mu\boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)(\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\theta)) + \mathbf{x}_t{}'\boldsymbol{\pi}_0\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) \right\}\right)\right) \\
&= \beta_1 E\left(\Upsilon'\mathbf{x}_t \left(\mathbf{x}_t{}'\boldsymbol{\pi}(\theta) - (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1} \left\{ \mu\boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)(\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\theta)) + \mathbf{x}_t{}'\boldsymbol{\pi}_0\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) \right\}\right)\right) \\
&= \beta_1 \Upsilon' \left\{ (\Phi + \mu^2 J_m)\boldsymbol{\pi}(\theta) - (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\mu\boldsymbol{\pi}_0'(\Phi + \mu^2 J_m)(\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\theta))\mu\mathbf{j} \right. \\
&\qquad\qquad \left. -(\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}(\Phi + \mu^2 J_m)\boldsymbol{\pi}_0\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) \right\}.
\end{aligned}
\tag{A.9}
$$

After simplification, (A.9) becomes

$$E\left(\mathbf{z}_t\varepsilon_t^A\right) = \beta_1\Upsilon'\left(\Phi\boldsymbol{\pi}(\theta) - (\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta)\Phi\boldsymbol{\pi}_0\right). \tag{A.10}$$

Note that let $\pi_{0,i}$ be the $i$-th element of $\boldsymbol{\pi}_0$, $(\Phi\boldsymbol{\pi}_0)_k$ be the $j$-th element of $\Phi\boldsymbol{\pi}_0$ for $k = 1, \cdots, m$, $\sigma_x^2$ be the variance of $x_{t-j/m}$ for any $t = 1, \cdots, T$, $j = 0, \cdots, m-1$. Suppose the parameter in the HF AR(1) process is $d$ such that $0 < |d| < 1$ (for i.i.d. case, let $d = 0$ and define $0^0 = 1$), then we have

$$\begin{aligned}
\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) &= \sum_{j=1}^{m}\sum_{i=1}^{m}\pi_{0,i}\phi_{i,j}\pi_j(\theta) = \sum_{j=1}^{m}\sum_{i=1}^{m}\pi_{0i}d^{|i-j|}\sigma_x^2\pi_j(\theta), \\
\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0 &= \sum_{j=1}^{m}\sum_{i=1}^{m}\pi_{0,i}\phi_{i,j}\pi_{0,j} = \sum_{j=1}^{m}\sum_{i=1}^{m}\pi_{0i}d^{|i-j|}\sigma_x^2\pi_{0j}, \\
(\Phi\boldsymbol{\pi}_0)_k &= \sum_{j=1}^{m}d^{|k-j|}\sigma_x^2\pi_{0,j}.
\end{aligned} \tag{A.11}$$

As we mentioned above, the weighted matrix $\Upsilon = [\Upsilon_1\,\Upsilon_2]$ is defined in (2.10). Let $S_\pi = \sum_{i=1}^{m}(2 - i/m)^{4\theta}$, $S_{\Upsilon_1} = \sum_{i=1}^{m}0.9^{i-1}$, $S_{\Upsilon_2} = \sum_{i=1}^{m}(m+1-i)$. $\boldsymbol{\pi}(\theta) = (\pi_1(\theta), \cdots, \pi_m(\theta))'$, here $\pi_j(\theta) = (2-j/m)^{4\theta}/\sum_{i=1}^{m}(2-i/m)^{4\theta}$ for $j = 1, 2, \cdots, m$. Consider two cases separately: (i) $\mathbf{x}_t$ is an i.i.d. sequence ($\Phi = \sigma_x^2 I$ where $I$ is the identity matrix); (ii) $\mathbf{x}_t$ is an AR(1) process with parameter $d$ where $0 < |d| < 1$.

(i) When $\mathbf{x}_t$ is an i.i.d. sequence, then we can easily derive the following equations from (A.39).

$$E\left(\mathbf{z}_t\varepsilon_t^A\right) = \beta_1\sigma_x^2\Upsilon'\left(\boldsymbol{\pi}(\theta) - \frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0}\boldsymbol{\pi}_0\right) = \beta_1\sigma_x^2\Upsilon'\boldsymbol{\pi}(\theta) - \beta_1\sigma_x^2\left(\frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0}\Upsilon'\boldsymbol{\pi}_0\right). \tag{A.12}$$

Since $\Upsilon_r'\boldsymbol{\pi}(\theta)$ does not depend on the null $\boldsymbol{\pi}_0$, then we consider the first term for both the flat aggregation and the general case of end-of-period sampling. Since $\theta > 0$, $S_\pi = O(m)$ and $1 \leq (2 - i/m)^{4\theta} \leq 2^{4\theta}$ for $i = 1, \cdots, m$, then

$$\Upsilon_r'\boldsymbol{\pi}(\theta) = (S_\pi S_{\Upsilon_1})^{-1}\sum_{i=1}^{m} a_{i,r}(2 - i/m)^{4\theta} \in [(S_\pi)^{-1}, 2^{4\theta}(S_\pi)^{-1}] = O(m^{-1}). \tag{A.13}$$

Consider the time-averaging weights $\boldsymbol{\pi}_0$ with two cases respectively: (a) the flat aggregation weights $\boldsymbol{\pi}_0 = (1/m, \cdots, 1/m)'$; (b) $\boldsymbol{\pi}_0 = (\pi_{0,1}, \cdots, \pi_{0,n}, 0, \cdots, 0)'$ for any fixed integer $n \in [0, m)$ independent of $m$ such that $\pi_{0,i}$ is positive constants independent of $m$ for all $i = 1, \cdots, n$ and $\sum_{i=1}^{n}\pi_{0,i} = 1$. In particular, when $n = 1$, it is the end-of-period sampling. Note that for case (b), we can assumed that $\boldsymbol{\pi}_0 = (0, \cdots, 0, \pi_{0,m-n+1}, \cdots, \pi_{0,m})'$ or any fixed $n$ element with positive values of $\boldsymbol{\pi}_0$ with the property $\sum_{i=1}^{m}\pi_{0,i} = 1$. The proof will be straightforward by following similar processes shown below. Without loss of generality, we only show the proof with the aggregating weight as $\boldsymbol{\pi}_0 = (\pi_{0,1}, \cdots, \pi_{0,n}, 0, \cdots, 0)'$.

For case (a),

$$\frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0} = \frac{\sum_{i=1}^m \pi_{0,i}\pi(\theta)}{\sum_{i=1}^m \pi_{0,i}^2} = \frac{1/m\sum_{i=1}^m \pi(\theta)}{m\cdot(1/m^2)} = 1, \quad \Upsilon_r'\boldsymbol{\pi}_0 = 1/m, \text{ for } r=1,2. \quad \text{(A.14)}$$

Then, it follows that the second term $\frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0}\Upsilon_r'\boldsymbol{\pi}_0 = O(m^{-1})$.

Hence, $E\left(\mathbf{z}_t\varepsilon_t^A\right) = (O(m^{-1}),\ O(m^{-1}))'$*.

For case (b),

$$\frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0} = \frac{\sum_{i=1}^n \pi_{0,i}\pi(\theta)}{\sum_{i=1}^n \pi_{0,i}^2} \leq \frac{(2-1/m)^{4\theta}\sum_{i=1}^n \pi_{0,i}}{S_\pi\sum_{i=1}^n \pi_{0,i}^2} = O(m^{-1}). \quad \text{(A.15)}$$

$$\begin{aligned} |\Upsilon_1'\boldsymbol{\pi}_0| &\leq \sigma_x^2\frac{\sum_{i=1}^n 0.9^{i-1}}{S_{\Upsilon_1}}\max_{1\leq i\leq n}(\pi_{0,i}) \leq \sigma_x^2\frac{1-0.9^n}{1-0.9^m} \leq 0.1\sigma_x^2 = O(1), \\ |\Upsilon_2'\boldsymbol{\pi}_0| &\leq \sigma_x^2\frac{(m+m+1-n)n}{(m+1)m}\max_{1\leq i\leq n}(\pi_{0,i}) \leq \frac{(2m+1-n)n}{(m+1)m}\sigma_x^2 = O(m^{-1}). \end{aligned} \quad \text{(A.16)}$$

It implies that the second term follows

$$\left|\frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0}\Upsilon_1'\boldsymbol{\pi}_0\right| = O(m^{-1}), \quad \left|\frac{\boldsymbol{\pi}_0'\boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0'\boldsymbol{\pi}_0}\Upsilon_2'\boldsymbol{\pi}_0\right| = O(m^{-2}). \quad \text{(A.17)}$$

Since the first term dominantly determine the order of $E\left(\mathbf{z}_t\varepsilon_t^A\right)$, then we can derive that $E\left(\mathbf{z}_t\varepsilon_t^A\right) = (O(m^{-1}),\ O(m^{-1}))'$.

*The notation $\left(O(m^{-1}),\ O(m^{-1})\right)'$ indicates that each element of this vector is equal to $O(m^{-1})$.

We have proved that with the i.i.d. HF regressor, our choice of IVs satisfies Assumption 2.1(f) asymptotically in case (i). In case (ii) where the HF regressor is an AR(1) process, similar results can be drawn with either the flat aggregation or the end-of-period sampling in the more general scenario.

(ii) When $\mathbf{x}_t$ is an AR(1) sequence with the parameter $|d| \in (0,1)$, recall (A.39),

$$\begin{aligned} E\left(\mathbf{z}_t \varepsilon_t^A\right) &= \beta_1 \Upsilon' \left(\Phi \boldsymbol{\pi}(\theta) - (\boldsymbol{\pi}_0' \Phi \boldsymbol{\pi}_0)^{-1} \boldsymbol{\pi}_0' \Phi \boldsymbol{\pi}(\theta) \Phi \boldsymbol{\pi}_0\right) \\ &= \beta_1 \Upsilon' \Phi \boldsymbol{\pi}(\theta) - \beta_1 \Upsilon' \left(\frac{\boldsymbol{\pi}_0' \Phi \boldsymbol{\pi}(\theta)}{\boldsymbol{\pi}_0' \Phi \boldsymbol{\pi}_0} \Phi \boldsymbol{\pi}_0\right). \end{aligned} \tag{A.18}$$

Similar to the i.i.d. case, the first term $\Phi\boldsymbol{\pi}(\theta)$ does not depend on the form of $\boldsymbol{\pi}_0$, then let $(\Phi\boldsymbol{\pi}(\theta))_k$ be the $k$-th element $\Phi\boldsymbol{\pi}(\theta)$ for $k = 1, \cdots, m$,

$$(\Phi \boldsymbol{\pi}(\theta))_k = \sigma_x^2 \sum_{j=1}^{m} d^{|k-j|} \pi_j = \sigma_x^2 \left( \sum_{i=k}^{m} d^{i-k} \pi_i + \sum_{j=1}^{k-1} d^{k-j} \pi_j{}^{\dagger} \right). \tag{A.19}$$

Note that when $k = 1$, let $\sum_{j=1}^{k-1} d^j \pi_{k-j} = 0$.

Recall that in (2.10), we define $\Upsilon_1$ and $\Upsilon_2$ as

$$\begin{aligned} \Upsilon_1 &= (f_1(1), f_1(2), \cdots, f_1(m))', \text{ where } f_1(j) = 0.9^{j-1} / \sum_{i=1}^{m} 0.9^{i-1}, \\ \Upsilon_2 &= (f_2(1), f_2(2), \cdots, f_2(m))', \text{ where } f_2(j) = 2(m+1-j)/\{m(m+1)\}, \end{aligned} \tag{A.20}$$

$^{\dagger}$To simplify the notation, we will use $\pi_j$ as $j$-th element of $\boldsymbol{\pi}(\theta)$ instead of $\pi_j(\theta)$.

for $j = 1, \cdots, m$.

Since $S_\pi = \sum_{i=1}^m \pi_i = \sum_{i=1}^m (2 - i/m)^{4\theta} \in [m, 2^{4\theta} m]$, for $r = 1, 2$,

$$\begin{aligned}
|\Upsilon_r' \Phi \boldsymbol{\pi}(\theta)| &= \sigma_x^2 \left| \sum_{k=1}^m f_r(k) (\Phi \boldsymbol{\pi}(\theta))_k \right| = \sigma_x^2 \left| \sum_{k=1}^m f_r(k) \left( \sum_{i=k}^m d^{i-k} \pi_i + \sum_{j=1}^{k-1} d^{k-j} \pi_j \right) \right| \\
&\leq \sigma_x^2 \sum_{k=1}^m f_r(k) \frac{2^{4\theta}}{S_\pi} \left( \sum_{i=k}^m |d|^{i-k} + \sum_{j=1}^{k-1} |d|^{k-j} \right) \\
&= \sigma_x^2 \cdot \frac{2^{4\theta}}{S_\pi} \cdot \frac{\sum_{k=1}^m f_r(k) \left(1 + |d| - |d|^{m-k+1} - |d|^k\right)}{1 - |d|} \\
&< \sigma_x^2 \cdot \frac{2^{4\theta}}{S_\pi} \cdot \frac{\sum_{k=1}^m f_r(k) \left(1 + |d|\right)}{1 - |d|} \leq m^{-1} \sigma_x^2 C_1(d, \theta),
\end{aligned} \tag{A.21}$$

where $C_1(d, \theta) = \dfrac{2^{4\theta}(1 + |d|)}{1 - |d|}$ depends on $d$ and $\theta$, but is independent of $m$. Therefore, the first term $\Upsilon_r' \Phi \boldsymbol{\pi}(\theta) = O(m^{-1})$ for $r = 1, 2$.

Consider case (a) and (b) mentioned above.

For case (a),

$$\begin{aligned}
\boldsymbol{\pi}_0' \Phi \boldsymbol{\pi}_0 &= \sigma_x^2 \frac{m(1 - d^2) - 2d + 2d^{m+1}}{m^2 (1 - d)^2}, \\
\boldsymbol{\pi}_0' \Phi \boldsymbol{\pi}(\theta) &= \sigma_x^2 \frac{(1 + d) - \sum_{i=1}^m (d^i + d^{m+1-i}) \pi_i}{m(1 - d)}, \\
(\Phi \boldsymbol{\pi}_0)_k &= \sigma_x^2 \frac{1 + d - d^{m-k+1} - d^k}{m(1 - d)},
\end{aligned} \tag{A.22}$$

where $(\Phi \boldsymbol{\pi}_0)_k$ is the $k$-th element of $\Phi \boldsymbol{\pi}_0$.

Based on (A.22), the second term of (A.18) follows

$$
\begin{aligned}
&\left|\Upsilon_r'(\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta)\Phi\boldsymbol{\pi}_0\right| = |(\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\|\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta)\|\Upsilon_r'\Phi\boldsymbol{\pi}_0| \\
=&\frac{\sigma_x^2}{m(1-d^2)-2d+2d^{m+1}}\left|(1+d)-\sum_{i=1}^{m}(d^i+d^{m+1-i})\pi_i\right| \\
&\left|\sum_{k=1}^{m}(1+d-d^{m-k+1}-d^k)f_r(k)\right| \\
\leq&\frac{\sigma_x^2}{|m(1-d^2)-2d|}\left(1+|d|+\left|\sum_{i=1}^{m}(d^i+d^{m+1-i})\pi_i\right|\right) \\
&\left(\sum_{k=1}^{m}(1+|d|+|d|^{m-k+1}+|d|^k)f_r(k)\right) \\
\leq&\frac{\sigma_x^2}{m(1-d^2)-2|d|}\left(1+|d|+(|d|+|d|)\sum_{i=1}^{m}\pi_i\right)\left((1+|d|+|d|+|d|)\sum_{k=1}^{m}f_r(k)\right) \\
\leq&\frac{\sigma_x^2(1+3|d|)^2}{m(1-d^2)-2|d|}=O(m^{-1}).
\end{aligned}
\tag{A.23}
$$

Hence, both the first term and the second term of (A.18) are $O(m^{-1})$ for two IVs. It follows that $E(\mathbf{z}_t\varepsilon_t^A) = (O(m^{-1}),\ O(m^{-1}))'$.

Now, consider case (b), the general case of the end-of-period sampling. We still assume that $\boldsymbol{\pi}_0 = (\pi_{0,1}, \cdots, \pi_{0,n}, 0, \cdots, 0)'$ for any integer $n \in [0, m)$ independent of $m$ such that $\pi_{0,i}$ is positive constants independent of $m$ for all $i = 1, \cdots, n$ and $\sum_{i=1}^{n}\pi_{0,i} = 1$. Since we assume that only the first $n$ elements can be assigned with positive values which are no greater than 1, then the $k$-th element of $\boldsymbol{\pi}_0'\Phi$ is

$$(\boldsymbol{\pi}_0'\Phi)_k = \begin{cases} \sigma_x^2 \left(\sum_{i=k}^n \pi_{0,i} d^{i-k} + \sum_{j=1}^{k-1} \pi_{0,j} d^j\right), & 1 \le k \le n, \\ \sigma_x^2 d^{k-n} \sum_{p=1}^n d^{n-p} \pi_{0,p}, & n < k \le m. \end{cases} \tag{A.24}$$

Then, similar to the i.i.d. case, we can derive the followings for $r = 1, 2$.

$$\begin{aligned}
\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0 =& \sigma_x^2 \sum_{k=1}^n \left(\sum_{i=k}^n \pi_{0,i} d^{i-k} + \sum_{j=1}^{k-1} \pi_{0,j} d^j\right) \pi_{0,k} = \sigma_x^2 D_0(d, n; \boldsymbol{\pi}_0), \\
\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta) =& \sigma_x^2 \sum_{k=1}^n \left(\sum_{i=k}^n \pi_{0,i} d^{i-k} + \sum_{j=1}^{k-1} \pi_{0,j} d^j\right) \pi_k \\
&+ \sigma_x^2 \sum_{k=n+1}^m \left(d^{k-n} \sum_{p=1}^n d^{n-p} \pi_{0,p}\right) \pi_k \\
\le& \sigma_x^2 \frac{2^{4\theta}}{S_\pi} \left(D_1(d, n; \boldsymbol{\pi}_0) + \left(\sum_{k=n+1}^m d^{k-n} \pi_k \cdot \sum_{p=1}^n d^{n-p} \pi_{0,p}\right)\right) \\
\le& \sigma_x^2 \frac{2^{4\theta}}{S_\pi} \left(D_1(d, n; \boldsymbol{\pi}_0) + \frac{1 - d^{m-n+1}}{1-d} D_2(d, n; \boldsymbol{\pi}_0)\right), \\
\Upsilon_r'\Phi\boldsymbol{\pi}_0 =& \sigma_x^2 \sum_{k=1}^n \left(\sum_{i=k}^n \pi_{0,i} d^{i-k} + \sum_{j=1}^{k-1} \pi_{0,j} d^j\right) f_r(k) \\
&+ \sigma_x^2 \sum_{k=n+1}^m \left(d^{k-n} \sum_{p=1}^n d^{n-p} \pi_{0,p}\right) f_r(k) \\
\le& \sigma_x^2 \max_{1 \le k \le m} f_r(k) \cdot \left(D_1(d, n; \boldsymbol{\pi}_0) + \frac{1 - d^{m-n+1}}{1-d} D_2(d, n; \boldsymbol{\pi}_0)\right),
\end{aligned} \tag{A.25}$$

where $D_1(d, n; \boldsymbol{\pi}_0) = \sum_{k=1}^n \left(\sum_{i=k}^n \pi_{0,i} d^{i-k} + \sum_{j=1}^{k-1} \pi_{0,j} d^j\right)$ and $D_2(d, n; \boldsymbol{\pi}_0) =$

$\sum_{p=1}^{n} d^{n-p}\pi_{0,p}$ relies on $d,\ n$ and $\boldsymbol{\pi}_0$. Therefore, we can derive that

$$
\begin{aligned}
&\left|\Upsilon_r'(\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}_0)^{-1}\boldsymbol{\pi}_0'\Phi\boldsymbol{\pi}(\theta)\Phi\boldsymbol{\pi}_0\right| \\
&\leq \frac{\sigma_x^2\cdot\max_{1\leq k\leq m} f_r(k)}{|D_0(d,n;\boldsymbol{\pi}_0)|}\cdot\frac{2^{4\theta}}{S_\pi}\cdot\left(D_1(d,n;\boldsymbol{\pi}_0)+\frac{1-d^{m-n+1}}{1-d}D_2(d,n;\boldsymbol{\pi}_0)\right)^2 = O(m^{-1}).
\end{aligned} \tag{A.26}
$$

Hence, both the first term and the second term of (A.18) are $O(m^{-1})$ for two IVs. It follows that $E(\mathbf{z}_t\varepsilon_t^A) = (O(m^{-1}),\ O(m^{-1}))'$.

Therefore, for either the i.i.d. or the AR(1) HF regressor, $E(z_{r,t}\varepsilon_t^A) = O(m^{-1})$ for $r = 1,2$ can be satisfied with either the flat aggregation $\boldsymbol{\pi}_0 = (1/m,\cdots,1/m)'$ or the general case of the end-of-period sampling $\boldsymbol{\pi}_0 = (\pi_{0,1},\cdots,\pi_{0,n},0,\cdots,0)'$.

□

## A.2 Proof of Theorems in Chapter 3

### A.2.1 Convergence of the Clustering Algorithm

*Proof of Theorem 3.1.* The proof of Theorem 3.1 can be separated into two parts.

$\|\boldsymbol{\kappa}^{s+1}\|_2^2 \xrightarrow{s\to\infty} 0$ can be shown similarly to the proof of Proposition 1 in Ma and Huang [55].

Refer to the proof of Theorem 3.1 in Zhu and Qu [74], the proof of $\|\boldsymbol{\tau}^{s+1}\|_2^2 \xrightarrow{s\to\infty} 0$ can be done by simply ignoring the penalty term in the objective function. The rest of the proof will be similar. □

### A.2.2 Convergence of the Estimators

Before we start the proof of Theorem 3.2 and 3.3, we shall prove some lemmas in advance.

**Lemma A.1.** *Suppose a random vector* $\boldsymbol{\varepsilon} = (\varepsilon_{1,1}, \varepsilon_{1,2}, \ldots, \varepsilon_{n,T})'$ *of length* $nT$ *as in (3.12) satisfies Assumption 3.7. Let* $A \in \mathbf{R}^{a\times nT}$ *be a nonrandom matrix with a positive integer* $a \leq nT$*. Let* $\Sigma = A'A$*. For any* $\zeta > 0$*,*

$$P\left[\|A\boldsymbol{\varepsilon}\|_2^2 > 2\tilde{c}\{\mathrm{tr}(\Sigma) + 2\sqrt{\mathrm{tr}(\Sigma^2)}\zeta + 2\|\Sigma\|_2\zeta\}\right] \leq e^{-\zeta}.$$

*Proof of Lemma A.1.* When $a = nT$, this lemma is a special case of Theorem 2.1 in [45]. This can be easily seen by recognizing their $\mu$, $\sigma^2$, and $\alpha$ are 0, $2\tilde{c}$, and $(\nu_{1,1}, \nu_{1,2}, \ldots, \nu_{n,T})'$, respectively.

If $a < nT$, a similar argument can still be used. Consider a singular value decomposition of $A = USV'$, where $U$ and $V$ are $a \times a$ and $nT \times nT$ orthogonal matrices, respectively. Let $\rho = (\rho_1, \ldots, \rho_a)'$ denote the nonzero eigenvalues of $A'A$ and $AA'$. $S$

is an $a \times nT$ matrix, where its diagonal elements are equal to $\sqrt{\rho_i}$ for $i = 1, \ldots, a$ and all other entries are zero. Let $z$ be a vector of $a$ independent standard Gaussian random variables. Since $U$ is orthogonal, $y = U'z$ is also an $a \times 1$ vector of $a$ independent standard Gaussian random variables. Let $y = (y_1, \ldots, y_a)'$. Applying Lemma 2.4 of [45] on $\|A'z\|^2 = z'AA'z = z'USV'VS'U'z = ySS'y' = \sum_{i=1}^{a} \rho_i y_i^2$, we have

$$E\left\{\exp\left(\gamma\|A'z\|^2\right)\right\} \leq \exp\left(\|\rho\|_1\gamma + \frac{\|\rho\|_2^2\gamma^2}{1 - 2\|\rho\|_\infty\gamma}\right) \tag{A.27}$$

for any $0 \leq \gamma < 1/(2\|\rho\|_\infty)$. For any $\lambda \in \mathbf{R}$ and $\delta \geq 0$, using a similar argument in (2.3) and (2.4) of [45], Assumption 3.7, and (A.27),

$$P(\|A\boldsymbol{\varepsilon}\|^2 > \delta) \leq \exp\left(-\frac{\lambda^2\delta}{2}\right)\exp\left\{\|\rho\|_1(\lambda^2\tilde{c}) + \frac{\|\rho\|_2^2(\lambda^2\tilde{c})^2}{1 - 2\|\rho\|_\infty(\lambda^2\tilde{c})}\right\}.$$

Let $\delta = 2\tilde{c}(\|\rho\|_1 + \tau)$, $\lambda^2 = \frac{1}{\tilde{c}}\frac{1}{2\|\rho\|_\infty}\left(1 - \sqrt{\frac{\|\rho\|_2^2}{\|\rho\|_2^2 + 2\|\rho\|_\infty\tau}}\right)$, and $\tau = 2\sqrt{\|\rho\|_2^2\zeta} + 2\|\rho\|_\infty\zeta$. The desired proof is concluded by using similar arguments as [45] and observing $\|\rho\|_1 = \sum_{i=1}^{a}\rho_i = \mathrm{tr}(\Sigma)$, $\|\rho\|_2^2 = \sum_{i=1}^{a}\rho_i^2 = \mathrm{tr}(\Sigma^2)$, and $\|\rho\|_\infty = \max_i \rho_i = \|\Sigma\|_2$. $\square$

**Lemma A.2.** *Suppose Assumptions 3.5 and 3.7 hold. Then given any matrix $W$, $\zeta^* > 0$ and $\zeta > 0$,*

$$P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > 2\tilde{c}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)\|W'W\|_2 \,\middle|\, W\right] \leq e^{-\zeta^*},$$
$$P\left[\|\Gamma'W'\boldsymbol{\varepsilon}\|_2^2 > 2\tilde{c}(Gp + 2\sqrt{Gp\zeta} + 2\zeta)\|\Gamma'W'W\Gamma\|_2 \,\middle|\, W\right] \leq e^{-\zeta}.$$

*Proof of Lemma A.2.* Given any matrix $W$ with the conditions in Lemma A.1, for any $\zeta^* > 0$, by Theorem 2.1 in Hsu et al. [45],

$$P\left[\|W'\varepsilon\|_2^2 > 2\tilde{c}(\mathrm{tr}(WW') + 2\sqrt{\mathrm{tr}((WW')^2)\zeta^*} + 2\|WW'\|_2\zeta^*) \mid W\right] \le e^{-\zeta^*},$$

$$P\left[\|\Gamma'W'\varepsilon\|_2^2 > 2\tilde{c}(\mathrm{tr}(\Gamma WW'\Gamma') + 2\sqrt{\mathrm{tr}((\Gamma WW'\Gamma')^2)\zeta} + 2\|\Gamma WW'\Gamma'\|_2\zeta) \mid W\right] \le e^{-\zeta}.$$

Since $\|WW'\|_2$ is the maximum eigenvalue of $WW'$ and using the fact that $WW'$ is symmetric and positive definite with rank $np$, then $\lambda_{max}(WW') = \lambda_{max}(W'W)$, and

$$\|WW'\|_2 = \|W'W\|_2 = \|diag(W_1'W_1, \cdots, W_n'W_n)\|_2 \le \max_i \|W_i'W_i\|_2,$$

$$\mathrm{tr}(WW') = tr(W'W) \le np\|W'W\|_2, \quad \mathrm{tr}((WW')^2) = tr((W'W)^2) \le np\|W'W\|_2^2,$$

then

$$\mathrm{tr}(WW') + 2\sqrt{\mathrm{tr}[(WW')^2]\zeta^*} + 2\|WW'\|_2\zeta^* \le (np + 2\sqrt{np\zeta^*} + 2\zeta^*)\|W'W\|_2.$$

Similarly, $\|W\Gamma\Gamma'W'\|_2 = \|\Gamma'W'W\Gamma\|_2$, and

$$\mathrm{tr}(W\Gamma\Gamma'W') = \mathrm{tr}(\Gamma'W'W\Gamma) \le Gp\lambda_{\max}(\Gamma'W'W\Gamma) = Gp\|\Gamma'W'W\Gamma\|_2, \tag{A.28}$$

$$\mathrm{tr}\{(W\Gamma\Gamma'W')^2\} = \mathrm{tr}\{(\Gamma'W'W\Gamma)^2\} \le Gp\{\lambda_{\max}(\Gamma'W'W\Gamma)\}^2 = Gp\|\Gamma'W'W\Gamma\|_2^2. \tag{A.29}$$

Therefore we have for any $\zeta > 0$,

$$\begin{aligned}
&\mathrm{tr}(\Gamma' W' W \Gamma) + 2\sqrt{\mathrm{tr}\{(\Gamma' W' W \Gamma)^2\}}\sqrt{\zeta} + 2\|\Gamma' W' W \Gamma\|_2 \zeta \\
\leq &(Gp + 2\sqrt{Gp\zeta} + 2\zeta)\|\Gamma' W' W \Gamma\|_2.
\end{aligned} \tag{A.30}$$

As a result, we have shown the inequalities in the statement given any matrix $W$. $\square$

**Lemma A.3.** *Suppose Assumptions 3.5 and 3.7 hold, let*

$$\begin{aligned}
S_\zeta :=& 2\tilde{c}(Gp + 2\sqrt{Gp\zeta} + 2\zeta) g_{\max} m \tilde{M} \sqrt{GpT} B_{q,m}, \\
S_{\zeta^*} :=& 2\tilde{c}(np + 2\sqrt{np\zeta^*} + 2\zeta^*) m \tilde{M} \sqrt{T}\sqrt{p} B_{q,m},
\end{aligned}$$

*where* $B_{q,m} = (q^{1/2} + m^{1/2}(L + 1 + 2K))$, $p = q + L + 1 + 2K$, $\tilde{M} = \max(M_1, M_2, M_3, M_4)$ *and* $\tilde{c}$ *given in Assumption 3.5 and 3.7, then* $P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*}\right] \leq e^{-\iota^*}$ *and* $P\left[\|\Gamma' W'\boldsymbol{\varepsilon}\|_2^2 > S_\zeta\right] \leq e^{-\iota}$ *where* $\iota = \min(\zeta, -\log(\epsilon)) - \log(2)$ *and* $\iota^* = \min(\zeta^*, -\log(\epsilon)) - \log(2)$ *for any* $\zeta$ *and* $\zeta^*$ *in Lemma A.2.*

*Proof of Lemma A.3.* Based on the iteration expectation, we have

$$\begin{aligned}
&E\left[P\left(\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*} \;\middle|\; W\right)\right] = P\left[\|W'\boldsymbol{\varepsilon}\|_2 > S_{\zeta^*}\right] \\
=&E\left[I_{\{\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*}\}} \mid \|WW'\|_2 \leq M^*\right] P(\|WW'\|_2 \leq M^*) \\
&+ E\left[I_{\{\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*}\}} \mid \|WW'\|_2 > M^*\right] P(\|WW'\|_2 > M^*) \\
=&P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*} \;\middle|\; \|WW'\|_2 \leq M^*\right] P(\|WW'\|_2 \leq M^*) \\
&+ P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*} \;\middle|\; \|WW'\|_2 > M\right] P(\|WW'\|_2 > M^*).
\end{aligned}$$

Since $\|M\|_\infty \leq m$ and $\|M'\|_\infty \leq L+1+2K$ as all elements of $M$ in (3.4) smaller than 1 in magnitude, then for any $\epsilon > 0$ defined in Assumption 3.5, with probability at least $1-\epsilon$,

$$
\begin{aligned}
\left\|\sum_{i\in\mathcal{G}_g} U_i'U_i\right\|_\infty &= \sum_{i\in\mathcal{G}_g} \|U_i'U_i\|_\infty \leq M_1|\mathcal{G}_g|\sqrt{qT}, \\
\left\|\sum_{i\in\mathcal{G}_g} U_i'\tilde{X}_i\right\|_\infty &\leq \sum_{i\in\mathcal{G}_g} \|U_i'X_i\|_\infty \|M'\|_\infty \leq M_3|\mathcal{G}_g|\sqrt{mT}(L+1+2K), \\
\left\|\sum_{i\in\mathcal{G}_g} \tilde{X}_i'U_i\right\|_\infty &\leq \|M\|_\infty \sum_{i\in\mathcal{G}_g} \|U_i'X_i\|_\infty \leq M_4|\mathcal{G}_g|m\sqrt{qT}, \\
\left\|\sum_{i\in\mathcal{G}_g} \tilde{X}_i'\tilde{X}_i\right\|_\infty &\leq \|M\|_\infty \sum_{i\in\mathcal{G}_g} \|X_i'X_i\|_\infty \|M'\|_\infty \leq M_2|\mathcal{G}_g|m\sqrt{mT}(L+1+2K).
\end{aligned}
\tag{A.31}
$$

The inequalities further imply that with probability at least $1-\epsilon$, take $\tilde{M} = \max\{M_1, M_2, M_3, M_4\}$,

$$
\left\|\sum_{i\in\mathcal{G}_g} U_i'U_i\right\|_\infty \leq \tilde{M}|\mathcal{G}_g|\sqrt{qT}, \quad \left\|\sum_{i\in\mathcal{G}_g} U_i'\tilde{X}_i\right\|_\infty \leq \tilde{M}|\mathcal{G}_g|\sqrt{mT}(L+1+2K),
$$
$$
\left\|\sum_{i\in\mathcal{G}_g} \tilde{X}_i'U_i\right\|_\infty \leq \tilde{M}|\mathcal{G}_g|m\sqrt{qT}, \quad \left\|\sum_{i\in\mathcal{G}_g} \tilde{X}_i'\tilde{X}_i\right\|_\infty \leq \tilde{M}|\mathcal{G}_g|m\sqrt{mT}(L+1+2K).
$$

Therefore, with probability at most $1-\epsilon$,

$$\begin{aligned}\|WW'\|_2 = \|W'W\|_2 = \|diag(W_1'W_1,\cdots,W_n'W_n)\|_2 &\le \sup_i \|W_i'W_i\|_2 \\ &\le \sqrt{p}\sup_i \|W_i'W_i\|_\infty = \sqrt{p}\sup_i \left\| \begin{matrix} U_i'U_i & U_i'\tilde{X}_i \\ \tilde{X}_i'U_i & \tilde{X}_i'\tilde{X}_i \end{matrix} \right\|_\infty \\ &\le \tilde{M}m\sqrt{T}B_{q,m}\sqrt{p},\end{aligned}$$

and we have

$$tr(WW') = tr(W'W) \le np\|W'W\|_2, \quad tr((WW')^2) = tr((W'W)^2) \le np\|W'W\|_2^2.$$

As a result,

$$tr(WW') + 2\sqrt{tr[(WW')^2]\zeta^*} + 2\|WW'\|_2\zeta^* \le (np + 2\sqrt{np\zeta^*} + 2\zeta^*)\|WW'\|_2$$

Since for any $\epsilon > 0$, there exists some $M^* = \tilde{M}m\sqrt{T}B_{q,m}\sqrt{p}$ such that $P[\|WW'\|_2 > M^*] \le \epsilon$, then

$$P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*} \,\middle|\, W, \|WW'\|_2 \le M^*\right] \le e^{-\zeta^*}, \;\; 1-\epsilon < P(\|WW'\|_2 \le M^*) \le 1,$$

$$P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*} \,\middle|\, W, \|WW'\|_2 > M^*\right] \le 1, \quad P(\|WW'\|_2 > M^*) \le \epsilon.$$

Therefore, $P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*}\right] \le e^{-\zeta^*} + \epsilon$ where $S_{\zeta^*} = 2\tilde{c}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)M^*$.

Without loss of generality, for large $\zeta^* > 1$ as well as $\epsilon \leq 1$, let $\tilde{\zeta}^* = \min\{\zeta^*, -\log(\epsilon)\}$, then

$$e^{-\zeta^*} + \epsilon = e^{-\zeta^*} + e^{\log(\epsilon)} = e^{-\tilde{\zeta}^*}(1 + e^{-|\zeta^* + \log(\epsilon)|}) \leq 2e^{-\tilde{\zeta}^*} = e^{\log(2) - \tilde{\zeta}^*}.$$

Take $\iota^* = \tilde{\zeta}^* - \log(2)$, then $P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > S_{\zeta^*}\right] \leq e^{-\iota^*}$. For large enough $\tilde{\zeta}^*$, $\log(2)$ is negligible. Similarly, we can find $S_\zeta$ in $P\left[\|\Gamma' W'\boldsymbol{\varepsilon}\|_2^2 > S_\zeta\right] \leq e^{-\iota}$ as the following.

A straightforward calculation derives that

$$\Gamma' W' W \Gamma = \operatorname{diag}\left(\sum_{i \in \mathcal{G}_1} W_i' W_i, \ldots, \sum_{i \in \mathcal{G}_G} W_i' W_i\right).$$

It follows that, with probability $1 - \epsilon$,

$$\begin{aligned}\|\Gamma' W' W \Gamma\|_\infty &= \max_{1 \leq g \leq G} \left\|\sum_{i \in \mathcal{G}_g} W_i' W_i\right\|_\infty \leq \max_{1 \leq g \leq G} \sum_{i \in \mathcal{G}_g} \|W_i' W_i\|_\infty \leq g_{max} \sup_{1 \leq i \leq n} \|W_i' W_i\|_\infty \\ &\leq g_{\max} m \tilde{M} \sqrt{T} B_{q,m},\end{aligned}$$

and therefore,

$$\|\Gamma' W' W \Gamma\|_2 \leq \sqrt{Gp} \|\Gamma' W' W \Gamma\|_\infty \leq g_{\max} m \tilde{M} \sqrt{GpT} B_{q,m}. \tag{A.32}$$

For any $\epsilon > 0$, there exists some $M = g_{\max} m \tilde{M} \sqrt{T} B_{q,m}$, such that $P[\|W\Gamma\Gamma' W'\|_2 >$

$M] \leq \epsilon$, then

$$P\left[\|\Gamma' W'\boldsymbol{\varepsilon}\|_2^2 > S_\zeta \;\middle|\; W, \|W\Gamma\Gamma' W'\|_2^2 \leq M\right] \leq e^{-\zeta}, \;\; 1-\epsilon < P(\|W\Gamma\Gamma' W'\|_2 \leq M) \leq 1,$$

$$P\left[\|W'\boldsymbol{\varepsilon}\|_2 > S_\zeta \;\middle|\; W, \|W\Gamma\Gamma' W'\|_2 > M\right] \leq 1, \;\; P(\|W\Gamma\Gamma' W'\|_2 > M) \leq \epsilon.$$

Therefore, $P\left[\|\Gamma' W'\boldsymbol{\varepsilon}\|_2^2 > S_\zeta\right] \leq e^{-\zeta}+\epsilon$ where $S_\zeta = 2\tilde{c}(Gp+2\sqrt{Gp\zeta}+2\zeta)M$. Similarly, take $\iota = \min\{\zeta, -\log(\epsilon)\} - \log(2)$, then $P\left[\|\Gamma' W'\boldsymbol{\varepsilon}\|_2^2 > S_\iota\right] \leq e^{-\iota}$. □

#### A.2.2.1 Convergence of the Oracle Estimator

With the help of Lemma A.1 – Lemma A.3, we further prove Theorem 3.2.

*Proof of Theorem 3.2.* The definition of $\Gamma$ and $\mathbf{y} = W\boldsymbol{\gamma}^{or} + \boldsymbol{\varepsilon}$ lead to

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0 &= \Gamma(\Gamma' W' W \Gamma)^{-1}\Gamma' W'\boldsymbol{\varepsilon} \\
&= \Gamma\left\{\operatorname{diag}\left(\textstyle\sum_{i\in\mathcal{G}_1} W_i' W_i, \ldots, \sum_{i\in\mathcal{G}_G} W_i' W_i\right)\right\}^{-1} \begin{pmatrix} \sum_{i\in\mathcal{G}_1} W_i'\boldsymbol{\varepsilon}_i \\ \vdots \\ \sum_{i\in\mathcal{G}_G} W_i'\boldsymbol{\varepsilon}_i \end{pmatrix},
\end{aligned}$$

where for any $g \in \{1, \ldots, G\}$,

$$\sum_{i\in\mathcal{G}_g} W_i' W_i = \begin{pmatrix} \sum_{i\in\mathcal{G}_g} U_i' U_i & (\sum_{i\in\mathcal{G}_g} U_i' X_i)\mathbf{M}' \\ \mathbf{M}(\sum_{i\in\mathcal{G}_g} X_i' U_i) & \mathbf{M}(\sum_{i\in\mathcal{G}_g} X_i' X_i)\mathbf{M}' \end{pmatrix}$$

and

$$\sum_{i \in \mathcal{G}_g} W_i' \boldsymbol{\varepsilon}_i = \begin{pmatrix} \sum_{i \in \mathcal{G}_g} U_i' \boldsymbol{\varepsilon}_i \\ \mathbf{M}(\sum_{i \in \mathcal{G}_g} X_i' \boldsymbol{\varepsilon}_i) \end{pmatrix}.$$

Assumption 3.5 implies that

$$\lambda_{\min}(\Gamma' W' W \Gamma) \geq c g_{\min} T,$$

so that

$$\|(\Gamma' W' W \Gamma)^{-1}\|_\infty \leq \sqrt{Gp} \|(\Gamma' W' W \Gamma)^{-1}\|_2 \leq \sqrt{Gp} (c g_{\min} T)^{-1}. \tag{A.33}$$

For all $p$-norms, $\|A \otimes B\| = \|A\| \|B\|$ holds (for example, see p. 433 of Langville and Stewart [50]),

$$\|\Gamma\|_\infty \leq \|\Pi\|_\infty \|I_p\|_\infty = 1. \tag{A.34}$$

Lemma A.3, equations (A.33) and (A.34), and the triangle inequality imply that for any $\iota > 0$,

$$\begin{aligned} \|\widehat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0\|_\infty &\leq \|\Gamma\|_\infty \|(\Gamma' W' W \Gamma)^{-1}\|_\infty \|\Gamma' W' \boldsymbol{\varepsilon}\|_\infty \\ &\leq (Gp)^{1/2} (c g_{\min} T)^{-1} \|\Gamma' W' \boldsymbol{\varepsilon}\|_2 \leq (Gp)^{1/2} (c g_{\min} T)^{-1} S_\zeta^{1/2}, \end{aligned}$$

with probability at least $1 - e^\iota$.

It results in

$$\phi_{n,T,G,\zeta} := \frac{\sqrt{2\tilde{c}}}{c} \frac{(m\tilde{M}g_{\max})^{1/2}(Gp)^{3/4}}{g_{\min}T^{3/4}} B_{q,m}^{1/2}(Gp + 2\sqrt{Gp}\sqrt{\zeta} + 2\zeta)^{1/2}, \tag{A.35}$$

where $B_{q,m}$ is defined in Lemma A.3. Therefore, with probability at least $1 - e^{-\iota}$,

$$\|\widehat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0\|_\infty \leq \phi_{n,T,G,\zeta}.$$

In the following proof, let $m$ and $q$ be fixed for simplification. It further indicates that $p$ is fixed. Let $C_{q,m} = \frac{\sqrt{2\tilde{c}}}{c} m^{1/2} p^{3/4} B_{q,m}^{1/2}$, (A.35) can be simplified as

$$\phi_{n,T,G} = C_{q,m} \frac{g_{\max}^{1/2} G^{3/4}}{g_{\min} T^{3/4}} (Gp + 2\sqrt{Gp}\sqrt{\zeta} + 2\zeta)^{1/2}. \tag{A.36}$$

1. Consider $T \to \infty$ with $n$ fixed. Let $\zeta \to \infty$ and $\zeta = o(T^{3/2})$. Since $G \leq n \ll \zeta$, then $(Gp + 2\sqrt{Gp}\sqrt{\zeta} + 2\zeta)^{1/2} = O(2\zeta^{1/2})$. Therefore,

   $$\phi_{n,T,G} = C_1 T^{-3/4} O(\zeta^{1/2}) \overset{T\to\infty}{\longrightarrow} 0,$$

   where $C_1 = 2C_{q,m} \frac{g_{\max}^{1/2} G^{3/4}}{g_{\min}}$, which is free of $T$.

   In other cases that we presented as the following, the inequalities of $\phi_{n,T,G}$ are derived as a result of $g_{max} \leq n$ and $G \leq n/g_{min}$.

2. Consider $n \to \infty$ with $T$ fixed.

(a) Consider $G \ll \zeta \to \infty$.

i. When $G$ is fixed, then $(Gp + 2\sqrt{Gp}\sqrt{\zeta} + 2\zeta)^{1/2} = O(2\zeta^{1/2})$. For some constant $\tilde{\alpha}_0 < 1/2$, let $g_{min} = O(n^{1/2+\tilde{\alpha}_0})$, $\zeta = o(n^{2\tilde{\alpha}_0})$ and $\zeta \to \infty$, then

$$\phi_{n,T,G} \leq C_3 \frac{n^{1/2}}{g_{min}} O(\zeta^{1/2}) \stackrel{n\to\infty}{\longrightarrow} 0,$$

where $C_3 = 2C_{q,m}\frac{G^{3/4}}{T^{3/4}}$, which is free of $n$.

ii. When $G \to \infty$, for some constant $\tilde{\alpha}_2 < 2/7$, let $g_{min} = O(n^{5/7+\tilde{\alpha}_2})$, $\zeta = o(n^{7\tilde{\alpha}_2/2})$ and $\zeta \to \infty$, then $(Gp+2\sqrt{Gp\zeta}+2\zeta)^{1/2} = O((p+2\sqrt{p}+2)^{1/2}\zeta^{1/2})$. Since $G \leq n/g_{min}$, then

$$\phi_{n,T,G} \leq C_4 \frac{n^{1/2}G^{3/4}}{g_{min}} O(\zeta^{1/2}) \leq C_4 \frac{n^{5/4}}{g_{min}^{7/4}} O(\zeta^{1/2}) \stackrel{n,G\to\infty}{\longrightarrow} 0,$$

where $C_4 = C_{q,m}\frac{1}{T^{3/4}}(p + 2\sqrt{p} + 2)^{1/2}$, which is free of $n$ and $G$.

(b) Consider $G \to \infty$. Let $g_{min} = O(n^{7/9+\tilde{\alpha}_1})$ for some $\tilde{\alpha}_1 < 2/9$, $\zeta = O(G)$ and $\zeta \to \infty$, then $Gp + 2\sqrt{Gp}\sqrt{\zeta} + 2\zeta = O((p + 2\sqrt{p} + 2)G) = O(G)$. Therefore,

$$\phi_{n,T,G} \leq C_2 \frac{n^{1/2}G^{3/4}}{g_{min}} O(G^{1/2}) \stackrel{n\to\infty}{\longrightarrow} 0,$$

where $C_2 = C_{q,m}\frac{1}{T^{3/4}}(p + 2\sqrt{p} + 2)^{1/2}$, which is free of $n$.

3. Consider $T, n \to \infty$.

(a) Consider $G \ll \zeta \to \infty$,

i. When $G$ is fixed, then $(Gp + 2\sqrt{Gp\zeta} + 2\zeta)^{1/2} = O(2\zeta^{1/2})$. Let $g_{min} = O(n^{1/2+\tilde{\alpha}_0})$ for some positive constant $\tilde{\alpha}_0 < 1/2$ and $\zeta = o(n^{2\tilde{\alpha}_0}T^{3/2})$, $\zeta \to \infty$, then

$$\phi_{n,T,G} \leq C_6 \frac{n^{1/2}}{g_{min}T^{3/4}} O(\zeta^{1/2}) \overset{n,T\to\infty}{\longrightarrow} 0,$$

where $C_6 = 2C_{q,m}G^{3/4}$.

ii. When $G \to \infty$, for some positive constant $\tilde{\alpha}_2 < 2/7$, let $g_{min} = O(n^{5/7+\tilde{\alpha}_2})$ and $G \leq n/g_{min}$, $\zeta = o(n^{7\tilde{\alpha}_2/2}T^{3/2})$ and $\zeta \to \infty$, then $(Gp + 2\sqrt{Gp\zeta} + 2\zeta)^{1/2} = O((p + 2\sqrt{p} + 2)^{1/2}\zeta^{1/2})$. Since $G \leq n/g_{min}$, then

$$\phi_{n,T,G} \leq C_7 \frac{n^{1/2}G^{3/4}}{g_{min}T^{3/4}} O(\zeta^{1/2}) \leq C_7 \frac{n^{5/4}}{g_{min}^{7/4}T^{3/4}} O(\zeta^{1/2}) \overset{n,T,G\to\infty}{\longrightarrow} 0,$$

where $C_7 = C_{q,m}(p + 2\sqrt{p} + 2)^{1/2}$, which is freen of $n, T$ and $G$.

(b) Consider $G \to \infty$. Let $g_{min} = O(n^{7/9+\tilde{\alpha}_1})$ for some constant $\tilde{\alpha}_1 < 2/9$, $\zeta = O(G)$ and $\zeta \to \infty$, then $Gp+2\sqrt{Gp\zeta}+2\zeta = O((p+2\sqrt{p}+2)G) = O(G)$. Since $G \leq n/g_{min}$,

$$\phi_{n,T,G} \leq C_5 \frac{n^{1/2}G^{3/4}}{g_{min}T^{3/4}} O(G^{1/2}) \leq C_5 \frac{n^{7/4}}{g_{min}^{9/4}T^{3/4}} O(1) \overset{n,T,G\to\infty}{\longrightarrow} 0,$$

where $C_5 = C_{q,m}(p + 2\sqrt{p} + 2p)^{1/2}$, which is free from $n, T$ and $G$.

Let $V_i = W_i(\Pi_{i\cdot} \otimes I_p)$ be a $T \times Gp$ matrix, where $\Pi_{i\cdot}$ is the $i$-th row of the matrix $\Pi$, $V = W\Gamma = (V_1', \cdots, V_n')'$. Then, for any $c_n \in \mathbb{R}^{Gp}$ with $\|c_n\|_2 = 1$,

$$c_n'(\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0) = \sum_{i=1}^{n} c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i = \sum_{i=1}^{n} c_n'(V'V)^{-1}\sum_{t=1}^{T}\mathbf{v}_{it}'\varepsilon_{it}. \tag{A.37}$$

Since $\{\boldsymbol{\varepsilon}_i\}$ is assumed to be an i.i.d. subgaussian distributed sequence with mean 0 and variance proxy $2\tilde{c}$, then $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$. Hence,

$$E\left[c_n'(\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0)\right] = 0.$$

Suppose that Assumption 3.5 and 3.7 hold where $\lambda_{\max}(V'V) = \lambda_{\max}(\Gamma'W'W\Gamma) \leq c^*|\mathcal{G}_g|T \leq c^* g_{\max}T$ and $Var(\varepsilon_{it}) = O(2\tilde{c})$, then

$$\sigma_\gamma^2 := Var[c_n'(\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0)] \geq \frac{Var(\varepsilon_{it})}{c^* g_{\max}T}. \tag{A.38}$$

Moreover, for any $\epsilon > 0$, applying Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} E\left((c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i)^2 1\{|c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i| > \epsilon\sigma_\gamma\}\right) \\
\leq &\sum_{i=1}^{n} \left\{E(c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i)^4\right\}^{1/2} \left\{E\left(1\{|c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i| > \epsilon\sigma_\gamma\}^2\right)\right\}^{1/2} \\
= &\sum_{i=1}^{n} \left\{E(c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i)^4\right\}^{1/2} \left\{E\left(1\{|c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i| > \epsilon\sigma_\gamma\}\right)\right\}^{1/2} \\
= &\sum_{i=1}^{n} \left\{E(c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i)^4\right\}^{1/2} \left\{P(|c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i| > \epsilon\sigma_\gamma)\right\}^{1/2}.
\end{aligned}
\tag{A.39}
$$

The first term can be derived as

$$
\begin{aligned}
\left[E(c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i)^4\right]^{1/2} &= \left[E(c_n'(V'V)^{-1}V_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'V_i'(V'V)^{-1}c_n)^2\right]^{1/2} \\
&= \left[\{c_n'(V'V)^{-1}V_i\}^2 E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')^2\{V_i'(V'V)^{-1}c_n\}^2\right]^{1/2} \\
&= c_n'(V'V)^{-1}V_i[E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')^2]^{1/2}V_i'(V'V)^{-1}c_n \\
&\leq \|c_n'(V'V)^{-1}V_i\|_2^2 \left\|E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')^2\right\|_2^{1/2}.
\end{aligned}
\tag{A.40}
$$

For any $n \times n$ matrix $A$, $\|A\|_2 \leq \sqrt{n}\|A\|_\infty$. Since $E(\varepsilon_{it}^k) \leq (2\sigma^2)^{k/2}k\Gamma(k/2)$ for $k \geq 1$, then

$$
\begin{aligned}
\left\|E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')^2\right\|_2 &\leq \sqrt{T}\left\|E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')^2\right\|_\infty \\
&= \sqrt{T} \max_{\tau=1,\cdots,T} E\left(\varepsilon_{i\tau}\sum_{t=1}^{T}\varepsilon_{it}\sum_{t=1}^{T}\varepsilon_{it}^2\right) \leq \sqrt{T}(16+T)4\tilde{c}^2.
\end{aligned}
\tag{A.41}
$$

According to Assumption 5, $\|V_i\|_\infty$ is bounded and let the upper bound be some constant $c_2$, then $\|V_i\|_2 \leq \sqrt{Gp}c_2$. Following Assmuption 5, $\|(V'V)^{-1}\|_2 \geq (cg_{min}T)^{-1}$, we have

$$\begin{aligned}\left\{E(c_n'(V'V)^{-1}V_i\varepsilon_{it})^4\right\}^{1/2} &\leq \|c_n'\|_2^2\|(V'V)^{-1}\|_2^2\|V_i\|_2^2T^{1/4}(16+T)^{1/2}2\tilde{c}^2 \\ &\leq \frac{c_2^2Gp(16+T)^{1/2}2\tilde{c}}{c^2g_{min}^2T^{3/4}}.\end{aligned} \tag{A.42}$$

Then, by Chebyshev's inequality, the second term of (A.39) can be derived as

$$P(|c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i| > \epsilon\sigma_\gamma) \leq \frac{E[c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i]^2}{\epsilon^2\sigma_\gamma^2}, \tag{A.43}$$

where

$$\begin{aligned}E(c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i)^2 &= E(c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'V_i'(V'V)^{-1}c_n) \\ &\leq \|c_n\|_2^2\|(V'V)^{-1}\|_2^2\|V_i\|_2^2\|E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\|_2 \leq \frac{c_2^2Gp2\tilde{c}}{c^2g_{min}^2T^2},\end{aligned} \tag{A.44}$$

then, (A.43) becomes

$$P(|c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i| > \epsilon\sigma_\gamma) \leq \frac{c_2^2Gp2\tilde{c}}{c^2g_{min}^2T^2\epsilon^2\sigma_\gamma^2}. \tag{A.45}$$

Therefore, we have

$$
\begin{aligned}
&\sigma_{\gamma}^{-2}\sum_{i=1}^{n} E\left((c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i)^2 1\{|c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i| > \epsilon\sigma_{\gamma}\}\right) \\
\leq&\sigma_{\gamma}^{-2}\sum_{i=1}^{n} \frac{c_2^2 Gp(16+T)^{1/2}2\tilde{c}}{c^2 g_{min}^2 T^{3/4}} \frac{c_2(Gp)^{1/2}\sqrt{2\tilde{c}}}{c g_{min} T \epsilon \sigma_{\varphi}} = \frac{c_2^3 p^{3/2}(2\tilde{c})^{3/2} G^{3/2}(16+T)^{1/2} n}{c^3 \epsilon g_{min}^3 T^{7/4}\sigma_{\gamma}^3} \\
\leq&C\frac{(2\tilde{c})^{3/2}(n/g_{min})^{3/2} n(16+T)^{1/2}}{\sigma_{\gamma}^3 g_{min}^3 T^{7/4}} = C\frac{\tilde{c}^3 n^{5/2}(16+T)^{1/2}}{\sigma_{\varphi}^3 g_{min}^{9/2} T^{7/4}} \\
=&C\frac{n^{5/2}(16+T)^{1/2} c^{*3/2} g_{\max}^{3/2} T^{3/2}}{g_{min}^{9/2} T^{7/4}} = O\left(\frac{g_{\max}^{3/2} n^{5/2} T^{1/4}}{g_{min}^{9/2}}\right).
\end{aligned}
\tag{A.46}
$$

Suppose that $\dfrac{g_{\min}^3}{g_{\max}} \gg n^{5/3}T^{1/6}$, then (A.46) further implies that

$$
\sigma_{\gamma}^{-2}\sum_{i=1}^{n} E\left((c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i)^2 1\{|c_n'(V'V)^{-1}V_i\boldsymbol{\varepsilon}_i| > \epsilon\sigma_{\gamma}\}\right) = O(1).
$$

Following LindebergFeller Central Limit Theorem,

$$
c_n'(\hat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0) \rightarrow N(0, \sigma_{\gamma}^2).
$$

□

#### A.2.2.2 Convergence of the Calculated Estimator for Heterogeneous Model

*Proof of Theorem 3.3.* This can be done similarly to the proof of Theorem 4.2 in [54].

Define $\mathcal{M}_{\mathcal{G}} := \{\boldsymbol{\gamma} \in \mathbb{R}^{np} : \boldsymbol{\gamma}_i = \boldsymbol{\gamma}_j, \forall i, j \in \mathcal{G}_g, g = 1, \cdots, G\}$ and the least-squares objective function and the penalty function

$$\begin{aligned} L_n(\boldsymbol{\gamma}) &= \frac{1}{2}\|\mathbf{y} - W\boldsymbol{\gamma}\|_2^2, \quad P_n(\boldsymbol{\gamma}) = \lambda_1 \sum_{i<j} \rho(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2) \\ L_n^{\mathcal{G}}(\boldsymbol{\varphi}) &= \frac{1}{2}\|\mathbf{y} - W\Gamma\boldsymbol{\varphi}\|_2^2, \quad P_n^{\mathcal{G}}(\boldsymbol{\varphi}) = \lambda_1 \sum_{g<g'} |\mathcal{G}_g||\mathcal{G}_{g'}|\rho(\|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_{g'}\|_2). \end{aligned} \tag{A.47}$$

Let $Q_n(\boldsymbol{\gamma}) = L_n(\boldsymbol{\gamma}) + P_n(\boldsymbol{\gamma}), \quad Q_n^{\mathcal{G}}(\boldsymbol{\varphi}) = L_n^{\mathcal{G}}(\boldsymbol{\varphi}) + P_n^{\mathcal{G}}(\boldsymbol{\varphi})$. and define

- ⋄ $F : \mathcal{M}_{\mathcal{G}} \to \mathbb{R}^{Gp}$, $g$-th vector component of $F(\boldsymbol{\gamma})$ equals to the common value of $\boldsymbol{\gamma}_i$ for $i \in \mathcal{G}_g$.

- ⋄ $F^* : \mathbb{R}^{np} \to \mathbb{R}^{Gp}$, $F^*(\boldsymbol{\gamma}) = \{|\mathcal{G}_g|^{-1} \sum_{i \in \mathcal{G}_g} \boldsymbol{\gamma}_i', g = 1, \cdots, G\}'$, average of each cluster vectors.

It results in that $F(\boldsymbol{\gamma}) = F^*(\boldsymbol{\gamma})$ if $\boldsymbol{\gamma} \in \mathcal{M}_{\mathcal{G}}$. For every $\boldsymbol{\gamma} \in \mathcal{M}_{\mathcal{G}}$, $P_n(\boldsymbol{\gamma}) = P_n^{\mathcal{G}}(F(\boldsymbol{\gamma}))$, for every $\boldsymbol{\varphi} \in \mathbb{R}^{Gp}$, $P_n(F^{-1}(\boldsymbol{\varphi})) = P_n^{\mathcal{G}}(\boldsymbol{\varphi})$. Hence,

$$Q_n(\boldsymbol{\gamma}) = Q_n^{\mathcal{G}}(F(\boldsymbol{\gamma})), \quad Q_n^{\mathcal{G}}(\boldsymbol{\varphi}) = Q_n(F^{-1}(\boldsymbol{\varphi})). \tag{A.48}$$

Theorem 3.2 results in

$$P(\sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_2 \leq p \sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_\infty = p\|\widehat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0\|_\infty \leq p\phi_{n,T,G,\zeta}) \geq 1 - e^{\iota},$$

there exists an event $E_1$ in which $\sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_2 \leq p\phi_{n,T,G} = \tilde{\phi}_{n,T,G}$, and $P(E_1^C) \leq e^{-\iota}$. $\sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_2 \leq \phi_{n,T,G}$, and $P(E_1^C) \leq e^{-\iota}$.

Consider the neighborhood of the true parameter $\boldsymbol{\gamma}^0$,

$$\Theta := \{\boldsymbol{\gamma} \in \mathbb{R}^{np} : \sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0\|_2 \leq \tilde{\phi}_{n,T,G}\}.$$

It implies that $\widehat{\boldsymbol{\gamma}}^{or} \in \Theta$ on the event $E_1$. For any $\boldsymbol{\gamma} \in \mathbb{R}^{np}$, let $\boldsymbol{\gamma}^* = F^{-1}(F^*(\boldsymbol{\gamma}))$, then $\boldsymbol{\gamma}_i^* = \frac{1}{|\mathcal{G}_g|}\sum_{i \in \mathcal{G}_g} \boldsymbol{\gamma}_i$ which implies that $\boldsymbol{\gamma}^*$ is a vector with duplicated group average of $\boldsymbol{\gamma}_i$. Through two steps as the following, we can show that with probability approximating to 1, $\widehat{\boldsymbol{\gamma}}^{or}$ is a strictly local minimizer of the objective function.

i. In $E_1$, $Q_n(\boldsymbol{\gamma}^*) > Q_n(\widehat{\boldsymbol{\gamma}}^{or})$ for any $\boldsymbol{\gamma} \in \Theta$ and $\boldsymbol{\gamma}^* \neq \widehat{\boldsymbol{\gamma}}^{or}$. This indicates that the oracle estimator $\widehat{\boldsymbol{\gamma}}^{or}$ is the minimizer over all duplicated group average $\boldsymbol{\gamma}^*$.

ii. There exists an event $E_2$ such that for large enough $\iota^*$, $P(E_2^C) \leq e^{-\iota^*}$. In $E_1 \cap E_2$, there exists a neighborhood $\Theta_n$ of $\widehat{\boldsymbol{\gamma}}^{or}$ such that $Q_n(\boldsymbol{\gamma}) \geq Q_n(\boldsymbol{\gamma}^*)$ for all $\boldsymbol{\gamma}^* \in \Theta_n \cap \Theta$ for sufficiently large $n$. It means that for all $\boldsymbol{\gamma}$, the duplicated group average $\boldsymbol{\gamma}^*$ is the minimizer.

Then, it results in $Q_n(\boldsymbol{\gamma}) > Q_n(\widehat{\boldsymbol{\gamma}}^{or})$ for any $\boldsymbol{\gamma} \in \Theta_n \cap \Theta$ and $\boldsymbol{\gamma} \neq \widehat{\boldsymbol{\gamma}}^{or}$ in $E_1 \cap E_2$. Hence, for large enough $n$, $\widehat{\boldsymbol{\gamma}}^{or}$ is a strictly local minimizer of $Q_n(\boldsymbol{\gamma})$ over $E_1 \cap E_2$ with $P(E_1 \cap E_2) \geq 1 - e^{-\iota} - e^{-\iota^*}$.

First, show $P_n^{\mathcal{G}}(F^*(\boldsymbol{\gamma})) = C_n$ for any $\boldsymbol{\gamma} \in \Theta$, where $C_n$ is a constant which does not depend on $\boldsymbol{\gamma}$. It means that when $\boldsymbol{\gamma}$ is close enough to the true parameter $\boldsymbol{\gamma}^0$, the penalty term won't affect the objective function with respect to different values of $\boldsymbol{\gamma}$.

Let $F^*(\boldsymbol{\gamma}) = \boldsymbol{\varphi}$. It suffices to show that $\|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_{g'}\|_2 > a\lambda$ for all $g \neq g'$ and some constant $a > 0$. Then by Assumption 6, $\rho(\|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_{g'}\|_2)$ is a constant, and as a result $P_n^{\mathcal{G}}(F^*(\boldsymbol{\varphi}))$ is a constant.

Consider the triangular inequality $\|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_{g'}\|_2 \geq \|\boldsymbol{\varphi}_g^0 - \boldsymbol{\varphi}_{g'}^0\|_2 - 2\sup_g \|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_g^0\|_2$.

Since $\boldsymbol{\gamma} \in \Theta$, then

$$\begin{aligned}
&\sup_g \|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_g^0\|_2^2 = \sup_g \left\| |\mathcal{G}_g|^{-1} \sum_{i \in \mathcal{G}_g} \boldsymbol{\gamma}_i - \boldsymbol{\varphi}_g^0 \right\|_2^2 \\
=&\sup_g \left\| |\mathcal{G}_g|^{-1} \sum_{i \in \mathcal{G}_g} (\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0) \right\|_2^2 = \sup_g |\mathcal{G}_g|^{-2} \left\| \sum_{i \in \mathcal{G}_g} (\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0) \right\|_2^2 \\
\leq& |\mathcal{G}_g|^{-1} \sup_g \sum_{i \in \mathcal{G}_g} \left\| (\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0) \right\|_2^2 \leq \sup_i \left\| (\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0) \right\|_2^2 \leq \tilde{\phi}_{n,T,G}^2,
\end{aligned} \tag{A.49}$$

Since $b_{n,T,G} := \min_{g \neq g'} \|\boldsymbol{\varphi}_g^0 - \boldsymbol{\varphi}_{g'}^0\|$, then for all $g \neq g'$ and $b_{n,T,G} > a\lambda + 2\tilde{\phi}_{n,T,G}$, we have

$$\|\boldsymbol{\varphi}_g^0 - \boldsymbol{\varphi}_{g'}^0\|_2 \geq \|\boldsymbol{\varphi}_g^0 - \boldsymbol{\varphi}_{g'}^0\|_2 - 2 \sup_g \|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_g^0\|_2 \geq b_{n,T,G} - 2\tilde{\phi}_{n,T,G} > a\lambda.$$

Therefore, $P_n^{\mathcal{G}}(F^*(\boldsymbol{\gamma})) = C_n$, and hence $Q_n^{\mathcal{G}}(F^*(\boldsymbol{\gamma})) = L_n^{\mathcal{G}}(T^*(\boldsymbol{\gamma})) + C_n$ for all $\boldsymbol{\gamma} \in \Theta$.

Since $\widehat{\boldsymbol{\varphi}}^{or}$ is the unique global minimizer of $L_n^{\mathcal{G}}(\boldsymbol{\varphi})$, then $L_n^{\mathcal{G}}(T^*(\boldsymbol{\gamma})) > L_n^{\mathcal{G}}(\widehat{\boldsymbol{\varphi}}^{or})$ for all $T^*(\boldsymbol{\gamma}) \neq \widehat{\boldsymbol{\varphi}}^{or}$ and hence $Q_n^{\mathcal{G}}(T^*(\boldsymbol{\gamma})) > Q_n^{\mathcal{G}}(\widehat{\boldsymbol{\varphi}}^{or})$ for all $T^*(\boldsymbol{\gamma}) \neq \widehat{\boldsymbol{\varphi}}^{or}$.

By the property of the clustering algorithm, for the $g$-th group, $\widehat{\boldsymbol{\varphi}}_g^{or} = |\mathcal{G}_g|^{-1} \sum_{i \in \mathcal{G}_g} \widehat{\boldsymbol{\gamma}}_i^{or}$. Along with the definition of operation $T$, it implies that $\widehat{\boldsymbol{\varphi}}_g^{or}$ equals to the $g$-th component of $T(\widehat{\boldsymbol{\gamma}}^{or})$ for all $i \leq g \leq G$. Then, by (A.48)

$$Q_n^{\mathcal{G}}(\widehat{\boldsymbol{\varphi}}^{or}) = Q_n^{\mathcal{G}}(T(\widehat{\boldsymbol{\gamma}}^{or})) = Q_n(\widehat{\boldsymbol{\gamma}}^{or}).$$

Furthermore, we can easily derive that $Q_n^{\mathcal{G}}(T^*(\boldsymbol{\gamma})) = Q_n(T^{-1}(T^*(\boldsymbol{\gamma}))) = Q_n(\boldsymbol{\gamma}^*)$. Therefore, $Q_n(\boldsymbol{\gamma}^*) > Q_n(\widehat{\boldsymbol{\gamma}}^{or})$ for all $\boldsymbol{\gamma}^* \neq \widehat{\boldsymbol{\gamma}}^{or}$. The result in step i. is proved.

Second, for a positive sequence $t_n$, let $\Theta_n := \{\boldsymbol{\gamma}_i : \sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2 \leq t_n\}$. For $\boldsymbol{\gamma} \in \Theta_n \cap \Theta$, by the first order Taylor's expansion,

$$Q_n(\boldsymbol{\gamma}) - Q_n(\boldsymbol{\gamma}^*) = \frac{dQ_n(\boldsymbol{\gamma}^m)}{d\boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) = \frac{dL_n(\boldsymbol{\gamma}^m)}{d\boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) + \sum_{i=1}^{n} \frac{\partial P_n(\boldsymbol{\gamma}^m)}{\partial \boldsymbol{\gamma}_i'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*),$$

and let $S_1 = \dfrac{dL_n(\boldsymbol{\gamma}^m)}{d\boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)$ and $S_2 = \sum_{i=1}^{n} \dfrac{\partial P_n(\boldsymbol{\gamma}^m)}{\partial \boldsymbol{\gamma}_i'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)$.

Since

$$\begin{aligned}
\frac{dL_n(\boldsymbol{\gamma})}{\boldsymbol{\gamma}_i} &= \frac{1}{2}(-2\mathbf{y}'W + 2\boldsymbol{\gamma}'W'W) = -(\mathbf{y}' - \boldsymbol{\gamma}'W)W, \\
\frac{\partial P_n(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i} &= \lambda_1 \sum_{i=1}^{n} \rho'(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2) \frac{1}{2\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2} 2(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j) \\
&= \lambda_1 \sum_{i=1}^{n} \rho'(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2) \frac{\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j}{\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2}
\end{aligned}$$

then

$$S_1 = -(\mathbf{y}' - \boldsymbol{\gamma}^{m\prime}W)W(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*), \quad S_2 = \sum_{i=1}^{n} \frac{\partial P_n(\boldsymbol{\gamma}^m)}{\partial \boldsymbol{\gamma}_i'}(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*).$$

Let $\boldsymbol{\gamma}^m = \vartheta\boldsymbol{\gamma} + (1-\vartheta)\boldsymbol{\gamma}^*$ for some constant $\vartheta \in (0,1)$.

$$
\begin{aligned}
S_2 =& \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) \\
&+ \lambda_1 \sum_{i>j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) \\
=& \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) \qquad \text{(A.50)} \\
&+ \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_j^m - \boldsymbol{\gamma}_i^m\|_2)\|\boldsymbol{\gamma}_j^m - \boldsymbol{\gamma}_i^m\|_2^{-1}(\boldsymbol{\gamma}_j^m - \boldsymbol{\gamma}_i^m)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*) \\
=& \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'[(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) - (\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*)].
\end{aligned}
$$

Consider to separate $S_2$ into two parts, $i, j \in \mathcal{G}_g$, and $i \in \mathcal{G}_g$, $j \in \mathcal{G}_{g'}$ for $g \neq g'$. When $i, j \in \mathcal{G}_g$, since $\boldsymbol{\gamma}^* = T^{-1}(T^*(\boldsymbol{\gamma})) \in \mathcal{M}_{\mathcal{G}}$, then $\boldsymbol{\gamma}_i^* = \boldsymbol{\gamma}_j^*$. Thus, the RHS of (A.50) becomes $S_2 = \lambda_1(S_{21} + S_{22})$ where

$$
\begin{aligned}
S_{21} =& \sum_{g=1}^{G} \sum_{i,j\in\mathcal{G}, i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j), \\
S_{22} =& \sum_{g<g'} \sum_{i\in\mathcal{G}_g, j\in\mathcal{G}_{g'}} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'[(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) - (\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*)].
\end{aligned}
\qquad \text{(A.51)}
$$

Moreover, by (A.49), for any $\boldsymbol{\gamma} \in \Theta_n \cap \Theta$, since $F^*(\boldsymbol{\gamma}) = \boldsymbol{\varphi}$, then for all $i \in \mathcal{G}_g$, $\boldsymbol{\gamma}_i^* = \boldsymbol{\varphi}_g$. So we have

$$
\sup_i \|\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^0\|_2^2 = \sup_g \|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_g^0\|_2^2 \leq \tilde{\phi}_{n,T,G}^2, \qquad \text{(A.52)}
$$

and the inequality of (A.52) is obtained by (A.49).

Since $\boldsymbol{\gamma}_i^m = \vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^*$ and the triangular inequality,

$$\begin{aligned}
\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^0\|_2 &= \sup_i \|\vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^0\|_2 \\
&= \sup_i \|\vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - (\vartheta + 1 - \vartheta)\boldsymbol{\gamma}_i^0\|_2 \\
&\leq \vartheta \sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0\|_2 + (1-\vartheta)\sup_i \|\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^0\|_2 \\
&\leq \vartheta\tilde{\phi}_{n,T,G} + (1-\vartheta)\tilde{\phi}_{n,T,G} = \tilde{\phi}_{n,T,G}. \qquad \text{(A.53)}
\end{aligned}$$

Hence, for $g \neq g'$, $i \in \mathcal{G}_g$, $j \in \mathcal{G}_{g'}$,

$$\begin{aligned}
\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2 &= \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^0 - \boldsymbol{\gamma}_j^m + \boldsymbol{\gamma}_j^0\|_2 \geq \|\boldsymbol{\gamma}_i^0 - \boldsymbol{\gamma}_j^0\|_2 - 2\max_{1\leq k\leq n}\|\boldsymbol{\gamma}_k^m - \boldsymbol{\gamma}_k^0\|_2 \\
&\geq \min_{i\in\mathcal{G}_g, j'\in\mathcal{G}_{g'}} \|\boldsymbol{\gamma}_i^0 - \boldsymbol{\gamma}_j^0\|_2 - 2\max_{1\leq k\leq n}\|\boldsymbol{\gamma}_k^m - \boldsymbol{\gamma}_k^0\|_2 \geq b_{n,T,G} - 2\tilde{\phi}_{n,T,G} > a\lambda.
\end{aligned}$$

Since $\rho(x)$ is constant for all $x \geq a\lambda$, then $\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2) = 0$. Therefore, following $\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m = \vartheta(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)$ for $i, j \in \mathcal{G}_g$, (A.51) becomes

$$\begin{aligned}
S_2 =& \lambda_1 \sum_{g=1}^{G} \sum_{i,j\in\mathcal{G},i<j} \frac{\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)}{\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j) \\
&+ \lambda_1 \sum_{g<g'} \sum_{i\in\mathcal{G}_g, j\in\mathcal{G}_{g'}} \frac{\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)}{\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'[(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) - (\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*)] \\
=& \lambda_1 \sum_{g=1}^{G} \sum_{i,j\in\mathcal{G}_g,i<j} \frac{\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)}{\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)
\end{aligned}$$

$$
\begin{aligned}
=&\lambda_1 \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g, i<j} \frac{\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)}{\|\vartheta(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)\|_2} \vartheta(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j) \\
=&\lambda_1 \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g, i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2 \qquad \text{(A.54)}
\end{aligned}
$$

.

Furthermore, by the same reasoning as (A.49) and for all $i \in \mathcal{G}_g$, $\boldsymbol{\gamma}_i^* = \boldsymbol{\varphi}_g$,

$$
\sup_i \|\boldsymbol{\gamma}_i^* - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2^2 = \sup_g \|\boldsymbol{\varphi}_g - \widehat{\boldsymbol{\varphi}}_g^{or}\|_2^2 \leq \sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2^2. \qquad \text{(A.55)}
$$

Then, since $\boldsymbol{\gamma}_i^* = \boldsymbol{\gamma}_j^*$,

$$
\begin{aligned}
\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2 &= \sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_j^m + \boldsymbol{\gamma}_j^*\|_2 \\
&\leq \|\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_j^*\|_2 + 2\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^*\|_2 \leq 2\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^*\|_2 \\
&= 2\sup_i \|\vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^*\|_2 \\
&= 2\vartheta \sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\|_2 \leq 2\sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\|_2 \\
&\leq 2(\sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2 + \sup_i \|\boldsymbol{\gamma}_i^* - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2) \\
&\leq 4\sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2 \leq 4t_n. \qquad \text{(A.56)}
\end{aligned}
$$

Hence, $\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2) \geq \rho'(4t_n)$ since $\rho(x)$ is nondecreasing and concave. Then,

$$S_2 \geq \lambda_1 \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_k, i<j} \rho'(4t_n) \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2. \tag{A.57}$$

Let $Q = (Q_1', \cdots, Q_n')' = [(\mathbf{y} - W\boldsymbol{\gamma}^m)'W]'$, then

$$\begin{aligned}
S_1 =& - Q'(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) = -(Q_1', \cdots, Q_n')' \begin{pmatrix} \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_1^* \\ \boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_2^* \\ \vdots \\ \boldsymbol{\gamma}_n - \boldsymbol{\gamma}_n^* \end{pmatrix} \\
=& - \sum_{i=1}^{n} Q_i'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) \\
=& - \sum_{g=1}^{G} \sum_{i \in \mathcal{G}_g} \frac{1}{|\mathcal{G}_g|} Q_i' \left( |\mathcal{G}_g| \boldsymbol{\gamma}_i - \sum_{j \in \mathcal{G}_g} \boldsymbol{\gamma}_j \right) \\
=& - \sum_{g=1}^{G} \sum_{i \in \mathcal{G}_g} \frac{1}{|\mathcal{G}_g|} Q_i' \sum_{j \in \mathcal{G}_g} \left( \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j \right) = - \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g} \frac{Q_i'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)}{|\mathcal{G}_g|} \\
=& - \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g} \frac{Q_i'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)}{2|\mathcal{G}_g|} + \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g} \frac{Q_j'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)}{2|\mathcal{G}_g|} \\
=& - \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g} \frac{(Q_j - Q_i)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i)}{2|\mathcal{G}_g|} \\
=& - \sum_{g=1}^{G} \sum_{i,j \in \mathcal{G}_g, i<j} \frac{(Q_j - Q_i)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i)}{|\mathcal{G}_g|}.
\end{aligned} \tag{A.58}$$

Moreover,

$$Q_i = W_i'(\mathbf{y}_i - W_i\gamma_i^m) = W_i'(W_i\gamma_i^0 + \boldsymbol{\varepsilon}_i - W_i\gamma_i^m) = W_i'(\boldsymbol{\varepsilon}_i + W_i(\gamma_i^0 - \gamma_i^m)),$$

and then,

$$\begin{aligned}
\sup_i \|Q_i\|_2 &\leq \sup_i\{\|W_i'(\boldsymbol{\varepsilon}_i + W_i(\gamma_i^0 - \gamma_i^m))\|_2\} \\
&\leq \sup_i\{\|W_i'\boldsymbol{\varepsilon}_i\|_2 + \|W_i'W_i(\gamma_i^0 - \gamma_i^m)\|_2\} \\
&\leq \sup_i \|W_i'\boldsymbol{\varepsilon}_i\|_2 + \sup_i \|W_i'W_i\|_2\|(\gamma_i^0 - \gamma_i^m)\|_2 \\
&\leq \sup_i \|W_i'\boldsymbol{\varepsilon}_i\|_2 + \sup_i \sqrt{p}\|W_i'W_i\|_\infty\tilde{\phi}_{n,T,G} \\
&\leq \sup_i \|W_i'\boldsymbol{\varepsilon}_i\|_2 + m\sqrt{pT}(q^{1/2} + m^{1/2}(L+1+2K))\tilde{\phi}_{n,T,G} \\
&\leq \sup_i \sqrt{p}\|W_i'\boldsymbol{\varepsilon}_i\|_\infty + m\sqrt{pT}(q^{1/2} + m^{1/2}(L+1+2K))\tilde{\phi}_{n,T,G} \\
&\leq \sqrt{p}\|W'\boldsymbol{\varepsilon}\|_2 + m\sqrt{pT}(q^{1/2} + m^{1/2}(L+1+2K))\tilde{\phi}_{n,T,G} \\
&= \sqrt{p}\|W'\boldsymbol{\varepsilon}\|_2 + m\sqrt{pT}B_{q,m}\tilde{\phi}_{n,T,G}, \qquad \text{(A.59)}
\end{aligned}$$

where $B_{q,m} = (q^{1/2} + m^{1/2}(L+1+2K))$.

By Lemma A.3, $P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > 2\tilde{c}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)m\tilde{M}\sqrt{T}B_{q,m}\sqrt{p}\right] \leq e^{-\iota^*}$, where $B_{q,m} = (q^{1/2} + m^{1/2}(L+1+2K))$, $p = q + L + 1 + 2K$, $\tilde{M} = \max(M_1, M_2, M_3, M_4)$ and $\tilde{c}$ given in 3.5 and 3.7. $\iota^*$ is defined in Lemma A.3.

Then, over the event $E_2$,

$$\left|\frac{(Q_j - Q_i)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i)}{|\mathcal{G}_g|}\right| \leq g_{min}^{-1}\|Q_j - Q_i\|_2\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i\|_2 \leq g_{min}^{-1}2\sup_i \|Q_i\|_2\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2$$

$$\leq 2g_{min}^{-1}T^{1/4}(mp)^{1/2}\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2$$

$$\left(p^{1/4}\tilde{B}_{q,m}^{1/2}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} + T^{1/4}m^{1/2}B_{q,m}\tilde{\phi}_{n,T,G}\right) \tag{A.60}$$

Therefore, by (A.57), (A.58) and (A.60),

$$\begin{aligned}
&Q_n(\boldsymbol{\gamma}) - Q_n(\boldsymbol{\gamma}^*)\\
\geq &\sum_{g=1}^{G}\sum_{i,j\in\mathcal{G}_g, i<j}\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2\\
&\Big\{\lambda\rho'(4t_n) - 2g_{min}^{-1}T^{1/4}(mp)^{1/2}(p^{1/4}\tilde{B}_{q,m}^{1/2}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}\\
&+T^{1/4}m^{1/2}B_{q,m}\tilde{\phi}_{n,T,G})\Big\}\\
\geq &\sum_{g=1}^{G}\sum_{i,j\in\mathcal{G}_g, i<j}\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2\\
&\Big\{\lambda\rho'(4t_n) - B_1 g_{min}^{-1}T^{1/4}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} - B_2 g_{min}^{-1}T^{1/2}\tilde{\phi}_{n,T,G}\Big\},
\end{aligned} \tag{A.61}$$

where $B_1 = 2(mp\tilde{B}_{q,m})^{1/2}p^{1/4}$ and $B_2 = 2mp^{1/2}B_{q,m}$.

Let $t_n = o(1)$, then $\rho'(4t_n) \to 1$. Suppose that the following condition is true over the event $E_1 \cap E_2$,

$$B_1 g_{min}^{-1}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}T^{1/4} \to 0, \quad B_2 p g_{min}^{-1} T^{1/2}\phi_{n,T,G} \to 0, \tag{A.62}$$

then $P\left(Q_n(\boldsymbol{\gamma}) - Q_n(\boldsymbol{\gamma}^*) \geq 0\right) \geq 1 - e^{\iota} - e^{\iota^*}$. Once (A.62) holds, $Q_n(\gamma) - Q_n(^*) \geq 0$ with probability approaching to 1 as $n \to \infty$.

Now we show (A.62). In the following context, we focus on deriving the conditions only for Theorem 3. To show that our estimator converges to the oracle estimator, which converges to the true parameter as well, we need to consider the conditions in both Theorem 2 and Theorem 3.

1. As $T \to \infty$ with $n$ fixed, our estimator cannot be proved to converge to the oracle estimator.

2. As $n \to \infty$ with $T$ fixed, when conditions in Theorem 2 are satisfied, the second part of (A.62) is true. So, here we discuss about the conditions for first part of (A.62).

   (a) Consider $\zeta^* \leq n$ and $g_{min} \gg (p + 2\sqrt{p} + 2)^{1/2} n^{1/2}$. Let $\zeta^* \to \infty$, since $(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} = (p + 2\sqrt{p} + 2)^{1/2} O(n^{1/2})$, then

$$\begin{aligned}&B_1 g_{min}^{-1}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}T^{1/4}\\ \leq &B_1 T^{1/4} g_{min}^{-1}(p + 2\sqrt{p} + 2)^{1/2} O(n^{1/2}) \to 0.\end{aligned}$$

(b) Consider $\zeta^* > n$ and $g_{min} \gg (p+2\sqrt{p}+2)^{1/2}\zeta^{*1/2} > (p+2\sqrt{p}+2)^{1/2}n^{1/2}$. Let $\zeta^* \to \infty$, since $(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} = (p+2\sqrt{p}+2)^{1/2}O(\zeta^{*1/2})$, then

$$B_1 g_{min}^{-1}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}T^{1/4}$$
$$\leq B_1 T^{1/4} g_{min}^{-1}(p+2\sqrt{p}+2)^{1/2}O(\zeta^{*1/2}) \to 0.$$

3. As $T, n \to \infty$. Consider the first part of (A.62).

(a) Consider $\zeta^* \leq n$ and $g_{min} \gg (p+2\sqrt{p}+2)^{1/2}n^{1/2}T^{1/4}$. Let $\zeta^* \to \infty$, then

$$B_1 g_{min}^{-1}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}T^{1/4}$$
$$\leq B_1 g_{min}^{-1}(p+2\sqrt{p}+2)^{1/2}n^{1/2}T^{1/4} \to 0.$$

(b) Consider $\zeta^* \geq n$ and $(p+2\sqrt{p}+2)^{1/2}n^{1/2}T^{1/4} \leq (p+2\sqrt{p}+2)^{1/2}\zeta^{*1/2}T^{1/4} \ll g_{min} \leq n$. Let $\zeta^* \to \infty$, then

$$B_1 g_{min}^{-1}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}T^{1/4}$$
$$\leq B_1 g_{min}^{-1}(p+2\sqrt{p}+2)^{1/2}\zeta^{*1/2}T^{1/4} \to 0.$$

Now, consider the second part of (A.62) as $n, T \to \infty$.

(a) Consider $G \to \infty$. Let $\frac{n^{7/13}}{T^{1/13}} \ll g_{min} < n/G$, $\zeta \leq G$ and $\zeta \to \infty$, then $G \ll \frac{T^{1/13}}{n^{6/13}}$ and $Gp + 2\sqrt{Gp\zeta} + 2\zeta \leq (p+2\sqrt{p}+2)G = O(G)$. Since

$G \leq n/g_{min}$,

$$\begin{aligned} B_2 p g_{min}^{-1} T^{1/2} \phi_{n,T,G} &\leq B_2 p C_5 \frac{n^{1/2} G^{3/4} T^{1/2}}{g_{min}^2 T^{3/4}} O(G^{1/2}) \\ &\leq B_2 p C_5 \frac{n^{7/4}}{g_{min}^{13/4} T^{1/4}} O(1) \overset{n,T,G\to\infty}{\longrightarrow} 0, \end{aligned}$$

where $C_5 = C_{q,m}(p + 2\sqrt{p} + 2p)^{1/2}$, which is free from $n, T$ and $G$.

(b) When $G \ll \zeta \to \infty$

i. When $G$ is fixed, then $(Gp + 2\sqrt{Gp\zeta} + 2\zeta)^{1/2} = O(2\zeta^{1/2})$. Let $g_{min} = O(n^{1/4+\tilde{\alpha}_1})$ for some positive constant $\tilde{\alpha}_1 < 3/4$ and $\zeta = o(n^{4\tilde{\alpha}_1} T^{1/2})$, $\zeta \to \infty$, then

$$B_2 p g_{min}^{-1} T^{1/2} \phi_{n,T,G} \leq B_2 p C_6 \frac{n^{1/2}}{g_{min}^2 T^{1/4}} O(\zeta^{1/2}) \overset{n,T\to\infty}{\longrightarrow} 0,$$

where $C_6 = 2C_{q,m} G^{3/4}$.

ii. When $G \to \infty$, for some positive constant $\tilde{\alpha}_2 < 6/11$, let $g_{min} = O(n^{5/11+\tilde{\alpha}_2})$ and $G \leq n/g_{min}$, $\zeta = o(n^{11\tilde{\alpha}_2/2} T^{1/2})$ and $\zeta \to \infty$, then $(Gp + 2\sqrt{Gp\zeta} + 2\zeta)^{1/2} = O((p + 2\sqrt{p} + 2)^{1/2} \zeta^{1/2})$. Since $G \leq n/g_{min}$, then

$$B_2 p g_{min}^{-1} T^{1/2} \phi_{n,T,G} \leq B_2 p C_7 \frac{n^{5/4}}{g_{min}^{11/4} T^{1/4}} O(\zeta^{1/2}) \overset{n,T,G\to\infty}{\longrightarrow} 0,$$

where $C_7 = C_{q,m}(p + 2\sqrt{p} + 2)^{1/2}$, which is freen of $n, T$ and $G$.

$\square$

#### A.2.2.3 Convergence of the Calculated Estimator for Homogenous Model

The proof of the homogenous model is similar to the proof of Theorem 3.3. We shall present the whole process.

*Proof.* When the true model only contains only one group, then model (11) becomes

$$\mathbf{y} \approx W^* \boldsymbol{\varphi} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (\mathbf{y}_1', \cdots, \mathbf{y}_n')'$, $W^* = (W_1', \cdots, W_n')'$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \cdots, \boldsymbol{\varepsilon}_n')$. We also have $\boldsymbol{\gamma}_1 = \cdots = \boldsymbol{\gamma}_n = \boldsymbol{\varphi}$ and $G = 1$. The estimator $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \cdots, \boldsymbol{\gamma}_n')'$ also has the oracle property. Define the oracle estimator of $\boldsymbol{\gamma}$ as

$$\widehat{\boldsymbol{\varphi}}^{or} = \text{argmin}_{\boldsymbol{\varphi} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - W^* \boldsymbol{\varphi}\|_2^2 = (W^{*\prime} W^*)^{-1} W^{*\prime} \mathbf{y}. \tag{A.63}$$

Let $\widehat{\boldsymbol{\gamma}}^{or} = (\widehat{\boldsymbol{\gamma}}_1^{or\prime}, \cdots, \widehat{\boldsymbol{\gamma}}_n^{or\prime})'$ where $\widehat{\boldsymbol{\gamma}}_1^{or} = \cdots = \widehat{\boldsymbol{\gamma}}_n^{or} = \widehat{\boldsymbol{\varphi}}^{or}$.

Define $\mathcal{M} := \{\boldsymbol{\gamma} \in \mathbb{R}^{np} : \boldsymbol{\gamma}_1 = \cdots = \boldsymbol{\gamma}_n\}$. For any $\boldsymbol{\gamma} \in \mathcal{M}$, $\boldsymbol{\gamma}_i = \boldsymbol{\alpha}$ for all $i$. Take the least-squares objective function and the penalty function.

$$L_n(\boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{y} - W\boldsymbol{\gamma}\|_2^2, \quad P_n(\boldsymbol{\gamma}) = \lambda \sum_{i<j} \rho(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2)$$

$$L_n^{\mathcal{G}}(\boldsymbol{\varphi}) = \frac{1}{2}\|\mathbf{y} - W^*\boldsymbol{\varphi}\|_2^2, \quad P_n^{\mathcal{G}}(\boldsymbol{\varphi}) = \lambda \sum_{g<g'} |\mathcal{G}_g||\mathcal{G}_{g'}|\rho(\|\boldsymbol{\varphi}_g - \boldsymbol{\varphi}_{g'}\|_2).$$

Let $Q_n(\boldsymbol{\gamma}) = L_n(\boldsymbol{\gamma}) + P_n(\boldsymbol{\gamma}), \quad Q_n^{\mathcal{G}}(\boldsymbol{\varphi}) = L_n^{\mathcal{G}}(\boldsymbol{\varphi}) + P_n^{\mathcal{G}}(\boldsymbol{\varphi})$ and

- ⋄ $F : \mathcal{M} \to \mathbb{R}^p$, $g$-th vector component of $T(\boldsymbol{\gamma})$ equals to the common value of $\boldsymbol{\gamma}_i$ for $i \in \mathcal{G}_g$.

- ⋄ $F^* : \mathbb{R}^{np} \to \mathbb{R}^p$, $T^*(\boldsymbol{\gamma}) = \{|\mathcal{G}_g|^{-1} \sum_{i \in \mathcal{G}_g} \boldsymbol{\gamma}_i', g = 1, \cdots, G\}'$, average of each cluster vectors.

For every $\boldsymbol{\gamma} \in \mathcal{M}_{\mathcal{G}}$, $P_n(\boldsymbol{\gamma}) = P_n^{\mathcal{G}}(T(\boldsymbol{\gamma}))$ and for every $\boldsymbol{\varphi} \in \mathbb{R}^{Gp}$, $P_n(F^{-1}(\boldsymbol{\varphi})) = P_n^{\mathcal{G}}(\boldsymbol{\varphi})$. Hence,

$$Q_n(\boldsymbol{\gamma}) = Q_n^{\mathcal{G}}(F(\boldsymbol{\gamma})), \quad Q_n^{\mathcal{G}}(\boldsymbol{\varphi}) = Q_n(F^{-1}(\boldsymbol{\varphi})). \tag{A.64}$$

By Theorem 3.2,

$$P(\sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_2 \leq p \sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_\infty = p\|\widehat{\boldsymbol{\gamma}}^{or} - \boldsymbol{\gamma}^0\|_\infty \leq p\phi_{n,T,G,\zeta}) \geq 1 - e^{\iota},$$

there exists an event $E_1$ in which $\sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_2 \leq p\phi_{n,T,G} = \tilde{\phi}_{n,T,G}$, and $P(E_1^C) \leq e^{-\iota}$. $\sup_i \|\widehat{\boldsymbol{\gamma}}_i^{or} - \boldsymbol{\gamma}_i^0\|_2 \leq \phi_{n,T,G}$, and $P(E_1^C) \leq e^{-\iota}$.

Consider the neighborhood of the true parameter $\boldsymbol{\gamma}^0$,

$$\Theta := \{\boldsymbol{\gamma} \in \mathbb{R}^{np} : \sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0\|_2 \leq \tilde{\phi}_{n,T,G}\}.$$

It implies that $\widehat{\boldsymbol{\gamma}}^{or} \in \Theta$ on the event $E_1$. For any $\boldsymbol{\gamma} \in \mathbb{R}^{np}$, let $\boldsymbol{\gamma}^* = F^{-1}(F^*(\boldsymbol{\gamma}))$, then $\boldsymbol{\gamma}_i^* = \frac{1}{n}\sum_{i\in n}\boldsymbol{\gamma}_i$ which implies that $\boldsymbol{\gamma}^*$ is a vector with duplicated group average of $\boldsymbol{\gamma}_i$. So, $\boldsymbol{\gamma}_1^* = \cdots = \boldsymbol{\gamma}_n^*$. Through two steps as the following, we can show that with probability approximating to 1, $\widehat{\boldsymbol{\gamma}}^{or}$ is a strictly local minimizer of the objective function.

i. In $E_1$, $Q_n(\boldsymbol{\gamma}^*) > Q_n(\widehat{\boldsymbol{\gamma}}^{or})$ for any $\boldsymbol{\gamma} \in \Theta$ and $\boldsymbol{\gamma}^* \neq \widehat{\boldsymbol{\gamma}}^{or}$. This indicates that the oracle estimator $\widehat{\boldsymbol{\gamma}}^{or}$ is the minimizer over all duplicated group average $\boldsymbol{\gamma}^*$.

ii. There is an event $E_2$ such that $P(E_2^C) \leq e^{-\zeta^*}$ for large enough $\iota^*$. In $E_1 \cap E_2$, there is a neighborhood of $\widehat{\boldsymbol{\gamma}}^{or}$ denoted by $\Theta_n$ such that $Q_n(\boldsymbol{\gamma}) \geq Q_n(\boldsymbol{\gamma}^*)$ for any $\boldsymbol{\gamma}^* \in \Theta_n \cap \Theta$ for sufficiently large $n$. It means that for all $\boldsymbol{\gamma}$, the duplicated group average $\boldsymbol{\gamma}^*$ is the minimizer.

Then, it results in $Q_n(\boldsymbol{\gamma}) > Q_n(\widehat{\boldsymbol{\gamma}}^{or})$ for any $\boldsymbol{\gamma} \in \Theta_n \cap \Theta$ and $\boldsymbol{\gamma} \neq \widehat{\boldsymbol{\gamma}}^{or}$ in $E_1 \cap E_2$, hence $\widehat{\boldsymbol{\gamma}}^{or}$ is a strictly local minimizer of $Q_n(\boldsymbol{\gamma})$ over $E_1 \cap E_2$ with $P(E_1 \cap E_2) \geq 1 - e^{-\zeta} - e^{-\zeta^*}$ for large enough $n$.

First, we show the result in step i.. By definition of $\widehat{\boldsymbol{\gamma}}^{or}$, we have $\frac{1}{2}\|\mathbf{y} - W\boldsymbol{\gamma}^*\|_2^2 \geq$

$\frac{1}{2}\|\mathbf{y} - W\widehat{\boldsymbol{\gamma}}^{or}\|_2^2$ for any $\boldsymbol{\gamma} \in \Theta$ and $\boldsymbol{\gamma}^* \neq \widehat{\boldsymbol{\gamma}}^{or}$. Moreover, since $\boldsymbol{\gamma}_1^* = \cdots = \boldsymbol{\gamma}_n^*$ and $\widehat{\boldsymbol{\gamma}}_1^{or} = \cdots = \widehat{\boldsymbol{\gamma}}_n^{or}$, then $\rho_\gamma(\|\widehat{\boldsymbol{\gamma}}_i^{or} - \widehat{\boldsymbol{\gamma}}_j^{or}\|_2, \lambda) = \rho_\gamma(\|\widehat{\boldsymbol{\gamma}}_i^* - \widehat{\boldsymbol{\gamma}}_j^*\|_2, \lambda) = 0$ for all $i, j$. So,

$$Q_n(\boldsymbol{\gamma}^*) = \frac{1}{2}\|\mathbf{y} - W\boldsymbol{\gamma}^*\|_2^2 \geq \frac{1}{2}\|\mathbf{y} - W\widehat{\boldsymbol{\gamma}}^{or}\|_2^2 = Q_n(\widehat{\boldsymbol{\gamma}}^{or}).$$

Therefore, $Q_n(\boldsymbol{\gamma}^*) \geq Q_n(\widehat{\boldsymbol{\gamma}}^{or})$.

Second, we focus on the result in step ii.. For a positive sequence $t_n$, let $\Theta_n := \{\boldsymbol{\gamma}_i : \sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2 \leq t_n\}$. For $\boldsymbol{\gamma} \in \Theta_n \cap \Theta$, by the first order Taylor's expansion,

$$Q_n(\boldsymbol{\gamma}) - Q_n(\boldsymbol{\gamma}^*) = \frac{dQ_n(\boldsymbol{\gamma}^m)}{d\boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) = \frac{dL_n(\boldsymbol{\gamma}^m)}{d\boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) + \sum_{i=1}^n \frac{\partial P_n(\boldsymbol{\gamma}^m)}{\partial \boldsymbol{\gamma}_i'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*),$$

and let $S_1 = \dfrac{dL_n(\boldsymbol{\gamma}^m)}{d\boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)$ and $S_2 = \sum_{i=1}^n \dfrac{\partial P_n(\boldsymbol{\gamma}^m)}{\partial \boldsymbol{\gamma}_i'}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)$.

Since

$$\begin{aligned}
\frac{dL_n(\boldsymbol{\gamma})}{\boldsymbol{\gamma}_i} &= \frac{1}{2}(-2\mathbf{y}'W + 2\boldsymbol{\gamma}'W'W) = -(\mathbf{y}' - \boldsymbol{\gamma}'W)W, \\
\frac{\partial P_n(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i} &= \lambda_1 \sum_{i=1}^n \rho'(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2)\frac{1}{2\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2}2(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j) \\
&= \lambda_1 \sum_{i=1}^n \rho'(\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2)\frac{\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j}{\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2}
\end{aligned}$$

then

$$S_1 = -(\mathbf{y}' - \boldsymbol{\gamma}^{m\prime}W)W(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*), \qquad S_2 = \sum_{i=1}^n \frac{\partial P_n(\boldsymbol{\gamma}^m)}{\partial \boldsymbol{\gamma}_i'}(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*).$$

Let $\boldsymbol{\gamma}^m = \vartheta\boldsymbol{\gamma} + (1-\vartheta)\boldsymbol{\gamma}^*$ for some constant $\vartheta \in (0,1)$. Since $\boldsymbol{\gamma}_i^* = \boldsymbol{\gamma}_j^*$, then

$$\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m = \vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - \vartheta\boldsymbol{\gamma}_j - (1-\vartheta)\boldsymbol{\gamma}_j^* = \vartheta(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j).$$

Therefore,

$$\begin{aligned}
S_2 =& \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2^{-1}(\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m)'[(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) - (\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*)] \\
=& \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\vartheta(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)\|_2^{-1}\vartheta(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j) \qquad \text{(A.65)} \\
=& \lambda_1 \sum_{i<j} \rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2)\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2.
\end{aligned}$$

Since $\boldsymbol{\gamma} \in \Theta$, then

$$\begin{aligned}
\|\boldsymbol{\varphi} - \boldsymbol{\varphi}^0\|_2^2 =& \left\| n^{-1}\sum_{i\in n}\boldsymbol{\gamma}_i - \boldsymbol{\varphi}^0 \right\|_2^2 = \left\| n^{-1}\sum_{i\in n}(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0) \right\|_2^2 = n^{-2}\left\| \sum_{i\in n}(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0) \right\|_2^2 \\
\le& n^{-1}\sum_{i\in n}\left\|(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0)\right\|_2^2 \le \sup_i \left\|(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0)\right\|_2^2 \le \tilde{\phi}_{n,T,G}^2.
\end{aligned} \qquad \text{(A.66)}$$

Thus, for any $\boldsymbol{\gamma} \in \Theta_n \cap \Theta$, since $F^*(\boldsymbol{\gamma}) = \boldsymbol{\varphi}$, then for all $i$, $\boldsymbol{\gamma}_i^* = \boldsymbol{\varphi}$. So we have

$$\sup_i \|\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^0\|_2^2 = \|\boldsymbol{\varphi} - \boldsymbol{\varphi}^0\|_2^2 \le \sup_i \left\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0\right\|_2^2 \le \tilde{\phi}_{n,T,G}^2. \qquad \text{(A.67)}$$

Since $\boldsymbol{\gamma}_i^m = \vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^*$ and the triangular inequality,

$$
\begin{aligned}
\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^0\|_2 &= \sup_i \|\vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^0\|_2 \\
&= \sup_i \|\vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - (\vartheta + 1 - \vartheta)\boldsymbol{\gamma}_i^0\|_2 \\
&\leq \vartheta \sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0\|_2 + (1-\vartheta)\sup_i \|\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^0\|_2 \\
&\leq \sup_i \left\|(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^0)\right\|_2^2 \leq \tilde{\phi}_{n,T,G}^2. \qquad \text{(A.68)}
\end{aligned}
$$

Furthermore, by the same reasoning as (A.78) and for all $i$, $\boldsymbol{\gamma}_i^* = \boldsymbol{\varphi}$,

$$
\|\boldsymbol{\gamma}_i^* - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2^2 = \|\boldsymbol{\varphi} - \widehat{\boldsymbol{\varphi}}^{or}\|_2^2 \leq \sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2^2. \qquad \text{(A.69)}
$$

Then, since $\boldsymbol{\gamma}_i^* = \boldsymbol{\gamma}_j^*$,

$$
\begin{aligned}
\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2 &= \sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_j^m + \boldsymbol{\gamma}_j^*\|_2 \\
&\leq \|\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_j^*\|_2 + 2\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^*\|_2 \leq 2\sup_i \|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_i^*\|_2 \\
&= 2\sup_i \|\vartheta\boldsymbol{\gamma}_i + (1-\vartheta)\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i^*\|_2 \\
&= 2\vartheta \sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\|_2 \leq 2\sup_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\|_2 \\
&\leq 2(\sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2 + \sup_i \|\boldsymbol{\gamma}_i^* - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2) \\
&\leq 4\sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^{or}\|_2 \leq 4t_n. \qquad \text{(A.70)}
\end{aligned}
$$

Hence, $\rho'(\|\boldsymbol{\gamma}_i^m - \boldsymbol{\gamma}_j^m\|_2) \geq \rho'(4t_n)$ since $\rho(x)$ is nondecreasing and concave. Then,

$$S_2 \geq \lambda \sum_{i<j} \rho'(4t_n)\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2. \tag{A.71}$$

Let $Q = (Q_1', \cdots, Q_n')' = [(\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}^m)'\mathbf{W}]'$, then

$$\begin{aligned}
S_1 =& - Q'(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) = -(Q_1', \cdots, Q_n')' \begin{pmatrix} \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_1^* \\ \boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_2^* \\ \vdots \\ \boldsymbol{\gamma}_n - \boldsymbol{\gamma}_n^* \end{pmatrix} \\
&= - \sum_{i=1}^n Q_i'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*) \\
&= - \sum_{i=1}^n \frac{1}{n} Q_i' \left( n\boldsymbol{\gamma}_i - \sum_{j=1}^n \boldsymbol{\gamma}_j \right) \\
&= - \sum_{i=1}^n \frac{1}{n} Q_i' \sum_{j=1}^n \left(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\right) = - \sum_{i=1}^n \sum_{j=1}^n \frac{Q_i'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)}{n} \\
&= - \sum_{i=1}^n \sum_{j=1}^n \frac{Q_i'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)}{2n} + \sum_{i=1}^n \sum_{j=1}^n \frac{Q_j'(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j)}{2n} \\
&= - \sum_{i=1}^n \sum_{j=1}^n \frac{(Q_j - Q_i)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i)}{2n} \\
&= - \sum_{i<j} \frac{(Q_j - Q_i)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i)}{n}.
\end{aligned} \tag{A.72}$$

Moreover,

$$Q_i = W_i'(\mathbf{y}_i - W_i\gamma_i^m) = W_i'(W_i\gamma_i^0 + \boldsymbol{\varepsilon}_i - W_i\gamma_i^m) = W_i'(\boldsymbol{\varepsilon}_i + W_i(\gamma_i^0 - \gamma_i^m)),$$

and then,

$$\begin{aligned}
\sup_i \|Q_i\|_2 &\le \sup_i\{\|W_i'(\boldsymbol{\varepsilon}_i + W_i(\gamma_i^0 - \gamma_i^m))\|_2\} \\
&\le \sup_i\{\|W_i'\boldsymbol{\varepsilon}_i\|_2 + \|W_i'W_i(\gamma_i^0 - \gamma_i^m)\|_2\} \\
&\le \sup_i \|W_i'\boldsymbol{\varepsilon}_i\|_2 + \sup_i \|W_i'W_i\|_2\|(\gamma_i^0 - \gamma_i^m)\|_2 \\
&\le \sup_i \|W_i'\boldsymbol{\varepsilon}_i\|_2 + \sup_i \sqrt{p}\|W_i'W_i\|_\infty\tilde{\phi}_{n,T,G} \\
&\le \sup_i \|W_i'\boldsymbol{\varepsilon}_i\|_2 + m\sqrt{pT}(q^{1/2} + m^{1/2}(L+1+2K))\tilde{\phi}_{n,T,G} \\
&\le \sup_i \sqrt{p}\|W_i'\boldsymbol{\varepsilon}_i\|_\infty + m\sqrt{pT}(q^{1/2} + m^{1/2}(L+1+2K))\tilde{\phi}_{n,T,G} \\
&\le \sqrt{p}\|W'\boldsymbol{\varepsilon}\|_2 + m\sqrt{pT}(q^{1/2} + m^{1/2}(L+1+2K))\tilde{\phi}_{n,T,G} \\
&= \sqrt{p}\|W'\boldsymbol{\varepsilon}\|_2 + m\sqrt{pT}B_{q,m}\tilde{\phi}_{n,T,G}, \qquad \text{(A.73)}
\end{aligned}$$

where $B_{q,m} = (q^{1/2} + m^{1/2}(L+1+2K))$.

By Lemma A.3, $P\left[\|W'\boldsymbol{\varepsilon}\|_2^2 > 2\tilde{c}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)m\tilde{M}\sqrt{T}B_{q,m}\sqrt{p}\right] \le e^{-\iota^*}$, where $B_{q,m} = (q^{1/2} + m^{1/2}(L+1+2K))$, $p = q + L + 1 + 2K$, $\tilde{M} = \max(M_1, M_2, M_3, M_4)$

and $\tilde{c}$ given in 3.5 and 3.7. $\iota^*$ is defined in Lemma A.3. Then, over the event $E_2$,

$$\left|\frac{(Q_j - Q_i)'(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i)}{n}\right| \tag{A.74}$$

$$\leq n^{-1}\|Q_j - Q_i\|_2\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_i\|_2 \leq n^{-1}2\sup_i \|Q_i\|_2\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2$$

$$\leq 2n^{-1}T^{1/4}(mp)^{1/2}\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2$$

$$\left(p^{1/4}\tilde{B}_{q,m}^{1/2}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} + T^{1/4}m^{1/2}B_{q,m}\tilde{\phi}_{n,T,G}\right). \tag{A.75}$$

Therefore, by (A.71), (A.72) and (A.75),

$$Q_n(\boldsymbol{\gamma}) - Q_n(\boldsymbol{\gamma}^*)$$

$$\geq \sum_{i<j} \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2$$

$$\Big\{\lambda\rho'(4t_n) - 2n^{-1}T^{1/4}(mp)^{1/2}(p^{1/4}\tilde{B}_{q,m}^{1/2}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2}$$

$$+T^{1/4}m^{1/2}B_{q,m}\tilde{\phi}_{n,T,G})\Big\}$$

$$\geq \sum_{i<j} \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2$$

$$\Big\{\lambda\rho'(4t_n) - B_1 n^{-1}T^{1/4}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} - B_2 n^{-1}T^{1/2}\tilde{\phi}_{n,T,G}\Big\}, \tag{A.76}$$

where $B_1 = 2(mp\tilde{B}_{q,m})^{1/2}p^{1/4}$ and $B_2 = 2mp^{1/2}B_{q,m}$.

Let $t_n = o(1)$, then $\rho'(4t_n) \to 1$. Suppose that the following condition is true over the event $E_1 \cap E_2$,

$$B_1 n^{-1} T^{1/4}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} \to 0, \quad B_2 p n^{-1} T^{1/2} \phi_{n,T,G} \to 0, \tag{A.77}$$

then $P\left(Q_n(\boldsymbol{\gamma}) - Q_n(\boldsymbol{\gamma}^*) \geq 0\right) \geq 1 - e^{\iota} - e^{\iota^*}$. Once (A.77) holds, $Q_n(\gamma) - Q_n(^*) \geq 0$ with probability approaching to 1 as $n \to \infty$.

Now we show (A.77). Note that in this case, $g_{min} = g_{max} = n$ and $G = 1$, then

$$\phi_{n,T,G} = C_{q,m} \frac{g_{\max}^{1/2} G^{3/4}}{g_{\min} T^{3/4}} (Gp + 2\sqrt{Gp}\sqrt{\zeta} + 2\zeta)^{1/2} = C_{q,m} \frac{1}{n^{1/2} T^{3/4}} (p + 2\sqrt{p}\sqrt{\zeta} + 2\zeta)^{1/2}. \tag{A.78}$$

We further derive the second part of (A.77) as

$$B_2 p n^{-1} T^{1/2} \phi_{n,T,G} = B_2 p C_{q,m} \frac{1}{n^{3/2} T^{1/4}} (p + 2\sqrt{p}\sqrt{\zeta} + 2\zeta)^{1/2}. \tag{A.79}$$

As $\zeta \to \infty$, $(p + 2\sqrt{p}\sqrt{\zeta} + 2\zeta)^{1/2} = O(2\zeta^{1/2})$.

1. As $n \to \infty$ with $T$ fixed. Consider the first part of (A.77).

   (a) Consider $\zeta^* = O(n)$. Let $\zeta^* \to \infty$, since $(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} = (p + 2\sqrt{p} + 2)^{1/2} O(n^{1/2})$, then

   $$B_1 n^{-1}(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} T^{1/4} \leq B_1 T^{1/4} n^{-1}(p + 2\sqrt{p} + 2)^{1/2} O(n^{1/2}) \to 0.$$

(b) Consider $\zeta^* \gg n$ and $\zeta^* = o(n^2)$. Let $\zeta^* \to \infty$, then

$$B_1 n^{-1}(np+2\sqrt{np\zeta^*}+2\zeta^*)^{1/2}T^{1/4} \leq B_1 T^{1/4} n^{-1}(p+2\sqrt{p}+2)^{1/2} O(\zeta^{*1/2}) \to 0.$$

Now, consider the second part of (A.77) as $n \to \infty$. Let $\zeta = o(n^3)$ and $\zeta \to \infty$, then

$$B_2 p n^{-1} T^{1/2} \phi_{n,T,G} = \tilde{C}_1 \frac{1}{n^{3/2}} O(\zeta^{1/2}) \overset{n\to\infty}{\longrightarrow} 0,$$

where $\tilde{C}_1 = 2B_2 p T^{-1/4} C_{q,m}$.

2. As $T, n \to \infty$. Consider the first part of (A.77).

(a) Consider $\zeta^* = O(n)$. Let $\zeta^* \to \infty$ and $T = o(n^2)$. Since $(np + 2\sqrt{np\zeta^*} + 2\zeta^*)^{1/2} = (p + 2\sqrt{p} + 2)^{1/2} O(n^{1/2})$, then

$$B_1 n^{-1}(np+2\sqrt{np\zeta^*}+2\zeta^*)^{1/2}T^{1/4} \leq B_1 (p+2\sqrt{p}+2)^{1/2} T^{1/4} n^{-1} O(n^{1/2}) \to 0.$$

(b) Consider $n \ll \zeta^* = o(n^2/T^{1/2})$. Let $\zeta^* \to \infty$, then $T = o(n^4)$, and

$$B_1 n^{-1}(np+2\sqrt{np\zeta^*}+2\zeta^*)^{1/2}T^{1/4} \leq B_1 (p+2\sqrt{p}+2)^{1/2} T^{1/4} n^{-1} O(\zeta^{*1/2}) \to 0.$$

Now, consider the second part of (A.77) as $n, t \to \infty$. Let $\zeta = o(n^3 T^{1/2})$ and

$\zeta \to \infty$, then

$$B_2 p n^{-1} T^{1/2} \phi_{n,T,G} = \tilde{C}_2 \frac{1}{n^{3/2} T^{1/4}} O(\zeta^{1/2}) \stackrel{n\to\infty}{\longrightarrow} 0,$$

where $\tilde{C}_1 = 2B_2 p C_{q,m}$.

□

# Appendix B

# Full Simulation Results

## B.1 On the Choice of IVs

All three methods perform similar sizes close to 0.05. By choosing our choice of instruments, larger powers are presented generally for large frequency ratios. However, our method does not perform larger powers for small frequency ratio, especially with small alternatives.

**Table B.1**

Empirical Sizes and Powers for the Simulation Model: $T = 125$, $m = 4$

| d | c | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.5 | -0.5 | Miller | 6.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.5** | **0.1** | **0** | **0.4** | **22** | **83.1** | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.8** | **0.4** | **0.1** | **0.8** | **22.6** | **80.1** | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5** | **0.8** | **0.2** | **1.5** | **22.7** | **76.9** | 98.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.9** | **1.5** | **0.6** | **2.7** | **22.6** | **72.4** | 97.7 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 7.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.9 | 95.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.6** | **3.1** | **1.8** | **4.4** | **18.9** | **56.7** | **89.5** | 99.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.0 | -0.5 | Miller | 5.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.5** | **0.1** | **0** | **0.3** | **9** | **67.6** | 97.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5** | **0.3** | **0** | **0.4** | **9.9** | **65.8** | 96.5 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.5** | **0.9** | **0.1** | **0.7** | **10.6** | **60.9** | 95.7 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.6** | **1.2** | **0.2** | **0.9** | **11.9** | **58** | 93.5 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 7 | 99.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.2 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.3** | **3.5** | **1.3** | **3.8** | **13.8** | **45.5** | **81.3** | 97.4 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | -0.5 | Miller | 6.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.4** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.5 | **0** | **0** | **0** | **0.2** | **7.5** | **47** | **86.5** | 97.8 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.9** | **0.2** | **0** | **0** | **0.4** | **7.8** | **46.5** | **85.7** | 97.7 | 99.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 5.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.1** | **0.4** | **0** | **0.1** | **0.6** | **8.5** | **45.9** | **84.1** | 97.1 | 99.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | **5.1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.8** | **0.8** | **0.1** | **0.1** | **0.9** | **9.8** | **43.9** | **81.6** | 96.4 | 99.4 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | **5.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 3.7 | 99.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 6.5 | **3.4** | **1.2** | **1.7** | **3.8** | **12.6** | **37.3** | **70.4** | 90.4 | 98.4 | 99.5 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Table B.2**

Empirical Sizes and Powers for the Simulation Model: $T = 125$, $m = 150$

| d | | c | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.5 | -0.5 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.3** | **31.4** | **45.3** | **52.4** | **58.7** | **62.1** | **66.2** | **69.5** | **73** | **76.1** | **78.2** | **79.8** | **81** | **82.4** | **83.5** | **83.9** | **84.8** | **85.3** | **85.7** | **86.4** | **86.9** |
| | | New | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **23.1** | **39.4** | **49.3** | **56.5** | **60.7** | **65** | **68.4** | **71.9** | **75.5** | **77.7** | **79.2** | **81.1** | **81.9** | **82.7** | **83.6** | **84.1** | **85** | **85.6** | **86.1** | **86.5** |
| | | New | 5.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.3** | **16.6** | **32.6** | **43.7** | **52.4** | **58** | **62.4** | **66.9** | **70.3** | **73.3** | **76.2** | **78.4** | **80** | **81.4** | **82.2** | **82.9** | **83.7** | **84.4** | **85.2** | **85.6** | **86** |
| | | New | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.4** | **12.4** | **24.8** | **36.2** | **45.3** | **52.9** | **58.6** | **63.3** | **67.2** | **71.1** | **73.8** | **76.6** | **78.5** | **80.3** | **81.2** | **82.1** | **83** | **83.7** | **84.1** | **85** | **85.6** |
| | | New | 5.6 | 98.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.7 | **40.8** | 91.7 | 99.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.1** | **6.5** | **10.8** | **15.2** | **20.9** | **27.7** | **33.2** | **39.2** | **44.8** | **49.3** | **53.4** | **58.3** | **62.8** | **66.3** | **69.5** | **72.6** | **74.2** | **75.7** | **77.1** | **78.2** | **79.6** |
| | | New | 5.5 | **41.4** | 90.3 | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.0 | -0.5 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | **32.2** | **41.5** | **47.4** | **52.3** | **57.2** | **61** | **64.1** | **67.2** | **70.1** | **73** | **74.7** | **76.6** | **77.8** | **79** | **80** | **81** | **81.9** | **82.8** | **83.6** | **84.1** |
| | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | **26.4** | **39.1** | **45.9** | **51.2** | **55.8** | **60** | **63.1** | **66.8** | **69.7** | **72.3** | **74.4** | **76** | **77.3** | **78.8** | **79.9** | **80.6** | **82** | **82.9** | **83.8** | **84.2** |
| | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.4 | **21.5** | **34.6** | **43.3** | **49.1** | **54.4** | **58.6** | **62.3** | **66.2** | **68.8** | **71.6** | **73.7** | **75.3** | **77** | **78.5** | **79.5** | **80.2** | **81.6** | **82.7** | **83.4** | **84.1** |
| | | New | **5.4** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | **15** | **28.7** | **38.5** | **45.7** | **51.2** | **56.1** | **60.4** | **64.6** | **67.5** | **70.4** | **72.7** | **74.3** | **76.2** | **77.5** | **79** | **79.9** | **81** | **82.3** | **83.1** | **83.5** |
| | | New | **5.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | **5.7** | **72.1** | 99.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.7** | **8.4** | **12.9** | **19.8** | **27.7** | **34.7** | **40** | **45.2** | **50.5** | **55.5** | **59.6** | **63.6** | **65.9** | **68.9** | **71** | **72.8** | **74.8** | **76.1** | **77.4** | **78.9** | **80.1** |
| | | New | **5.7** | **73.1** | 99.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | -0.5 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.6 | **44.1** | **49.5** | **54.2** | **58.8** | **63.1** | **66** | **70.6** | **73.9** | **75.9** | **78.1** | **80** | **81.7** | **82.9** | **84.1** | **84.9** | **85.5** | **86.2** | **86.9** | **87.4** | **87.8** |
| | | New | **5.4** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.6** | **41.5** | **48.5** | **53.5** | **58.6** | **62.9** | **66** | **70.4** | **73.7** | **75.7** | **78.3** | **80** | **81.5** | **82.8** | **84.1** | **84.8** | **85.5** | **86.2** | **86.7** | **87.3** | **87.7** |
| | | New | **5.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.8 | **37.6** | **47.6** | **52.9** | **58.3** | **62.4** | **65.9** | **70.3** | **73.5** | **75.9** | **78.2** | **80** | **81.3** | **82.9** | **84.2** | **84.8** | **85.3** | **86.3** | **86.7** | **87.4** | **87.7** |
| | | New | **5.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | **32.5** | **45.3** | **52.1** | **56.8** | **61.8** | **65.5** | **69.2** | **73.3** | **75.9** | **77.8** | **79.9** | **81.4** | **82.8** | **83.9** | **84.7** | **85.2** | **86** | **86.7** | **87.2** | **87.8** |
| | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.4 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.3** | **15.2** | **29.8** | **40.9** | **49.3** | **55** | **60.5** | **64.6** | **68.2** | **71.6** | **75.2** | **77** | **79.2** | **81.2** | **82.5** | **83.5** | **84.5** | **85.3** | **85.9** | **86.5** | **87** |
| | | New | 5.7 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

## Table B.3

Empirical Sizes and Powers for the Simulation Model: $T = 125$, $m = 365$

| d | c | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.5 | -0.5 | Miller | 7.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.7 | **15.5** | **19.6** | **23.1** | **26.8** | **30** | **33.2** | **36.1** | **39.1** | **41.4** | **43** | **45.2** | **47.8** | **50.1** | **51.6** | **54** | **55.9** | **57.5** | **59.3** | **60.8** | **61.5** |
| | | New | **4.7** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 7.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.5 | **13** | **17.7** | **21.1** | **25.1** | **28.7** | **32.6** | **35** | **38** | **40.7** | **42.5** | **44.6** | **46.8** | **49.5** | **51.5** | **53.4** | **55.6** | **57.1** | **58.8** | **60.1** | **61.1** |
| | | New | **4.8** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 7.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.6 | **11.1** | **15.4** | **19** | **22.8** | **26.8** | **30.2** | **33.1** | **36.2** | **38.8** | **41.4** | **43.4** | **46** | **48.4** | **50.4** | **52.3** | **54.7** | **56.4** | **58** | **59.2** | **60.5** |
| | | New | **4.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 7.2 | 98.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.7 | **8.8** | **13.5** | **17.2** | **20.2** | **23.7** | **27.4** | **30.7** | **33** | **36.3** | **39** | **41.5** | **44** | **46.8** | **48.5** | **50.6** | **52.7** | **54.2** | **56** | **57.9** | **59.3** |
| | | New | **4.7** | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 7 | **39.9** | **89.7** | 99.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.8 | **7.5** | **8.6** | **9.8** | **11.7** | **13.9** | **16.3** | **18.1** | **20.3** | **22.7** | **25.1** | **27.8** | **30.1** | **32** | **34.3** | **35.9** | **37.7** | **39** | **40.4** | **42.7** | **44.5** |
| | | New | 4.6 | **44.8** | 91.5 | 99.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.0 | -0.5 | Miller | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | **15.2** | **19** | **20.5** | **22.6** | **24.9** | **27.3** | **30.1** | **33.3** | **35** | **37.3** | **39.5** | **41.3** | **43.8** | **46.2** | **48** | **50.2** | **52.3** | **53.7** | **55.1** | **56.4** |
| | | New | **4.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.8 | **13.8** | **17.8** | **20.4** | **22.6** | **24.5** | **27** | **29.7** | **32.7** | **35.2** | **37.3** | **39.2** | **41** | **42.9** | **45.3** | **47.4** | **49.4** | **51.7** | **53.5** | **54.8** | **56.1** |
| | | New | **4.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.9 | **11.8** | **16.1** | **19.3** | **22.3** | **23.9** | **26.4** | **29** | **31.8** | **34.3** | **36.8** | **38.4** | **40.8** | **42.8** | **44.6** | **46.9** | **48.9** | **51** | **52.4** | **54.1** | **55.7** |
| | | New | **4.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | **9.5** | **15** | **18** | **20.7** | **23** | **25.7** | **27.9** | **30.4** | **33.3** | **35.4** | **37.5** | **39.7** | **41.6** | **43.7** | **46** | **48.2** | **50.3** | **51.5** | **53.1** | **54.4** |
| | | New | **4.7** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 6.5 | **70** | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.8 | **7** | **9.2** | **11.5** | **14.5** | **16.4** | **18.4** | **20.7** | **23.2** | **26.5** | **28.3** | **30** | **32.3** | **34.9** | **36.9** | **38.6** | **40.9** | **42.6** | **44.8** | **46.9** | **48.4** |
| | | New | **4.6** | **75.9** | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | -0.5 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.1 | **20.7** | **23.8** | **25.1** | **27.7** | **30.7** | **33.2** | **35.7** | **38.5** | **41.3** | **43.4** | **45.7** | **48.1** | **50.2** | **51.9** | **54.2** | **56.4** | **58** | **59.2** | **61.1** | **62.6** |
| | | New | **4.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | **19.8** | **23.8** | **25.6** | **27.6** | **30.3** | **33.2** | **35.5** | **38.5** | **41.2** | **43.6** | **45.9** | **47.9** | **50.3** | **51.9** | **54.2** | **56.3** | **57.9** | **59.5** | **61.1** | **62.3** |
| | | New | **4.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6 | **18.9** | **23** | **25.7** | **28.1** | **30.3** | **33** | **35.8** | **38.5** | **41** | **43.8** | **45.7** | **47.9** | **50.3** | **51.8** | **54** | **56.1** | **57.8** | **59.7** | **61** | **62.4** |
| | | New | **4.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.9 | **16.1** | **21.5** | **25.3** | **27.6** | **30.1** | **32.9** | **35.6** | **38.2** | **40.6** | **43** | **45.5** | **47.8** | **50** | **51.8** | **54** | **56.2** | **57.6** | **59.7** | **60.8** | **62** |
| | | New | **4.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 6 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.8 | **9.3** | **15.1** | **19.4** | **24** | **27.5** | **29.8** | **32.9** | **36.3** | **38.6** | **41.3** | **43.8** | **45.8** | **48.3** | **50.2** | **52.8** | **54.7** | **56.4** | **58.2** | **59.2** | **60.4** |
| | | New | **4.7** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

## Table B.4

Empirical Sizes and Powers for the Simulation Model: $T = 512$, $m = 4$

| d | | c | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.5 | -0.5 | Miller | 5.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.4** | **6.2** | **4.4** | **1.3** | **0.1** | **0** | **0** | **0.1** | **0.2** | **0.5** | **5.2** | **27.1** | **71.5** | 96.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.2 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.9** | **5.4** | **4.5** | **2.2** | **0.6** | **0.1** | **0.1** | **0.1** | **0.5** | **1.4** | **7.7** | **27** | **64.8** | 92.4 | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.6 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.3 | 99.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.2** | **5.1** | **4.5** | **3** | **1.3** | **0.4** | **0.1** | **0.2** | **0.9** | **2.7** | **9.7** | **26.4** | **56.5** | **85.8** | 97.7 | 99.7 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 7.1 | 97.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | 93.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.7** | **5** | **4.5** | **3.7** | **2** | **0.9** | **0.4** | **0.6** | **1.5** | **3.7** | **10.6** | **24.6** | **48.8** | **76.6** | 94.1 | 98.7 | 99.9 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 6.9 | **57.2** | 98.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 6.3 | **49.8** | 96.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.3** | **4.6** | **4.5** | **3.9** | **3.1** | **2.5** | **1.9** | **2** | **2.6** | **4.5** | **9.4** | **17.1** | **29.8** | **48.6** | **70.3** | **86** | 95.1 | 98.5 | 99.9 | 100 | 100 |
| 0.0 | -0.5 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.9** | **6.7** | **3.9** | **1.1** | **0** | **0** | **0** | **0** | **0** | **0.1** | **1.8** | **11.1** | **45.5** | **84.6** | 98.4 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.4 | **5.7** | **4.6** | **2.6** | **0.5** | **0** | **0** | **0.1** | **0.3** | **0.8** | **3** | **13.7** | **42.4** | **78.6** | 96.2 | 99.4 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.6 | **5.1** | **4.7** | **3.3** | **1.2** | **0.4** | **0.2** | **0.3** | **0.5** | **1.6** | **5.3** | **14.8** | **39.2** | **70.4** | 91.1 | 98.2 | 99.9 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.1 | 98.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.3 | 97.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.2** | **5.9** | **5.4** | **3.9** | **2.1** | **1.1** | **0.5** | **0.6** | **1.1** | **3** | **7.1** | **15.6** | **35.3** | **61.6** | **83.9** | 95.5 | 99 | 99.9 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.6 | **61.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.1** | **60.6** | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.4 | **5.5** | **5.6** | **4.6** | **3.8** | **3.1** | **2.2** | **2.8** | **3.6** | **5.4** | **8.6** | **14.4** | **23.9** | **39.6** | **57.4** | **75** | **88.4** | 96 | 98.9 | 99.7 | 100 |
| 0.5 | -0.5 | Miller | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **4.8** | **4.6** | **2** | **0.2** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0.1** | **1** | **8.7** | **32.8** | **68.9** | 92.7 | 99.2 | 100 | 100 | 100 |
| | 0.0 | Miller | 5.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | **5.1** | **4.9** | **3.1** | **1** | **0.1** | **0** | **0** | **0** | **0** | **0** | **0.1** | **0.4** | **2.2** | **10.7** | **33.5** | **64.9** | **89.2** | 97.9 | 99.9 | 100 | 100 |
| | 0.3 | Miller | 5.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.3 | **5.3** | **3.7** | **1.9** | **0.8** | **0.1** | **0.1** | **0.1** | **0.1** | **0.1** | **0.4** | **1.6** | **3.5** | **13.6** | **33** | **59** | **83** | 94.9 | 99.2 | 99.9 | 100 |
| | 0.5 | Miller | **4.8** | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **4.8** | 99.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.9 | **5.8** | **4.7** | **3.2** | **1.2** | **0.6** | **0.2** | **0.2** | **0.3** | **0.7** | **1.4** | **2.5** | **5.8** | **15.1** | **31.5** | **52.9** | **75.8** | 90.8 | 97.1 | 99.6 | 100 |
| | 0.8 | Miller | **4.3** | **77.6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 3.8 | **74.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | New | 5.9 | **5.8** | **5.6** | **4.7** | **3.6** | **2.9** | **2.4** | **2.2** | **2.4** | **3.1** | **4.3** | **6.7** | **10.9** | **16.1** | **25.9** | **39.3** | **53.5** | **69.2** | **83.1** | 91.9 | 97.1 |

**Table B.5**

Empirical Sizes and Powers for the Simulation Model: $T = 512$, $m = 150$

| d c | | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.5 | -0.5 | Miller | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **23.9** | **61.2** | **81.7** | 90.5 | 94.8 | 96.2 | 97.3 | 98 | 98.2 | 98.4 | 98.4 | 98.5 | 98.5 | 98.7 | 98.7 | 98.7 | 98.9 | 98.9 | 98.9 | 99 |
| | | New | 5.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.6 | 98.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **13.2** | **37.5** | **62.5** | **78.2** | **86.6** | 92.2 | 94.9 | 96.1 | 97.1 | 97.8 | 98.1 | 98.3 | 98.5 | 98.5 | 98.6 | 98.6 | 98.7 | 98.9 | 98.9 | 98.9 |
| | | New | 5.8 | 98.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.8 | **85.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.2 | **9.7** | **22.1** | **40.6** | **58.6** | **72.8** | **81.9** | **87.4** | 91.8 | 94.2 | 95.6 | 96.7 | 97.3 | 97.9 | 98.2 | 98.3 | 98.6 | 98.6 | 98.7 | 98.7 | 98.8 |
| | | New | 5.9 | **86.1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.6 | **55** | 98.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **7.3** | **13.6** | **25** | **39** | **52.5** | **65.3** | **74.3** | **81.1** | **85.8** | **89.5** | 92.3 | 94.3 | 95.9 | 96.5 | 97.3 | 97.8 | 98 | 98.2 | 98.4 | 98.6 |
| | | New | 5.9 | **58.5** | 98.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.9 | **12.9** | **37.8** | **72.0** | 92.2 | 98.2 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.3** | **5.5** | **6.3** | **8.6** | **11.1** | **14.2** | **18.9** | **23.7** | **30.1** | **36.1** | **42.7** | **48.9** | **54.4** | **60.2** | **66.2** | **70.4** | **74.7** | **78** | **81.1** | **83.9** | **86.1** |
| | | New | 5.5 | **14.9** | **40.8** | **73.9** | 93 | 98.1 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.0 | -0.5 | Miller | 6.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **37.1** | **73.9** | **86.9** | 91.3 | 93.5 | 95 | 95.8 | 96.3 | 96.6 | 97 | 97.1 | 97.4 | 97.6 | 97.8 | 97.9 | 98 | 98.1 | 98.2 | 98.3 | 98.3 |
| | | New | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.5** | **21.8** | **52.9** | **75.3** | **85.1** | 90.3 | 92.5 | 94.3 | 95.4 | 95.9 | 96.4 | 96.7 | 97.1 | 97.3 | 97.5 | 97.6 | 98 | 98.1 | 98.1 | 98.2 | 98.3 |
| | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.3** | **13.6** | **34.1** | **56.6** | **73.5** | **83** | **87.8** | 91.1 | 93 | 94.2 | 95.5 | 96 | 96.5 | 96.9 | 97.3 | 97.5 | 97.7 | 97.7 | 98 | 98 | 98.1 |
| | | New | 5.5 | 99.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.2 | **89** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **9.4** | **22.5** | **38.5** | **55.3** | **68.2** | **78.3** | **84.3** | **87.8** | 90.8 | 92.4 | 93.8 | 95.1 | 96 | 96.4 | 96.7 | 97 | 97.4 | 97.7 | 97.7 | 97.9 |
| | | New | 5.4 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 6.1 | **23.4** | **71.3** | 96.5 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.1** | **5.5** | **7.6** | **12.1** | **17.2** | **24.1** | **31** | **37.8** | **45.6** | **53.2** | **60.1** | **66.5** | **72** | **76.8** | **80.3** | **83.7** | **85.9** | **88.5** | **89.9** | 91.3 | 92.2 |
| | | New | 5.4 | **25.4** | **72.6** | 96.5 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | -0.5 | Miller | 6.0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | **79.9** | 94.6 | 96.8 | 97.4 | 97.5 | 98 | 98.1 | 98.4 | 98.5 | 98.7 | 98.9 | 99 | 99.1 | 99.1 | 99.2 | 99.2 | 99.2 | 99.2 | 99.2 | 99.2 |
| | | New | **5.4** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.7 | **59.6** | **89.3** | 95 | 96.7 | 97.5 | 97.7 | 98 | 98.1 | 98.4 | 98.6 | 98.8 | 98.9 | 99 | 99.1 | 99.1 | 99.2 | 99.2 | 99.2 | 99.2 | 99.2 |
| | | New | 5.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.5** | **38.8** | **78.4** | 90.7 | 95 | 96.6 | 97.4 | 97.8 | 98 | 98.2 | 98.4 | 98.7 | 98.9 | 98.9 | 99 | 99.1 | 99.1 | 99.2 | 99.2 | 99.3 | 99.3 |
| | | New | **5.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 6.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.5 | **23.5** | **60.8** | **82** | 90.7 | 94 | 95.9 | 97.1 | 97.7 | 98 | 98.2 | 98.4 | 98.7 | 98.8 | 98.9 | 99 | 99.1 | 99.1 | 99.1 | 99.2 | 99.3 |
| | | New | **5.3** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.8 | **69.6** | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.4** | **8.9** | **18.5** | **33** | **49.2** | **64.3** | **74.4** | **82.8** | **87.9** | 91.2 | 93.1 | 94.9 | 96.2 | 97.1 | 97.7 | 97.9 | 98.2 | 98.6 | 98.8 | 98.9 | 98.9 |
| | | New | **5.4** | **72.1** | 99.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Table B.6**

Empirical Sizes and Powers for the Simulation Model: $T = 512$, $m = 365$

| d | c | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.5 | -0.5 | Miller | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.4 | **11** | **25.1** | **40.3** | **50.5** | **57.1** | **62.4** | **66.5** | **68.5** | **70.7** | **73.3** | **74.7** | **76.9** | **78.2** | **79.5** | **80.3** | **81.5** | **82.2** | **83** | **83.8** | **84.7** |
| | | New | **5.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 5.9 | 98.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.3 | **8.1** | **15.3** | **25.7** | **36.6** | **45.2** | **52.1** | **57.7** | **62.3** | **65.5** | **68.6** | **70.7** | **73.1** | **75.1** | **77** | **78.4** | **79.9** | **81** | **81.9** | **82.7** | **83.3** |
| | | New | **5.1** | 99.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 5.8 | **84.5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **6.3** | **10.5** | **16.7** | **24.4** | **32.1** | **40** | **46.6** | **51.9** | **56.4** | **60.3** | **64.2** | **67** | **69.9** | **72.5** | **74.2** | **75.9** | **77.6** | **78.7** | **80.4** | **81.3** |
| | | New | **5.2** | **89.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 5.8 | **54.7** | 98.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.1** | **5.4** | **8.2** | **11.4** | **15.9** | **21.7** | **27.5** | **33.5** | **39.6** | **44.8** | **49.4** | **53.1** | **57.4** | **60.6** | **63.7** | **66.7** | **68.8** | **71.6** | **73.8** | **75.3** | **76.6** |
| | | New | 5.3 | **63.2** | 99.3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.7 | **14.3** | **38.6** | **71.9** | 92.1 | 98.9 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5** | **5.1** | **5.4** | **5.9** | **7.3** | **8.4** | **9.7** | **11.3** | **13** | **15** | **17.2** | **19.7** | **22.4** | **25.2** | **28.6** | **31.5** | **33.9** | **36.8** | **38.8** | **41.4** | **44.1** |
| | | New | 5.5 | **16** | **45** | **77.8** | 94.4 | 99.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.0 | -0.5 | Miller | 5.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 4.7 | **15.5** | **34.3** | **47.6** | **54** | **57.6** | **61** | **63.2** | **65.5** | **67.8** | **69.1** | **70.4** | **71.6** | **72.5** | **73.3** | **74** | **74.8** | **75.5** | **76.5** | **77.4** | **78.5** |
| | | New | **5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 5.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 4.7 | **9.7** | **22.8** | **35.3** | **44.8** | **51.6** | **55.6** | **59** | **61.7** | **64.5** | **67** | **68.5** | **70.1** | **71.2** | **72.5** | **73** | **74.2** | **74.7** | **75.6** | **76.6** | **77.8** |
| | | New | **5.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 5.7 | 99.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 4.6 | **7.2** | **14.6** | **24.8** | **33.7** | **41.6** | **47.7** | **52.9** | **56.5** | **59.5** | **62.5** | **64.8** | **67.1** | **68.6** | **70.1** | **71.3** | **72.6** | **73.5** | **74.4** | **75.4** | **76.9** |
| | | New | **5.2** | 99.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 5.7 | 88.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **4.6** | **6.2** | **9.9** | **16.5** | **23.8** | **30.7** | **37.5** | **43.1** | **48.2** | **52** | **56.1** | **58.9** | **61.7** | **63.9** | **66** | **67.9** | **69.3** | **71.2** | **72.4** | **73.4** | **75** |
| | | New | **5.4** | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.6 | **23.6** | **71.3** | 96.8 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **4.7** | **4.9** | **5.4** | **6.6** | **8.2** | **10.5** | **12.3** | **15.8** | **19** | **22.6** | **26.2** | **30** | **32.8** | **35.7** | **39.5** | **41.8** | **44.8** | **47.3** | **50.6** | **53.3** | **55.3** |
| | | New | **5.3** | **29.3** | **77.9** | 97.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.5 | -0.5 | Miller | 5.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **38.2** | **57.6** | **64.2** | **68.5** | **70.4** | **72.3** | **74.1** | **75.1** | **76.1** | **77.4** | **78.3** | **79** | **79.9** | **81** | **81.6** | **82.3** | **82.9** | **83.4** | **84.7** | **85.5** |
| | | New | **5.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.0 | Miller | 5.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.2** | **24.8** | **49.1** | **58.8** | **64.9** | **68.3** | **70.4** | **72.3** | **74.4** | **75.8** | **76.8** | **78** | **78.8** | **79.5** | **80.3** | **81.4** | **82** | **82.4** | **83.4** | **84.3** | **85.3** |
| | | New | **5.2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.3 | Miller | 5.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | 5.3 | **15.6** | **36.6** | **52** | **58.7** | **64.4** | **67.7** | **70.5** | **72.4** | **74.3** | **76** | **77.1** | **78.2** | **79** | **79.9** | **80.8** | **81.4** | **82.3** | **83.1** | **83.9** | **85** |
| | | New | **5.1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.5 | Miller | 5.5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5.3** | **11.1** | **25.4** | **40.5** | **51.6** | **57.8** | **62.1** | **66.8** | **69.1** | **71.5** | **73.3** | **75.6** | **76.9** | **78.1** | **79** | **80** | **80.7** | **81.6** | **82.6** | **83.5** | **84.2** |
| | | New | **5.3** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | Miller | 5.4 | **70.9** | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | | AGK | **5** | **6.7** | **8.9** | **14** | **19.9** | **26.9** | **33.7** | **41.3** | **47.4** | **52** | **56.7** | **60** | **63.2** | **66.6** | **69.3** | **71** | **72.8** | **74.7** | **76.4** | **78.3** | **79.5** |
| | | New | 5.2 | **78** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |