A SEMIPARAMETRIC CAPITAL ASSET PRICING MODEL WITH LIQUIDITY AND ITS APPLICATION

BY LEE, JIYON

A dissertation submitted to the Graduate School - New Brunswick Rutgers, The State University of New Jersey in partial fulfillment of the requirements for the degree of Doctor of Philosophy Graduate Program in Economics Written under the direction of Dr. Roger Klein and approved by

New Brunswick, New Jersey October, 2011
ABSTRACT OF THE DISSERTATION

A Semiparametric Capital Asset Pricing Model with Liquidity and Its Application

by Lee, Jiyon

Dissertation Director: Dr. Roger Klein

In this paper, I propose a model that flexibly accounts for liquidity. Extending conventional asset pricing equations, I construct a semiparametric model that depends on an unknown liquidity function. This function depends on the standard set of liquidity proxies such as bid-ask spread, turnover rate, number of transactions and number of non-trading days. This approach neither imposes any restriction on the structure of liquidity nor limits the number of proxies which can be associated with liquidity function. The model is estimated by using semiparametric least squares in a time series context and the estimator is shown to be consistent and asymptotically normal.

In applying this model to stock data we find that the estimated coefficients of bid-ask spread, turnover rate and number of non-trading days are similar in magnitude and are statistically significant. That is, bid-ask spread, turnover rate and number of non-trading days have similar weights in affecting asset returns. Such an outcome implies that missing any of these three variables may significantly change the estimated value of liquidity thus distort the relationship between asset returns and liquidity. The signs of the estimates also support the common belief that bid-ask spread and
number of non-trading days are related with the illiquidity cost thus increasing re-
quired returns while the other two capture the flexibility in trading. Liquidity impact
parameters in construction, manufacturing, wholesale trade and retail industries are
statistically significant. The two industries which have the largest liquidity impacts
are agriculture, forestry and fishing, and public administration. The companies in
these two industries are deemed to have little incentive to smooth out the liquidity
level, which induces the second order impacts of the liquidity to be high.
Acknowledgements

Without the help of faculty, friends, and family members, I would not have been able to complete this dissertation. I would like to give particular mention to some of those who have made this work possible.

First and foremost, I want to express my utmost gratitude to my advisor, Dr. Roger Klein. It has been an honor and a privilege to be his student. His enthusiasm, unsurpassed insight and profound knowledge in his research were motivational and inspiring. He has also been a supportive and patient advisor throughout my graduate school career. At each stage of my research over the past 6 years, especially whenever I was confronted with difficult problems, he not only has given me invaluable guidance to solve those problems but he has encouraged and supported me through them. Such valuable lessons have ultimately expanded my perspective on how to perform my research, which I daresay is the greatest achievement of my Ph.D. experience.

I am also grateful to Dr. John Landon-Lane for his incisive advice. He offered me different angles of looking at the issues in my thesis and provided guidance how to effectively express an idea and communicate with an audience during a presentation. His keen insight on the econometrical problems, especially in connection with empirical applications, was truly inspiring.

Additionally, I would like to thank my other two dissertation committee members, Dr. Ira Gang and Dr. Chan Shen, for their time and helpful comments. I also thank Dr. Hiroki Tsurumi and Dr. Norman Swanson for the profitable advice they gave me during and after the job market seminar. I appreciate Dr. Martin Perry’s practical help in preparing for job market interviews and seminars. I would also like to thank
all the faculty members who gave me invaluable comments during my practice inter-
views; I am particularly indebted to Dr. Richard McLean for his insightful guidance
in conceptualizing the theoretical merits of my research.
I want to thank all of my fellow graduate students at Rutgers for inspiring me with
their unfailing enthusiasm in their research. The encouragements and support from
my good old friends, Hyejin, Jinwon, Sowon, Jiwon, Wonkyung, Hansook, Woojung
and Youra also sustained me during the toughest time of my Ph.D. years.
Lastly, I would like to thank my parents, who taught me the greatest value of a life
and supported me in all my pursuits, and my sister, Seyon, who has an unconventional
sense of humor and makes me laugh all the time.
Dedication

For my family
# Table of Contents

Abstract .............................................................................................................. ii  
Acknowledgements ................................................................................................. iv  
Dedication ............................................................................................................... vi  
List of Tables .......................................................................................................... ix  
List of Figures ......................................................................................................... x  

1. The Model and Estimator .................................................................................. 1  
   1.1. Introduction ................................................................................................. 2  
   1.2. Model ........................................................................................................... 5  
   1.3. Estimator .................................................................................................... 6  
   1.4. Assumptions ............................................................................................... 8  
   1.5. Conclusion ................................................................................................. 11  

2. Properties of the Estimator ............................................................................ 13  
   2.1. Introduction ................................................................................................. 14  
   2.2. Finite Sample Properties .......................................................................... 14  
   2.3. Large Sample Properties ........................................................................... 17  
      2.3.1. Consistency of the Estimator ................................................................. 17  
      2.3.2. Asymptotic Distribution .................................................................... 19  
   2.4. Conclusion ................................................................................................. 22  

3. Application ..................................................................................................... 24  
   3.1. Introduction ................................................................................................. 25
3.2. Estimation Results ............................................. 26
3.3. Conclusion ...................................................... 30

Bibliography .......................................................... 41

Appendix A. Proofs of Theorems ................................. 42
A.1. Consistency Lemmas ............................................. 42
A.2. Proof of Theorem 1 .............................................. 44
A.3. Proof of Theorem 2 .............................................. 46
  A.3.1. Step 1: First part of Hessian converges to a positive definite
         matrix .......................................................... 47
  A.3.2. Step 2: Remainder of Hessian converges to zero in probability. 50
  A.3.3. Step 3: Gradient converges to a multivariate normal random
         vector in distribution ......................................... 52
         R converges to zero in probability .......................... 52
         (A) Term R_1 converges to zero in probability ......... 53
         (B) Term R_2 converges to zero in probability ......... 56
         (C) Term R_3 converges to zero in probability ......... 58
         N converges to a multivariate normal random variable. 59
A.4. Proof of Theorem 3 .............................................. 68
A.5. Detailed Discussion for Cross Product Term in R_1 .......... 71
A.6. Range for Bandwidth and Mixing Rate ...................... 82
List of Tables

3.1. Monte-Carlo Simulation Results ........................................... 32
3.2. Estimation Results .............................................................. 33
3.3. Data Description ............................................................... 34
3.4. Standard Deviation of Estimated Liquidity ............................. 35
3.5. Sample Cross Correlations of Error Terms ............................ 36
3.6. Marginal Effects ............................................................... 37
A.1. Rate Comparison .............................................................. 70
A.1. Lower bound of $M_\tau$ ........................................ 84
Chapter 1

The Model and Estimator
1.1 Introduction

In conventional asset pricing models, investors can trade assets without frictions. In such a world, any two assets with the same cash flows have the same price. However, it is not hard to find evidence that assets with the same payoff process do not share the same price in the real world. As shown in Amihud, Mendelson, and Pedersen (2005)'s list of the existing studies, for instance, identical Treasuries, notes and certain type of restricted securities often trade at different prices. These observations suggest the possibility that the transactions are subject to the trading frictions which form different pricing environment for different assets. In order to examine how these frictions or transaction costs affect asset prices, the literature often assumes that investors have limited ability to participate in trading due to different liquidity levels in stocks.

Hicks (1967) defines liquidity as an ability to trade large quantities quickly at low cost with little impact on price. Theoretical approaches hypothesize various aptitudes of investors in trading such as different investment horizons, stochastic arrivals or limited ability to collect information and study how price of an asset is determined in such framework. Amihud and Mendelson (1986), Constantinides (1986) and Huang (2003) study the effect of liquidity by assuming exogenously given transaction costs in their equilibrium models. Amihud and Mendelson (1986), for example, argues that the assets with different transaction costs are priced differently based on trading horizons. The assets with greater transaction costs must offer higher returns to the investors who own those assets in equilibrium. Duffie, Garleanu, and Pedersen (2005), Vayanos and Wang (2007) and Weill (2008) develop search-based models of asset trading where investors should incur costs to find buyers or sellers. They examine the model with endogenously determined transaction costs and discuss that in equilibrium, more liquid assets have higher prices reflecting a liquidity premium. While search theory provides a direct way to model the channels through
which market frictions work, recent work of Guvenen (2009) studies asset price under a two-agent macroeconomic model. Illiquidity of an asset is associated with trading frictions driven by limited participation in this case and the author shows that the model generates a higher equity premium when investors have limited access to the stock market. These theoretical approaches also rationalize the standard measures of liquidity often used in empirical applications.

In line with the theoretical work, many of the empirical studies have reported that after controlling for risk premia, asset returns are related with variations in liquidity measured by various proxies. Amihud (2002), Datar, Naik, and Radcliffe (1998), Pastor and Stambaugh (2003), Liu (2006) and many others use market microstructure variables such as bid-ask spread, turnover rate, trading volume and non-trading days as a measure of liquidity. However, the main empirical difficulty in asset pricing is that liquidity is a latent variable. Numerous applications in this area have dedicated a great deal of effort to finding the measure that best fits Hick’s definition. Researchers have devised a measurement by using indicators related to stock prices which are deemed to co-move with the trading cost reasonably well and market microstructure variables became the standard measures of liquidity in this regard. However as Amihud states in his 2002 paper, all the aspects of liquidity are rarely captured by a single indicator, thus such measures inevitably lose their power to explain liquidity. Despite the existence of a set of standard variables, searches for a better measure are still in progress and there seems to be no consensus on the standard functional form of existing measures as an approximation of liquidity.

By employing a semiparametric single index model, this paper attempts to overcome such limitation and improve the performance of the liquidity estimator. Regarding the goal, we mainly focus on two things: first, systematic risk of an asset is estimated by the relationship between individual excess return and market excess return as suggested by the capital asset pricing model. Secondly, we construct a liquidity measure that is an unknown function of an index of liquidity indicators such
as bid-ask spread, turnover rate, number of transaction and non-trading days.

We may view this model as an extension of the capital asset pricing model where the variability in excess return is explained by the variability in market excess return and a liquidity factor. Liquidity additively enters into the model as an unknown function and is estimated by a semiparametric conditional expectation estimator. The distinctive feature of the liquidity measure here is that the function depends on an index of standard set of market microstructure indicators. Thus this measure does not limit the number of variables that can be associated with the liquidity function and allow us to consider many dimensions of liquidity. By estimating an index of liquidity indicators, we can learn the relative importance of these indicators in contributing to aggregate liquidity. The functional form of liquidity is unknown and assumed to be the same for all assets. The aggregate liquidity varies over assets not because the function itself changes but because the individual liquidity variables change. We allow the impact of aggregate liquidity to be asset-specific.

This model can be considered as a variation of a partially linear model, which is in fact a generalization of the regression model. The estimation method is called “semiparametric least square (SLS)” and a semiparametric analogue of nonlinear least square. Ichimura (1993) studies asymptotic properties of SLS estimator with a single index. The framework of this paper is closest to Ichimura and Lee (1991) which extends Ichimura (1993) to the case of multiple indices. We will extend their approach to the panel data setup under time dependence and heterogeneity.

The first chapter focuses mainly on presenting basic structure of the model and the environment on which the theoretical features of the model are established. Section 2 introduces the model and the estimator and Section 3 discusses the necessary assumptions.
1.2 Model

Assume there are $i = 1, \ldots, n$ assets and $t = 1, \ldots, T$ time periods. Let $l_{it} = \left( l_{it}^{(1)} \cdots l_{it}^{(b)} \right)$ be a vector of individual liquidity measures such as the trading cost (bid ask spread), the trading quantity (turnover ratio), the impact on price (price/trading volume) and the trading speed (number of non-trading days). Suppose that liquidity is an unknown function, $\mathfrak{L}$, of these individual measures. To reliably estimate $\mathfrak{L}$, without specifying its functional form, assume that it can be written as a function of an unknown linear index. That is,

$$\mathfrak{L}(v_{it}(\rho_0)) \equiv L_{it}$$

where $\mathfrak{L}$ is an unknown function and $v_{it}(\rho_0)$ is a vector of individual liquidity measures such as the trading cost (bid ask spread), the trading quantity (turnover ratio), the impact on price (price/trading volume) and the trading speed (number of non-trading days). Without a parametric form for $\mathfrak{L}$, Ichimura (1993) shows that the constant term in the index cannot be identified and the remaining slope coefficients are identified up to scale (with $l_{it}^{(1)}$ as a continuous variable that belongs in the model as required for identification under current sufficient conditions in the literature). Intuitively, when $\mathfrak{L}$ is unknown, we can always write

$$L(v_{it}(\alpha_0)) = \mathfrak{L}(v_{it}(\rho_0)) = L_{it}$$

while $v_{it}(\rho_0)$ and $v_{it}(\alpha_0)$ differ by $\rho_{0,1}$ and $\rho_{0,b+1}$, they generate the same value for liquidity as above. Since we cannot differentiate between $\mathfrak{L}$ and $L$, we can only identify ratios of slope parameters $\alpha$. It is worth noting that all marginal effects with respect to liquidity are identified.

By allowing heterogeneity in the channels through which liquidity works, the excess return of security $i$ at time $t$ is modeled as

$$R_{it}^e = \beta_{0i} R_{mt}^e + \theta_{0i} L_{it} + u_{it}. \quad (1.1)$$
where $R_{it}^e$ and $R_{mt}^e$ are the excess return of asset $i$ at time $t$ and market portfolio at time $t$ respectively. The $u_{it}$ are error terms with $E(u_{it} \mid R_{mt}^e, I_{it}) = 0$. Here the parameter vector in the liquidity index, $\alpha$, is assumed to be the same for all the assets. With the functional form of $L$ is the same for all the assets, $L$ varies over assets due to variability in the liquidity components. As is traditional in asset pricing models, the $\beta_i$’s vary over securities. Similarly we allow for the impact of liquidity to vary over securities. We view $\theta_{0i}$ as summarizing the an asset-specific liquidity impact.

Following conventional notation, let any greek letter with a hat on it denote estimated parameters and with subscript 0 denote the truth in the following discussions. Finally assume that the single index assumption holds:

$$E(w_{it}(\beta_{0i}) \mid I_{it}) = E(w_{it}(\beta_{0i}) \mid v_{it}(\alpha_0))$$

where $w_{it}(\beta_{0i}) = R_{it}^e - \beta_{0i}R_{mt}^e$.

### 1.3 Estimator

Then model (1.1) is estimated by semiparametric least square (SLS) (Ichimura (1993)). Ichimura and Lee (1991) extend the SLS estimator of Ichimura (1993) by including a parametric part and a more general multiple indices in the model. They study this SLS estimator with a single equation and (1.1) with $t = 1$ and with $\beta_{0i}$ and $\theta_{0i}$ constant over $i$. Therefore our model differs from theirs in two respects. First, they assume that observations are i.i.d while we take the time dimension into account. Second, here the $\beta$’s and $\theta$’s are not the same over securities.

Wang, Linton, and Hardle (2004) also examines a similar structure of i.i.d data but with missing responses. They suggest a way to perform an inference on the mean of the dependent variable. They consider a regression model with a linear part as in Ichimura and Lee (1991) and an unknown function which depends on a multiple dimensional object. In the time series context, especially in the application on asset
pricing model, Connor and Linton (2007) uses nonparametric kernel method to construct Fama-French portfolios (Fama and French (1993)). Lu and Linton (2007) study a nonparametric regression with near epoch dependent data. In their model, the dependent variable is modeled as an unknown function of explanatory variables possibly including auto-regression. Linton, Nielsen, and Nielsen (2009) study a model where the dependent variable is explained by an unobserved time series and an unknown function of observed covariates.

It may appear that (1.1) is a model with multiple equations and $\beta'$s and $\theta'$s are estimated equation by equations. However, note that $L_{it}$ is a function of an index of individual liquidity measures and the index parameter $\alpha$ is associated with its functional form. We assume $\alpha$ is the same for all securities, that is, all assets weight the individual liquidity measures in aggregate liquidity $L_{it}$ in a similar fashion. Therefore we employ data from all assets to estimate $\alpha$. With this formulation, we will be able to allow the impact of individual liquidity indicators to differ for different securities in that we will recover $\theta'$s up to a scaled factor. To see this, from (1.1) and (1.2):

$$E(w_{it}(\beta_{0i}) | v_{it}(\alpha_0)) = \theta_{0i}L(v_{it}(\alpha_0))$$

where $L(\cdot) \equiv \bar{\theta}_0L(\cdot)$. As in the case of $\alpha$, we cannot tell the difference between $L$ and $\mathcal{L}$ when they are estimated by a semiparametric estimator. Accordingly, below we will show that what we can recover is $\frac{\theta_{0i}}{\theta_0}$, but not $\theta_{0i}$. For the purpose of further use, define this recoverable parameter as $\delta_{0i} \equiv \frac{\theta_{0i}}{\theta_0}$ which is a normalized version of $\theta'_0$s.

Let $\gamma$ denote the full parameter vector of $\beta'_i$s, the $\delta'_i$s ($n$ of them), and $\alpha$ ($\alpha_1, \ldots, \alpha_b$). Then the SLS estimator for the model in (1.1) is given as:

$$\hat{\gamma} = \arg \min \hat{Q}(\gamma)$$

where

$$\hat{Q}(\gamma) = \frac{1}{2(n \times T)} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\tau}_{it} \left[ w_{it}(\beta_i) - \delta_i \cdot \hat{M}(v(l_{it}; \alpha), \gamma) \right]^2,$$
where $\hat{M}(v(l_t; \alpha), \gamma)$ is a semiparametric estimator for a conditional expectation:

$$\hat{M}(v(l_t; \alpha), \gamma) = \frac{\sum_{j=1}^{n} \sum_{s \neq t}^{T} \frac{\sum_{j=1}^{T} \frac{w_{js}(\beta_j)}{n(T-1)h} K((v(l_t; \alpha) - v(l_{js}; \alpha))/h)}{n(T-1)h} K((v(l_t; \alpha) - v(l_{js}; \alpha))/h)}{n(T-1)h} \equiv \hat{f}(v_{it}(\alpha)) \cdot \hat{g}(v_{it}(\alpha)).$$  \hspace{1cm} (1.4)

Here, $K(\cdot)$ is a symmetric bounded function which integrates over the support to one and has finite first, ..., $p^*th$ derivatives for some $p^* > 0$.

### 1.4 Assumptions

To deal with dependence, the data is assumed to have a decaying memory property in addition to strict stationarity, that is, the correlation between the observations at different time points decreases at a certain rate as the lag becomes larger. Thus our asymptotics will be based on growing number of period ($T$) while number of assets ($n$) is fixed. To present the assumptions, first define $g(z)$ as the true density of the liquidity index at the point $z$. Define $M(z; \gamma)$ as the probability limit of $\hat{M}(z; \gamma)$ and $f(z)$ is the probability limit of $\hat{f}(z)$. We will show below that:

$$f(z) = g(z) \cdot M(z; \gamma).$$  \hspace{1cm} (1.5)

$\hat{f}$ and $\hat{g}$ are corresponding nonparametric estimators of $g$ and $f$ respectively as defined in (1.4). Denote $\nabla_{z}^{k}F(z)$ as the $k^{th}$ partial derivative with respect to $z$ for a function $F$, where $\nabla_{z}^{0}F(z) \equiv F(z)$. Then we assume:

1. **Weakly Dependent Data.**

   (a) **Definition of mixing processes.** A sequence of random variable, e.g. $\{z_t\}_{t \in \mathbb{N}}$ is called $\beta-$mixing\(^1\) (absolutely regular) if

$$\beta_m \equiv \sup [\beta((z_1, ..., z_k), (z_{k+m}, z_{k+m+1}, ...)) \mid k \in \mathbb{N}] \xrightarrow{m \to \infty} 0$$

\(^1\)Note that absolute regularity implies strong mixing (Bradley (2005)).
where \( \beta \) is the mixing coefficient defined as

\[
\beta(z, z') = E \left[ \sup_{A \in \sigma(z)} |P(A \mid B) - P(A)| \right]
\]

and \( \sigma(z) \) denotes a sigma algebra generated by random variable (or random vector) \( z \). Additionally, a sequence of random variable is called \( \alpha \)-mixing (strong mixing) if

\[
\alpha_m \equiv \sup \{ \alpha ((z_1, \ldots, z_k), (z_{k+m}, z_{k+m+1}, \ldots)) \mid k \in \mathbb{N} \} \xrightarrow{m \to \infty} 0
\]

where \( \alpha \) is the strong mixing coefficient defined as

\[
\alpha(z, z') = \sup_{A \in \sigma(z)} \sup_{B \in \sigma(z')} |P(A \cap B) - P(A)P(B)|.
\]

(b) Assumptions from Hansen (2008). Let \( x_{it} \equiv \{R_{mt}^e, l_{it}\} \in R \times R^b \) be a vector of explanatory variables of asset \( i \) at time \( t \) with liquidity indicators \( l_{it} = (l_{it}^{(1)}, \ldots, l_{it}^{(b)}) \). Additionally let \( y_{it} \equiv \{R_{it}^e, x_{it}\} \in R \times R^{b+1} \) be a vector of dependent and all the explanatory variables of asset \( i \) at time \( t \). Assume that \( \{y_{1t}, \ldots, y_{nt}\}_{t \in T} \) is strictly stationary and absolutely regular process\(^2\) with mixing coefficients \( \beta_m \leq Bm^{-c} \) for some finite \( B \). Then the process is also a strong mixing process with \( \alpha_m \) :

\[
\alpha_m \leq Am^{-c}
\]

where \( A < \infty \). Assume that Assumption 2-(2), (3) and (4) in Hansen (2008) are satisfied for some \( s^* > 2 \) as follows

\[
E |R_{i1}^e|^{s^*} < \infty \text{ and } E |R_{m1}^e|^{s^*} < \infty
\]

and

\[
c > \frac{2s^* - 2}{s^* - 2}.
\]

\(^2\)Absolute regularity is required for the U-statistic argument in the proof of asymptotic normality. Consistency and most of the normality argument go through with strong mixing condition only. It seems to be possible to further weaken the assumption down to strong mixing and this is still in progress.
2. Moments and Strict Exogeneity\textsuperscript{3} Conditions. Define $w_{it} := w_{it}(\beta_0) = R_{it}^e - \beta_0 R_{mt}^e$.

\begin{itemize}
  \item[(i)] $E \left( |w_{it}|^s \right) < C_0$; $E \left( |u_{it}|^s \right) < C_0$,
  \item[(ii)] $E \left( |w_{it}|^s | \{l_{it}\}_{i=1}^\infty \right) \leq C_0$ and $E \left( |u_{it}|^s | \{l_{it}\}_{i=1}^\infty \right) \leq C_0$,
  \item[(iii)] $x_{it}$ are strictly exogenous ($E(u_{it} | x_{i't'}) = 0$ for all $i', t'$).
\end{itemize}

3. Kernel Function. The function $K(z) : R \rightarrow R$ is a symmetric $p^\ast$th-order kernel which satisfies\textsuperscript{4}

\begin{itemize}
  \item[(i)] $|K(z)| \leq \bar{k} < \infty$,
  \item[(ii)] $\int_R |K(z)|^3 dz \leq \mu_3 < \infty$ for $3 \geq 1$,
  \item[(iii)] $\int_R z^{k-1}K(z)dz = 0$ for $k = 1, \ldots, p^\ast - 1$, and
  \item[(iv)] $\int_R z^KK(z)dz = O(1)$ for $k = p^\ast$,
\end{itemize}

where

$$p^\ast = \max \left\{ 3, \frac{1}{2} \frac{(c + 2)(s^* - 1) - (1 + c)}{(c - 2)(s^* - 1) - (1 + c)} \frac{1}{1 - 2d} \right\}.$$ 

4. Liquidity Function and Index Density. For $l_{it} \leq |B|$ for finite $B$, assume the liquidity $L$ is bounded. Let nonnegative Index density $g(z)$ satisfies

$$\left| \nabla_z^k g(z) \right| = O(1) \text{ for } k = 0, \ldots, 2p^\ast,$$

$$\left| \nabla_z^k f(z) \right| = O(1) \text{ for } k = 1, \ldots, 2p^\ast$$

and Assumption 2-(5), (6) and (7) in Hansen (2008).

\textsuperscript{3}This is in fact a stronger version of strict exogeneity.

\textsuperscript{4}For the simulation and estimation purpose, a twicing kernel of Newey, Hsieh, and Robins (2004)

$$K(z) = 2\phi(z) - \frac{1}{\sqrt{2}} \phi \left( \frac{z}{\sqrt{2}} \right).$$

is used, where $\phi(z)$ is a standard normal density.
5. **Trimming Function.** Define a smoothed trimming following Klein (1993) as:

\[
\tau(z, \pi_0) \equiv [1 + \exp \{- \ln T(\pi_0 - z)\}]^{-1}.
\]  

(1.7)

\(\tau(z, \pi_0)\) is a continuous approximation to an indicator function \(1 \{ z \leq \pi_0 \}\) where \(\pi_0 = (a_0, b_0)\) denotes the upper or lower quantile of \(z\). Let \(\hat{\pi} = (\hat{a}, \hat{b})\) be the estimated quantile and \(\Pi\) denote a parameter space for quantile estimator \(\pi \in \Pi\). The estimated trimming is different from the true one only by \(\pi_0\) being replaced by \(\hat{\pi}\). Then the trimming function for the support is formulated as \(\tau_t \equiv \tau(l_t, b) - \tau(l_t, a)\).

6. **Window Parameter.** The bandwidth \(h = O((n \times T)^{-r})\) is selected by

\[
\frac{1}{2 \cdot p^*} < r < \frac{1}{2 \cdot (p^* - 1)}.
\]

1.5 Conclusion

The first chapter aims to propose a model to reliably estimate an unknown liquidity measure. In particular, liquidity is formed as an unknown function of an index of liquidity factors where the index weights are estimated. Therefore this approach facilitates an estimation tool robust to the structural assumption of a liquidity measure. Additionally, the systematic risk of an asset is estimated by the relationship between individual excess return and market excess return as suggested by the capital asset pricing model (CAPM). Thus the unknown liquidity function will be added as an additional variable as a way of extending CAPM. The framework of this work is motivated by the liquidity asset pricing; however, the proposed model is applicable to any economic problems of a similar structure.

The model is estimated by semiparametric least squares (SLS) developed by Ichimura. Ichimura and Lee (1991) extended the SLS estimator of Ichimura (1993)
by including a parametric part and a more general multiple index in the model. They study this SLS estimator with a single equation, and the model in this paper with one asset is in fact identical to theirs. However the estimation technique here differs from the existing theories largely in two respects: first, the data have not only a cross-sectional dimension but also a time-series dimension. Secondly, this model assumes the partial effects of the unobserved variables are individual specific. In the following section we examine the theoretical features of the estimator under the existence of time dependence and individual-specific coefficients of an unknown function. We accordingly show that the estimator behaves well in finite samples as well as in large samples.
Chapter 2

Properties of the Estimator
2.1 Introduction

The model introduced in the previous chapter can be considered as a variation of partially linear model, which is in fact a generalization of linear regression. Thus one may view this model as a regression with two explanatory variables where the second regressor is replaced by an unknown liquidity measure. We additionally let the coefficients of the unknown liquidity function vary over securities. These parameters summarize the asset specific impacts of liquidity while the unknown function captures a common channel through which the index variables affect asset returns. Then the estimator is obtained by using a semiparametric analogue of least squares.

The structure of this model is straightforward, however, the large sample properties of the estimator are not immediate. The interaction of the asset specific coefficients with the unknown liquidity function requires some work to establish consistency and identification results. Additionally, since the data of this model have a time series elements as well as a cross-sectional elements, time dependence requires that the proof here differ from the i.i.d. case.

In this context, we establish asymptotic properties of the estimator when the number of time periods is large. Prior to delving into large sample properties, we first check the finite sample performance of the estimator by investigating Monte-Carlo simulation results in Section 2. Then section 3 presents the large sample properties arguing that the estimator is consistent and asymptotically normal. The appendix contains the details of the proofs.

2.2 Finite Sample Properties

In this section, the finite sample performance of the estimator is evaluated by a Monte-Carlo simulation. The simulations were designed in 3 different ways represented by Sim 1-3. The distinctions among different simulations lie in the structures of individual liquidity measures and the existence of time dependence. In Sim 1 and 2,
individual liquidity measures are assumed to have no serial correlation while they are constructed to depend on time in Sim 3. For all simulations, true model is given by

\[ R_{it}^e = \beta_i R_{mt}^e + \theta_i L_{it} + u_{it} \]  

(2.1)

for \( i = 1, \ldots, 5 \), \( t = 1, \ldots, 100 \). \( R_{mt}^e \) denotes market excess return which is the same for all assets but varies over time. For simplicity we employ two liquidity indicators which are denoted by \( l_{it}^{(1)} \) and \( l_{it}^{(2)} \). These liquidity indicators are assumed to be independent with each other, however, we do not necessarily posit independence between \( R_{mt}^e \) and \((l_{it}^{(1)}, l_{it}^{(2)})\). To check the robustness to the correlation among explanatory variables, \( R_{mt}^e \) is designed as a linear function of \( l_{it}^{(1)} \times l_{it}^{(2)} \) with standard normal disturbances:

\[ R_{mt}^e = \sum_{i=1}^{5} (l_{it}^{(1)} \times l_{it}^{(2)}) + \epsilon_t \text{ where } \epsilon \sim N(0,1). \]  

(2.2)

Note that this structure, however, prevents linear dependency among explanatory variables.

\( l_{it}^{(1)} \) and \( l_{it}^{(2)} \) are generated to have uniform distributions in positive interval for Sim 1 and chi-square distributions for Sim 2 and 3:

Sim 1 : \( l_{it}^{(1)} = 4 \times \xi_{it}^{(1)} \); \( l_{it}^{(2)} = \xi_{it}^{(2)} \)

where \( \xi_{it}^{(j)} \sim \text{Unif}[0,1] \) for \( j = 1, 2 \).

Sim 2 : \( l_{it}^{(1)} = 2 \times (\xi_{it}^{(1)})^2 \); \( l_{it}^{(2)} = (\xi_{it}^{(2)})^2 \)

where \( \xi_{it}^{(j)} \sim N(0,1) \) for \( j = 1, 2 \)

Thus liquidity indicators are generated to be bounded in Sim 1 while they are unbounded in the rest. Additionally note that the liquidity indicators are nonnegative by construction. This setup is reasonable since individual liquidity indicators such as bid-ask spread, trading volume, outstanding share of individual stock, share price or number of non-trading days take nonnegative values. In Sim 3, the i.i.d. assumption is released and we let the data be time dependent. What distinguishes Sim 3 from Sim 1-2 is that liquidity indicators, market excess returns are now auto-regressive
processes and serially correlated. $l_{it}^{(1)}$ and $l_{it}^{(2)}$ are generated to have two components, one is the chi-square part which has the same structure as in Sim 1-3 and the other is AR(1) disturbances.

\begin{align*}
\text{Sim 3} & : l_{it}^{(1)} = (2 \righttimes (\xi_{it}^{(1)})^2) + \zeta_{it}^{(1)} \\
& : l_{it}^{(2)} = (\xi_{it}^{(2)})^2 + \zeta_{it}^{(2)} \\
& : \text{where } \xi_{it}^{(j)} \sim N(0, 1); \\
& : \zeta_{it}^{(j)} = \rho_i^{(j)} \zeta_{i,t-1}^{(j)} + \epsilon_{it}^{(j)}; \epsilon_{it}^{(j)} \sim N(0, 1) \text{ for } j = 1, 2. 
\end{align*}

(2.3)

Note that AR(1) process with Gaussian disturbance satisfies strong mixing condition. The auto-regressive coefficients $\rho_i$ is randomly chosen from the interval $(0, 0.9]$. As in the independent case, $R_{mt}^e$ and $(l_{it}^{(1)}, l_{it}^{(2)})$ are assumed to be correlated but $R_{mt}^e$ also contains a stationary Gaussian AR(1) disturbance now. Hence $\epsilon_i$ in (2.2) is replaced by an error process designed same as $\zeta_{it}$. However no time dependence involves in $u_{it}$ hence serial correlation in $R_{it}^e$ is determined solely by that in $R_{mt}^e$ and $(l_{it}^{(1)}, l_{it}^{(2)})$.

To check that estimators work well for the nonlinear case, $L_{it}$ is chosen to be a quadratic function of $(l_{it}^{(1)}, l_{it}^{(2)})$ such that $(\frac{1}{2}l_{it}^{(1)} + l_{it}^{(2)})^2$. Then the function is continuous and increasing in $l_{it}^{(1)}$ and $l_{it}^{(2)}$ with true $\alpha = 2$. Parameter $\{\beta_i\}$ is a sequence of numbers from 5 to 20 equally spaced. When either $l_{it}^{(1)}$ or $l_{it}^{(2)}$ represents turnover ratio, its relationship with excess return is different from those of other indicators, more specifically lower turnover ratio tends to require higher excess return while other indicators such as bid-ask spread or number of days without trade have positive relationship with excess returns. Hence if we want to maintain $L_{it}$ as an increasing function of $l_{it}^{(1)}$ and $l_{it}^{(2)}$ while setting either of them to be turnover ratio, the work can be done by putting a negative sign on that variable.

Once the objective function is designed to generate a vector of residuals for each asset, then the estimation is to find the parameter values which minimize the sum of elements of this vector. As mentioned previously, the coefficients of this conditional expectations $\delta$’s are designed to capture the impact of aggregate liquidity measure. In
order to do this, the optimizing process needs a step to compare estimated liquidity over assets. In other words, the aggregate liquidity cannot be obtained by equation-by-equation estimation. When we estimate an aggregate liquidity of a certain asset and use only the data from that specific asset, the model we pursue is in fact identical to Ichimura and Lee (1991)’s.

Table 3.1 exhibits simulation results of Sim 1-3 with higher order kernel and smoothed trimming function. In all simulations, the bandwidth is chosen as $N^{-\frac{1}{4}}$. As discussed in normality argument we need to make the bias converge to zero fast enough, which can be done by employing higher order kernels. The number of assets and time periods are chosen in accordance with the sample size of the application; 10 assets and 150 time periods. The simulation results shown in Table 3.1 are from 100 replications. Third, sixth and ninth column exhibit the mean of parameter estimates from these 100 replications. In all simulations, the average of the estimates are very close to the truth which indicate that the estimator performs well with finite sample. The following columns report the standard deviation of parameter estimates and mean of square root of estimated standard errors of 100 replications respectively. Estimates of standard errors are obtained by evaluating (2.8) and (2.9) at parameter estimates. These estimates are deemed to roughly approximate the standard deviation of parameter estimates. Except the standard errors of single index parameter $\alpha$, the estimates are close to the standard deviation of the estimated parameter values.

2.3 Large Sample Properties

2.3.1 Consistency of the Estimator

A standard consistency argument (see Amemiya (1985)) requires two conditions: first, the objective function is close uniformly in the parameters to a fixed function in probability and second, this fixed function must have a unique minimum at the truth. This result formalizes the intuition that when two functions are appropriately close,
their minimizers or maximizers will be appropriately close. Uniform consistency of our estimator can be studied by following steps:

\[(a) : \sup_{\gamma} \left| \hat{Q}(\gamma) - Q(\gamma) \right| \xrightarrow{p} 0, \text{ and} \]

\[(b) : \sup_{\gamma} |Q(\gamma) - E[Q(\gamma)]| \xrightarrow{p} 0. \]

where $Q(\gamma)$ be the corresponding objective function in which the estimated expectation is replaced by its probability limit. Define

\[ M(z; \gamma) = \frac{1}{n} \sum_{j=1}^{n} E(w_{js}(\beta_j) \mid v_{js}(\alpha) = z). \quad (2.4) \]

With $Q(\gamma)$ defined as above, $(a)$ is an immediate result of the uniform convergence of $\hat{M}(z; \gamma)$ to $M(z; \gamma)$ in \{z, \gamma\}. Once $\hat{M}(z; \gamma)$ satisfies all the necessary conditions then Theorem 2 in Hansen (2008) provides the uniform convergence of $\hat{f}(z)$ and $\hat{g}(z)$ to their expectations. Additionally, Lemma 4 in the appendix proves $E[\hat{f}(z)]$ converges to its true counterpart, $f(z)$. Combining these results, we can show that $\hat{Q}(\gamma)$ uniformly tends to $Q(\gamma)$. For $(b)$, first note that its summands are stationary process with expectation zero. Given the assumptions, the result can be proved following the strategy in Theorem 4.2.1. and 4.2.2. of Amemiya (1985) and by applying an appropriate law of large numbers. Having dependent but stationary observations, we can use a law of large numbers which is attributed to McLeish (1975) (Theorem 2.10). Now $(a)$ and $(b)$ prove that our objective function tends to the expectation of its true counterpart and the convergence does not depend on its arguments. This provides sufficient conditions for an estimator uniquely determined by the minimizer of true objective function in limit. The detailed argument is presented in the proof of following consistency theorem:

**Theorem 1 (Consistency)** With $\hat{Q}(\gamma)$ and $Q(\gamma)$ defined as above, assume the convergence results in $(a)$ and $(b)$ hold (see Lemma 5-6 of the Appendix). Further assume the following identification conditions:

\[(i) : E(R_{mt}^e \mid I_{st}) \neq R_{mt}^e, \]
\( \theta_0 \neq 0 \),

\( L(z) \geq 0 \) for all \( z \) and \( L(z) > 0 \) with positive probability,

Ichimura’s (1993) identification condition.

Then, \( \hat{\gamma} \xrightarrow{p} \gamma_0 \).

2.3.2 Asymptotic Distribution

Expanding the gradient to the objective function:

\[
\sqrt{T}(\hat{\gamma} - \gamma_0) = -\hat{H}(\bar{\gamma})^{-1}\sqrt{T}\hat{G}(\gamma_0)
\]

where \( \bar{\gamma} = (\bar{\alpha}, \bar{\beta}, \bar{\delta})' \) lies between \( \hat{\gamma} \) and \( \gamma_0 \),

\[
\hat{H}(\bar{\gamma}) = \left\{ \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\tau}_{it}\bar{\chi}_{it}(\bar{\gamma})'(\bar{\chi}_{it}(\bar{\gamma}))' \right\} - \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\tau}_{it}\left[ w_{it}(\bar{\beta}_i) - \bar{\delta}_i \cdot \hat{M}(v_{it}(\alpha); \gamma) \right] \nabla_{\gamma}\bar{\chi}_{it}(\bar{\gamma}) \tag{2.5}
\]

and

\[
\sqrt{T}\hat{G}(\gamma_0) = \frac{1}{n\sqrt{T}} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\tau}_{it}\left[ w_{it}(\beta_{0i}) - \delta_{0i} \cdot \hat{M}(v_{it}(\alpha_0); \gamma_0) \right] \hat{\chi}_{it}(\gamma_0) \tag{2.7}
\]

with

\[
\hat{\chi}_{it}(\gamma) = \nabla_{\gamma}w_{it}(\beta_i) - \nabla_{\gamma}\left[ \delta_i \cdot \hat{M}(v_{it}(\alpha); \gamma) \right].
\]

Let \( \chi_{it}(\gamma) \) be the same function in which the estimated conditional expectation of \( w_{it} \) given \( v_{it} \) is replaced by the true expectation. The asymptotic distribution is studied in three steps: (a) show that (2.5) converges to a positive definite matrix, (b) (2.6) converges to zero in probability, and (c) (2.7) converges in distribution to a multivariate normal random vector, then the desired result follows.

To introduce the proof strategy\(^1\), the following discussion considers an analogous OLS case. First recall that the OLS analogue of the function \( H(\gamma) \) is \( \frac{X'X}{N} \) where

\(^1\)The appendix contains detailed argument.
$X$ is the usual matrix of explanatory variables while $\frac{X'X}{N}$ is the OLS gradient where $\varepsilon$ is the vector of error terms. Then $\frac{X'X}{N}$ tends to a fixed positive definite matrix due to the law of large numbers and $\sqrt{N} \frac{X'\varepsilon}{N}$ is asymptotically normally distributed by a central limit theorem. In our problem, the unknown functions and parameter estimates are involved in the analogues of $X$ and $\varepsilon$, thus the problem becomes showing $\frac{\hat{X}'\hat{X}}{N}$ and $\sqrt{N} \frac{\hat{X}'\hat{\varepsilon}}{N}$ respectively converge to a fixed matrix, $A = E(\frac{X'X}{N})$ and a normal random variable $B$, where a variable with the conventional notation, hat, denotes an estimate. In $\hat{H}(\hat{\gamma})$ above (2.5) plays a role of $\frac{\hat{X}'\hat{X}}{N}$ and (2.6) acts as a remainder in $\hat{H}(\hat{\gamma})$ which vanishes as $N$ gets larger. Also note that $\frac{\hat{X}'\hat{X}}{N}$ and remainder term are evaluated at arbitrary parameter values thus the convergence required in this part is uniform. The unknown functions inside $\hat{X}$ are the first and second derivatives of estimated conditional expectation besides the function itself, thus we need uniform convergence of first and second derivative of the density estimator. The lemmas for uniform convergences of estimated functions will let us show that $\frac{\hat{X}'\hat{X}}{N}$ converges to $\frac{X'X}{N}$ which has limiting functions instead of estimated functions in it. Finally $\frac{X'X}{N}$ converges to $A$ by the (uniform) law of large numbers and consistency of estimators.

To show $\sqrt{N} \frac{X'\varepsilon}{N}$ converges to a normal random variable $B$, we need to divide it into parts and study them separately. Rewrite the gradient as

$$\sqrt{N} \frac{X'\varepsilon}{N} + \sqrt{N} \frac{(\hat{X} - X)'\varepsilon}{N} + \sqrt{N} \frac{(\hat{X} - X)'(\hat{\varepsilon} - \varepsilon)}{N},$$

where the first part is asymptotically equivalent to a normal random variable and the second and the third disappear as sample size gets larger. The asymptotic distribution of $\sqrt{N} \frac{X'\varepsilon}{N}$ is examined by first splitting this term into two parts: $\sqrt{N} \frac{X'\varepsilon}{N}$ and $\sqrt{N} \frac{X'(\hat{\varepsilon} - \varepsilon)}{N}$. $\sqrt{N} \frac{X'(\hat{\varepsilon} - \varepsilon)}{N}$ is then approximated by a sample mean form of U-statistics, $\sqrt{N} \hat{U}_T$. With $\sqrt{N} \frac{X'\varepsilon}{N}$ again combined with $\sqrt{N} \hat{U}_T$, this sum is nothing but a sample average of a mixing process to which we can apply an appropriate central limit theorem. The caveat here is that since $\sqrt{N} \frac{X'(\hat{\varepsilon} - \varepsilon)}{N}$ is a part of the gradient, its expectation must be zero exactly or in the limit at least. In this case, bias generates slower convergence rate preventing the expectation from descending to zero. In order
to make the argument work, we will use a bias reducing kernel which sends the bias
to zero faster so that the expectation of $\sqrt{N}X^t(\hat{\varepsilon} - \varepsilon)$ converges to zero faster. Once
the statistic is properly centered, theorem for the asymptotic distribution of a strong
mixing process is readily available which completes the proof. Next, the second term
requires some work since $\sqrt{N}$ depresses the speed of convergence in $\frac{(\hat{X} - X)'\varepsilon}{N}$ and, as
a matter of fact, $\hat{X} \xrightarrow{D} X$ alone does not immediately give a desirable rate. Notice
that $(\hat{X} - X)'\varepsilon$ is a product of two terms across time (and across assets as well) and
time dependence is the factor which slows down the convergence rate comparing with
i.i.d. case. We cannot remove the time dependence entirely and we will require the
decaying memory property of data. If $(\hat{X} - X)$ and $\varepsilon$ happened at distinct points
of time (the lag between them is larger than a certain threshold), we may safely
conclude that there is little correlation between the two. In other words, decaying
memory downweights their product, thus gives a faster convergence rate of $(\hat{X} - X)'\varepsilon$
than what we are able to get when only the behavior of $\hat{X}$ is considered. Then its
convergence rate overpowers $\sqrt{N}$ and gives the desired result for the second part.
Finally it is relatively straightforward to argue that the third term converges to zero
by asymptotic behavior of $\hat{X}$ and $\hat{\varepsilon}$. $\hat{X}$ and $\hat{\varepsilon}$ converge to their true counterparts at
a slower rate than $\sqrt{N}$ thus cannot make the argument work individually. However
the convergence rate of their product subdues the speed of expansion of the entire
term which completes the proof. Now the asymptotic normality is presented in the
theorem below. Note that with $n$, the number of asset, being fixed, the asymptotic
depends only on the growing time dimension.

**Theorem 2 (Normality)** If Assumptions 1-6, (10) - (13) in Hansen (2008), identi-
fication conditions hold then all the conditions of Theorem 1.8 in Dehling and Wendler
(2010) are satisfied, consequently as $T \to \infty$

$$\sqrt{T}(\hat{\gamma} - \gamma_0) \xrightarrow{D} N \left(0, H_0^{-1}\Sigma_0 H_0^{-1}\right)$$

where $H_0 \equiv \frac{1}{n} \sum_{i=1}^{n} E \left[\tau_i\chi_{i1}(\gamma_0)(\chi_{i1}(\gamma_0))'\right]$,
\[ \Sigma_T(\gamma_0) \equiv \frac{1}{T} \sum_{t=1}^{T} \sum_{\ell=1}^{T} E [h_t(\gamma_0)h_{\ell}(\gamma_0)'], \quad \text{and} \quad \Sigma_0 \equiv \lim_{T \to \infty} \Sigma_T(\gamma_0) \]

with

\[ h_t(\gamma_0) \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ \tau_{it}\hat{\chi}_{it} - \frac{1}{n(T-1)} \sum_{j=1}^{n} \sum_{s \neq t} \delta_{0j} \cdot E [\tau_{js}\hat{\chi}_{js} \mid v_{js} = v_{it}] \right] \cdot u_{it}. \]

From Lemma 8 in the appendix, the matrix \( H_0 \) can be consistently estimated by

\[ \hat{H} \equiv \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\tau}_{it}\hat{\chi}_{it}(\hat{\gamma})\hat{\chi}_{it}(\hat{\gamma})'. \quad (2.8) \]

We can construct the sample analogue of \( \Sigma_0 \) by replacing the parameters and unknown functions with their estimators :

\[ \hat{\Sigma}_T(\gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{h}_t(\gamma)\hat{h}_t(\gamma)' + \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{m\Sigma(T)} \hat{h}_t(\gamma)\hat{h}_{t-m}(\gamma)' \quad (2.9) \]

where

\[ \hat{h}_t(\gamma) \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\tau}_{it}\hat{\chi}_{it} - \frac{1}{n(T-1)} \sum_{j=1}^{n} \sum_{s \neq t} \delta_{j} \hat{E} [\hat{\tau}_{js}\hat{\chi}_{js} \mid \hat{v}_{js} = v_{it}] \right] \cdot \hat{u}_{it}. \]

Here \( \hat{E} \) denotes a semiparametric conditional expectation estimator and \( m\Sigma(T) \) is the number of sample auto-covariance. Given \( \hat{\Sigma}_T(\hat{\gamma}) \) converges to \( \Sigma_0 \), \( \hat{H}^{-1}\hat{\Sigma}_T(\hat{\gamma})\hat{H}^{-1} \) consistently estimates the covariance-variance matrix.

**Theorem 3** Suppose Assumptions 1-6 and (10) - (13) in Hansen (2008) hold, then \( \hat{\Sigma}_T(\hat{\gamma}) \) is a consistent estimate of \( \Sigma_0 \).

### 2.4 Conclusion

Consistency and identification can be studied first by specifying the uniform probability limit for the estimator of the objective function. Then if we can show that the limit has a unique minimum at the truth, then the desirable results follow. The
convergence of the sample analogue of the objective function to its true counterpart is immediate once we have the result that the estimator of the unknown function converges to the truth. Also by deducing inherent attributes of the stock data and postulating them as additional assumptions, we are able to show that the potential minimizers of the true objective function must coincide with the true parameter values at the minimum.

Additionally, we can prove the parameter estimators have asymptotic normal distribution by showing that the sample analogue of the Hessian uniformly converges to a semi-positive definite matrix and the estimator of the gradient evaluated at the true parameter values converges in distribution to a normal random variable at root $T$ rate. The convergence of the Hessian follows that of the first and second derivative of semiparametric conditional expectation estimator. On the other hand, the correlation between the variables at different time periods complicates the argument for the gradient part. However, we can show that the time dependence decays at a rate fast enough to eliminate the bias from the limiting random variable. A similar argument is employed to construct an estimator for covariance matrix.

In addition to the consistency and asymptotic normality, the Monte-Carlo simulation results show us that the estimator performs well in the finite sample. Such theoretical features entail several extensions of existing empirical studies as well, which will be discussed in the following chapter.
Chapter 3
Application
3.1 Introduction

Given that the estimator is asymptotically well-behaving, the proposed model is applied to stock data in this chapter. By analyzing the application results, we aim to find out: first, which liquidity indicators play the most important role, second, how these liquidity proxies affect asset returns and finally, in which industries liquidity has greater impacts on asset returns.

Based on the proposed structure, monthly excess returns from 1993 to 2006 are modeled as a linear function of the monthly market excess returns and an unknown liquidity measure. The liquidity indicators comprise bid-ask spreads, turnover rates, number of trades, and number of non-trading days. The assets are grouped by industry; therefore, the systematic risks and liquidity impacts differ by industries and not by individual securities. The industry classification is based on Standard Industrial Classification (SIC). All data are from the CRSP/WRDS database. Individual liquidity indicators are originally collected on a daily frequency and averaged over each month. The S&P 500 is employed as a market portfolio and 3-month Treasury yields as a risk-free rate. Note that for application purposes, assets are grouped by industry; therefore, systematic risks and liquidity impacts vary over industries.

The application results suggest that bid-ask spread, turnover rate and number of non-trading days have similar weight in estimating liquidity. This fact suggests that missing any of these three variables may significantly change the estimated value of liquidity, thus distorting the relationship between excess returns and liquidity. Moreover, bid-ask spread and number of non-trading days affect liquidity differently than the turnover rate and the number of transactions. These results are consistent with the common belief that bid-ask spread and number of non-trading days are related with the illiquidity cost while the other two proxies capture the flexibility in trading. We can find further evidence for such conjecture in the estimates of marginal effects. Marginal effects of bid-ask spread and number of non-trading days indicate
that they are positively related with excess returns of assets. On the other hand, a larger turnover rate and number of transactions depress required returns.

The parameter estimates additionally show that liquidity impacts are greater for assets in the agriculture and public administration industries. Industry-specific liquidity impacts signal how investors react to an issuer’s tendency to smooth out the liquidity level in stocks. As discussed in Amihud and Mendelson (1988), companies can enhance the liquidity level in their stocks by using "liquidity-enhancing projects" such as disclosing inside information or splitting stocks. Improved liquidity leads to benefits such as a lower cost of capital and higher value of firms.

### 3.2 Estimation Results

Table 3.2 shows the results from the application of the proposed model. The index parameters in the upper part of the table show the size and direction of the liquidity indicators’ impacts on excess returns which are general to all securities. The systematic risks capture the comovement of the securities’ excess returns with a well-diversified market portfolio. Finally, the liquidity impacts summarize the asset-specific effects of liquidity on returns.

Note that the individual liquidity indicators are standardized first for the efficiency of the estimation. Then the estimated coefficients are multiplied back to the ratios of sample standard deviations so that the magnitudes of these rescaled coefficients represent the relative importance of different indicators in measuring liquidity. Based on the estimates in Table 3.2 first, bid-ask spread and non-trading days work toward a different direction than turnover rate and number of transactions do. This result is consistent with our intuition, since larger bid-ask spread and non-trading days are related with greater frictions in trades while larger turnover rate and number of transactions signal an ease in trading. The magnitudes of index coefficients demonstrate that the impacts of a one unit change in bid-ask spread, turnover rate
and non-trading days are similar. On an average $1 increase in bid-ask spread, a decrease in trading volume by outstanding shares and one additional day without trade have similar impact in increasing the level of illiquidity (or decreasing the level of liquidity). However, one additional transaction has considerably smaller impact compared to the other variables, presumably due to the large quantity of average number of transactions. This result suggests that any single indicator may fail to incorporate all the dimensions of liquidity. Missing any of those three indicators may significantly change the value of liquidity, thus distorting the relationship between excess returns and liquidity. However, we cannot directly interpret these coefficients as the ceteris paribus impacts of liquidity proxies on excess returns since the index variables are associated with a nonlinear function. The marginal effects are estimated separately for this purpose and reported in the following tables. The results demonstrated in Table 3.6 indicate that larger bid-ask spreads and non-trading days tend to demand higher average excess returns. On the other hand, larger number of transactions and turnover rates moderate the required returns on assets. Consequently, not only do these two groups of variables work toward different directions in affecting excess returns, but the directions are consistent with our projection as well.

Parameter estimates in Table 3.2 also reveal that assets in agriculture, forestry and fishing and public administration categories have considerably higher liquidity impacts on average excess returns compared to other industries. In order to see why liquidity in a certain industry has higher impact, we need to think about the role that firms play in determining liquidity levels in stocks. Even if the relationship between asset price and liquidity level is driven by investors’ optimizing behavior, the different impacts of liquidity on various industries may stem from managerial behaviors or firm-specific reactions to liquidity variations in stock trading. That is, maintaining the liquidity level in stocks and alleviating illiquidity costs are managerial concerns. As Amihud and Mendelson (1988) states, firms gain benefits by performing "liquidity-enhancing projects"; they can not only increase the value of their firms but can also
reduce their cost of capital. They argue that those firms which have already enjoyed high liquidity levels can gain larger benefit from maintaining liquidity in stocks and thus tend to more actively use liquidity-enhancing projects. This is because more liquid assets are held by investors with shorter investment horizons and illiquidity costs are depreciated over shorter time periods. Therefore, a reduction in illiquidity costs has a greater impact for such companies. They can then mitigate illiquidity in stocks at a lower cost than other companies. They also argue that the effect of liquidity-enhancing projects is amplified for larger firms as opposed to smaller firms. Hence, larger companies have greater incentive to balance out liquidity levels in stocks. In this regard, we can conjecture that those firms which are less inclined to reduce the cost of capital or face a lower default risk tend to invest less in alleviating illiquidity costs. It is also possible that an investor’s reaction differs by the firms’ propensity to maintain liquidity levels of stocks. As investors gain knowledge about the issuers, they will also be more informed about whether a certain stock is more vulnerable to a liquidity shock or not. We can find the justification for the higher liquidity impacts in agriculture industries in the fact that the average size of firms is relatively small and the competitive nature of the industry encourages companies to care less about the firms’ values. More importantly, the existence of government subsidies may depress the firms’ incentive to lower capital costs even further. Meanwhile, entities in the public administration industry are owned or sponsored by the government, and thus tend to be less concerned about gains from improvements in operational performances. Companies in these two industries therefore have a lower incentive to smooth out liquidity levels in stocks. Thus, different liquidity impacts for different industries may suggest that investors consider the firms’ interests in mitigating illiquidity costs. That is, liquidity has the second order impact on investors’ decision making. Table 3.4 exhibits the standard errors of estimated aggregate liquidity. We can find that the variability in liquidity is much greater in agriculture and public administration industries. This result provides evidence that a positive relationship exists between
volatility in liquidity and asset returns.

Table 3.2 also reports the standard errors of estimates. However, it should be noted that they require special review due to the existence of serial correlation. In panel data models, serial correlation in the error terms causes no serious problem since the time horizon is fixed. The asymptotic in this study, however, depends on the growing number of time periods; thus, the serial correlation magnifies the standard errors and damages the efficiency of the estimator. In order to obtain a serial correlation robust variance-covariance matrix, we may apply a method studied in Newey and West (1987). The method argues that even after specifying the maximum number of sample auto-covariances, the middle matrix in the variance-covariance matrix needs not be positive semi-definite if any auto-correlation exists. In order to make the matrix positive semi-definite, sample auto-covariances are weighted in a way that those terms not many periods apart play a greater role in determining auto-covariance. Newey and West (1987) shows that the estimator obtained in this way consistently estimates a true covariance-variance matrix. We may directly apply this result to our covariance estimator when $i = j$. Note, however, that there are additional covariances induced by the cross-correlations between different industries over time. It is tempting to practice a similar strategy for cross-covariances, yet once we depart from a single cross-sectional dimension, there seems to be no guarantee that the estimated covariance-variance matrix is positive semi-definite. Unless a large value is chosen for the maximum number of auto-covariances, even the diagonal terms often fail to generate positive values. Ignoring cross-covariances underestimates standard errors, while including all cross-covariances may preclude us from obtaining the matrix at all. However, the asymptotic of the estimator in this paper depends on the growing number of time periods, while the cross-sectional dimension is fixed. This allows us to apply Newey and West (1987)'s results to the cross-covariances between industries. Theoretical bases for this argument are provided in theorem 3. Meanwhile, the standard errors in Table 3.2 are computed following the estimator constructed based on
Empirical work in this area has focused largely on three aspects: the level effect of liquidity on asset returns, commonality in liquidity or impact of market liquidity level, and the impact of liquidity risk (volatility in liquidity) on asset returns. The proposed model summarizes these three impacts in one model and provides a convenient way to estimate them all together. Lastly, even though the model is applied to financial data based on the asset pricing equation, the application can be made to other economic problems that give rise to many proxies for some unobserved variable of interest. The method employed here can be viewed as combining these in a relatively flexible manner.

While the model offers a more flexible way to estimate a liquidity measure, it also postulates several parametric assumptions. In this paper, the liquidity is assumed to have the same functional form for all assets and the intrinsic difference in its impacts on asset returns are captured by asset-specific coefficients. One may argue that this is a strong assumption and require further flexibility in the model. Such concerns in fact suggest potential extensions of this approach in several phases. First, by relaxing relevant assumptions, we can let the functional form of liquidity and single index parameter vary over assets or industries. This is the most general form of the model regarding the structure of liquidity. This application allows us to make further conjectures based on its estimation results. One possibility is that since this added flexibility requires a larger sample size, the parameter estimates of liquidity indicators might be no longer statistically significant. Then we can inquire into another variation of the model where the liquidity function varies over assets while the index parameters are held the same. This approach clearly expands the sample size used in estimation of index parameters, however, it requires additional theoretical work. Another possible
outcome is that the estimates are statistically significant but sizes or signs of the impacts of liquidity variables are not consistent with the existing empirical evidence. This suggests that the potential inconsistency of the estimator may be caused by an incorrect assumption and the single index structure should be tested. There are several existing papers which study the tests for single index assumption, and in the future I plan to apply their approaches to the model. Lastly, in the proposed model the liquidity function and the asset-specific impact parameters are assumed to be constant over time. Since liquidity is a constraint on investors’ ability to trade, the scope and the form of its impact on asset returns may evolve as the environment of trade changes. Thus it is possible that both general and asset specific liquidity impacts alter after a shock to the economy. The structural break in liquidity can also be tested and estimated, which is another direction for future research.
Table 3.1: Monte-Carlo Simulation Results

(100 Replications)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Truth</th>
<th>Simulation 1</th>
<th>Simulation 2</th>
<th>Simulation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean of</td>
<td>Standard</td>
<td>Mean of</td>
<td>Standard</td>
</tr>
<tr>
<td></td>
<td>Estimates</td>
<td>Deviation of</td>
<td>Estimates</td>
<td>Deviation of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Estimates</td>
<td></td>
<td>Estimates</td>
</tr>
</tbody>
</table>

**Liquidity Index Parameter**

<table>
<thead>
<tr>
<th>alpha</th>
<th>2.0000</th>
<th>2.0153</th>
<th>0.0205</th>
<th>2.0029</th>
<th>0.0763</th>
<th>2.0461</th>
<th>0.2347</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.0000</td>
<td>2.0153</td>
<td>0.0205</td>
<td>2.0029</td>
<td>0.0763</td>
<td>2.0461</td>
<td>0.2347</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Systematic Risks**

<table>
<thead>
<tr>
<th>beta_1</th>
<th>5.0000</th>
<th>5.0277</th>
<th>0.0385</th>
<th>4.9228</th>
<th>0.0653</th>
<th>4.9484</th>
<th>0.0874</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta_2</td>
<td>6.6667</td>
<td>6.6971</td>
<td>0.0318</td>
<td>6.6225</td>
<td>0.0645</td>
<td>6.6389</td>
<td>0.0957</td>
</tr>
<tr>
<td>beta_3</td>
<td>8.3333</td>
<td>8.3547</td>
<td>0.0357</td>
<td>8.2802</td>
<td>0.0916</td>
<td>8.3234</td>
<td>0.1087</td>
</tr>
<tr>
<td>beta_4</td>
<td>10.0000</td>
<td>10.0241</td>
<td>0.0367</td>
<td>9.9690</td>
<td>0.0829</td>
<td>9.9833</td>
<td>0.1055</td>
</tr>
<tr>
<td>beta_5</td>
<td>11.6667</td>
<td>11.6842</td>
<td>0.0345</td>
<td>11.6341</td>
<td>0.0993</td>
<td>11.6698</td>
<td>0.1054</td>
</tr>
<tr>
<td>beta_6</td>
<td>13.3333</td>
<td>13.3558</td>
<td>0.0371</td>
<td>13.3363</td>
<td>0.0906</td>
<td>13.3541</td>
<td>0.1504</td>
</tr>
<tr>
<td>beta_7</td>
<td>15.0000</td>
<td>15.0160</td>
<td>0.0362</td>
<td>15.0264</td>
<td>0.0910</td>
<td>15.0253</td>
<td>0.1849</td>
</tr>
<tr>
<td>beta_8</td>
<td>16.6667</td>
<td>16.6867</td>
<td>0.0386</td>
<td>16.7032</td>
<td>0.1000</td>
<td>16.6949</td>
<td>0.1452</td>
</tr>
<tr>
<td>beta_9</td>
<td>18.3333</td>
<td>18.3498</td>
<td>0.0459</td>
<td>18.3963</td>
<td>0.1226</td>
<td>18.3878</td>
<td>0.1985</td>
</tr>
<tr>
<td>beta_10</td>
<td>20.0000</td>
<td>20.0126</td>
<td>0.0469</td>
<td>20.1049</td>
<td>0.1676</td>
<td>20.0920</td>
<td>0.2073</td>
</tr>
</tbody>
</table>

**Liquidity Impacts**

<table>
<thead>
<tr>
<th>delta_1</th>
<th>0.4000</th>
<th>0.3996</th>
<th>0.0116</th>
<th>0.4174</th>
<th>0.0267</th>
<th>0.4250</th>
<th>0.0392</th>
</tr>
</thead>
<tbody>
<tr>
<td>delta_2</td>
<td>0.5333</td>
<td>0.5352</td>
<td>0.0116</td>
<td>0.5415</td>
<td>0.0266</td>
<td>0.5592</td>
<td>0.0454</td>
</tr>
<tr>
<td>delta_3</td>
<td>0.6667</td>
<td>0.6728</td>
<td>0.0125</td>
<td>0.6801</td>
<td>0.0383</td>
<td>0.6817</td>
<td>0.0595</td>
</tr>
<tr>
<td>delta_4</td>
<td>0.8000</td>
<td>0.8084</td>
<td>0.0153</td>
<td>0.8061</td>
<td>0.0351</td>
<td>0.8228</td>
<td>0.0631</td>
</tr>
<tr>
<td>delta_5</td>
<td>0.9333</td>
<td>0.9460</td>
<td>0.0153</td>
<td>0.9434</td>
<td>0.0461</td>
<td>0.9513</td>
<td>0.0606</td>
</tr>
<tr>
<td>delta_6</td>
<td>1.0667</td>
<td>1.0794</td>
<td>0.0182</td>
<td>1.0583</td>
<td>0.0441</td>
<td>1.0749</td>
<td>0.0878</td>
</tr>
<tr>
<td>delta_7</td>
<td>1.2000</td>
<td>1.2172</td>
<td>0.0166</td>
<td>1.1819</td>
<td>0.0502</td>
<td>1.2015</td>
<td>0.1037</td>
</tr>
<tr>
<td>delta_8</td>
<td>1.3333</td>
<td>1.3501</td>
<td>0.0191</td>
<td>1.3078</td>
<td>0.0583</td>
<td>1.3418</td>
<td>0.0910</td>
</tr>
<tr>
<td>delta_9</td>
<td>1.4667</td>
<td>1.4871</td>
<td>0.0243</td>
<td>1.4300</td>
<td>0.0685</td>
<td>1.4612</td>
<td>0.1350</td>
</tr>
<tr>
<td>delta_10</td>
<td>1.6000</td>
<td>1.6230</td>
<td>0.0220</td>
<td>1.5486</td>
<td>0.0912</td>
<td>1.5701</td>
<td>0.1644</td>
</tr>
</tbody>
</table>
Table 3.2: Estimation Results

<table>
<thead>
<tr>
<th>Parameter (Industry)</th>
<th>Estimates</th>
<th>Std Err</th>
<th>Parameter (Industry)</th>
<th>Estimates</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Liquidity Index Parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bid ask spread</td>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnover rate</td>
<td>-0.9682***</td>
<td>0.2613</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of transactions</td>
<td>-0.0021***</td>
<td>0.0007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-trading days</td>
<td>0.9562***</td>
<td>0.0373</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Systematic Risks</strong></td>
<td></td>
<td></td>
<td><strong>Liquidity Impacts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.8651***</td>
<td>0.1813</td>
<td>Agriculture</td>
<td>1.7821</td>
<td>1.4192</td>
</tr>
<tr>
<td>Mining</td>
<td>0.9026***</td>
<td>0.0987</td>
<td>Mining</td>
<td>0.9328</td>
<td>0.6319</td>
</tr>
<tr>
<td>Construction</td>
<td>0.8355***</td>
<td>0.1114</td>
<td>Construction</td>
<td>1.0067*</td>
<td>0.5516</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.8512***</td>
<td>0.1191</td>
<td>Manufacturing</td>
<td>1.0039*</td>
<td>0.5599</td>
</tr>
<tr>
<td>Transp/Com</td>
<td>0.8389***</td>
<td>0.0938</td>
<td>Transp/Com</td>
<td>0.8411</td>
<td>0.5497</td>
</tr>
<tr>
<td>Wholesale Trade</td>
<td>0.8615***</td>
<td>0.1029</td>
<td>Wholesale Trade</td>
<td>1.2176**</td>
<td>0.5918</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>0.8797***</td>
<td>0.1082</td>
<td>Retail Trade</td>
<td>0.8835*</td>
<td>0.5252</td>
</tr>
<tr>
<td>Finance, Insur/RE</td>
<td>0.8041***</td>
<td>0.1264</td>
<td>Finance, Insur/RE</td>
<td>1.2267</td>
<td>0.7611</td>
</tr>
<tr>
<td>Services</td>
<td>0.8718***</td>
<td>0.1141</td>
<td>Services</td>
<td>0.7116</td>
<td>0.6917</td>
</tr>
<tr>
<td>Public Admin</td>
<td>0.9067***</td>
<td>0.2239</td>
<td>Public Admin</td>
<td>2.3276</td>
<td>1.7538</td>
</tr>
<tr>
<td>No</td>
<td>Industry</td>
<td>Average Number of Companies</td>
<td>Monthly Average Gross Return</td>
<td>Average No of Transactions per Day</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
<td>------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Agriculture, Forestry and Fishing</td>
<td>19</td>
<td>0.01142</td>
<td>212.97</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Mining</td>
<td>202</td>
<td>0.01144</td>
<td>445.92</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Construction</td>
<td>326</td>
<td>0.01300</td>
<td>528.73</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Manufacturing</td>
<td>496</td>
<td>0.01253</td>
<td>543.59</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Transportation, Communications, Electric, Gas and Sanitary Services</td>
<td>218</td>
<td>0.01141</td>
<td>578.16</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Wholesale Trade</td>
<td>190</td>
<td>0.01325</td>
<td>495.32</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Retail Trade</td>
<td>349</td>
<td>0.01270</td>
<td>442.67</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Finance, Insurance and Real Estate</td>
<td>184</td>
<td>0.01476</td>
<td>436.71</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Services</td>
<td>136</td>
<td>0.01286</td>
<td>401.74</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Public Administration</td>
<td>4</td>
<td>0.01102</td>
<td>276.45</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Standard Deviation of Estimated Liquidity

<table>
<thead>
<tr>
<th>No</th>
<th>Industry</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agriculture, Forestry and Fishing</td>
<td>0.01055</td>
</tr>
<tr>
<td>2</td>
<td>Mining</td>
<td>0.00625</td>
</tr>
<tr>
<td>3</td>
<td>Construction</td>
<td>0.00594</td>
</tr>
<tr>
<td>4</td>
<td>Manufacturing</td>
<td>0.00588</td>
</tr>
<tr>
<td>5</td>
<td>Transportation, Communications, Electric, Gas and Sanitary Services</td>
<td>0.00763</td>
</tr>
<tr>
<td>6</td>
<td>Wholesale Trade</td>
<td>0.00669</td>
</tr>
<tr>
<td>7</td>
<td>Retail Trade</td>
<td>0.00586</td>
</tr>
<tr>
<td>8</td>
<td>Finance, Insurance and Real Estate</td>
<td>0.00596</td>
</tr>
<tr>
<td>9</td>
<td>Services</td>
<td>0.00583</td>
</tr>
<tr>
<td>10</td>
<td>Public Administration</td>
<td>0.01570</td>
</tr>
</tbody>
</table>
Table 3.5: Sample Cross Correlations of Error Terms

<table>
<thead>
<tr>
<th></th>
<th>Ind 1</th>
<th>Ind 2</th>
<th>Ind 3</th>
<th>Ind 4</th>
<th>Ind 5</th>
<th>Ind 6</th>
<th>Ind 7</th>
<th>Ind 8</th>
<th>Ind 9</th>
<th>Ind 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contemporaneous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind 1</td>
<td>1</td>
<td>0.58</td>
<td>0.606</td>
<td>0.588</td>
<td>0.53</td>
<td>0.603</td>
<td>0.592</td>
<td>0.628</td>
<td>0.51</td>
<td>0.245</td>
</tr>
<tr>
<td>Ind 2</td>
<td>0.58</td>
<td>1</td>
<td>0.897</td>
<td>0.902</td>
<td>0.851</td>
<td>0.835</td>
<td>0.88</td>
<td>0.85</td>
<td>0.804</td>
<td>0.266</td>
</tr>
<tr>
<td>Ind 3</td>
<td>0.606</td>
<td>0.897</td>
<td>1</td>
<td>0.948</td>
<td>0.892</td>
<td>0.881</td>
<td>0.916</td>
<td>0.911</td>
<td>0.851</td>
<td>0.276</td>
</tr>
<tr>
<td>Ind 4</td>
<td>0.588</td>
<td>0.902</td>
<td>0.948</td>
<td>1</td>
<td>0.904</td>
<td>0.906</td>
<td>0.93</td>
<td>0.909</td>
<td>0.859</td>
<td>0.281</td>
</tr>
<tr>
<td>Ind 5</td>
<td>0.53</td>
<td>0.851</td>
<td>0.892</td>
<td>0.904</td>
<td>1</td>
<td>0.83</td>
<td>0.856</td>
<td>0.879</td>
<td>0.771</td>
<td>0.282</td>
</tr>
<tr>
<td>Ind 6</td>
<td>0.603</td>
<td>0.835</td>
<td>0.881</td>
<td>0.906</td>
<td>0.83</td>
<td>1</td>
<td>0.88</td>
<td>0.855</td>
<td>0.831</td>
<td>0.293</td>
</tr>
<tr>
<td>Ind 7</td>
<td>0.592</td>
<td>0.88</td>
<td>0.916</td>
<td>0.93</td>
<td>0.856</td>
<td>0.88</td>
<td>1</td>
<td>0.874</td>
<td>0.853</td>
<td>0.3</td>
</tr>
<tr>
<td>Ind 8</td>
<td>0.628</td>
<td>0.85</td>
<td>0.911</td>
<td>0.909</td>
<td>0.879</td>
<td>0.855</td>
<td>0.874</td>
<td>1</td>
<td>0.833</td>
<td>0.301</td>
</tr>
<tr>
<td>Ind 9</td>
<td>0.51</td>
<td>0.804</td>
<td>0.851</td>
<td>0.859</td>
<td>0.771</td>
<td>0.831</td>
<td>0.853</td>
<td>0.833</td>
<td>1</td>
<td>0.264</td>
</tr>
<tr>
<td>Ind 10</td>
<td>0.245</td>
<td>0.266</td>
<td>0.276</td>
<td>0.281</td>
<td>0.282</td>
<td>0.293</td>
<td>0.3</td>
<td>0.301</td>
<td>0.264</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ind 1</th>
<th>Ind 2</th>
<th>Ind 3</th>
<th>Ind 4</th>
<th>Ind 5</th>
<th>Ind 6</th>
<th>Ind 7</th>
<th>Ind 8</th>
<th>Ind 9</th>
<th>Ind 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Period Apart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind 1</td>
<td>0.023</td>
<td>0.177</td>
<td>0.164</td>
<td>0.191</td>
<td>0.146</td>
<td>0.19</td>
<td>0.189</td>
<td>0.13</td>
<td>0.186</td>
<td>0.134</td>
</tr>
<tr>
<td>Ind 2</td>
<td>0.067</td>
<td>0.173</td>
<td>0.163</td>
<td>0.172</td>
<td>0.179</td>
<td>0.173</td>
<td>0.146</td>
<td>0.15</td>
<td>0.152</td>
<td>0.078</td>
</tr>
<tr>
<td>Ind 3</td>
<td>0.076</td>
<td>0.206</td>
<td>0.19</td>
<td>0.195</td>
<td>0.216</td>
<td>0.188</td>
<td>0.192</td>
<td>0.175</td>
<td>0.165</td>
<td>0.041</td>
</tr>
<tr>
<td>Ind 4</td>
<td>0.068</td>
<td>0.226</td>
<td>0.202</td>
<td>0.214</td>
<td>0.231</td>
<td>0.217</td>
<td>0.19</td>
<td>0.181</td>
<td>0.167</td>
<td>0.026</td>
</tr>
<tr>
<td>Ind 5</td>
<td>0.063</td>
<td>0.209</td>
<td>0.187</td>
<td>0.178</td>
<td>0.193</td>
<td>0.172</td>
<td>0.153</td>
<td>0.172</td>
<td>0.141</td>
<td>0.021</td>
</tr>
<tr>
<td>Ind 6</td>
<td>0.065</td>
<td>0.204</td>
<td>0.179</td>
<td>0.194</td>
<td>0.19</td>
<td>0.192</td>
<td>0.163</td>
<td>0.159</td>
<td>0.122</td>
<td>0.035</td>
</tr>
<tr>
<td>Ind 7</td>
<td>0.115</td>
<td>0.228</td>
<td>0.22</td>
<td>0.225</td>
<td>0.232</td>
<td>0.22</td>
<td>0.207</td>
<td>0.19</td>
<td>0.19</td>
<td>0.047</td>
</tr>
<tr>
<td>Ind 8</td>
<td>0.074</td>
<td>0.237</td>
<td>0.226</td>
<td>0.24</td>
<td>0.242</td>
<td>0.232</td>
<td>0.221</td>
<td>0.21</td>
<td>0.188</td>
<td>0.072</td>
</tr>
<tr>
<td>Ind 9</td>
<td>0.058</td>
<td>0.149</td>
<td>0.132</td>
<td>0.128</td>
<td>0.141</td>
<td>0.158</td>
<td>0.136</td>
<td>0.127</td>
<td>0.135</td>
<td>0.024</td>
</tr>
<tr>
<td>Ind 10</td>
<td>-0.008</td>
<td>0.001</td>
<td>0.029</td>
<td>0.01</td>
<td>0.028</td>
<td>-0.012</td>
<td>0.014</td>
<td>0.029</td>
<td>-0.039</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ind 1</th>
<th>Ind 2</th>
<th>Ind 3</th>
<th>Ind 4</th>
<th>Ind 5</th>
<th>Ind 6</th>
<th>Ind 7</th>
<th>Ind 8</th>
<th>Ind 9</th>
<th>Ind 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Period Apart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind 1</td>
<td>-0.042</td>
<td>-0.07</td>
<td>-0.049</td>
<td>-0.01</td>
<td>-0.019</td>
<td>0.045</td>
<td>-0.005</td>
<td>-0.042</td>
<td>-0.029</td>
<td>-0.056</td>
</tr>
<tr>
<td>Ind 2</td>
<td>0.054</td>
<td>0.028</td>
<td>0.104</td>
<td>0.12</td>
<td>0.148</td>
<td>0.158</td>
<td>0.066</td>
<td>0.129</td>
<td>0.074</td>
<td>-0.029</td>
</tr>
<tr>
<td>Ind 3</td>
<td>0.076</td>
<td>0.082</td>
<td>0.143</td>
<td>0.167</td>
<td>0.17</td>
<td>0.198</td>
<td>0.1</td>
<td>0.162</td>
<td>0.103</td>
<td>-0.017</td>
</tr>
<tr>
<td>Ind 4</td>
<td>0.073</td>
<td>0.06</td>
<td>0.118</td>
<td>0.148</td>
<td>0.152</td>
<td>0.175</td>
<td>0.093</td>
<td>0.156</td>
<td>0.084</td>
<td>-0.032</td>
</tr>
<tr>
<td>Ind 5</td>
<td>0.044</td>
<td>0.049</td>
<td>0.101</td>
<td>0.132</td>
<td>0.15</td>
<td>0.168</td>
<td>0.087</td>
<td>0.127</td>
<td>0.078</td>
<td>-0.036</td>
</tr>
<tr>
<td>Ind 6</td>
<td>0.041</td>
<td>0.062</td>
<td>0.131</td>
<td>0.169</td>
<td>0.159</td>
<td>0.203</td>
<td>0.118</td>
<td>0.195</td>
<td>0.094</td>
<td>-0.006</td>
</tr>
<tr>
<td>Ind 7</td>
<td>0.097</td>
<td>0.099</td>
<td>0.135</td>
<td>0.157</td>
<td>0.167</td>
<td>0.199</td>
<td>0.113</td>
<td>0.18</td>
<td>0.112</td>
<td>0</td>
</tr>
<tr>
<td>Ind 8</td>
<td>0.098</td>
<td>0.113</td>
<td>0.16</td>
<td>0.183</td>
<td>0.207</td>
<td>0.214</td>
<td>0.122</td>
<td>0.18</td>
<td>0.091</td>
<td>-0.026</td>
</tr>
<tr>
<td>Ind 9</td>
<td>0.085</td>
<td>0.054</td>
<td>0.113</td>
<td>0.144</td>
<td>0.13</td>
<td>0.186</td>
<td>0.08</td>
<td>0.17</td>
<td>0.064</td>
<td>-0.001</td>
</tr>
<tr>
<td>Ind 10</td>
<td>0.064</td>
<td>0.169</td>
<td>0.168</td>
<td>0.137</td>
<td>0.163</td>
<td>0.224</td>
<td>0.194</td>
<td>0.193</td>
<td>0.173</td>
<td>0.152</td>
</tr>
</tbody>
</table>
### Table 3.6: Marginal Effects

<table>
<thead>
<tr>
<th>Industry</th>
<th>Liquidity Index</th>
<th>1 Unit Change</th>
<th>2 Unit Change</th>
<th>3 Unit Change</th>
<th>4 Unit Change</th>
<th>5 Unit Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricu</td>
<td>Bid ask spread</td>
<td>0.00215</td>
<td>0.00447</td>
<td>0.00679</td>
<td>0.00880</td>
<td>0.01020</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00384</td>
<td>-0.00636</td>
<td>-0.01039</td>
<td>-0.00507</td>
<td>0.01054</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00003</td>
<td>-0.00004</td>
<td>-0.00005</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00424</td>
<td>0.00846</td>
<td>0.01074</td>
<td>0.01037</td>
<td>0.00777</td>
</tr>
<tr>
<td>Mining</td>
<td>Bid ask spread</td>
<td>0.00074</td>
<td>0.00175</td>
<td>0.00315</td>
<td>0.00488</td>
<td>0.00675</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00147</td>
<td>-0.00444</td>
<td>-0.01063</td>
<td>-0.00716</td>
<td>-0.03388</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00001</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00164</td>
<td>0.00453</td>
<td>0.02382</td>
<td>0.00968</td>
<td>0.00943</td>
</tr>
<tr>
<td>Constr</td>
<td>Bid ask spread</td>
<td>0.00123</td>
<td>0.00269</td>
<td>0.00441</td>
<td>0.00630</td>
<td>0.00819</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00211</td>
<td>-0.00508</td>
<td>-0.02069</td>
<td>-0.02372</td>
<td>-0.01274</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00254</td>
<td>0.00593</td>
<td>0.00938</td>
<td>0.01069</td>
<td>0.01007</td>
</tr>
<tr>
<td>Manufa</td>
<td>Bid ask spread</td>
<td>0.00117</td>
<td>0.00261</td>
<td>0.00431</td>
<td>0.00621</td>
<td>0.00813</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00184</td>
<td>-0.00336</td>
<td>-0.00403</td>
<td>-0.02153</td>
<td>-0.00629</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00246</td>
<td>0.00583</td>
<td>0.00938</td>
<td>0.01175</td>
<td>0.01006</td>
</tr>
<tr>
<td>Trasnp</td>
<td>Bid ask spread</td>
<td>0.00086</td>
<td>0.00187</td>
<td>0.00307</td>
<td>0.00446</td>
<td>0.00600</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00166</td>
<td>-0.00443</td>
<td>-0.11251</td>
<td>0.00189</td>
<td>0.00378</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00177</td>
<td>0.00418</td>
<td>0.00715</td>
<td>0.00973</td>
<td>0.00962</td>
</tr>
<tr>
<td>Wholes</td>
<td>Bid ask spread</td>
<td>-0.00005</td>
<td>0.00280</td>
<td>0.00445</td>
<td>0.00614</td>
<td>0.00776</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>0.00024</td>
<td>-0.00521</td>
<td>-0.00826</td>
<td>-0.01664</td>
<td>0.02645</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00003</td>
<td>-0.00003</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00265</td>
<td>0.00582</td>
<td>0.00893</td>
<td>0.00965</td>
<td>0.00911</td>
</tr>
<tr>
<td>Retail</td>
<td>Bid ask spread</td>
<td>0.00101</td>
<td>0.00234</td>
<td>0.00402</td>
<td>0.00589</td>
<td>0.00772</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00159</td>
<td>-0.00422</td>
<td>-0.00699</td>
<td>-0.00783</td>
<td>0.11977</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00219</td>
<td>0.00553</td>
<td>0.00886</td>
<td>0.01048</td>
<td>0.00947</td>
</tr>
<tr>
<td>Financ</td>
<td>Bid ask spread</td>
<td>0.00132</td>
<td>0.00290</td>
<td>0.00472</td>
<td>0.00668</td>
<td>0.00859</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00284</td>
<td>-0.01247</td>
<td>-0.00715</td>
<td>-0.00318</td>
<td>0.00073</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00274</td>
<td>0.00629</td>
<td>0.00979</td>
<td>0.01163</td>
<td>0.01107</td>
</tr>
<tr>
<td>Servic</td>
<td>Bid ask spread</td>
<td>0.00120</td>
<td>0.00274</td>
<td>0.00458</td>
<td>0.00656</td>
<td>0.00850</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00169</td>
<td>-0.00352</td>
<td>-0.00630</td>
<td>-0.14738</td>
<td>-0.00341</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>0.00000</td>
<td>-0.00001</td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.00258</td>
<td>0.00618</td>
<td>0.00973</td>
<td>0.01170</td>
<td>0.01100</td>
</tr>
<tr>
<td>PubAdm</td>
<td>Bid ask spread</td>
<td>0.00206</td>
<td>0.00288</td>
<td>0.00768</td>
<td>0.00528</td>
<td>0.00746</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td>-0.00294</td>
<td>0.01191</td>
<td>-0.00143</td>
<td>-0.00996</td>
<td>0.00211</td>
</tr>
<tr>
<td></td>
<td>Number of transactions</td>
<td>-0.00004</td>
<td>-0.00008</td>
<td>-0.00012</td>
<td>-0.00016</td>
<td>-0.00020</td>
</tr>
<tr>
<td></td>
<td>Nontrading days</td>
<td>0.16754</td>
<td>0.00719</td>
<td>0.00836</td>
<td>0.00721</td>
<td>0.00789</td>
</tr>
</tbody>
</table>
References


Appendix A
Proofs of Theorems

A.1 Consistency Lemmas

Lemma 4 Suppose that Assumption 1, 3-5 hold then,
\[ |E[\hat{f}(z)] - f(z)| = O\left(h^{2p^*}\right). \]

Proof. Note that
\[ E[\hat{f}(z)] = E\left[ \sum_{j=1}^{n} \sum_{s=1}^{T} \frac{w_{js}(\beta_j)}{(n \times T)h} K\left(\frac{z - v_{js}(\alpha)}{h}\right) \right] \]
\[ = \frac{1}{n} \sum_{j=1}^{n} E\left[ \frac{w_{j1}(\beta_j)}{h} K\left(\frac{z - v_{j1}(\alpha)}{h}\right) \right]. \]  
(A.1)

for fixed $z$. Time subscript $s$ in the above expression is suppressed due to stationarity.

With $u \equiv \frac{z - v_{j1}(\alpha)}{h}$, the expectation in (A.1) becomes:
\[ \int_R \int_R w_{j1}(\beta_j) \frac{1}{h} K\left(\frac{z - v_{j1}(\alpha)}{h}\right) g(w_{j1}, v_{j1}) dwdz \]
\[ = \int_R \int_R w_{j1}(\beta_j) K(u) \tilde{g}_j(z + hu | w_{j1}) g(w_{j1}) dwdu \]
where $\tilde{g}_j$ denote a conditional density of $v$ given $w$. By Taylor expanding $\tilde{g}_j(z + hu | w_{j1})$ around $h = 0$ with $K(u)$ which satisfies Assumption 3 (iii) and (iv), we can show that the above equation reduces to:
\[ \int_R w_{j1}(\beta_j) \tilde{g}_j(z | w_{j1}) g_j(w_{j1}) dw + O(h^{2p^*})E[w_{j1}(\beta_j)] \]
\[ = g(z)E[w_{j1}(\beta_j) | v_{j1}(\alpha) = z] + O(h^{2p^*}). \]
where $E[w_j(\beta_j)]$ are bounded as provided in Assumption 2. Given $n$ fixed, (A.1) becomes

$$E[\hat{f}(z)] = \frac{g(z)}{n} \sum_{j=1}^{n} E[w_j(\beta_j) | v_j(\alpha) = z] + O(h^{2p^*})$$

$$= f(z) + O(h^{2p^*}).$$

\[\Box\]

**Lemma 5** Suppose that Assumption 1, 3-5 and (10) - (13) in Hansen (2008) hold with compact $\Gamma \in \mathbb{R}^{2n+b-1}$ then, $\sup_{\gamma} \left| \hat{Q}(\gamma) - Q(\gamma) \right| \overset{p}{\to} 0$.

**Proof.** The uniform convergence of $\hat{Q}(\gamma) - Q(\gamma)$ is an immediate result of uniform convergence of $\hat{M}(z; \gamma)$ to $M(z; \gamma)$ in $\{z, \gamma\}$. Hansen (2008) shows the consistency of SLS estimator under the assumption of weak dependence of data. Under the Assumption 1,3, 4 and 5, $\hat{M}$ satisfies the conditions of the Theorem 2 in Hansen (2008) and the result follows Lemma 4 above. \[\Box\]

**Lemma 6** Suppose that Assumption 1, 3-5 and (10) - (13) in Hansen (2008) hold with

$$\vartheta_{it}(\gamma) \equiv (w_{it}(\beta_i) - \delta_i \cdot M(z; \gamma))^2 - E \left[ (w_{it}(\beta_i) - \delta_i \cdot M(z; \gamma))^2 \right]$$

and let $\vartheta_{it}(\gamma)$ satisfy (i) all the assumptions in Theorem 4.2.2. of Amemiya (1985) for each $\gamma \in \Gamma$, a compact subset of $\mathbb{R}^{2n+b-1}$ (Euclidean K-space) and (ii) assumptions in Corollary 3.48 in White (2001). Then,

$$\sup_{\gamma} |Q(\gamma) - E[Q(\gamma)]| \overset{p}{\to} 0.$$

**Proof.** Note that $\vartheta_{it}(\gamma)$ is strictly stationary and strong mixing with mixing coefficients $\alpha_m$ since it is a function of $w_{it}(\beta_i)$ and $v_{it}(\alpha)$ for all $i$. Note also that
\( E[\partial_{it}(\gamma)] = 0 \). Given assumptions above, desired result follows law of large numbers of McLeish (1975) and the strategy in Theorem 4.2.1. and 4.2.2. of Amemiya (1985). Theorem 2.10 McLeish (1975) provides conditions applicable to weakly dependent data and replaces Hoadley (1971)’s theorem in Amemiya (1985)’s proof. Theorem 3.48 in White (2001) is corollary of Theorem 2.10 McLeish (1975) with simpler version of conditions. ■

A.2 Proof of Theorem 1

Note that \( v_{it}(\alpha) = i_{it}^{(1)} + \cdots + \alpha_b \cdot i_{it}^{(b)} \) and

\[
E[Q(\gamma)] = \frac{1}{2} \cdot \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} E \left[ \left( w_{it}(\beta_i) - \delta_i \cdot M(z; \gamma) \right)^2 \right].
\]  

(A.2)

for an arbitrary \( z \). Now by substituting the right-hand side of model (1.1) for \( R_{it}^e \) in \( w_{it}(\beta_i) \) and using the structure of \( M(z; \gamma) \), the expectation inside double summation can be rewritten as:

\[
E((u_{it} + \Delta_{it,1} + \Delta_{it,2})^2)
\]

where

\[
\begin{align*}
u_{it} &= w_{it}(\beta_{0i}) - \delta_{0i}\bar{\theta}_0 L(z) \\
\Delta_{it,1} &= (\beta_{0i} - \beta_i)R_{mt}^e - \delta_i E(R_{ms}^e \mid v_{js}(\alpha) = z) \sum_{j=1}^{n} \frac{(\beta_{0j} - \beta_j)}{n}, \text{ and} \\
\Delta_{it,2} &= \delta_{0i}\bar{\theta}_0 L(z) - \delta_i \bar{\theta}_0 E(L_{js} \mid v_{js}(\alpha) = z).
\end{align*}
\]

where \( L_{js} \equiv L(v_{js}(\alpha_0)) \). Note that \( \Delta_{it,1} \) is a function of \( R_{mt}^e \) and \( z \), and \( \Delta_{it,2} \) is a function of \( z \). Thus \( u_{it} \) and \( \Delta_{it,1} + \Delta_{it,2} \) are orthogonal to each other, then \( E[Q(\gamma)] \) becomes

\[
\frac{1}{2} \cdot \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ E(u_{it}^2) + E((\Delta_{it,1} + \Delta_{it,2})^2) \right]
\]

Notice that there are no unknown parameters in \( u_{it} \) and \( E(u_{it}^2) \geq 0 \). It can be shown that \( \Delta_{it,1} + \Delta_{it,2} = 0 \) at the true parameter values. Therefore, at any other candidate
for a minimizer, we must have
\[ \Delta_{it,1} + \Delta_{it,2} = 0 \quad (A.3) \]
for almost all \( t \). If we take the conditional expectation \( \Delta_{it,2} \) given \( l_{it} \), it fixes the value of the argument for \( \Delta_{it,2} \):
\[ E(\Delta_{it,2} \mid \{l_{it}\}) = \Delta_{it,2}. \]
Thus
\[ E(\Delta_{it,1} + \Delta_{it,2} \mid \{l_{it}\}) = E(\Delta_{it,1} \mid \{l_{it}\}) + \Delta_{it,2} = 0. \quad (A.4) \]
Subtracting (A.4) from (A.3) gives
\[ (\beta_{0i} - \beta_i)(R_{mt} - E(R_{mt} \mid \{l_{it}\})) = 0. \]
Recall that condition (i) \( E(R_{mt} \mid \{l_{it}\}) \neq R_{mt} \) excludes the case when the market return is completely explained by liquidity indicators. Therefore the above equation holds for all \( i \) and \( t \) if and only if \( \beta_i = \beta_{0i} \) for all \( i \). We can apply this result to our objective function and obtain:
\[ \frac{1}{n} \cdot \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} E \left[ [w_{it}(\beta_{0i}) - \delta_i \cdot M(z; (\alpha, \beta_0, \delta))]^2 \right] \quad (A.5) \]
where \( \alpha, \beta_0, \delta \) are vectors of \( \alpha' \)'s, \( \beta_{0i}' \)s and \( \delta_i' \)'s, respectively. Then again by replacing \( R_{it}^e \) in \( w_{it}(\beta_{0i}) \) with the right-hand side of (1.1), we have
\[ E[[w_{it}(\beta_{0i}) - \delta_i \cdot M(z; (\alpha, \beta_0, \delta))]^2] = E((u_{it} + \Delta_{it})^2) \]
where
\[ \Delta_{it} = \delta_{0i} \left[ \tilde{\theta}_0 L(z) \right] - \delta_i \cdot M(z; (\alpha, \beta_0, \delta)). \quad (A.6) \]
Note that
\[ M(z; (\alpha, \beta_0, \delta)) = \tilde{\theta}_0 E(L_{js} \mid v_{js}(\alpha) = z) \]
and \( E(L_{js} | v_{js}(\alpha_0) = z) = L(z) \). \( u_{it} \) and \( \Delta_{it} \) are orthogonal to each other for all \( i \) and \( t \), then (A.6) becomes:

\[
\frac{1}{2} \cdot \frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ E((u_{it})^2) + E((\Delta_{it})^2) \right].
\]

Again if we can find parameter values for each \( i \) which make \( \Delta_{it} = 0 \) for almost all \( t \), \( Q(\gamma) \) achieves its minimum. Here, \( \Delta_{it} \) is rearranged into a term which depends on the gap between \( \delta_{0i} \) and \( \delta_i \) and the other which arises from the difference between \( \alpha_0 \) and \( \alpha \). Once the problems of \( \alpha' \)s and \( \delta' \)s are separated, \( \Delta_{it} \) is shown to be zero only when \( \alpha = \alpha_0 \) and \( \delta_i = \delta_{0i} \) under condition (ii). Now

\[
\Delta_{it} = (\delta_{0i} - \delta_i) \left[ \bar{\theta}_0 L(z) \right] \\
+ \left[ \delta_i \left[ \bar{\theta}_0 L(z) \right] - \delta_i \left[ \bar{\theta}_0 E(L_{js} | v_{js}(\alpha) = z) \right] \right]. \tag{A.7}
\]

From the single index assumption, the expectation of the second term in (A.7) conditioned on \( v_{js}(\alpha) \) is zero. Thus if for each \( i \), \( \Delta_{it} = 0 \) for almost all \( t \),

\[
(\delta_{0i} - \delta_i) \left[ \bar{\theta}_0 E(L_{js} | v_{js}(\alpha) = z) \right] = 0 \tag{A.8}
\]

for almost all \( t \). Provided that \( E(L_{js} | v_{js}(\alpha) = z) \) is not zero for only one period in (A.8), \( \delta_i \) can be identified. Given condition (ii) \( \bar{\theta}_0 \neq 0 \) and \( E(L_{js} | v_{js}(\alpha) = z) \neq 0 \) which is a result of condition (iii), it follows that \( \delta_i = \delta_{0i} \) for all \( i \). Now (A.7) becomes:

\[
L(z) - E(L_{js} | v_{js}(\alpha) = z) = 0.
\]

Identification of the liquidity index parameters follow from Ichimura (1993).

\section*{A.3 Proof of Theorem 2}

As stated in previous section, our asymptotic is based on fixed \( n \) and growing \( T \). Once any claim is proved with one asset, the argument can be easily extended to the case of \( n \) number of assets. Thus, index for assets are suppressed in the following discussions unless noted otherwise.
We will further study the asymptotics of estimated functions which are involved in our estimator. In many cases, we examine the sample average of product of multiple estimated functions. Such object is divided into terms where the estimated functions are replaced by the gaps between themselves and their true counterparts. Since the trimming should be based on the estimated quantiles, this argument will also work through the estimated trimming as well. Trimming function is originally employed to keep the index density inside the support but obviously there is no guarantee that the replacement performs as one. In order to resolve this conflict, we use a function which has a special structure given by Assumption 5.

Lemma A.1. and A.2. of Klein (1993) show that \( \tau(z, \pi_0) \) can be approximated by the product of \( M^{th} \)-order polynomials of \( \hat{\pi} - \pi_0 \) and \( M^{th} \)-order derivative of trimming function with \( M_{\tau} \) to be specified later. By virtue of special structure of (1.7), the derivatives (of any order) of \( \tau \) act as a trimming function, namely, successfully prevent the index density from being close to zero. Thus whenever it is necessary to study the convergence of estimated functions in the following discussion and \( \hat{\tau}_t - \tau_t \) appears in the subsequent argument, we replace it with its approximation.

A.3.1 Step 1: First part of Hessian converges to a positive definite matrix.

By adding and subtracting true functions of \( \hat{\tau}_t \) and \( \hat{\chi}_t(\bar{\gamma}) \) in turn, (2.5) is expressed as a sum of different sample means whose elements converge to zero in probability. Hence the argument here requires the estimated trimming function and the first derivative of the kernel regression estimator to go to their true counterparts. Note that the convergence must be uniform since (2.5) is evaluated at \( \bar{\gamma} \) which depends on \( \hat{\gamma} \). Here, \( \hat{\chi}_t(\bar{\gamma}) \) differs from \( \chi_t(\bar{\gamma}) \) only in that it has estimated conditional expectation instead of true one, its convergence solely depends on the behavior of first derivative of \( \delta \cdot \bar{M}(z; \gamma) \).
Lemma 7 Suppose that Assumption 1, 3-5 and (10) - (13) in Hansen (2008) hold with compact $V \in \mathbb{R}$ and $\Gamma \in \mathbb{R}^{2n+b-1}$, then

$$\sup_{(z,\gamma) \in V \times \Gamma} \left| \nabla_{\gamma} \hat{M}(z;\gamma) - \nabla_{\gamma} M(z;\gamma) \right| = O_p \left( \left( \frac{\ln T}{Th^3} \right)^{\frac{1}{2}} + h^{2p^*} \right).$$  \hfill (A.9)

**Proof.** First derivative of estimated conditional expectation can be divided into parts by using $\hat{f}(z)$, $\hat{g}(z)$ and their first derivatives, thus (A.9) relies on the convergence of those functions. They can be proved essentially in the same manner, namely, split the difference between estimated function and true function into variance and bias parts and study them separately. Theorem 2 in Hansen (2008) gives the uniform result for variance part and show that it tends to zero at the rate of $\left[ \frac{\ln T}{Th^3} \right]^{\frac{1}{2}}$. It also can be shown that the bias is $O(h^{2p^*})$ and the convergence is in fact uniform by using exactly the same strategy as in Lemma (4). Finally, trimming function keeps $\hat{g}(z)$ (the denominator in $\hat{M}(z;\gamma)$ and $\nabla_{\gamma} \hat{M}(z;\gamma)$) away from zero or let it converge to zero slower than the numerator, which completes the proof for (A.9). \qed

Provided (A.9) and uniform convergence of trimming function which is to be discussed later, the limiting behavior of estimated functions removes the remainder leaving only (A.10) below.

Lemma 8 Suppose that Assumption 1, 3-5 and (10) - (13) in Hansen (2008) hold, then

$$\frac{1}{T} \sum_{t=1}^{T} \hat{\tau}_t \hat{\chi}_t(\bar{\gamma})\hat{\chi}_t(\bar{\gamma})' \overset{p}{\to} H_0$$  \hfill (A.10)

for matrix $H_0$, where $H_0 = E(\tau_1 \cdot \chi_1(\gamma_0)\chi_1(\gamma_0))$.

**Proof.** In order to show (A.10), first we need to show $\frac{1}{T} \sum_{t=1}^{T} \hat{\tau}_t \hat{\chi}_t(\bar{\gamma})(\hat{\chi}_t(\bar{\gamma}))'$ converges to the corresponding function where $\hat{\chi}_t(\bar{\gamma})$ is replaced by $\chi_t(\bar{\gamma})$. This can be done by expressing each estimating function as the sum of the true counterpart and the remainder. This strategy is firstly applied to $\hat{\tau}_t$ to get

$$\frac{1}{T} \sum_{t=1}^{T} \hat{\tau}_t \hat{\chi}_t(\bar{\gamma})\hat{\chi}_t(\bar{\gamma})'$$
\[ = \frac{1}{T} \sum_{t=1}^{T} \tau_t \tilde{\chi}_t(\tilde{\gamma})\tilde{\chi}_t(\tilde{\gamma})' + \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \tilde{\chi}_t(\tilde{\gamma})\tilde{\chi}_t(\tilde{\gamma})'. \]

Then employing similar approach for each \( \hat{\chi}_t(\tilde{\gamma}) \) gives:

\[
\frac{1}{T} \sum_{t=1}^{T} \tau_t \chi_t(\gamma)\chi_t(\gamma)' + \frac{1}{T} \sum_{t=1}^{T} \tau_t (\hat{\chi}_t(\tilde{\gamma}) - \chi_t(\gamma))\chi_t(\gamma)' + \frac{1}{T} \sum_{t=1}^{T} \tau_t \chi_t(\gamma)(\hat{\chi}_t(\tilde{\gamma}) - \chi_t(\gamma))' + \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \chi_t(\gamma)(\hat{\chi}_t(\tilde{\gamma}) - \chi_t(\gamma))' + \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t)(\hat{\chi}_t(\tilde{\gamma}) - \chi_t(\gamma))(\hat{\chi}_t(\tilde{\gamma}) - \chi_t(\gamma))'.
\]

By the uniform convergence theorem of Lemma 6, (A.11) converges in probability to

\[ E \left[ \tau_t \left( \nabla_\gamma w_t(\tilde{\beta}) - \nabla_\gamma \left[ \delta \cdot M(\tilde{z}; \tilde{\gamma}) \right] \right) \left( \nabla_\gamma w_t(\tilde{\beta}) - \nabla_\gamma \left[ \delta \cdot M(\tilde{z}; \tilde{\gamma}) \right] \right)' \right] \]

It can also be shown that the rest of the terms in the above equation tend to zero uniformly. For example,

\[
\frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \chi_t(\gamma)(\hat{\chi}_t(\tilde{\gamma}) - \chi_t(\gamma))' = \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \left( \nabla_\gamma w_t(\tilde{\beta}) - \nabla_\gamma \left[ \delta \cdot M(\tilde{z}; \tilde{\gamma}) \right] \right) \times \left( \nabla_\gamma \left[ \delta \cdot M(\tilde{z}; \tilde{\gamma}) \right] - \nabla_\gamma \left[ \delta \cdot \hat{M}(\tilde{z}; \tilde{\gamma}) \right] \right)' \]

is bounded by

\[
\sqrt{\frac{1}{T} \sum_{t=1}^{T} (\nabla_\gamma w_t(\tilde{\beta}) - \nabla_\gamma \left[ \delta \cdot M(\tilde{z}; \tilde{\gamma}) \right])^2} \times \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \left( \nabla_\gamma \left[ \delta \cdot M(\tilde{z}; \tilde{\gamma}) \right] - \nabla_\gamma \left[ \delta \cdot \hat{M}(\tilde{z}; \tilde{\gamma}) \right] \right)^2} \]
by Cauchy-Schwarz inequality. Given smoothed trimming function as suggested in (1.7)

\[ \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left( \nabla_{\gamma} w_t(\beta) - \nabla_{\gamma} \left[ \delta \cdot M(\bar{z}; \bar{\gamma}) \right] \right)^2} \]

(A.12)

\[ \times \sqrt{\frac{1}{T} \sum_{t=1}^{T} ((\pi - \hat{\pi})^{2M} \left( \nabla_{\pi}^{M} \tau_t(\nabla_{\gamma} \left[ \delta \cdot M(\bar{z}; \bar{\gamma}) \right]) - \nabla_{\gamma}[\delta \cdot \hat{M}(\bar{z}; \bar{\gamma})])^2 \right)^2}. \]

(A.13)

where \( \hat{\pi} \) and \( \pi \) are quantile estimator and true quantile of \( \{x_{it}\} \), respectively. Note that in (A.13), \( \nabla_{\pi}^{k} \tau_t \) plays a role of trimming function and \( (\pi - \hat{\pi})^{2M} \) is bounded with finite \( M_r \). By an argument similar to Lemma 6, (A.12) converges to its finite expectation and (A.13) converges to zero by Lemma 7. The arguments for the rest of the items are analogous. Finally using the uniform law of large numbers Lemma 6 and stationarity,

\[ \sup_{z, \gamma} \left| \frac{1}{T} \sum_{t=1}^{T} [\tau_t \chi_t(\bar{\gamma})(\chi_t(\bar{\gamma}))'] - E \left[ \tau \cdot \chi(\bar{\gamma})(\chi(\bar{\gamma}))' \right] \right| \]

converges to 0 in probability. Due to consistency of \( \gamma \),

\[ \lim_{\bar{\gamma} \to \gamma_0} E \left[ \tau \cdot \chi(\bar{\gamma})(\chi(\bar{\gamma}))' \right] = E \left[ \tau \cdot \chi(\gamma_0)(\chi(\gamma_0))' \right] \]

and the result follows.

A.3.2 Step 2: Remainder of Hessian converges to zero in probability.

Step 2 follows essentially the same argument introduced in Step 1. However we have one more term to consider, \( \nabla_{\gamma} \hat{\chi}_t(\bar{\gamma}) \), in which the second derivative of estimated conditional expectation involves. Thus a stronger smoothness assumption is required for the objective function and the limiting function in the neighborhood of \( \gamma_0 \).

Lemma 9 Suppose that Assumption 1, 3-5 and (10) - (13) in Hansen (2008) hold with compact \( V \in R \) and \( \Gamma \in R^{2n+b-1} \), then

\[ \sup_{(z, \gamma) \in V \times \Gamma} \left| \nabla^2 \hat{M}(z; \gamma) - \nabla^2 M(z; \gamma) \right| = O_p \left( \left( \frac{\ln T}{Th^5} \right)^{\frac{1}{2}} + h^{2p^*} \right). \]

(A.14)
Proof. By the same argument in Lemma 7, we can show (A.14).

Lemma 10 Suppose that Assumption 1, 3-5 and (10) - (13) in Hansen (2008) hold, then
\[
\frac{1}{T} \sum_{t=1}^{T} \hat{\tau}_t \left[ w_t(\tilde{\beta}) - \bar{\delta} \cdot \hat{M}(\bar{z}; \bar{\gamma}) \right] \nabla_{\gamma} \hat{\chi}_t(\bar{\gamma}) \xrightarrow{p} 0. \tag{A.15}
\]

Proof. Again by decomposing an estimating function into the true function and the remainder, (A.15) gives
\[
\frac{1}{T} \sum_{t=1}^{T} \hat{\tau}_t \left[ w_t(\tilde{\beta}) - \bar{\delta} \cdot \hat{M}(\bar{z}; \bar{\gamma}) \right] \nabla_{\gamma} \hat{\chi}_t(\bar{\gamma}) = \frac{1}{T} \sum_{t=1}^{T} \tau_t \left[ w_t(\tilde{\beta}) - \bar{\delta} \cdot \hat{M}(\bar{z}; \bar{\gamma}) \right] \nabla_{\gamma} \hat{\chi}_t(\bar{\gamma}) + \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \left[ w_t(\tilde{\beta}) - \bar{\delta} \cdot \hat{M}(\bar{z}; \bar{\gamma}) \right] \nabla_{\gamma} \hat{\chi}_t(\bar{\gamma}).
\]

A similar approach for each estimated function, let us divide the above equation into parts where the convergence is studied separately (to simplify notations, drop the arguments, which are evaluated at \( \bar{\gamma} \), in this equation):
\[
\frac{1}{T} \sum_{t=1}^{T} \tau_t \left[ w_t - \bar{\delta} M_t \right] \nabla_{\gamma} \chi_t \tag{A.16}
\]
\[
+ \frac{1}{T} \sum_{t=1}^{T} \tau_t \left[ w_t - \bar{\delta} M_t \right] (\nabla_{\gamma} \hat{\chi}_t - \nabla_{\gamma} \chi_t) + \frac{1}{T} \sum_{t=1}^{T} \tau_t \bar{\delta} \left[ \hat{M}_t - M_t \right] \nabla_{\gamma} \chi_t
\]
\[
+ \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \left[ w_t - \bar{\delta} M_t \right] \nabla_{\gamma} \hat{\chi}_t + \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \bar{\delta} \left[ \hat{M}_t - M_t \right] \nabla_{\gamma} \hat{\chi}_t
\]
\[
+ \frac{1}{T} \sum_{t=1}^{T} \tau_t \bar{\delta} \left[ \hat{M}_t - M_t \right] (\nabla_{\gamma} \hat{\chi}_t - \nabla_{\gamma} \chi_t) + \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \bar{\delta} \left[ \hat{M}_t - M_t \right] (\nabla_{\gamma} \hat{\chi}_t - \nabla_{\gamma} \chi_t)
\]
\[
+ \frac{1}{T} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \left[ w_t - \bar{\delta} M_t \right] (\nabla_{\gamma} \hat{\chi}_t - \nabla_{\gamma} \chi_t)
\]
By Lemma 4, 9, 10, the remainder terms other than (A.16) in the above expression uniformly tend to zero. Thus we have shown that (A.15) is asymptotically equivalent to (A.16). To see further that (A.16) converges to zero, note that by a uniform law of large numbers given as a result of Lemma 6 and consistency of parameter estimates,

\[
\frac{1}{T} \sum_{t=1}^{T} \tau_t \left[ w_t(\bar{\beta}) - \delta M(\bar{z}; \bar{\gamma}) \right] \left[ \nabla_\gamma^2 w_t(\bar{\beta}) - \nabla_\gamma^2 \left[ \delta M(\bar{z}; \bar{\gamma}) \right] \right]
\]

converges uniformly to

\[
E \left[ \tau_t [w_t(\beta_0) - \delta_0 \cdot M(z; \gamma_0)] \left[ \nabla_\gamma^2 w_t(\beta_0) - \nabla_\gamma^2 \left[ \delta_0 \cdot M(z; \gamma_0) \right] \right] \right].
\]

(A.17)

Since \( u_t \) is orthogonal to \( x_t \equiv \{ R_{mt}, 1 \} \) and \( \nabla_\gamma \chi_t(\gamma) |_{\gamma = \gamma_0} \) does not depend on \( R_{it} \), taking iterated expectation conditional on \( x_t \) shows that the (A.17) is exactly zero.

\[\blacksquare\]

A.3.3 Step 3: Gradient converges to a multivariate normal random vector in distribution.

By adding and subtracting \( \tau_t \chi_t(\gamma_0) \hat{u}_t \) from the summands of \( \sqrt{T} \hat{G}(\gamma_0) \):

\[
\sqrt{T} \hat{G}(\gamma_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \hat{\tau}_t \chi_t(\gamma_0) - \tau_t \chi_t(\gamma_0) \right] \hat{u}_t + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t \chi_t(\gamma_0) \hat{u}_t = R + N.
\]

Then the claim is that \( R \) converges to zero and \( N \) goes to a multivariate normal random variable. Since the gradient term is evaluated at the truth, the parameter values will be omitted the relevant functions of following subsections.

\( R \) converges to zero in probability.

Note that \( R \) can be again divided into several terms:

\[
\sum_{t=1}^{T} \frac{\tau_t(\hat{\chi}_t(\gamma_0) - \chi_t(\gamma_0))u_t}{\sqrt{T}} + \sum_{t=1}^{T} \frac{(\hat{\tau}_t - \tau_t)\chi_t(\gamma_0)u_t}{\sqrt{T}} + \text{Remainder}
\]
\[ R = R_1 + R_2 + R_3, \]

where the remainder consists of
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t)(\hat{\chi}_t(\gamma_0) - \chi_t(\gamma_0))u_t \\
+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t)\hat{\chi}_t(\gamma_0)\delta_0 \left[ M(z; \gamma_0) - \hat{M}(z; \gamma_0) \right] \\
+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t(\hat{\chi}_t(\gamma_0) - \chi_t(\gamma_0))\delta_0 \left[ M(z; \gamma_0) - \hat{M}(z; \gamma_0) \right].
\]

Note that at least one of the functions \((\hat{\tau}_t, \hat{\chi}_t, \hat{u}_t)\) in the summand of \(R\) has been replaced by the difference between the function and its true counterpart in the above expression. This decomposition reveals that the convergence of \(R\) partially or entirely depends on the convergence of each estimated functions. If it can be shown that for each case, the summand converges to zero at a rate faster than \(\sqrt{T}\), the desired result follows. In doing this, we can easily see that the convergence of \(R_1\) and \(R_2\) are bound to be slower than that of any term in \(R_3\). In other words, the convergence of \(R_3\) follows as a consequence of the convergence of first two terms. Therefore we study \(R_1\) and \(R_2\) first in the next two subsections and the results for \(R_3\) will be briefly summarized in the following subsection.

(A) Term \(R_1\) converges to zero in probability.

In the following lemma we will use a similar strategy in Ichimura and Lee (1991) and show that \(R_1\) is dominated by a term that we subsequently show converges in probability to zero.

**Lemma 11** \(R_1\) is bounded by:
\[
\sum_{b_t} B_{b_t} \left| \frac{\delta_0}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t b_t \right| \tag{A.18}
\]
where \(B_{b_t}\) is a positive constant and \(b_t \in \{f_{0t} - \hat{f}_{0t}, \hat{f}_{0t}' - f_{0t}', \hat{g}_{0t} - g_{0t}, \hat{g}_{0t}' - \hat{g}_{0t}'\}\) with \(g_{0t} \equiv g(z), \hat{g}_{0t} \equiv \hat{g}(z), f_{0t} \equiv M(z; \gamma_0) \cdot g_{0t}, \hat{f}_{0t} \equiv \hat{M}(z; \gamma_0) \cdot \hat{g}_{0t}, g_{0t}' \equiv \nabla_{\gamma} g_{0t}, \hat{g}_{0t}' \equiv \nabla_{\gamma} \hat{g}_{0t}, f_{0t}' = \nabla_{\gamma} f_{0t}\) and \(\hat{f}_{0t}' = \nabla_{\gamma} \hat{f}_{0t}\).
Proof. We can rearrange $R_1$ to be linear in $\hat{f}_t$ and $\hat{g}_t$:

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t \left( \nabla_{\gamma} \left[ \delta_0 M(z; \gamma_0) - \nabla_{\gamma} \left[ \delta_0 M(z; \gamma_0) \right] \right] \right)
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t \left[ \left( \hat{f}_0' - \hat{g}_0' \hat{f}_0 \hat{g}_0 \right) \frac{1}{\hat{g}_0} - \left( f_0' - g_0' \hat{f}_0 \hat{g}_0 \right) \frac{1}{g_0} \right]
$$

By mean value theorem, the above term is equivalent to:

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t \times \left[ \frac{\hat{g}_0'(f_0 - \hat{f}_0)}{\hat{g}_0^2} + \frac{1}{g_0} (f_0' - f_0') + (\hat{f}_0' + 2\hat{g}_0' \hat{f}_0 \hat{g}_0) \frac{1}{\hat{g}_0^2} (g_0 - g_0') + \frac{\hat{f}_0}{\hat{g}_0^2} (g_0' - \hat{g}_0') \right],
$$

where $g_0$ is a value between $\hat{g}_0$ and $g_0$ and other terms are similarly defined. Hence we can get the bound for $R_1$:

$$
|R_1| \leq \sup_{x \in X} \left| \frac{\hat{g}_0'}{\hat{g}_0^2} \right| \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t (f_0 - \hat{f}_0) \right| + \sup_{x \in X} \left| \frac{1}{g_0} \right| \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t (f_0' - f_0') \right| \\
+ \sup_{x \in X} \left| \hat{f}_0' + 2\hat{g}_0' \hat{f}_0 \hat{g}_0 \right| \left| \frac{1}{\hat{g}_0^2} \right| \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t (g_0 - g_0') \right| \\
+ \sup_{x \in X} \left| \frac{\hat{f}_0}{\hat{g}_0^2} \right| \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t (g_0' - \hat{g}_0') \right|.
$$

Note that $\hat{g}_0$ converges uniformly in $x \in X$ and its limiting function is bounded away from zero on $X$. Thus we can show that the factor $\sup_{x \in X} \left| \frac{\hat{g}_0'}{\hat{g}_0^2} \right|$ is bounded in probability. Also, $\hat{g}_0$ converges to $g_0$ uniformly in $x \in X$. Boundedness of parallel factors in the other terms can be proved in a similar way.

Now we will show that (A.18) vanishes away by applying mean squared error argument. First, it is split into squared part and cross product part:

$$
E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t u_t b_t \right)^2 \right] = \frac{1}{T} \sum_{t=1}^{T} \tau_t E(u_t^2 b_t^2) + \frac{1}{T} \sum_{t=1}^{T} \sum_{t \neq t} \tau_t \tau_t E(u_t u_t b_t b_t) \quad (A.19)
$$
\[ S \equiv S + C, \]

and we study the convergence separately.

**Lemma 12** If Assumption 1-6 and (10) - (13) in Hansen (2008) hold then,

\[ S = O \left( \frac{1}{Th} \right). \]

**Proof.** By taking an iterated expectation of \( S \),

\[ E ( E (u_t^2 b_t^2 \mid y_t)) = E (u_t^2 E (b_t^2 \mid y_t)) \]

and \( E(b_t^2 \mid y_t) \) is \( O \left( \frac{1}{Th} + h^{2p'} \right) \) by Theorem 1 in Hansen (2008). By Assumption 6, the speed of convergence is dominated by variance. Therefore,

\[ S \leq O \left( \frac{1}{Th} \right) E(u_t^2) \]

Since \( E(u_t^2) \) is bounded by Assumption 2,

\[ S \leq O \left( \frac{1}{Th} \right). \]

\[ \blacksquare \]

**Lemma 13** Suppose Assumption 1-6 and (10) - (13) in Hansen (2008) hold. Then

\[ C \leq O \left( \frac{1}{T^{1-2dh^2}} \right) \]

for \( d \in (0, 0.5) \).

**Proof.** To show that the cross product term also tends to zero, it will be divided into several parts again. Due to the time dependence in the data, the correlation between \( u_t \) and \( b_\ell \) (similarly between any combination of \( u_t \), \( u_\ell \), \( b_t \) and \( b_\ell \)) is much larger than that in the i.i.d. case. However our mixing condition in Assumption 1 which postulates the characteristic of dependence leads us to an asymptotic uncorrelatedness property of data or any function of data depending on finite time periods. The
uncorrelatedness is "asymptotic" in the sense that the covariance of two random variables at different time is not exactly zero but goes to zero as the gap between those time periods becomes large. Hence the problem boils down to separating terms into two parts where (a) the time indices of random variables in the product term are far apart and (b) they are close together. Here, for example $\ell$ being "close" to $t$ means that $\ell$ is $t + 1, ..., t + k$ for some finite $k$, more precisely $k(T)$ since we let $k$ grows as $T$ increases. The expected value of cross product term under (a) can be shown to vanish fast enough by virtue of the weak correlation between random variables in the product. Also the expected value of cross product term under (b) disappears as $T$ increases since the magnitude of correlation among the random variables whose time indices are close together will be nevertheless "small" in the sense that we have only finite number of such terms. Note that even if $k(T)$ grows as $T$ gets larger, we can let it expand much slower than $T$ does, making the above argument plausible. This works through the quadruple sum in cross product part and conclusion follows. Lemmas 18 - 24 in next section provides a detailed argument about the proof. Given parameter values, $O\left(\frac{1}{T^{1-2d}b^2}\right)$ gives the slowest convergence rate among the cases in the discussion which defines the convergence rate for $C$. ■

(B) Term $R_2$ converges to zero in probability.

As studied in Lemma A.1. and A.2. of Klein (1993), Taylor expanding the estimated trimming around $\hat{\pi} = \pi_0$ approximates $\hat{\tau}_t - \tau_t$ by a $M^{th}$ order polynomial of $\hat{\pi} - \pi_0$. It can be shown that the polynomials converge to zero at a certain rate due to the consistency of quantile estimator. Once $\hat{\tau}_t - \tau_t$ is replaced by its approximation, the derivatives of trimming function evaluated at the true quantile converge to zero in probability in the first $M_r - 1$ terms. Taking out $\hat{\pi} - \pi_0$ outside the summation, we can then apply a central limit theorem to show the remainder part tends to a normal random variable which is bounded in probability. Thus the first $M_r - 1$ terms disappear at the same rate as the quantile estimator $\hat{\pi}$ tends to $\pi_0$. The
convergence rate of quantile estimator for weakly dependent data can be given by the preexisting studies. However, we cannot employ the same strategy for the $M_\tau$-order term since the derivative of trimming function is evaluated at an arbitrary value between $\hat{\pi}$ and $\pi_0$. Yet, by setting $M_\tau$ sufficiently large ("sufficiently" should be defined appropriately), we can let $(\hat{\pi} - \pi_0)^{M_\tau}$ converge fast enough to make the desired result follow.

Lemma 14 Let $\tau_t = \tau(l_t, b) - \tau(l_t, a)$ where $\tau(\cdot, \cdot)$ is defined as in (1.7). If Assumption 1-5 hold,

$$R_2 = O_p(T^{-\left(r_\pi - \varepsilon\right)})$$

where $\varepsilon > 0$ and $|\hat{\pi} - \pi_0| = O_p(T^{-r_\pi})$.

Proof. Taylor expanding ($M_\tau$th order) the trimming function around $\hat{a} = a$ and employing approaches in Lemma A.1. and A.2. of Klein (1993) with central limit theorem (Ibragimov (1962)), our trimming is approximated by:

$$(M_\tau - 1)O_p\left(T^{-r_\pi + \varepsilon}\right) O_p\left(1\right)$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{M_\tau!} \left[\nabla_{M_\tau} \tau(a^+ - l_t)\right] (\hat{a} - a)^{M_\tau} u_t \chi_t.$$  

Then the convergence of the remainder depends on $[\ln T(\hat{a} - a)]^{M_\tau}$. Since $|\hat{a} - a| = O_p(T^{-r_\pi})$ for $r_\pi \leq \frac{1}{2}$, we have

$$[\ln T(\hat{a} - a)]^{M_\tau} = O_p\left(\left(\frac{T^{r_\pi}}{\ln T}\right)^{-M_\tau}\right).$$

If we can find $M_\tau$ which makes

$$\left(\frac{T^{r_\pi}}{\ln T}\right)^{-M_\tau} < T^{-r_\pi},$$

we get the desired result and this can be done by choosing $M_\tau$ such that

$$- M_\tau \ln \left(\frac{T^{r_\pi}}{\ln T}\right) < -r_\pi \ln T$$

$$M_\tau > \frac{r_\pi \ln T}{r_\pi \ln T - \ln \ln T}. \quad (A.20)$$
With the corresponding convergence rate, the remainder is bounded by:

\[ O_p(T^{-(\alpha - \varepsilon)}) \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |u_t \chi_t| \right] \]

and the convergence of \( R_2 \) follows.

The remaining task is summarized to finding the convergence rate of \((\hat{\pi} - \pi_0)\). This can be done by obtaining central limit theorem for quantile estimator which guarantees \( \sqrt{T} \) consistency of \( \hat{\pi} \). Note that our data has time dependence thus the quantile estimator should appropriately reflect the memory property. Sun and Cordero (2005) studies behavior of (smooth) quantile estimator for a sequence of stationary strong mixing random variables. They provide conditions for central limit theorem and the result can be applied to our case above. Mixing rates necessary for the theorem will be checked in the last section.

(C) Term \( R_3 \) converges to zero in probability.

Note that \( R_3 \) has terms:

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t)(\hat{\chi}_t - \chi_t) u_t \\
+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{\tau}_t - \tau_t) \hat{\chi}_t \left[ \delta_0 M(z; \gamma_0) - \delta_0 \hat{M}(z; \gamma_0) \right] \\
+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau_t(\hat{\chi}_t - \chi_t) \left[ \delta_0 M(z; \gamma_0) - \delta_0 \hat{M}(z; \gamma_0) \right].
\]

All three summands in the above expression in fact have the same structure, namely two components in the product of \( \tau_t \chi_t u_t \) have been replaced by the difference between the estimated function and true function. In order for all terms above to converge to zero, the summands need to disappear quickly enough to offset \( \sqrt{T} \). Here those two differences in each summand will contribute the convergence of the corresponding term. In the first expression, for example, the convergence rate of either \( \hat{\tau}_t - \tau_t \) or \( \hat{\chi}_t - \chi_t \) alone is not fast enough to overpower \( \sqrt{T} \), however, the product of these two
terms converges faster than $\sqrt{T}$ rate. This strategy can be applied to all the cases above which implies $R_3$ converges to zero in probability.

**N converges to a multivariate normal random variable.**

We need to construct a U-statistics which incorporates with symmetrization of terms for each time period, we will consider the full asset dimension in this subsection. Recall that $N$ is in the form of

$$
\frac{1}{n\sqrt{T}} \sum_{i=1}^{n} \sum_{t=1}^{T} \tau_{it} \chi_{it} \hat{u}_{it}
$$

$$
= \frac{\sqrt{T}}{T} \sum_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} \tau_{it} \chi_{it} u_{it} + \frac{\sqrt{T}}{T} \sum_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} \tau_{it} \chi_{it} \left[ \delta_{0i} M(v_{it}; \gamma_0) - \delta_{0i} \hat{M}(v_{it}; \gamma_0) \right]
$$

$$
\equiv \sqrt{T} (N_1 + N_2)
$$

(A.21)

The proof strategy in this part is (i) define (centered) U-statistics $U_T$ which approximates $N_2$, (ii) define $\hat{U}_T$ which is asymptotically equivalent to $U_T$ and (iii) apply central limit theorem to $\sqrt{T} (N_1 + \hat{U}_T)$. In order to make a centered U-statistics argument work, first we need to show that the mean of summand in $N_2$ converges to zero. For this purpose, Lemma 15 shows that we may simplify $N_2$ so that it is linear in the estimated components.

**Lemma 15** *Under Assumption 1-5,*

$$
\frac{\sqrt{T}}{T} \sum_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} \tau_{it} \chi_{it} \left[ \delta_{0i} M(v_{it}; \gamma_0) - \delta_{0i} \hat{M}(v_{it}; \gamma_0) \right] \left( \frac{\hat{g}_{it}}{g_{it}} - 1 \right) \xrightarrow{p} 0.
$$

**Proof.** Rearrange the above expression as

$$
\frac{\sqrt{T}}{T} \sum_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} \tau_{it} \chi_{it} \left( \frac{\delta_{0i} M(v_{it}; \gamma_0) - \delta_{0i} \hat{M}(v_{it}; \gamma_0)}{g_{it}} \right) (\hat{g}_{it} - g_{it}).
$$

This tends to zero by Cauchy-Schwarz inequality and convergence of $\hat{g}_{it}$ and $\hat{f}_{it}$. □
Thus we are allowed to study

$$\frac{\sqrt{T}}{T} \sum_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} \tau_{it} \chi_{it} \left[ \delta_{0i} M(v_{it}; \gamma_0) - \delta_{0j} M(v_{it}; \gamma_0) \right] \frac{\hat{g}_{it}}{g_{it}}$$

$$= \frac{\sqrt{T}}{T(T-1)} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \left[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (M_{it,0} \phi_{js} - w_{js} \phi_{js}) \lambda_{it} \right].$$

(A.22)

where $M_{it,0} \equiv M(v_{it}; \gamma_0)$, $\phi_{it} \equiv \frac{1}{h} K \left( \frac{v_{it} - v_{js}}{h} \right)$ and $\lambda_{it} \equiv \frac{\tau_{it} \chi_{it} \delta_{0i}}{g_{it}}$. Next, define

$$\rho^*(y_t, y_s) \equiv \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (M_{it,0} \phi_{js} - w_{js} \phi_{js}) \lambda_{it}$$

(A.23)

where $y_t = \{R_{it}^e, R_{mt}^e, l_{it} \}_{i=1}^{n}$. Additionally, let

$$\rho(y_t, y_s) = \frac{\rho^*(y_t, y_s) + \rho^*(y_s, y_t)}{2}.$$  

(A.24)

In the following lemma, we will show that the expectation of $\rho(y_t, y_s)$ converges to zero at $\sqrt{T}$ rate.

**Lemma 16** Under Assumption 1-5,

$$\sqrt{T} E \left[ \rho(y_t, y_s) \right] \rightarrow 0$$

(A.25)

uniformly in $x_t \in X$.

**Proof.** Note that exactly same argument works for the expectations of two $\rho^*$s, thus here we focus on :

$$E \left[ E \left[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (M_{it,0} \phi_{js} - w_{js} \phi_{js}) \lambda_{it} \mid x_t \right] \right].$$

If we replace $w_{js}$ with the right hand side of the model, then we get

$$E \left[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{it} E \left[ M_{it,0} \phi_{js} - \delta_{0j} M_{js,0} \phi_{js} \mid x_t \right] \right]$$

(A.26)

$$- E \left[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{it} E \left[ \phi_{js} E \left[ u_{js} \mid x_t, v_s \right] \mid x_t \right] \right]$$

(A.27)

$$= E \left[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{it} E \left[ M_{it,0} - \delta_{0j} M_{it,0} \mid x_t \right] \right] + O(h^{2\rho^*}).$$
(A.27) disappears right away due to the strict exogeneity. By Taylor expanding the conditional density associated with the first expectation, we can rearrange (A.26) as a sum of its limit and the remainder term which vanishes at $h^{2p^*}$ rate. This follows the argument of Lemma 8 in Ichimura and Lee (1991) and bias reducing kernel of order $p^*$. Note the expectation in the last expression vanishes immediately due to:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} \lambda_{it} E\left[\frac{1}{n}\sum_{j=1}^{n} (M_{it,0} - \delta_{0j}M_{it,0}) \mid x_t\right]\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} \lambda_{it} E\left[M_{it,0} - M_{it,0} \mid x_i\right]\right] = 0.$$  

Then (A.25) follows for $h = O(T^{-r})$, $r > \frac{1}{4p^*}$. □

By Lemma 16, the U-statistics defined by

$$U_T = \frac{2}{T(T-1)} \sum_{t=1}^{T} \sum_{s<t}^{T} \rho(y_t, y_s)$$  

(A.28)

is asymptotically equivalent to the centralized version. As a consequence of Lemma 3.4 and 3.6 of Dehling and Wendler (2010), $U_T$ can be replaced by an approximation in sample mean form:

$$\hat{U}_T = \frac{2}{T} \sum_{t=1}^{T} E\left[\rho(y_t, y_s) \mid y_t\right].$$

Asymptotic normality of $\sqrt{T}(N_1 + N_2)$ is ready at this point by Ibragimov (1962)'s central limit theorem and Assumption 1-3. However, to obtain an appropriate approximation of the variance-covariance matrix, we need to further study the limiting form of $\sqrt{T}\hat{U}_T$ so that we can construct a sample analogue of variance-covariance matrix. First, note that from the approximation above $\sqrt{T}\hat{U}_T$ can be examined in two separate parts:

$$\sqrt{T}\hat{U}_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} E\left[\rho^*(y_t, y_s) \mid y_t\right]$$  

(A.29)

$$+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} E\left[\rho^*(y_s, y_t) \mid y_t\right].$$  

(A.30)

The strategy here is first to argue that (A.29) disappears as $T$ goes to infinity and (A.30) can be rewritten in a simplified sample mean form to which central limit
theorem is applicable. Now, note that replacing \( w_{js} \) by \( \delta_{0j} M_{js,0} + u_{js} \) expands \( E[\rho^*(y_t, y_s) \mid y_t] \) in (A.29) as

\[
\frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \sum_{s \neq t} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{it} E[(M_{it,0} - \delta_{0j} M_{js,0}) \phi_{js} \mid y_t] \tag{A.31}
\]

\[
- \frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \sum_{s \neq t} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{it} E[u_{js} \phi_{js} \mid y_t]. \tag{A.32}
\]

Write the conditional expectation in (A.31) using integrals:

\[
\int \int [M(v_{it}, \gamma_0) - \delta_{0j} M(v_{js}, \gamma_0)] \frac{1}{h} K \left( \frac{v_{it} - v_{js}}{h} \right) g_{wv}(w_{js}, v_{js} \mid y_t) dv_{js} dw_{js} \tag{A.33}
\]

where \( g_{wv} \) is a (joint) conditional density of \((w_{js}, v_{js})\) given \( y_t \). Define \( z \equiv \frac{v - v_{it}}{h} \) where \( v_{it} = v \) and substitute \( \frac{v - v_{js}}{h} \) with \( z \) and \( v_{js} \) with \( v + hz \) in the above expression. Finally Taylor expanding \([M(v, \gamma_0) - \delta_{0j} M(v + hz, \gamma_0)] g_{wv}(w_{js}, v + hz \mid y_t) \) around \( h = 0 \) gives an approximation to \( E[(M_{it,0} - \delta_{0j} M_{js,0}) \phi_{js} \mid y_t] \):

\[
[M(v, \gamma_0) - \delta_{0j} M(v, \gamma_0)] g_v(v \mid y_t) \\
+ \frac{h^{2p^*}}{(2p^*)!} \int \nabla_{v}^{2p^*} [M(v, \gamma_0) - \delta_{0j} M(\bar{v}, \gamma_0)] \cdot g_v(\bar{v} \mid y_t) \cdot z^{2p^*} K(z) \ dz \\
+ \frac{h^{2p^*-1}}{(2p^*-1)!} \int \nabla_{v}^{2p^*-1} [M(v, \gamma_0) - \delta_{0j} M(\bar{v}, \gamma_0)] \cdot \nabla_{v} g_v(\bar{v} \mid y_t) \cdot z^{2p^*} K(z) \ dz \\
+ \frac{h^{2p^*}}{(2p^*)!} \int [M(v, \gamma_0) - \delta_{0j} M(\bar{v}, \gamma_0)] \cdot \nabla_{v}^{2p^*} g_v(\bar{v} \mid y_t) \cdot z^{2p^*} K(z) \ dz
\]

where \( \bar{v} \in (v, v + hz) \) and \( g_v \) is a conditional density of \( v \) given \( y_t \). The first term becomes zero once it is averaged over \( j \) since

\[
\frac{1}{n} \sum_{j=1}^{n} (M(v, \gamma_0) - \delta_{0j} M(v, \gamma_0)) = M(v, \gamma_0) - M(v, \gamma_0) = M(v, \gamma_0) - M(v, \gamma_0).
\]

Additionally note that \(|\nabla_{v}^{k} M(v, \gamma_0)| = O(1)| for \( k = 0, ..., 2p^* \) by Assumption 4.

Provided

\[
|\nabla_{v}^{k} g_v(w_{js}, \bar{v} \mid y_t)| = O(1)
\]
for $k = 0, \ldots, 2p^*$, this induces the rest of terms to reduce to
\[
\frac{O(\hbar^{2p^*})}{(2p^*)!} \cdot \int |z^{2p^*}K(z)| \, dz = O(\hbar^{2p^*})
\]
since $\int z^{2p^*}K(z)dz = O(1)$. Then (A.31) becomes $\sqrt{T}O(h^{2p^*})$ which is $o(1)$ given a window parameter $r > \frac{1}{4p^*}$. Therefore, (A.32) is the only remaining term in (A.29) which is to be shown to converge to zero. Note that $u_{js}$ has unconditional mean zero and $\chi_{js}$ relies only on $x_{js} = \{R_{ms}, 1_{js}\}$ thus it is strictly exogenous. Furthermore, due to weak dependence in data, time dependence between $u_{js}$ and $y_t$ diminishes as the time lag becomes large which allows us to claim that $E[u_{js}\phi_{js} | y_t]$ converges to an unconditional expectation. This is in fact true and the result is given by McLeish (1975). Then the convergence of (A.32) follows because the series $\sum_{s\neq t} E[u_{js}\phi_{js} | y_t]$ in (A.29) explodes at a rate slower than $T$. In order to see this, we need to examine terms in which $s$ and $t$ are close together and the others where $s$ and $t$ are relatively far apart separately:
\[
\frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \sum_{s\neq t} \lambda_{it} E[u_{js}\phi_{js} | y_t]
= \frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \lambda_{it} \sum_{|s-t| \leq T^d} E[u_{js}\phi_{js} | y_t] + \frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \lambda_{it} \sum_{|s-t| > T^d} E[u_{js}\phi_{js} | y_t]
\]
for $d < 1$. Note first that
\[
\sum_{|s-t| \leq T^d} E[u_{js}\phi_{js} | y_t] = O(T^d)
\]
since $E[u_{js}\phi_{js} | y_t]$ is finite and the number of finite summands grows at a rate of $T^d$. Also $\frac{1}{T} \sum_{t=1}^{T} \lambda_{it}$ converges to $E[\lambda_{it}]$ which is bounded by law of large numbers attributed to McLeish (1975). Then it is immediate that the first term in (A.34) is $O(T^{d-\frac{1}{2}})$. Thus as long as $d$ lies inside the interval $(0, \frac{1}{2})$, the first term converges to zero. Next, the second term explodes at rate $\sqrt{T}$ unless $E[u_{js}\phi_{js} | y_t]$ converges to zero faster than $\sqrt{T}$. In order to resolve this problem, rewrite $E[u_{js}\phi_{js} | y_t] :$
\[
E[E[u_{js} | y_t, v_s] \cdot \phi_{js} | y_t] .
\]
(A.35)
Here the strategy is first to claim that this conditional expectation converges to an unconditional mean due to weak dependence then to show the convergence rate obtained from this is fast enough to dominate $\sqrt{T}$. In his 1975 paper, McLeish provides the rate of convergence which in fact depends on how fast the dependence among the observations at different times decays. Lemma (2.1) of the same paper gives

$$\| E [u_{js} | y_t] - E [u_{js}] \|_{p_1} = E [E [u_{js} | y_t]_{p_1}]^{\frac{1}{p_1}} \leq 2 (2^{\frac{1}{p_1}} + 1) \alpha_{p_1}^{\frac{1}{p_1} - \frac{1}{p_1 + \epsilon}} \| u_{js} \|_{p_1}$$

$$\equiv A \cdot |t - s|^{-c \left(\frac{1}{p_1} - \frac{1}{p_1 + \epsilon}\right)}.$$  \hspace{1cm} (A.36)

for $1 \leq p_1 \leq \infty$, $E |u_{js}|^{p_1 + \epsilon} < \infty$, $\epsilon > 0$ and for finite $A$. To use the above result, first apply Cauchy-Schwarz inequality to (A.35),

$$E [E [u_{js} | y_t, v_s] \cdot \phi_{js} | y_t]$$

$$\leq E |E [u_{js} | y_t, v_s]^2 | y_t|^{\frac{1}{2}} E [|\phi_{js}|^2 | y_t|^{\frac{1}{2}}.$$  \hspace{1cm}

Since $E [|\phi_{js}|^2 | y_t]$ and $E [|u_{js}|^{2 + \epsilon_1} | y_t]$ for $\epsilon_1 > 0$ are finite, (A.36) provides an upper bound of (A.35):

$$A' \cdot |t - s|^{-c \left(\frac{1}{2} - \frac{1}{2 + \epsilon_1}\right)}.$$  \hspace{1cm}

Now return to (A.34) and employ the above result to show the second part also vanishes away. Note that

$$\sum_{|s - t| > T^d} E [u_{js} \phi_{js} | y_t] = \sum_{|s - t| > T^d} A' \cdot |s - t|^{-c \left(\frac{1}{2} - \frac{1}{2 + \epsilon_1}\right)}$$

$$\leq \int_{T^d}^{\infty} A' \cdot |m|^{-c \left(\frac{1}{2} - \frac{1}{2 + \epsilon_1}\right)} dm.$$  \hspace{1cm}

If $1 - c \left(\frac{1}{2} - \frac{1}{2 + \epsilon_1}\right) < 0$ i.e. $c > 2 + \frac{1}{\epsilon_1}$, then the last term becomes $\frac{2}{c \cdot \frac{1}{2 + \epsilon_1} - 2} \cdot T^d(1 - c \left(\frac{1}{2} \cdot \frac{1}{2 + \epsilon_1}\right))$. Then

$$\frac{1}{\sqrt{T} (T - 1)} \sum_{t=1}^{T} \lambda_{it} \sum_{|s - t| > T^d} E [u_{js} \phi_{js} | y_t]$$

$$\leq \frac{1}{\sqrt{T} (T - 1)} \sum_{t=1}^{T} |\lambda_{it}| \cdot \left[ \frac{2}{c \cdot \frac{1}{2 + \epsilon_1} - 2} \cdot T^d(1 - c \left(\frac{1}{2} \cdot \frac{1}{2 + \epsilon_1}\right)) \right]$$
\[
\leq \frac{1}{T} \sum_{t=1}^{T} |\lambda_{it}| \cdot O\left(T^{-\frac{1}{2}+d\left(1-c\left(\frac{1}{2} \cdot \frac{\epsilon_1}{1+\epsilon_1}\right)\right)}\right).
\]

Again \(\frac{1}{T} \sum_{t=1}^{T} |\lambda_{it}|\) converges to finite \(E[|\lambda_{it}|]\) due to McLeish (1975)'s LLN. Let \(\epsilon_1 > 2s^* - 4\) (Assumption 1.(b)), then \(c > 2 + \frac{4}{\epsilon_1}\) which results in
\[
c\left(\frac{1}{2} \cdot \frac{\epsilon_1}{2+\epsilon_1}\right) - 1 > 0.
\]

Hence the first term of (A.34) goes to zero and the convergence rate is
\[
O\left(T^{-\frac{1}{2}+d\left(1-c\left(\frac{1}{2} \cdot \frac{\epsilon_1}{1+\epsilon_1}\right)\right)}\right).
\]

Both of the first and second term in (A.34) vanish with \(d \in (0, \frac{1}{2})\) but the former does at a slower rate than the latter therefore the convergence rate is determined by \(O(T^{d-\frac{1}{2}})\).

So far we have proved that (A.29) does not contribute to the limiting function for large \(T\). In the following argument, we will see that (A.32) converges to something simpler in structure to which we can apply central limit theorem. Before delving into the discussion, note that (A.29) is also divided into parts, which are examined in steps. As done in previous arguments, we isolate the part which incorporates with the error term \(u_{it}\) in (A.32) and study the remainder separately:

\[
\frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ E[\lambda_{js} | y_t, v_{js}] \left[ M(v_{js}; \gamma_0) - \delta_0, M(v_{it}; \gamma_0) \right] \phi_{it} | y_t \right] \cdot u_{it} (A.37)
\]

\[
-\frac{1}{\sqrt{T(T-1)}} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ E[\lambda_{js} | y_t, v_{js}] \phi_{it} | y_t \right] \cdot u_{it} \quad (A.38)
\]

By using a similar approach as in (A.31), we can show that (A.37) becomes \(\sqrt{T}O(h^{2p^*})\) which is again \(o(1)\) with \(r > \frac{1}{4p^*}\). Now in order to simplify (A.38), we again separate the summands into two parts based on the limiting form of \(E[\lambda_{js} | y_t, v_{js}]\) and the remainder:

\[
E\left[ E[\lambda_{js} | y_t, v_{js}] \phi_{it} | y_t \right]
\]
Then the first term becomes $O(E)$ by applying Lemma (2.1) of McLeish (1975). Then, summands and the second term is of the series. Thus for $d$ the lag expands following $T$ is measured by $T$. Thus the claim asserted above can be translated as: the weak time dependence which for the remainder:

$$E [E [\lambda_{js} | y_t, v_{js}] | \phi_{it} | y_t] + E [(E [\lambda_{js} | y_t, v_{js}] - E [\lambda_{js} | v_{js}]) \phi_{it} | y_t].$$

Here the first claim is that sample average of the remainder goes to zero faster than $\sqrt{T}$. This can be shown by using an approach similar to that used for (A.36), namely, by applying Lemma (2.1) of McLeish (1975). Then,

$$\|E [\lambda_{js} | y_t, v_{js}] - E [\lambda_{js} | v_{js}]\|_{p_2} \leq A \cdot |t - s|^{-c \left( \frac{1}{p_2} - \frac{1}{p_2 + \epsilon} \right)}$$

for $0 < p_2 < \infty$ and $\epsilon > 0$. This result and Cauchy-Schwarz inequality form a bound for the remainder:

$$E [(E [\lambda_{js} | y_t, v_{js}] - E [\lambda_{js} | v_{js}]) \phi_{it} | y_t] \leq E [(|E [\lambda_{js} | y_t, v_{js}] - E [\lambda_{js} | v_{js}]|^2 | y_t])^{\frac{1}{2}} \cdot E [\phi_{it}^2 | y_t]^{\frac{1}{2}} \leq A'' \cdot |t - s|^{-c \left( \frac{1}{2} - \frac{1}{p_2 + \epsilon} \right)}$$

for $E |\lambda_{js}|^{2 + \epsilon_2} < \infty$. Accordingly, (A.38) is bounded by

$$\frac{2}{\sqrt{T} (T - 1)} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E [E [\lambda_{js} | v_{js}] \phi_{it} | y_t] \cdot u_{it}$$

and

$$+ \frac{2}{\sqrt{T} (T - 1)} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} A'' \cdot |t - s|^{-c \left( \frac{1}{2} - \frac{1}{p_2 + \epsilon} \right)} \cdot u_{it}.$$ (A.39)

Thus the claim asserted above can be translated as: the weak time dependence which is measured by $A'' \cdot |t - s|^{-c \left( \frac{1}{2} - \frac{1}{p_2 + \epsilon} \right)}$ lets (A.40) decay fast enough to dominate $\sqrt{T}$ as the lag expands following $T$. In order to show this, we use the same approach applied to (A.34) and divide the terms into two parts again:

$$\frac{1}{\sqrt{T} (T - 1)} \sum_{t=1}^{T} u_{it} \left[ \sum_{|s - t| \leq T^d} A'' \cdot |t - s|^{-c \left( \frac{1}{2} - \frac{1}{p_2 + \epsilon} \right)} + \sum_{|s - t| > T^d} A'' \cdot |t - s|^{-c \left( \frac{1}{2} - \frac{1}{p_2 + \epsilon} \right)} \right].$$

Then the first term becomes $O \left( T^{d - \frac{\epsilon_2}{2}} \right)$ due to slowly expanding number of finite summands and the second term is $O \left( T^{-\frac{1}{2} + d \left( 1 - \frac{1}{2} - \frac{1}{p_2 + \epsilon} \right)} \right)$ due to the tail behavior of the series. Thus for $d \in \left( 1, \frac{1}{2} \right)$ and $\epsilon_2 > 2s^* - 4$, (A.40) converges to zero at the rate of $T^{d - \frac{\epsilon_2}{2}}$. Now the remaining task is to show that

$$\frac{\sqrt{T}}{T (T - 1)} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E [E [\lambda_{js} | v_{js}] \phi_{it} | y_t] \cdot u_{it}$$
can be rearranged into a simpler form. Again an approach similar to that applied to (A.31) provides an approximation for (A.39) :  

\[- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{n} \left( \frac{1}{n(T-1)} \sum_{j=1}^{n} \sum_{s \neq t}^{T} \delta_{0j} \cdot E \left[ \tau_{js} \chi_{js} \mid v_{js} = v_{it} \right] \right) \cdot u_{it} \]  

(A.41)

for each \( v_{it} \), the realization of \( v_{it} \). We have shown that (A.29) diminishes to zero and the conditional expectation in (A.30) has a limit as above. Finally, \( \sqrt{T}(N_1 + N_2) \) is approximated by

\[
\sqrt{T}(N_1 + \hat{U}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{n} \left( \tau_{it} \chi_{it} - \frac{1}{n(T-1)} \sum_{j=1}^{n} \sum_{s \neq t}^{T} \delta_{0j} \cdot E \left[ \tau_{js} \chi_{js} \mid v_{js} = v_{it} \right] \right) \cdot u_{it}.
\]

9Theorem 1.7 of Ibragimov (1962) provides the central limit theorem for the processes under weak time dependence. All the conditions asserted in the theorem are automatically satisfied by Assumption 1 and 2. The below states this theorem in terms of our notation.

**Theorem 17 (Ibragimov (1962) 1.7)** If the random process \( \{y_t\} \) is stationary and satisfies the strong mixing condition with

\[ E |\chi_{it}|^{2+\delta} \leq \infty, \quad E \left[ |\chi_{it}|^{2+\delta} \mid y_t \right] \leq \infty \quad \text{and} \quad E |u_{it}|^{2+\delta} \leq \infty \]

for some \( \delta > 0 \), and

\[ \sum_{m=1}^{n} \alpha_{m}^{\frac{\delta}{2+\delta}} < \infty, \]

then

\[ \sqrt{T}(N_1 + U_T) \overset{\mathcal{D}}{\longrightarrow} N(0, \Sigma_0) \]

where

\[ h_t(\gamma_0) \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ \tau_{it} \chi_{it} - \frac{1}{n(T-1)} \sum_{j=1}^{n} \sum_{s \neq t}^{T} \delta_{0j} \cdot E \left[ \tau_{js} \chi_{js} \mid v_{js} = v_{it} \right] \right] \cdot u_{it}, \]

\[ \Sigma_T(\gamma_0) \equiv \frac{1}{T} \sum_{t=1}^{T} \sum_{\ell=1}^{T} E \left[ h_t(\gamma_0) h_{\ell}(\gamma_0)' \right], \quad \text{and} \quad \Sigma_0 \equiv \lim_{T \to \infty} \Sigma_T(\gamma_0). \]
### A.4 Proof of Theorem 3

Note that \( h_t \) is a function of \( y_t \) thus a mixing process by the definition. This allows us to apply Newey and West (1987)'s result and to construct an estimator for covariance-variance matrix. Define

\[
\tilde{\Sigma}_T(\gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} h_t(\gamma)h_t(\gamma)' + \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{m_\Sigma(T)} h_t(\gamma)h_t(\gamma)'
\]

with bounded \( m_\Sigma \) which is an upper bound of the number of sample auto-covariances. Then by Theorem 2 of Newey and West (1987), \( \tilde{\Sigma}_T(\gamma) \) converges to \( \Sigma_T(\gamma_0) \) in probability. Next, replace all the functions and parameters in \( h_t(\gamma) \) with their estimators and define \( \hat{h}_t(\gamma) \):

\[
\hat{\Sigma}_T(\gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{h}_t(\gamma)\hat{h}_t(\gamma)' + \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{m_\Sigma(T)} \hat{h}_t(\gamma)\hat{h}_{t-m}(\gamma)'
\]

uniformly approximates \( \tilde{\Sigma}_T \) in probability. For the notational convenience, \( \hat{h}_t(\gamma) \) is first rearranged as:

\[
\hat{h}_t(\gamma) = \frac{1}{n(T-1)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{T} \sum_{a=1}^{n} \sum_{b\neq a}^{n} \left[ \hat{\tau}_{it}\hat{\chi}_{it} - \delta_j \hat{E} \left[ \hat{\tau}_{js}\hat{\chi}_{js} \mid \hat{v}_{it} = v_{it} \right] \right] \cdot \hat{u}_{it}
\]

where \( \hat{E} \) denotes the semiparametric conditional expectation estimator. Now, we conjecture that the estimator formulated by substituting function \( h_t \) in \( \tilde{\Sigma}_T \) by \( \hat{h}_t \):

\[
\hat{\Sigma}_T(\gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{h}_t(\gamma)\hat{h}_t(\gamma)' + \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{m_\Sigma(T)} \hat{h}_t(\gamma)\hat{h}_{t-m}(\gamma)'
\]

is uniformly approximates \( \tilde{\Sigma}_T \) in probability. For the notational convenience, \( \hat{h}_t \) is first rearranged as:

\[
\hat{h}_t(\gamma) = \sum_{m=0}^{m_\Sigma(T)} \langle \hat{\tau}_{it}\hat{\chi}_{it}\hat{u}_{it}\hat{\tau}_{k,t-m}\hat{\chi}_{k,t-m}\hat{u}_{k,t-m} \rangle
\]

where \( \langle \rangle \) denotes the sample mean operator. Rewrite the square and cross products of \( \hat{h}_t \) using the above expression and rearrange the terms:

\[
\hat{\Sigma}_T(\gamma) = \sum_{m=0}^{m_\Sigma(T)} \sum_{m=0}^{m_\Sigma(T)} \left[ \hat{\tau}_{it}\hat{\chi}_{it}\hat{u}_{it}\hat{\tau}_{k,t-m}\hat{\chi}_{k,t-m}\hat{u}_{k,t-m} \right]
\]

\[
+ \sum_{m=0}^{m_\Sigma(T)} \sum_{m=0}^{m_\Sigma(T)} \left[ \hat{\tau}_{it}\hat{\chi}_{it}\hat{u}_{it}\hat{\tau}_{cd}\hat{\chi}_{cd}\hat{u}_{k,t-m} \right] \cdot \delta_{k,t} K \left( \frac{v_{k,t-m} - v_{cd}}{h} \right)
\]

where

\[
\delta_{k,t} K \left( \frac{v_{k,t-m} - v_{cd}}{h} \right)
\]

is a kernel function.
Then we can study the convergence of each term to its true counterpart separately, for example,

\[
\sum_{m=0}^{m_{\Sigma}(T)} \left\langle \frac{1}{g(v_{it})} \tilde{\tau}_{ab} \tilde{\chi}_{ab} \tilde{u}_{it} \tilde{\tau}_{k,t-m} \tilde{\chi}_{k,t-m} \tilde{u}_{k,t-m} \frac{\delta_j}{h} K \left( \frac{v_{it} - v_{ab}}{h} \right) \right\rangle.
\]

However, note that the structures of the above 4 lines are essentially the same, furthermore, with \( K(\cdot) \) being bounded, the problem here is not different from what we study for the Hessian part in Lemma 8 and 10. More correctly, with \( m_{\Sigma}(T) \) being finite and fixed, the convergence is immediate by the arguments of Lemma 8 and 10.

What we must show is basically that the sample mean of the products of the estimators of functions such as \( \tau, \chi \) and \( u \) converges to its true counterpart. The strategy is to divide the gap between these two into sub-parts each of which accounts for the gap between individual function and its estimator, such as the discrepancy between \( \chi \) and \( \hat{\chi} \). Thus the uniform convergence in (A.42) and the rest of the terms of \( \hat{\Sigma}_T \) depend upon the uniform convergence of \( \hat{\tau}, \hat{\chi} \) and \( \hat{u} \). The only possible distinction from the Hessian part here arises when we let \( m_{\Sigma}(T) \) grow as the sample size increases. Recall that \( m_{\Sigma}(T) \) is the number of sample auto-covariances thus the time dependence in data becomes more persistent as \( m_{\Sigma}(T) \) gets larger. Hence (A.42) now depends on whether the rate of convergence of the summands dominates time dependence or not. Notice that the rate of convergence of the summands is mainly determined by that of the first derivative of the conditional expectation estimator. Therefore as long as \( m_{\Sigma}(T) \) increases slower than \( \sqrt{T}h^2 \), (A.42) can be shown resulting

\[
\left| \hat{\Sigma}_T(\gamma) - \bar{\Sigma}_T(\gamma) \right| \xrightarrow{p} 0.
\]
Table A.1: Rate Comparison

<table>
<thead>
<tr>
<th>S Term</th>
<th>( O \left( \frac{1}{Th} \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C Term</td>
<td>((h = T^{-r}))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( t ) and ( \ell )</th>
<th>((A.46))</th>
<th>((A.47))</th>
<th>((A.48))</th>
</tr>
</thead>
<tbody>
<tr>
<td>close</td>
<td>( O \left( \frac{1}{T^{2-3d}h^2} \right) )</td>
<td>( O \left( \frac{1}{T^{1-2d}h^2} \right) )</td>
<td>( O \left( \frac{1}{T^{1+d(\frac{5}{6}c-1)}h^2} \right) )</td>
</tr>
<tr>
<td></td>
<td>(i) ( \frac{2}{3}(1 - r) \geq d )</td>
<td>(ii) ( \frac{1}{2} - r \geq d )</td>
<td>(iii) ( d \geq \frac{2(2r - 1)}{c - 2} )</td>
</tr>
<tr>
<td>far</td>
<td>( O \left( \frac{1}{T^{1-2d}h^2} \right) )</td>
<td>( O \left( \frac{1}{T^{1+d(\frac{5}{6}c-2)}h^2} \right) )</td>
<td>( O \left( \frac{1}{T^{d(\frac{5}{6}c-1)}h^2} \right) )</td>
</tr>
<tr>
<td></td>
<td>(iv) ( \frac{1}{2} - r \geq d )</td>
<td>(v) ( d \geq \frac{2(2r - 1)}{c - 4} ) if ( c &gt; 4 )</td>
<td>(vi) ( d \geq \frac{4r}{c - 2} ) if ( c &lt; 4 )</td>
</tr>
</tbody>
</table>
A.5 Detailed Discussion for Cross Product Term in $R_1$

First, divide $C$ into two parts where $t$ and $\ell$ are close together and where they are far apart. In order to do this, define a set of time indices $A_t := \{ t' : |t - t'| \leq T^d \}$ where $d > 0$. Then for each $t$, $A_t$ indicates the area "close" to $t$, where "close" means within a range of $T^d$ time period. Notice that a set $A_t$ grows as $T$ increases but at a slower rate. Now based on $A_t$, $C$ can be rearranged as:

$$
\frac{1}{T} \sum_{t=1}^{T} \left[ \sum_{\ell \in A_t} E(u_t u_\ell b_t b_\ell) + \sum_{\ell \notin A_t} E(u_t u_\ell b_t b_\ell) \right]. \tag{A.43}
$$

Note that if we could find orders of magnitude for two expectations inside the bracket, then (A.43) can be expressed as $O(1)O(a) + O(T)O(b)$ for some $a$ and $b$. Also note that $b_t$ contains a summation inside, where we divide terms again into three parts. Again define $A_\ell := \{ \ell' : |\ell - \ell'| \leq T^d \}$ which indicates the area close to $\ell$ for each $\ell$ analogous to $A_t$. Then,

$$
b_t = \frac{1}{Th} \sum_{s \in A_t} w_s K \left( \frac{v_t - v_s}{h} \right) + \frac{1}{Th} \sum_{s \in A_t} w_s K \left( \frac{v_t - v_s}{h} \right)
+ \left[ \frac{1}{Th} \sum_{s \notin A_t \cup A_\ell} w_s K \left( \frac{v_t - v_s}{h} \right) - f_{t,0} \right]
= b_{1t} + b_{2t} + b_{3t}, \tag{A.44}
$$

Similarly the summation inside $b_\ell$ can be divided into tree parts:

$$
b_\ell = \frac{1}{Th} \sum_{q \in A_\ell} w_q K \left( \frac{v_\ell - v_q}{h} \right) + \frac{1}{Th} \sum_{q \in A_\ell} w_q K \left( \frac{v_\ell - v_q}{h} \right)
+ \left[ \frac{1}{Th} \sum_{q \notin A_t \cup A_\ell} w_q K \left( \frac{v_\ell - v_q}{h} \right) - f_{\ell,0} \right]
= b_{1\ell} + b_{2\ell} + b_{3\ell}. \tag{A.45}
$$

By using (A.44) and (A.45), the expectation in (A.43) can be rearranged as:

$$
E(u_t u_\ell b_t b_\ell)
= E(u_t u_\ell b_{1t} b_{1\ell}) + E(u_t u_\ell b_{1t} b_{2\ell}) + E(u_t u_\ell b_{2t} b_{1\ell}) + E(u_t u_\ell b_{2t} b_{2\ell}) \tag{A.46}
$$
\[ + E(u_t u_t b_2 b_3) + E(u_t u_t b_2 b_1) + E(u_t u_t b_3 b_2) \quad (A.47) \]
\[ + E(u_t u_t b_3 b_3) \quad (A.48) \]

Lemmas 18 - 24 argue that the expectations in (A.46), (A.47) and (A.48) converge to zero in both cases where \( t \) and \( \ell \) are far apart and they are close together. The resulted rates of convergence are exhibited in Table A.1. Note that (i) and (iv) are the same, the upper bound in (ii) is always lower than that in (i) as long as \( r > 0 \) and (iii) and (v) are not effective given \( 0 < d < 1 \). Thus the feasible range for \( d \) is determined only by (ii) and (vi). With \( s^* > 4 \), \( d \) lies roughly in the interval \((0, 0.5)\) which implies \( O \left( \frac{1}{T^{1-2d} h^2} \right) \) gives the slowest convergence rate among the six cases.

**Lemma 18** Suppose Assumption 1-5 hold. (i) If \( E(u_t u_t b_t b_\ell) \) is one of the expectations in (A.46), then
\[ E(u_t u_t b_t b_\ell) = O \left( \frac{1}{T^{2(1-d)} h^2} \right). \quad (A.49) \]
(ii) If \( E(u_t u_t b_\ell b_\ell) \) is one of the expectations in (A.47), then
\[ E(u_t u_t b_\ell b_\ell) = O \left( \frac{1}{T^{1-d} h^2} \right). \]

**Proof.** Note that in (A.46) and (A.47), either \( b_t \) or \( b_\ell \) (or both) has finite number of summands inside and we can use this fact to prove the expectations in these terms are bounded. The result will be repeatedly discussed in the following arguments. We will present a proof of the first case for (A.46) and (A.47) but discussion below can be applied to all other expectations in (A.46) and (A.47). Note that
\[
\begin{align*}
E(u_t u_t b_1 b_1) &= \frac{1}{T^2 h^2} \sum_{s \in A_t} \sum_{q \in A_t} E \left[ u_t u_\ell w_s w_q K \left( \frac{v_t - v_s}{h} \right) K \left( \frac{v_\ell - v_q}{h} \right) \right] \\
&\leq \frac{1}{T^2 h^2} \sum_{s \in A_t} \sum_{q \in A_t} E \left[ K \left( \frac{v_t - v_s}{h} \right) K \left( \frac{v_\ell - v_q}{h} \right) \right] E \left[ |u_t u_\ell w_s w_q| | \{v_i\}_{i=1}^T \right]
\end{align*}
\]
The last inequality is due to Cauchy-Schwarz inequality and we can show that each conditional expectation is bounded:

\[
E \left[ w_s^2 w_q^2 \mid \{v_i\}_{i=1}^{T} \right] = E \left[ w_s^2 \mid \{v_i\}_{i=1}^{T} \right] E \left[ w_q^2 \mid \{v_i\}_{i=1}^{T} \right] + \text{Cov} \left[ w_s^2, w_q^2 \mid \{v_i\}_{i=1}^{T} \right]
\]

\[
\leq E \left[ w_s^2 \mid \{v_i\}_{i=1}^{T} \right]^2 + \text{Var} \left[ w_s^2 \mid \{v_i\}_{i=1}^{T} \right] \leq C_0.
\]

The second inequality results from the stationarity and the third inequality holds by Assumption 2 (ii). Similarly we can show that \(E \left[ u_t^2 u_\ell^2 \mid \{v_i\}_{i=1}^{T} \right] \leq C_0\). Then (A.50) is bounded by

\[
\frac{1}{T^2 h^2} \sum_{s \in A_t} \sum_{q \in A_\ell} C_0^2 E \left[ K \left( \frac{v_t - v_s}{h} \right) K \left( \frac{v_\ell - v_q}{h} \right) \right] \leq \frac{1}{T^2(1-d) h^2} C_0^2 k^2 \quad \text{(by Assumption 3 (i))}
\]

\[
= O \left( \frac{1}{T^2(1-d) h^2} \right).
\]

Now consider the first case in (A.47):

\[
E(u_t u_\ell b_1 b_\ell) = E \left[ (u_t u_\ell) \sum_{s \in A_t} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) \left[ \sum_{q \notin A_t \cup A_\ell} \frac{w_q}{Th} K \left( \frac{v_\ell - v_q}{h} \right) - f_0 \right] \right].
\]

Then we can rearrange the above equation:

\[
\sum_{s \in A_t} \sum_{q \notin A_t \cup A_\ell} \frac{1}{T^2 h^2} E \left[ K \left( \frac{v_t - v_s}{h} \right) K \left( \frac{v_\ell - v_q}{h} \right) u_t u_\ell w_s w_q \right]
\]

\[
- \sum_{s \in A_t} \sum_{q \notin A_t \cup A_\ell} \frac{1}{T^2 h^2} E \left[ K \left( \frac{v_t - v_s}{h} \right) f_0 u_t u_\ell w_s \right].
\]

By using a similar argument discussed for (A.46),

\[
E \left[ K \left( \frac{v_t - v_s}{h} \right) K \left( \frac{v_\ell - v_q}{h} \right) u_t u_\ell w_s w_q \right] \leq C_0^2 k^2 \quad \text{and}
\]

\[
E \left[ K \left( \frac{v_t - v_s}{h} \right) f_0 u_t u_\ell w_s \right] \leq C_0^2 k^2.
\]
\[ E \left[ K \left( \frac{v_t - v_s}{h} \right) f_{0t} u_t \ell w_s \right] \leq C_0^2 k C_1. \]

Recall that \( f_{0\ell} \leq C_1 \) for some finite constant \( C_1 \) by Assumption 4, which results in the second inequality. Then,

\[
E(u_t u_\ell b_t b_\ell) \leq \frac{1}{T^2 h^2} \sum_{s \in A_t} \sum_{q \notin A_t \cup A_\ell} \left( \delta_0^2 C_0^2 k^2 + \delta_0 C_0^2 k C_1 \right) = O \left( \frac{1}{T^{1-d} h^2} \right).
\]

- Next three lemmas provide convergence result of (A.46), (A.47) and (A.48) when \( t \) and \( \ell \) are close together.

**Lemma 19** Suppose Assumption 1-5 hold and \( E(u_t u_\ell b_t b_\ell) \) be one of the terms in (A.46). If the lag between \( t \) and \( \ell \) is no greater than \( T^d \), then

\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} E(u_t u_\ell b_t b_\ell) = O \left( \frac{1}{T^{2-3d} h^2} \right).
\]

**Proof.** Note that (A.49) is true for any expectation in form of \( E(u_t u_\ell b_t b_\ell) \) for \( b_t \) and \( b_\ell \) that contain only finite elements in them. \( b_t b_\ell \) is in fact a sum over \( 2T^d \times 2T^d \) terms and the expectation of each term multiplied by \( u_t u_\ell \) is bounded which eventually hits \( \frac{1}{T^2 h^2} \). Then the conclusion follows, for example, as in the first case:

\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} E(u_t u_\ell b_t b_\ell) \leq \frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} O \left( \frac{1}{T^{2(1-d)} h^2} \right) = O \left( \frac{1}{T^{2-3d} h^2} \right).
\]

- **Lemma 20** Suppose Assumption 1-5 hold and \( E(u_t u_\ell b_t b_\ell) \) be one of the terms in (A.47). If the lag between \( t \) and \( \ell \) is no greater than \( T^d \), then

\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} E(u_t u_\ell b_t b_\ell) = O \left( \frac{1}{T^{1-2d} h^2} \right).
\]
Proof. The argument applied to (A.47) is the same as the one we use for (A.46) above except now either $b_t$ or $b_\ell$ (but not both) has $T - T^d$ terms in it which grows at the rate of $T$. Again if we show that the expectation of elements inside the summation of $b_t b_\ell$ multiplied by $u_t u_\ell$ are bounded, it follows that $E(u_t u_\ell b_t b_\ell)$ is $O\left(\frac{1}{T^{1-h^2}}\right)$. Comparing with the previous case, here we lose $\frac{1}{T^{1-h^2}}$ in the order and that is the price for allowing the number of terms grow at rate $T$ either in $b_t$ or $b_\ell$. Then, for example, the first case of (A.47) becomes:

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} E(u_t u_\ell b_t b_\ell) \leq \frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} O\left(\frac{1}{T^{1-dh^2}}\right) = O\left(\frac{1}{T^{1-2dh^2}}\right).$$

Lemma 21 Suppose Assumption 1-5 hold. If the lag between $t$ and $\ell$ is no greater than $T^d$, then

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} E(u_t u_\ell b_{3t} b_{3\ell}) = O\left(\frac{1}{T^{1+d\left(\frac{1}{2}c-1\right)h^2}}\right).$$

Proof. Now both $b_{3t}$ and $b_{3\ell}$ have sums over the number of terms growing at the rate of $T$. We cannot adhere to the strategy employed for the previous cases, since the sum of bounded elements grows at a rate of $T$ in $b_{3t}$ (and analogously in $b_{3\ell}$) which costs $\frac{1}{T^d}$ in it and we are left with nothing to compensate $h$ in the denominator. Hence the strategy here is to show that the speed at which $u_t u_\ell b_{3t} b_{3\ell}$ grows is suppressed by asymptotic uncorrelatedness of data enough to make its expectation go to zero. Note that $b_{3t}$ and $b_{3\ell}$ contain only those terms which have weaker time dependence with both $u_t$ and $u_\ell$. Under Assumption 1 we use an inequality attributed to Davydov (1968), which provides conditions for asymptotic uncorrelatedness for strong mixing processes. $E(u_t u_\ell b_{3t} b_{3\ell})$ can be decomposed into $E(u_t u_\ell) E(b_{3t} b_{3\ell})$ and a remainder term which goes to zero at a rate of $\frac{1}{T^{b\ell h^2}}$ for some $a$ and $b$ which will be specified.
later. Then the next step is to show \( E(u_t u_e) E(b_{3t} b_{3e}) \) converges to zero and whichever slower between the rates of \( E(u_t u_e) E(b_{3t} b_{3e}) \) and remainder will determine the speed of convergence for (A.48). \( \frac{1}{T} \sum_{t=1}^{T} E(u_t u_e b_{3t} b_{3e}) \) then becomes either \( O \left( \frac{1}{T^{a/3}} \right) \) or \( O \left( \frac{1}{T^{b/3}} \right) \) whichever is slower. Now a detailed argument follows. Note that \((s, q)\) and \((t, \ell)\) are now far apart. Hence by using Assumption 1 and Lemma 2.1 and following Corollary in Davydov (1968) \(^1\), we can make

\[
|E(u_t u_e b_{3t} b_{3e})| \\
\leq |E(u_t u_e) E(b_{3t} b_{3e})| + \sum_{s \notin \mathcal{A}_t \cup \mathcal{A}_e} \sum_{q \notin \mathcal{A}_t \cup \mathcal{A}_e} \frac{10A^{-\frac{1}{2}}D_1}{T^2 h^2} \sum_{t, \ell, s, q} m_1^{-\frac{1}{2}c} \left[ \sup_{t, \ell, s, q} \left[ E(|u_t u_e|^4) \right]^{\frac{1}{2}} \left[ E(|b_{3t} b_{3e}|^4) \right]^{\frac{1}{2}} \right]
\]

where \( m_1 = \min \{|t-s|, |t-q|, |s-q|, |t-s|, |s-q|\} \). Without loss of generality, assume that \( m_1 = |t-q| \) and let \( \sup_{t, \ell, s, q} \left[ E(|u_t u_e|^4) \right]^{\frac{1}{2}} \left[ E(|b_{3t} b_{3e}|^4) \right]^{\frac{1}{2}} =: D_1 \), then the above equation can be further rearranged:

\[
|E(u_t u_e b_{3t} b_{3e})| \\
\leq |E(u_t u_e) E(b_{3t} b_{3e})| + \sum_{s \notin \mathcal{A}_t \cup \mathcal{A}_e} \sum_{q \notin \mathcal{A}_t \cup \mathcal{A}_e} \frac{10A^{-\frac{1}{2}}D_1}{T^2 h^2} \sum_{t, \ell, s, q} m_1^{-\frac{1}{2}c} \\
\leq |E(u_t u_e) E(b_{3t} b_{3e})| + O(T) \cdot \frac{10A^{-\frac{1}{2}}D_1}{T^2 h^2} \cdot \frac{1}{T^{d(\frac{1}{2}c-1)} h^2} - 1 \\
\leq |E(u_t u_e) E(b_{3t} b_{3e})| + O \left( \frac{1}{T^{1+d(\frac{1}{2}c-1)}} \right). \tag{A.51}
\]

\(^1\)Davydov’s lemma is in fact applied to each term inside the summation of \( b_t \) and \( b_t \) multiplied by \( u_t \) and \( u_e \) but it can be expressed in terms of expectation of \( u_t, u_e, b_t \) and \( b_e \) as follows:

\[
|E(u_t u_e b_{3t} b_{3e}) - E(u_t u_e) E(b_{3t} b_{3e})| = \left| E \left( u_t u_e \sum_s Z_s \sum_q Z_q \right) - E(u_t u_e) E \left( \sum_s Z_s \sum_q Z_q \right) \right| \\
= \left| \sum_s \sum_q E(u_t u_e Z_s Z_q) - \sum_s \sum_q E(u_t u_e) E(Z_s Z_q) \right| \\
\leq \sum_s \sum_q |E(u_t u_e Z_s Z_q) - E(u_t u_e) E(Z_s Z_q)| \\
\leq \left[ \sup_{s, q} \left[ E(|u_t u_e|^4) \right]^{\frac{1}{2}} \left[ E(|Z_s Z_q|^4) \right]^{\frac{1}{2}} \right] \sum_s \sum_q \frac{A m^{-\frac{1}{2}c}}{(T-1)^2 h^2}. \tag{Assumption 9}
\]

where \( Z_s = \frac{W_s}{(T-1)h} K \left( \frac{v_t(a_0) - v_s(a_0)}{h} \right) - \frac{f(v_t(a_0))}{S}. \)
The second inequality uses the fact that
\[ \sum_{|t-q|>T^d} |t-q|^{-\frac{1}{c}} \leq \int_{T^d}^{\infty} x^{-\frac{1}{c}} \, dx \leq \frac{T^{-d(\frac{1}{c}-1)}}{\frac{1}{c} - 1} \]
for any \( t, c > 2 \) and \( T^d \geq 1 \). Now we need to show that \( E(u_tu_\ell)E(b_{3t}b_{3\ell}) \) converges to zero and specify its convergence rate. Having \( E(u_tu_\ell) \) bounded this can be done by showing that \( E(b_{3t}b_{3\ell}) \) vanishes as \( T \) grows. By using Cauchy-Schwarz inequality,
\[
|E(b_{3t}b_{3\ell})| \leq \sqrt{E(b_{3t}^2)}\sqrt{E(b_{3\ell}^2)}.
\]
Then it can be shown that \(|E(u_tu_\ell)E(b_{3t}b_{3\ell})|\) disappears if \( E(b_{3t}^2) \) tends to zero which in fact is given by Theorem 1 in Hansen (2008). From that result, we already know that \( E(b_{3t}^2) \) is \( O(\frac{1}{Th}) \) pointwisely. Almost analogous argument can be applied to \( E(b_{3\ell}^2) \), however, we need to consider that at most \( 2T \) terms are missing in \( b_{3\ell} \). Yet, this fact does not harm the convergence of \( E(b_{3\ell}^2) \) since the number of terms missing is finite. If we add and subtract those finite terms and rearrange \( E(b_{3t}^2) \):
\[
E \left[ \left( \sum_{s \in A_t \cup A_\ell} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) - f_{0t} \right)^2 \right]
\]
\[
= E \left[ \sum_{s=1}^{T} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) - f_{0t} \right]^2 \quad (A.52)
\]
\[
\leq E \left[ \sum_{s=1}^{T} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) - f_{0t} \right]^2 \quad (A.53)
\]
\[
+ E \left[ \left( \sum_{s\in A_t \cup A_\ell} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) \right)^2 \right] \quad (A.54)
\]
Note that we have \( \text{Var} \left( \sum_{s=1}^{T} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) - f_{0t} \right) = O \left( \frac{1}{Th} \right) \). Then by Chebyshev’s inequality we can show (A.52) goes to zero at the rate of \( \frac{1}{Th} \). Next, an argument used for (A.50) can be applied to (A.54),
\[
E \left[ \left( \sum_{s\in A_t \cup A_\ell} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) \right)^2 \right]
\]
\[
\leq \frac{\bar{k}^2}{T^2 h^2} \sum_{s \in A_t \cup A_T} \sum_{q \in A_t \cup A_T} |E(w_s w_q | \{v_i\}_{i=1}^T)| \\
\leq \frac{\bar{k}^2}{T^2 h^2} T^2 d \left[ \text{Var}(w_s | \{v_i\}_{i=1}^T) + E(w_s | \{v_i\}_{i=1}^T)^2 \right] = O \left( \frac{1}{T^2 - 2d h^2} \right).
\]

Then by (A.52) and (A.54), (A.53) becomes,
\[
E \left[ \sum_{s=1}^T \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) - \delta_0 f_{0t} \right] \left[ \sum_{s \in A_t \cup A_T} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) \right] \\
\leq \sqrt{E \left[ \sum_{s=1}^T \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) - \delta_0 f_{0t} \right]^2} \sqrt{E \left[ \sum_{s \in A_t \cup A_T} \frac{w_s}{Th} K \left( \frac{v_t - v_s}{h} \right) \right]^2} \\
\leq O \left( \frac{1}{T^{1/2} h^{3/2}} \right) \cdot O \left( \frac{1}{T^{1-d} h} \right).
\]

Since \(T^{1-d} h \to \infty\) must hold for (A.49), (A.52) generates the slowest convergence rate resulting \(|E(b_{3t}b_{3\ell})| \leq O \left( \frac{1}{Th} \right)\). By applying this to (A.51) concludes
\[
\frac{1}{T} \sum_{t=1}^T \sum_{\ell \in A_t} E(u_t u_{\ell} b_{3t}b_{3\ell}) = O(1) \left[ O \left( \frac{1}{Th} \right) + O \left( \frac{1}{T^{1+d(\frac{1}{2} \varepsilon - 1)} h^2} \right) \right] \\
= O \left( \frac{1}{T^{1+d(\frac{1}{2} \varepsilon - 1)} h^2} \right).
\]

The next following three lemmas argue about the same cases where \(t\) and \(\ell\) are far apart.

**Lemma 22** Suppose Assumption 1-5 hold and \(E(u_t u_{\ell} b_t b_{\ell})\) be one of the terms in (A.46). If the lag between \(t\) and \(\ell\) is greater than \(T^d\), then
\[
\frac{1}{T} \sum_{t=1}^T \sum_{\ell \in A_t \setminus A_T} E(u_t u_{\ell} b_t b_{\ell}) = O \left( \frac{1}{T^{1-2d} h^2} \right).
\]

**Proof.** Recall that \(b_t\) and \(b_{\ell}\) in (A.46) contain only finite number of summands inside. Following exactly the same argument in the case when \(t\) and \(\ell\) are close together, we can show that \(E(u_t u_{\ell} b_t b_{\ell})\) is \(O \left( \frac{1}{T^{2d} a_{ij} h^2} \right)\). However, \(\frac{1}{T} \sum_{t=1}^T \sum_{\ell \notin A_t} E(u_t u_{\ell} b_t b_{\ell})\)
becomes $O\left(\frac{1}{T^{1-2d^2}}\right)$ since we consume $\frac{1}{T}$ for the sum over the terms whose time indices are not close to $t$. For example, for the first case of (A.46):

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \notin A_t} E(u_t u_{\ell} b_{t} b_{\ell}) \leq O(T) O\left(\frac{1}{T^{2(1-d)h^2}}\right) \leq O\left(\frac{1}{T^{1-2d^2}h^2}\right).$$

Lemma 23 Suppose Assumption 1-5 hold and $E(u_t u_{\ell} b_{t} b_{\ell})$ be one of the terms in (A.47). If the lag between $t$ and $\ell$ is greater than $T^d$, then

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \notin A_t} E(u_t u_{\ell} b_{t} b_{\ell}) = O\left(\frac{1}{T^{1+d(\frac{1}{2}c-2)}h^2}\right).$$

Proof. For (A.47), we cannot follow the same argument as in the previous case since now $t$ and $\ell$ are far apart even though only one of $b_{t}$ and $b_{\ell}$ has a growing number of terms in it. Then again we need an asymptotic uncorrelatedness to make the argument work. Here $u_t$ and $u_{\ell}$ are far apart from each other and both of them are distant from either of $b_{t}$ or $b_{\ell}$ (not from both). Assume that it is $b_{\ell}$. By using Davydov’s inequality, $E(u_t u_{\ell} b_{t} b_{\ell})$ can be shown to go to $E(u_t u_{\ell} b_{t}) E(b_{\ell})$ at a rate of $\frac{1}{T^{ab}}$ for some $a$ and $b$ which will be specified later. Here we may show $E(u_t u_{\ell} b_{t}) E(b_{\ell})$ goes to zero by using the results from previous section that $E(b_{\ell})$ is $O\left(\frac{1}{\sqrt{T h}}\right)$ and that $E(u_t u_{\ell} b_{t})$ is bounded. However, the rate is not fast enough to deal with the number of terms which increases at the rate of $T$. Hence we apply Davydov’s inequality once again to $E(u_t u_{\ell} b_{t})$. It is plausible since $u_{t} b_{t}$ ($u_{\ell} b_{\ell}$) and $u_{t}$ ($u_{\ell}$) have weak correlation due to their time indices which are far apart. Having $E(u_{\ell}) = 0$, $E(u_{t} b_{t}) E(u_{\ell})$ becomes zero and the remaining term multiplied by $E(b_{t})$ goes to zero at the rate of $\frac{1}{T^{a'b}}$ for some $a'$ and $b'$. Finally the convergence rate is given by either $\frac{1}{T^{ahb}}$ or $\frac{1}{T^{a'b}}$ whichever slower. For more complete discussion, note that $s$ and $t$ are close but
both ℓ and q are now far from s and t. This again means that we can separate \( u_\ell b_{1t} \) either from \( u_\ell \) or \( b_{3t} \) by using Davydov’s lemma,

\[
|E(u_\ell u_\ell b_{1t} b_{3t})| \leq |E(u_\ell u_\ell b_{1t}) E(b_{3t})| + \sum_{s \in A_\ell \cup A_\ell} \sum_{q \notin A_\ell \cup A_\ell} 10 A^{-\frac{1}{2}} D_2 \frac{T^2}{T^2 h^2} m_1^{-\frac{1}{2}c}
\]

where \( m_1 = \min\{|t - q|, |\ell - q|, |s - q|\} \) and

\[
D_2 = \sup \left[ E(|u_\ell u_\ell b_{1t}|^4) \right] \frac{1}{2} \left[ E(|b_{3t}|^4) \right] \frac{1}{2}.
\]

Without loss of generality, assume that \( m_1 = |t - q| \), then the above equation can be further rearranged as

\[
|E(u_\ell u_\ell b_{1t} b_{3t})| \leq |E(u_\ell u_\ell b_{1t}) E(b_{3t})| + O \left( \frac{1}{T^2 h^2} \right).
\]

Notice that we can apply Davydov’s lemma again to \( |E(u_\ell u_\ell b_{1t}) E(b_{3t})| \) then by using \( E(u_\ell) = 0 \) and \( E(b_{3t}) \leq O \left( \frac{1}{\sqrt{T h}} \right) \),

\[
|E(u_\ell u_\ell b_{1t}) E(b_{3t})| \leq |E(u_\ell b_{1t}) E(u_\ell) E(b_{3t})| + |E(b_{3t})| \left[ \sum_{s \in A_\ell \cup A_\ell} 10 A^{-\frac{1}{2}} D_3 \frac{T^2}{T^2 h^2} m_2^{-\frac{1}{2}c} \right]
\]

where \( m_2 = \min\{|t - \ell|, |s - \ell|\} \) and \( D_3 = \sup \left[ E(|u_\ell b_{1t}|^4) \right] \frac{1}{2} \left[ E(|u_\ell|^4) \right] \frac{1}{2} \). Finally,

\[
\frac{1}{T} \sum_{t=1}^T \sum_{\ell \notin A_t} E(u_\ell u_\ell b_{1t} b_{3t}) \leq \frac{1}{T} \sum_{t=1}^T O \left( \frac{1}{\sqrt{T h}} \right) \left[ \sum_{s \in A_\ell \cup A_\ell} \sum_{q \notin A_\ell} 10 A^{-\frac{1}{2}} D_3 \frac{T^2}{T^2 h^2} m_2^{-\frac{1}{2}c} \right] + O(T) O \left( \frac{1}{T^{1+d(\frac{1}{2}c-2)} h^2} \right)
\]

\[
\leq O \left( \frac{1}{\sqrt{T h}} \right) \cdot O \left( \frac{1}{T^{1+d(\frac{1}{2}c-2)} h^2} \right) + O \left( \frac{1}{T^{1+d(\frac{1}{2}c-2)} h^2} \right)
\]

\[
\leq O \left( \frac{1}{T^{\frac{1}{2}+d(\frac{1}{2}c-2)} h^2} \right) + O \left( \frac{1}{T^{1+d(\frac{1}{2}c-2)} h^2} \right)
\]

\[
= O \left( \frac{1}{T^{1+d(\frac{1}{2}c-2)} h^2} \right).
\]
Lemma 24 Suppose Assumption 1-5 hold. If the lag between $t$ and $\ell$ is no greater than $T^d$, then
\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \in A_t} E(u_t u_\ell b_{3t} b_{3\ell}) = O \left( \frac{1}{T^{d(\frac{1}{2}c-1)}h^2} \right).
\]

Proof. Here in (A.48), $u_t$, $u_\ell$, $b_{3t}$ and $b_{3\ell}$ are all far apart from each other in terms of time indices, thus we have more freedom in using asymptotic uncorrelatedness property. Applying the inequality for $u_t u_\ell$ and $b_{3t} b_{3\ell}$ results that $E(u_t u_\ell b_{3t} b_{3\ell})$ goes to $E(u_t u_\ell)E(b_{3t} b_{3\ell})$ by the remaining term being $O \left( \frac{1}{T^{3a^2}} \right)$ for some $a$ and $b$ which will be specified in the following discussion. By using the same argument again for $u_t$ and $u_\ell$, we can show that $E(u_t)E(u_\ell)E(b_{3t} b_{3\ell})$ becomes exactly zero. The remaining term from the second application is multiplied by $E(b_{3t} b_{3\ell})$ which itself goes to zero at the rate of $\frac{1}{T^h}$. For more complete discussion, first note that now $t$, $s$, $\ell$ and $q$ are far apart. Then,
\[
|E(u_t u_\ell b_{3t} b_{3\ell})| \leq |E(u_t u_\ell) E(b_{3t} b_{3\ell})| + \sum_{s \notin A_t \cup A_\ell} \sum_{q \notin A_t \cup A_\ell} \frac{10A^{-\frac{1}{2}}D_4}{T^2 h^2} m_1^{-\frac{1}{2}c}
\]
where $m_1 = \min\{|t-s|, |t-q|, |\ell-s|, |\ell-q|\}$ and
\[
D_4 = \sup \left[ E(|u_t u_\ell|^4) \right]^{\frac{1}{2}} \left[ E(|b_{3t} b_{3\ell}|^4) \right]^{\frac{1}{2}}.
\]
Without loss of generality, assume that $m_1 = |t-q|$,
\[
|E(u_t u_\ell b_{3t} b_{3\ell})| \leq |E(u_t u_\ell) E(b_{3t} b_{3\ell})| + \sum_{s \notin A_t \cup A_\ell} \frac{10A^{-\frac{1}{2}}D_4}{T^2 h^2} m_1^{-\frac{1}{2}c}
\]
\[
\leq |E(u_t u_\ell) E(b_{3t} b_{3\ell})| + O \left( \frac{1}{T^{1+d(\frac{1}{2}c-1)h^2}} \right).
\]
Notice that we can apply Davydov’s lemma again to $|E(u_t u_\ell) E(b_{3t} b_{3\ell})|$. Then by using $E(u_t) = 0$ and $E(b_{3t}^2) \leq O \left( \frac{1}{T^h} \right)$,
\[
|E(u_t u_\ell) E(b_{3t} b_{3\ell})| \leq |E(u_t) E(u_\ell) E(b_{3t} b_{3\ell})| + |E(b_{3t} b_{3\ell})| \left[ \sum_{q \notin A_t \cup A_\ell} \frac{10A^{-\frac{1}{2}}D_5}{T^2 h^2} m_2^{-\frac{1}{2}c} \right]
\]

\[ \leq O\left(\frac{1}{Th}\right) \left[ \sum_{q \notin A_t \cup A_\ell} \frac{10A^{-\frac{1}{2}}D_5}{T^2h^2} m_2^{-\frac{1}{2}c} \right] \]

where \( m_2 = |t - \ell| \) and \( D_5 = \sup [E(|u_t|^4)]^{\frac{1}{4}} [E(|u_\ell|^4)]^{\frac{1}{4}} \). Finally,

\[ \frac{1}{T} \sum_{t=1}^{T} \sum_{\ell \notin A_t} E(u_t u_\ell b_{3t} b_{3\ell}) \]

\[ \leq \frac{1}{T} \sum_{t=1}^{T} O\left(\frac{1}{Th}\right) \left[ \sum_{q \notin A_t \cup A_\ell} \frac{10A^{-\frac{1}{2}}D_5}{T^2h^2} \sum_{\ell \notin A_t} m_2^{-\frac{1}{2}c} \right] \]

\[ + O(T) O\left(\frac{1}{T^{1+d(\frac{1}{2}c-1)}h^2}\right) \]

\[ \leq O\left(\frac{1}{T^{2+d(\frac{1}{2}c-1)}h^2}\right) + O\left(\frac{1}{T^{d(\frac{1}{2}c-1)}h^2}\right) = O\left(\frac{1}{T^{d(\frac{1}{2}c-1)}h^2}\right). \]

\[ \square \]

A.6 Range for Bandwidth and Mixing Rate

At the beginning of the discussion for an asymptotic distribution, we expanded the problem as in (2.5), (2.6) and (2.7) and study them separately. The arguments for these problems are essentially in the same structure: the estimated function is replaced by a true one then we are left to show that the gap between these two functions converges to zero at a certain rate. Note that the window parameter, \( h = (n \times T)^{-r} \), is involved in convergence rates required at each step. Thus we must choose an \( r \) which works for all the cases. Note also that we use higher order kernel for the estimation which requires \( \sqrt{n \times T} \cdot h^{p^*} \) converge to zero. Since the convergence rate of bias in estimator is determined once \( p^* \) is specified, we need to find candidates for \( p^* \) first.

Note that the optimal window obtained for consistency is no longer applicable since the convergence of the terms in (2.5), (2.6) rely on uniform convergence of the first and second derivative of density estimator. Taking derivative of density estimator piles up \( h \) in the denominator, which induces slower convergence rates. Theorem 2 in
Hansen (2008) which will give us the uniform convergence rates of first and second derivative requires \( \sqrt{n \times T} \cdot h^3 \to \infty \) and \((n \times T) \cdot h^{1/\theta^*} \to \infty \) where

\[
\theta^* = \frac{c - 2 - \frac{1+c}{s-1}}{c + 2 - \frac{1+c}{s-1}}.
\]

The only convergence rate that matters in (2.7) is from cross product term (A.19) for which \((n \times T) \cdot h^{1/\theta^*} \to \infty \) must be satisfied. Thus if \( \sqrt{n \times T} \cdot h^{\theta^*} \) converges to zero, then \( p^* \) should be

\[
\max \left\{ 3, \frac{1}{2} \frac{(c-2)(s^*-1) - 1 + c}{(c+2)(s^*-1) - 1 + c} \frac{1}{1 - 2d} \right\}.
\]

As is clear in the requirements above, \( c \) must be specified first to obtain desirable \( r \) and \( p^* \). This can be done by studying conditions of theorems required for the arguments in this paper, which are presented below:

- Hansen (2008) : \( c > \frac{2s^* - 1}{s^* - 2} \),
- Dehling and Wendler (2010) : \( c > \frac{(3s^* + 1)(s^* - 2) + 5s^* + 2}{2s^*(s^* - 2)} \),
- Sun and Cordero (2005) : \( c > \frac{2s^* - 2}{s^* - 2} \)

for \( s^* > 0 \) such that

\[
E \left[ |\rho(y_t, y_s)|^{\bar{s}} \right] \leq O(1).
\]

Suppose \( s^* = 4 \) and \( \bar{s} = 4 \), Hansen (2008) gives the highest lower bound which is 3.5.

We use a mean squared error argument to show \( R_1 \) converges to zero. In order to do this, we need an asymptotic uncorrelatedness of data and study terms whose time periods are not far apart and the others whose time periods are remote separately. The criteria determining proximity of observations also grows as sample size gets larger, however we need it explode more slowly than the sample size does. The rate depends on bandwidth and the speed of diminishing time dependence which is given by the exponent \( c \) in (1.6) as above. Once we obtained values for \( r \) and \( c \), we can apply them to the convergence rates of individual terms in \( R_1 \) Big \( O \) terms in Table A.1 indicate the convergence rates and the inequalities below specify the ranges of \( d \)
which induce each term go to zero. With $r$ and $c$ computed as above, we can find the common area for $d$ which works through all the cases. For example, when we set $c = 4.5$, $r$ lies in the interval $(0.033, 0.036)$ and the range of $d$ roughly follows as $(0.05, 0.46)$.

Finally, even though none of $r$, $d$ and $c$ are involved in the argument, we need to specify a parameter $M_{\tau}$ to make $R_2$ converge to zero. $M_{\tau}$ is the order of approximation to the trimming function around true quantiles. Since the convergence of $R_2$ is closely related with the convergence of quantile estimators, possible range for $M_{\tau}$ also depends on the rate at which quantile estimator converges its true value. Given the range for $M_{\tau}$ as in (A.20), Figure A.1 illustrates the lower bounds when $r_{\pi}$ is slightly lower than $\frac{1}{2}$. When our sample size exceeds 500, Taylor expansion of order 3 would suffice for $R_2$ to converge.