NONPARAMETRIC PRICING KERNEL MODELS

by

Linman Sun

A dissertation submitted to the faculty of
the University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2011

Approved by:

____

Dr. Zongwu Cai

____

Dr. Weihua Zhou

____

Dr. Jiancheng Jiang

____

Dr. Lisa S.Walker

ABSTRACT

Linman Sun. Nonparametric Pricing Kernel Models.
(Under the direction of DR. ZONGWU CAI)

The capital asset pricing model (CAPM) and the arbitrage asset pricing theory (APT) have been the cornerstone in theoretical and empirical finance for the recent few decades. The classical CAPM usually assumes a simple and stable linear relationship between an asset's systematic risk and its expected return. However, this simple relationship assumption has been challenged and rejected by several recent studies based on empirical evidences of time variation in betas and expected returns.

It is well documented that large pricing errors could be due to the linear approach used in a nonlinear model and treating a non-linear relationship as a linear could lead to serious prediction problems in estimation. To overcome these problems, in the first part of this dissertation I would like to investigate a general nonparametric asset pricing model to avoid functional form misspecification of betas, risk premia, and the stochastic discount factor by considering estimating unknown functional involved in the nonparametric pricing kernel. To estimate the nonparametric functionals, I propose a new nonparametric estimation procedure, termed as nonparametric generalized estimation equations (NPGEE), which combines the local linear fitting and the generalized estimation equations. I establish the asymptotic properties of the resulting estimator. Also, as a rule of thumb, I propose a data-driven method to select the bandwidth and provide a consistent estimate of the asymptotic variance.

The nonparametric method may provide a useful insight for further parametric fitting, while parametric models for time-varying betas can be most efficient if the underlying betas are specified. However, a misspecification may cause serious bias

and model constraints may distort the betas in local area. Hence, to test whether the pricing kernel model has some specific parametric form becomes essentially important. In the second part of this dissertation, I propose a consistent nonparametric testing procedure to test whether the model is correctly specified and I establish the asymptotic properties of the test statistic using a U-statistic technique.

Finite sample results are investigated using Monte Carlo simulation studies in order to show the usefulness of the estimation method and the test statistics. The empirical applications using CRSP monthly returns are also implemented to illustrate our proposed models and methods.

ACKNOWLEDGEMENTS

First of all, I would like to express my deep appreciation to my advisor Dr. Zongwu Cai. His excellent guidance, caring and patience played an important role throughout my Ph.D. study. Besides providing me with an effective guidance on my research, his enthusiasm to work also inspired me to push myself for better. This dissertation would not be possible without his guidance and encouragement.

Appreciation also extends to Dr. Jiancheng Jiang, Dr. Weihua Zhou, and Dr. Lisa S. Walker for serving as very valuable members in my advisory committee.

Finally, my special gratitude goes to my p parents Zhenying Sun and Kunyu Yang for their endless love, encouragement, and support throughout my life. To them I dedicate this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

The capital asset pricing (CAP) model and the arbitrage asset pricing theory (APT) have been the cornerstone in theoretical and empirical finance for the recent few decades. The classical CAPM usually assumes a simple and stable linear relationship between an asset's systematic risk and its expected return; see the books by Campbell, Lo and MacKinlay (1997) and Cochrane (2001) for details. However, this simple relationship assumption has been challenged and rejected by several recent studies based on empirical evidences of time variation in betas and expected returns (as well as return volatilities). As with other models, one considers the conditional CAP or nonlinear APT models with time-varying betas to characterize the time variation in betas and risk premia.

The recent work, to name just a few, includes Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), Cochrane (1996), Jaganathan and Wang (1996, 2002), Reyes (1999), Ferson and Harvey (1991, 1993, 1998, 1999), Cho and Engle (2000), Wang (2002, 2003), Akdeniz, Altay-Salih and Caner (2003), Ang and Liu (2004), Fraser, Hamelink, Hoesli and MacGregor (2004), and the references therein. In particular, Fama and French (1992, 1993, 1995) used some instrumental (fundamental) variables like book-to-market equity ratio and market equity, as proxies for some unidentified risk factors to explain the time variation in returns, whereas Ferson (1989), Harvey (1989), Ferson and Harvey (1991, 1993, 1998, 1999), Ferson and Korajczyk (1995), and Jaganathan and Wang (1996) concluded that beta and market risk premium vary over time. Therefore, a static CAPM should incorporate time variation in beta in the model.

Although there is a vast amount of empirical evidences on time variation in betas and risk premia, there is no theoretical guidance on how betas and risk premia vary with time or variables that represent conditioning information.

Many recent studies focus on modelling the variation in betas using continuous approximation and under the theoretical framework of the conditional CAPM; see, for example, Cochrane (1996), Jaganathan and Wang (1996, 2002), Wang (2002, 2003) and Ang and Liu (2004) and the references therein. Recently, Ghysels (1998) discussed the problem in detail and stressed the impact of misspecification of beta risk dynamics on inference and estimation. Also, he argued that betas change through time very slowly and linear factor models like the conditional CAPM may have a tendency to overstate the time variation. Furthermore, he showed that among several well-known time-varying beta models, a serious misspecification produces time variation in beta that is highly volatile and leads to large pricing errors. Finally, he concluded that it is better to use a static CAPM in pricing when one does not have a proper model to capture time variation in betas correctly.

It is well documented that large pricing errors may be due to the linear approach used in a nonlinear model and treating a nonlinear relationship as a linear can lead to serious prediction problems in estimation.

To overcome these problems, some nonlinear models have been considered in the recent literature. For example, Bansal, Hsieh and Viswanathan (1993) and Bansal and Viswanathan (1993) were the first to advocate the idea of a flexible stochastic discount factor (SDF) model in empirical asset pricing and they focused on nonlinear arbitrage pricing theory models by assuming that the SDF is a nonlinear function of a few of state variables. Also, Akdeniz, Altay-Salih and Caner (2003) tested for the existence of significant evidence of nonlinearity in the time series relationship of industry returns with market returns using the heteroskedasticity consistent Lagrange multiplier test of Hansen (1996) under the framework of the threshold model and they found that there exists statistically significant nonlinearity in this relationship with respect to real interest rates. Furthermore, under the mean-covariance efficiency framework, Wang (2002, 2003) explored a nonparametric form of the SDF model and conducted a simple test based on the nonparametric pricing errors.

Gourieroux and Monfort (2007) considered a class of nonlinear parametric and semiparametric SDF models for derivative pricing by assuming that the stochastic discount factors are exponential-affine functions of underlying state variable. In particular, they discussed the conditionally Gaussian framework and introduced semiparametric pricing methods for models with path dependent drift and volatility.

A nonparametric modeling is appealing in these situations. One of the advantages for nonparametric modeling is that no or little restrictive prior information on betas and pricing kernel is needed. Moreover, it may provide useful insight for further parametric fitting. Parametric models for time-varying betas and nonlinear pricing kernel can be most efficient if the underlying models are correctly specified. However, a misspecification may cause serious bias and model constraints may distort the betas in local area.

In the following section, I give a brief introduction for famous asset pricing models.

1.2 Asset Pricing Models

Over the past decades, many studies have been conducted to examine the performance of SDF approach for econometric evaluation of asset-pricing models and CAP models in expected returns. As Jagannathan and Wang (2002) demonstrated, a classical beta method or CAPM can be expressed as a SDF form. Also, as Cochrane (2001) pointed out, a SDF method is sufficiently general that it can be used for analysis of linear as well as nonlinear asset-pricing models, including pricing models for derivative securities. Now I give a brief review about the CAP model and SDF form.

1.2.1 CAP Model

The basic theorem of capital asset pricing model (CAPM) and portfolio selection problem were proposed by Markovitz (1959). Investors would optimally hold a mean-variance efficient portfolio which is a portfolio with the highest expected return for a given level of variance. It was shown by Sharper (1964) and Lintner (1965a, 1965b) that without market frictions, if all investors have homogeneous expectations and optimally hold mean-variance efficient portfolio, then the market portfolio also becomes a mean-variance efficient portfolio.

The Sharpe-Lintner version of the CAPM can be expressed as the following statistical model:

$$E(R_i) = R_f + \beta_{im}(E(R_m) - R_f); \qquad \beta_{im} = \frac{Cov(R_i, R_m)}{Var(R_m)}, \quad 1 \leq i \leq N,$$

where R_i is the ith asset return and R_m is the market portfolio return. Also, one can express CAPM model in terms of excess returns $r_i = R_i - R_f$ and $r_m = R_m - R_f$, where R_f is the return on the risk-free asset,

$$E(r_i) = \frac{E(r_m)}{Var(r_m)} Cov(r_i, r_m). \tag{1.1}$$

Efficient-set mathematics plays an important role in the analysis of pricing models.

Portfolio p is the minimum-variance portfolio of all portfolios with means return μ_p if its portfolio weight vector is the solution to the following constrained optimization:

$$\min_{\omega} \omega' \Omega \omega$$

subject to $\omega'\mu = \mu_p$ and $\omega'\iota = 1$, where μ and Ω is the mean and covariance matrix of the N risky assets and ω is the vector of portfolio weights summing to unity.

For any risky portfolio R_p, one can calculate its Sharpe ratio defined as the mean excess return divided by the standard deviation of return

$$sr_p = \frac{\mu_p - R_f}{\sigma_p}.$$

Testing the mean-variance efficiency of a given portfolio can also be tested as whether Sharpe ratio of that portfolio is the maximum of the set of Sharpe ratios of all possible portfolios.

Empirical test of Sharpe-Litner CAPM usually focuses on the following implications: (1) The intercept is zero and the regression intercepts may be viewed as the pricing errors; (2) β captures the cross-sectional variation of expected excess returns; and (3) The market risk premium $E(r_m)$ is positive. The key testable implication of the CAPM is the first one which means the market portfolio of risky assets is a mean-variance efficient portfolio. One can run N time-series regressions:

$$R_{it} = \alpha_i + \beta_{im} R_{mt} + e_{it}, \quad i = 1, \cdots, N, \quad 1, \cdots, T,$$

where R_{it} is the ith risky asset and R_{mt} is the market portfolio. By using t-test, one can check whether the pricing error α_i is zero individually, and also one can use the following Wald-type χ^2 test discussed by Cochrane (2001) to test the pricing errors α_i are jointly aero,

$$T\left[1 + (\frac{\hat{\mu}_m}{\hat{\sigma}_m})^2\right]^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} \sim \chi^2_N,$$

where $\hat{\Sigma}$ is the residual covariance matrix, and $\hat{\mu}_m$ and σ_m are the mean and the standard deviation of R_{mt}, respectively. Under the normality assumption, a finite-sample $F-$test for the hypothesis that the pricing errors α_i are jointly zero:

$$\frac{T-N-1}{N}\left[1+(\frac{\hat{\mu}_m}{\hat{\sigma}_m})^2\right]^{-1}\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}\sim F_{N,T-N-1}.$$

Consider excess returns model of Sharpe and Lintner

$$E(r_i)=\beta_{im}\lambda, \quad \text{and} \quad E(r_m)=\beta_{im}\lambda,$$

where λ is the factor risk premium with only factor without intercept in the cross-sectional regression. Then, one could test whether all pricing errors are zero with the test statistic:

$$\hat{\alpha}'Var(\hat{\alpha})^{-1}\hat{\alpha}\sim\chi^2_{N-K}.$$

In the early years, CAPM was largely positive reporting the evidence of mean-variance efficiency of the market portfolio. However, in anomalies literature, less favorable evidence for the CAPM started to appear. Contrary to the prediction of the CAPM, the firm characteristics such as size, earning yield effect, leverage, ratio of a firm's book value of equity to its market value and ratio of earning to price are very important to predicting the asset return.

1.2.2 Valuation Theory

It is common in the literature to use a stochastic process to measure the probability of risky events which might occur over time. Meanwhile, financial security payoffs are functions of these events. Valuation theory or asset pricing theory describes how uncertainty evolves over time and try to figure out today's value of future, uncertain cash flows; see Duffie (1996). There are two general approaches to the valuation problem, no-arbitrage approach and equilibrium approach being complementary. SDF itself is a very nice equilibrium pricing approach.

1.2.3 Stochastic Discount Factor

In modern finance and economics, the stochastic discount factor model is rapidly emerging as the most popular way to price assets. Most existing asset pricing methods can be shown to be specific versions of SDF. For example, CAPM and the general equilibrium consumption-based inter-temporal capital asset pricing model (CCAPM) of Rubinsteain (1976) and Lucas (1978). Pricing kernel can be viewed as a mathematical term which represents an operator. The purpose of stochastic discount factor is to include adjustments for financial risk. Notice that the connection between SDF and pricing kernel is very strong and that two concepts are often used interchangeably.

SDF is a very nice equilibrium pricing approach. In financial economics, risk is measured by covariances. CAPM is the most obvious example, while SDF is much more general than CAPM. SDF essentially defines what risk is. Formulating term-structure models in terms of the SDF proves particularly useful when one wants to model interest-rate dynamics in the actual world.

The SDF asset pricing model is based on the following simple idea

$$P_t = E_t \sum_{s=1}^{T-t} \left[m_{t,t+s} \delta_{t+s} \right],$$

where P_t is the price of the asset in period t, δ_{t+s} is the pay-off of the asset in period $t+s$, m_{t+s} is the discount factor for period $t+s$ ($0 \leq m_{t+s} \leq 1$). By the valuation theorem, P_t is essentially the current value of the period $t+s$ income δ_{t+s} which is in general a random variable. The discount factor is a stochastic variable and is also called the pricing kernel. By no-arbitrage condition, one can derive the recursive representation $m_{t,t+2} = m_{t,t+1} m_{t+1,t+2}$. To be more generally, one has

$$m_{t,t+s} = \Pi_{k=1}^{s} m_{t+k-1,t+k}$$

Moreover, by iterated exsections,

$$P_t = E_t[m_{t+1}(P_{t+1} + \delta_{t+1})], \qquad m_{t+1} \equiv m_{t,t+1}$$

In term of the asset's gross return $R_{t+1} = \frac{P_{t+1}+\delta_{t+1}}{P_t}$, one has

$$E[m_{t+1}\, R_{t+1} \,|\, \Omega_t] = 1. \tag{1.2}$$

Suppose there are several assets, $i = 1, 2, \cdots, I$, by subtracting the risk-free return $R_{f,t}$, one can get

$$E[m_{t+1}\, r_{i,t+1} \,|\, \Omega_t] = 0, \tag{1.3}$$

where Ω_t denotes the information set at time t, m_{t+1} is the SDF or the marginal rate of substitution (MRS) or the pricing kernel, and $r_{i,t} = R_{i,t} - R_{f,t}$ is the excess return on the i-th asset or portfolio. This very simplified version of the SDF framework is universal and admits a basic pricing representation such as Sharper-Lintner CAPM. As $E_t(m_{t+1})E_t(r_{i,t+1}) = -Cov_t(m_{t+1}, r_{i,t+1})$, it is easy to see from the above equation that assets with returns whose covariance is positive with the SDF will pay a negative risk premium.

There are lots of nice properties for SDF. For example, one can do capital budgeting and pricing using SDF. If a project pays a random amount δ_{t+1} and costs P_t, the investment return thus is $(1 + r_{t+1} = \delta_{t+1}/P_t)$. By no-arbitrage assumption.

$$E_t[m_{t+1}(r_{t+1} - r_t)] = 0 \Rightarrow E_t(r_{t+1}) = r_t - Cov_t(m_{t+1}, r_{t+1})/E_t(m_{t+1}).$$

Thus, the expected return can be defined based on the investment as $1 + E_t r_{t+1} \equiv$

$E_t\delta_{t+1}/P_t$. The project is therefore priced as the expected discounted present value

$$P_t = \frac{E_t(\delta_{t+1})}{1 + r_t - Cov_t(r_{t+1}, m_{t+1})/E_t(m_{t+1})}.$$

In real world, since risk-aversion investors hope to be compensated for taking on risk, people use nonnegative risk premium to measure this. Risk premium essentially is extra return over the risk-free rate equals to the price of risk multiplied by the quantity of risk. If market portfolio is taken as the benchmark risky portfolio, one has

$$\frac{E(R_j - r_f)}{E(R_m - r_f)} = \frac{Cov(m, r_j)}{Cov(m, R_m)} \equiv \beta_j \rightarrow E(R_j - r_f) = \beta_j E(R_m - r_f).$$

If pricing kernel $m(r_m)$ has a linear form as $m(r_m) = a + br_m$ through the β-representation, one can derive CAPM

$$\beta_j = Cov(R_m, R_j)/Var(R_m); \qquad ER_j = r_f + \beta_j E(R_m - r_f),$$

where β_j measures the quantity of risk in asset j and the excess return $E(R_m - r_f)$ is the market price of risk.

1.2.4 Conditional Asset Pricing Models

In empirical finance, different models impose different constraints on the SDF. Particularly, the SDF is usually assumed to be a linear function of factors in various applications. Furthermore, when the SDF is fully parameterized such as linear form, the general method of moments (GMM) of Hansen (1982) can be used to estimate parameters and test the model; see Campbell, Lo and MacKinlay (1997) and Cochrane (2001) for details.

Because of different purposes in applications, different forms of (1.3) have been imposed in the finance literature. For example, Bansal, Hsieh and Viswanathan (1993) and Bansal and Viswanathan (1993) were the pioneers to propose nonlinear

APT models in empirical asset pricing by assuming that the SDF or MRS is a nonlinear function of a few of state variables. Under some assumptions, Bansal, Hsieh and Viswanathan (1993) re-expressed (1.3) as

$$E\left[\prod_{r=1}^{s} G(p_{t+r}^{b}) X_i(t, t+s) \,|\, \Omega_t\right] = \pi(X_i(t, t+s)); \tag{1.4}$$

see (8) in Bansal, Hsieh and Viswanathan (1993), where p_{t+1}^b is the low-dimensional ex post payoffs or prices at $t+1$ that do not contain non-factor risk, $G(\cdot)$ is an unknown function, and $X_i(t, t+s)$ is the payoff of the ith asset at time $t+s$ that has price $\pi(X_i(t, t+s))$ at time t (the maturity of the i-th payoff is s periods ahead). Here, the low-dimensional function $G(p_{t+r}^b)$ is the relevant pricing kernel implied by the nonlinear APT. In contrast to most APT models, as pointed out by Bansal and Viswanathan (1993), this pricing kernel can price dynamic trading strategies.

Indeed, equation (1.4) can be derived recursively by using

$$E\left[G(P_{t+1}^{b}) X_i(t, t+1) | \Omega_t\right] = \pi(X_i(t, t+1))$$

along with the law of iterated expectations; see (7) in Bansal, Hsieh and Viswanathan (1993). Moreover, equation (1.4) leads to a nonlinear arbitrage pricing kernel with nonnegativity restriction on the pricing kernel. To estimate the nonlinear model, Bansal, Hsieh and Visvanathan (1993) did not impose the no-arbitrage condition, however, they proposed the orthogonality conditions by the payoffs in the nonlinear APT as follows

$$E\left[\left(\prod_{r=1}^{s} G(p_{t+r}^{b}) X_i(t, t+s) - 1\right) Z_t\right] = 0, \tag{1.5}$$

where Z_t is an instrument that belongs to the information set Ω_t; see (7) in Bansal and Viswanathan (1993) or (14) in Bansal, Hsieh and Viswanathan (1993). While Bansal, Hsieh and Viswanathan (1993) suggested using the polynomial expansion to approximate it and then applied the GMM of Hansen (1982) for estimating and testing, Bansal and Viswanathan (1993) used neural networks to approximate the unknown pricing kernel. In addition to estimating the nonlinear model, Bansal, Hsieh and Viswanathan (1993) estimated the following conditional linear model which is not nested in the nonlinear model,

$$E\left[\left(\prod_{r=1}^{s}\eta_{t+r}^{T}p_{t+r+1}^{b}\right)X_i(t,t+s)Z_t - Z_t\right] = 0, \tag{1.6}$$

and they suggested estimating the conditional weights $\{\eta_{kt}\}$ in the above equation using a nonparametric method. As suggested by Bansal, Hsieh and Viswanathan (1993), the conditional weights, η_{kt}, are nonparametrically estimated by $\eta_{kt} = \eta_k(Z_{1t}) = \lambda_k^T Z_{1t} = \sum_{l=1}^{L}\lambda_{kl}Z_{1tl}$, where Z_{1t} might be exactly the same conditioning variables that are used as instruments; here, L is the number of instruments used in estimation. As the number of conditioning variables increases to infinity, we use all the relevant conditional information and this estimate of the conditional weight converges to the true conditional weight. Thus, this approach provides asymptotically consistent estimates without imposing the usual restrictive parametrization on the conditional mean process and the conditional covariance process of the (factor) payoffs.

As pointed out by Wang (2003), although the aforementioned approach is intuitive and general, one of shortcomings is that it is difficult to obtain the distribution theory and the effective assessment of finite sample performance. Instead of considering the nonparametric pricing kernel, Harvey (1991) focused on the nonlinear parametric model for conditional CAPM and used a set of moment conditions suitable for GMM estimation of parameters involved. More precisely, Harvey (1991) used conditional asset pricing restrictions that conditionally expected return on an asset is proportional to its covariance with the market portfolio, see Sharpe (1964) and Lintner (1965),

$$E[r_{j,t+1}|\Omega_t] = \frac{E[r_{m,t+1}|\Omega_t]}{\text{Var}(r_{m,t+1})}\text{Cov}(r_{j,t+1}, r_{m,t+1}|\Omega_t), \tag{1.7}$$

where $r_{j,t}$ is the return on jth equity from time t to $t+1$ in excess of a risk-free return and $r_{m,t+1}$ is the excess return on the market portfolio. Equation (1.7) is the so-called mean-variance efficient condition. Also, Hervey (1991) specified the model for the first conditional moments by assuming that

$$E[r_{j,t+1}|Z_t] = \delta_j^\top Z_t, \quad j = 1, \cdots, N, \text{ and } m, \tag{1.8}$$

where N is the number of assets, and Hervey (1991) showed that by plugging equation (1.8) to (1.7) and rewriting (1.7), one can obtain a set of moment conditions suitable for GMM estimation of $\delta^\top = (\delta_1, \cdots, \delta_N)$ and δ_m as follows:

$$E\begin{bmatrix} \begin{matrix} r_{t+1} - \delta Z_t \\ r_{m,t+1} - \delta_m^\top Z_t \\ u_{m,t+1}^2 \delta Z_t - u_{m,t+1} u_{t+1} \delta_m^T Z_t \end{matrix} & \otimes Z_t \end{bmatrix} = 0, \tag{1.9}$$

where $u_{j,t+1} = r_{j,t+1} - \delta_j^\top Z_t$, $u_{t+1} = (u_{1,t+1}, \cdots, u_{N,t+1})^\top$, and $\otimes$ is the Kronecker product of two matrices. Furthermore, Ferson and Harvey (1993) suggested another similar specification for the conditional CAPM, assuming time-varying betas as $\beta_{it} = \beta_{ic}^\top Z_t$, then the moment conditions become

$$E\begin{bmatrix} \begin{matrix} r_{t+1} - \delta Z_t \\ r_{m,t+1} - \delta_m^\top Z_t \\ \delta Z_t - \beta_c Z_t Z_t^\top \delta_m \end{matrix} & \otimes Z_t \end{bmatrix} = 0. \tag{1.10}$$

Moreover, Jagannathan and Wang (1996) examined the ability of the conditional

CAPM to explain the cross-sectional variation in average returns on a large collection of stock portfolios,

$$E[R_{i,t+1}|\Omega_t] = \gamma_{0,t} + \gamma_{1,t}\beta_{i,t}, \tag{1.11}$$

where $\beta_{i,t}$ is the conditional beta of asset i, defined as

$$\beta_{i,t} = \text{Cov}(R_{i,t+1}, R_{m,t+1}|\Omega_t)/\text{Var}(R_{m,t+1}|\Omega_t), \tag{1.12}$$

where $R_{i,t}$ is the gross (one plus the rate of) return on asset i in period t, $R_{m,t}$ is the gross return on the aggregate wealth portfolio of all assets in the economy in period t, $\gamma_{0,t}$ is the conditional expected return on a "zero-beta" portfolio, and $\gamma_{1,t}$ is the conditional market risk premium; see equations (2) and (3) in Jagannathan and Wang (1996). As pointed out by Hansen and Richard (1987) and Jagannathan and Wang (1996, 2002), the conditional CAPM given in (1.11) can be rewritten in terms of the conditional stochastic discount factor representation,

$$E[R_{i,t+1}m_{t+1}|\Omega_t] = 1;$$

see (26) in Jagannathan and Wang (1996), where m_{t+1} is generally referred to as SDF, defined as

$$m_{t+1} = \kappa_{0,t} + \kappa_{1,t}R_{m,t+1}$$

with

$$\kappa_{0,t} = \frac{1}{\gamma_{0,t}} + \left[\frac{\gamma_{1,t}}{\gamma_{0,t}\text{Var}(R_{m,t+1|\Omega_t})}\right]E(R_{m,t+1}|\Omega_t), \quad \text{and} \quad \kappa_{1,t} = -\frac{\gamma_{1,t}}{\gamma_{0,t}\text{Var}(r_{m,t+1}|\Omega_t)}.$$

Clearly, both $\kappa_{0,t}$ and $\kappa_{1,t}$ are a nonlinear function of state (conditioning) variables.

Furthermore, Ghysels (1998) tried to detect whether the beta risk is inherently misspecified and he found that pricing errors with constant traditional beta models are smaller than those with conditional CAPM. Therefore, Ghysels (1998) argued that based on evaluation of conditional asset pricing models, misidentification of functional forms is of first-order importance. To deal with this problem, Wang (2003) studied the nonparametric conditional CAPM and gave an explicit expression for the nonparametric form of conditional CAPM for the excess return. First, Wang (2003) started from the following mean-variance efficient model

$$E(r_{i,t+1}|\Omega_t) = E(r_{p,t+1}|\Omega_t)\frac{\mathrm{Cov}(r_{i,t+1}, r_{p,t+1}|\Omega_t)}{\mathrm{Var}(r_{p,t+1}|\Omega_t)} \tag{1.13}$$

$$\Longleftrightarrow E(r_{i,t+1}|\Omega_t) - E(r_{p,t+1}|\Omega_t)\frac{E(r_{i,t+1}r_{p,t+1}|\Omega_t)}{E(r_{p,t+1}^2|\Omega_t)} = E(m_{t+1}r_{i,t+1}|\Omega_t), \tag{1.14}$$

where $m_{t+1} = 1 - b(Z_t)\, r_{p,t+1}$, Z_t is an $L \times 1$ vector of conditioning variables from Ω_t, $b(z) = E(r_{p,t+1}|Z_t = z)/E(r_{p,t+1}^2|Z_t = z)$ is an unknown function, and $r_{p,t+1}$ is the return on the market portfolio in excess of the riskless rate. Equation (1.13) is the CAPM beta-pricing equation which leads to so called "cross-moment" representation in equation (1.14). Since the functional form of $b(\cdot)$ is unknown, Wang (2003) suggested estimating $b(\cdot)$ by using the Nadaraya-Watson method to two regression functions $E(r_{p,t+1}|Z_t = z)$ and $E(r_{p,t+1}^2|Z_t = z)$, respectively. Also, he conducted a simple nonparametric test about the pricing error. Furthermore, Wang (2003) extended this setting to multifactor models by allowing $b(\cdot)$ to change over time; that is, $b(Z_t) = b(t)$.

1.2.5 Flexible SDF Models

Recently, some more flexible SDF models have been studied by several authors. For example, Gourieroux and Monfort (2007) considered the problem of derivative pricing when the stochastic discount factors can be written under an exponential-affine form

$$m_{t+1} = \exp(\alpha_t^\top r_{t+1} + \beta_t), \tag{1.15}$$

where coefficients α_t and β_t are a function of history of $r_{t+1} = (r_{1,t+1}, ...r_{N,t+1})^\top$, which is a vector of geometric returns of the N risky assets. Clearly, the SDF specified in (1.15) is a nonlinear function of conditioning variables. Also, they suggested that the conditional historical distribution of the return is defined by means of its Laplace transform as

$$E[\exp(u^\top r_{t+1})|r_t] = \exp(\psi_t(u;\theta))$$

for some $\psi_t(u;\theta)$, which is indeed the conditional historical log-Laplace transformation (moment generating function). See Table 1 in Gourieroux and Monfort (2007) for the explicit expression of $\psi_t(u;\theta)$ for some specific examples. Thus, one can get $N+1$ restrictions on the SDF and historical distribution as follows

$$\begin{cases} E(m_{t+1}|r_t) = 1 \\ E[m_{t+1}\frac{p_{j,t+1}}{p_{j,t}}|r_t] = E[m_{t+1}\exp(r_{j,t+1})|r_t] = 1, \quad j = 1, \ldots, N, \end{cases}$$

where $p_{j,t}$ is the price of asset j and $r_{j,t} = \log(p_{j,t}/p_{j,t-1})$ is the log return of asset j,

$$\Longleftrightarrow \begin{cases} E[\exp(\alpha_t^\top r_{t+1} + \beta_t)|r_t] = 1 \\ E[\exp(\alpha_t^\top r_{t+1} + e_j^\top r_{t+1} + \beta_t)|r_t] = 1, \quad j = 1, \ldots, N, \end{cases}$$

where $e_j = (0, \cdots, 0, 1, 0, \cdots, 0)^\top$, with 1 as component of order j. Then, the system of $N+1$ equations generally admits a unique solution:

$$\Longleftrightarrow \begin{cases} \beta_t = -\psi_t(\alpha_t;\theta) \\ \psi_t(\alpha_t + e_j;\theta) - \psi_t(\alpha_t;\theta) = 0, \quad j = 1, \ldots, N. \end{cases}$$

For details, see Section 3.2 in Gourieroux and Monfort (2007). Furthermore, Gourieroux and Monfort (2007) provided some examples of stochastic discount factors which are exponential-affine functions of underlying state variables. One of those examples is consumption based CAPM at equilibrium,

$$p_t = E_t \left[p_{t+1} \frac{q_t}{q_{t+1}} \delta \frac{\frac{dU}{dc}(C_{t+1})}{\frac{dU}{dc}(C_t)} \right],$$

where $U(\cdot)$ is the utility function, δ is the infratemporal psychological discount rate, p_t is the vector of prices of financial assets, q_t is the price of the consumption good, and C_t is the quantity consumed at data t. Clearly, different choice of the utility function produces a different SDF. For example, for a power utility function $U(c) = c^{\gamma+1}/(\gamma+1)$, the SDF based on CAPM has the following form

$$m_{t+1} = \frac{q_t}{q_{t+1}} \delta \left(\frac{C_{t+1}}{C_t} \right)^{\gamma} = \exp \left[\log(\delta) - \log(q_{t+1}/q_t) + \gamma \log(C_{t+1}/C_t) \right],$$

which is actually an exponential-affine function of parameters γ, and $\log(\delta)$. For the constant absolute risk aversion (CARA) utility function $U(c) = -\exp(-Ac)/A$, the SDF is

$$m_{t+1} = \exp \left[\log(\delta) - \log(q_{t+1}/q_t) - A(C_{t+1} - C_t) \right],$$

which is an exponential-affine form. Thus, the foregoing examples imply that the choice of a power or CARA utility function is equivalent to the selection of an appropriate (parameter free) transformation of the consumption as state variable. For more examples, see Gourieroux and Monfort (2007).

1.3 Overview

In the first part of this dissertation, Chapter 2, I first describe a general nonparametric asset pricing model to avoid functional form misspecification of betas, risk premia, and the stochastic discount factor. I propose a new nonparametric estimation procedure to estimate unknown functional involved in the pricing kernel and derive the asymptotic properties of the proposed nonparametric estimator.

Furthermore, a simple bandwidth selector is suggested and a consistent estimate of the asymptotic variance is provided. Results based on the Monte Carlo simulation study and a real example are reported in Section 2.5 to illustrate the finite sample performance.

The nonparametric method may provide a useful insight for further parametric fitting. Parametric models for time-varying betas can be most efficient if the underlying betas are specified. Hence, to test whether the SDF model has a linear structure and whether some parametric form is correct is essentially important. In the second part of this dissertation, Chapter 3 I propose a consistent nonparametric testing procedure to test whether the model is correctly specified under a U-statistic framework. I adopt general GMM (Hansen 1982) method to estimate the assumed functional form inside SDF. Under fairly general stationarity, continuity, and the moment condition that the expectations of the pricing errors delivered by SDF equal to zero, the estimate inside SDF is consistent. An efficient and feasible estimation procedure is suggested and its asymptotic behavior is studied. Furthermore, to test a misspecification of functional forms, in Section 3.2, a nonparametric consistent test is proposed to test the pricing error using a U-statistic technique. Also, I establish the asymptotic properties of the test statistic. Finally, in Section 3.4, finite sample properties of the proposed estimators and testing procedures are investigated under both null and alternative hypothesis by the Monte Carlo simulations and the empirical examples.

CHAPTER 2: NONPARAMETRIC ASSET PRICING MODELS

In this chapter, first, I consider a general nonlinear pricing kernel model and propose a new nonparametric estimation procedure by combining local polynomial estimation technique and generalized estimation equations, termed as *nonparametric generalized estimation equations* (NPGEE). Secondly, I establish the asymptotic consistency and normality of the proposed nonparametric estimator. Moreover, I propose a rule of thumb method based on data-driven fashion to select a bandwidth and provide a consistent estimate for the asymptotic variance. Finally, finite sample properties of the proposed estimators are investigated by Monte Carlo simulation study and an empirical study.

2.1 The Model

To combine the models studied by Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), Ghysels (1998), Jagannathan and Wang (1996, 2002), Wang (2002, 2003), and some other models in the finance literature under a very general framework, I assume that the nonlinear pricing kernel has the form as $m_{t+1} = 1 - m(Z_t) r_{p,t+1}$, where $m(\cdot)$ is unspecified. My approach focuses on estimating the following nonparametric APT model

$$E[\{1 - m(Z_t)\, r_{p,t+1}\}\; r_{i,t+1} \,|\, \Omega_t] = 0, \tag{2.1}$$

where $m(\cdot)$ is an unknown function of Z_t, Z_t is an $L \times 1$ vector of conditioning variables from Ω_t, $r_{i,t+1}$ is the return on the asset of portfolio in excess of the risk free rate, and $r_{p,t+1}$ is the excess return on benchmark portfolio. Indeed, (2.1) can be regarded as a conditional moment (orthogonal) condition, and it, unlike Wang (2002, 2003) and others, is unnecessary to require the mean-variance efficiency. Hence, our interest is to identify and to estimate the nonlinear function $m(z)$. Clearly, an alternative expression for (2.1) when $m(\cdot)$ is a scalar function

$$m(Z_t) = \frac{E(r_{i,t+1}|Z_t)}{E(r_{p,t+1}r_{i,t+1}|Z_t)}, \tag{2.2}$$

and under the mean-variance efficiency, $m(Z_t)$ reduces to

$$b(Z_t) \equiv E(r_{p,t+1}|Z_t)/E(r_{p,t+1}^2|Z_t)$$

which was discussed by Wang (2002, 2003) in detail; see (1.14). Therefore, $m(Z_t) = b(Z_t)$ is equivalent to the mean-variance efficiency. In other words, testing the mean-variance efficiency is equivalent to testing the hypothesis $H_0 : m(\cdot) = b(\cdot)$.

Remark 1: (Extension to multiple market portfolios and multifactor models). It is easy to extend the model in (2.1) to cover multiple market portfolios. In such a case, the $r_{p,t+1}$ should be a vector. Then, the model (2.1) becomes

$$E[\{1 - m(Z_t)^\top r_{p,t+1}\}\, r_{i,t+1} \,|\, \Omega_t] = 0. \tag{2.3}$$

Moreover our model can be used in the case that a parametric structure is proposed for excess returns on the benchmark portfolio in terms of important factors. For example, in the famous Fama and French (1993)'s three-factor model, $r_{p,t+1}$ can be expressed as

$$\begin{aligned} r_{p,t+1} &= MKT_{t+1} + \theta_1 SMB_{t+1} + \theta_2 HML_{t+1} \\ &= \begin{pmatrix} 1 \\ \theta_1 \\ \theta_2 \end{pmatrix}^{\top} \begin{pmatrix} MKT_{t+1} \\ SMB_{t+1} \\ HML_{t+1} \end{pmatrix} \equiv \theta^{\top} r_{mf,t+1}. \end{aligned}$$

Then, the model becomes

$$E[\{1 - m^*(Z_t)^{\top}\, r_{mf,t+1}\}\; r_{i,t+1} \,|\, \Omega_t] = 0, \tag{2.4}$$

where $m^*(Z_t) = m(Z_t) * \theta(t)$. In our model θ is then allowed to vary over time and can be fully nonparametric.

The modeling procedure and its econometric theory developed in the next two sections for a single market portfolio in the model (2.1) continue to hold for the models in (2.3) and (2.4), and the details are omitted due to the similarity. Notice that unfortunately, the simple expression for the nonparametric pricing kernel in (2.2) does not hold for these cases.

2.2 Nonparametric Estimation Procedure

To ease notation, our focus in this section is only on the model (2.1) with a single market portfolio. Let I_t be a $q \times 1$ $(q \geq L)$ vector of conditioning variables from Ω_t, including Z_t, satisfying the following orthogonal condition

$$E[\{1 - m(Z_t) r_{p,t+1}\}\; r_{i,t+1} \,|\, I_t] = 0, \tag{2.5}$$

which can be regarded as an approximation of (2.1). It follows from the orthogonality condition in (2.5) that, for any vector function $Q(I_t) \equiv Q_t$ with a dimension d_q

specified later, we have

$$E\left[Q_t\left\{1-m(Z_t)r_{p,t+1}\right\}\, r_{i,t+1}\,|\, I_t\right]=0, \tag{2.6}$$

and its sample version is

$$\frac{1}{T}\sum_{t=1}^{T} Q_t\left\{1-m(Z_t)r_{p,t+1}\right\}\, r_{i,t+1}=0. \tag{2.7}$$

Therefore, this provides an estimation approach similar to the generalized method of moment of Hansen (1982) for parametric models and the estimation equations in Cai (2003) for nonparametric models. I propose a new nonparametric estimation procedure to combine the orthogonality conditions given in (2.5) with the local linear fitting scheme of Fan and Gijbels (1996) to estimate the unknown function $m(\cdot)$. This nonparametric estimation approach is termed as the nonparametric generalized estimation equations (NPGEE).

It is well known in the literature (see, e.g., Fan and Gijbels, 1996) that local linear fitting has several nice properties, over the classical Nadaraya-Watson (local constant) method, such as high statistical efficiency in an asymptotic minimax sense, design-adaptation, and automatic edge correction. I estimate $m(\cdot)$ using local linear fitting from observations $\{(r_{i,t+1},\, r_{p,t+1},\, Z_t)\}_{t=1}^{T}$. I assume throughout that $m(\cdot)$ is twice continuously differentiable. Then, for a given point z_0 and for $\{Z_t\}$ in a neighborhood of z_0, by the Taylor expansion, $m(Z_t)$ is approximated by a linear function $a+b^\top(Z_t-z_0)$ with $a=m(z_0)$ and $b=m'(z_0)$ (the derivative of $m(z)$), so that model (2.2) is approximated by the working orthogonality condition

$$E[Q_t\left\{1-(a+b^\top(Z_t-z_0))r_{p,t+1}\right\}\, r_{i,t+1}\,|\, Z_t]\approx 0. \tag{2.8}$$

Therefore, for $\{Z_t\}$ in a neighborhood of z_0, the orthogonality conditions in (2.5) can be approximated by the following locally weighted orthogonality conditions

$$\sum_{t=1}^{T} Q_t \left[1 - (a + b^\top (Z_t - z_0)) r_{p.t+1}\right] r_{i,t+1} K_h(Z_t - z_0) = 0, \tag{2.9}$$

where $K_h(\cdot) = h^{-L} K(\cdot/h)$, $K(\cdot)$ is a kernel function in R^L, and $h = h_T > 0$ is a bandwidth, which controls the amount of smoothing used in the estimation. (2.9) can be viewed as a generalization of the nonparametric estimation equations in Cai (2003) and the locally weighted version of (9.2.29) in Hamilton (1994, p.243) or (14.2.20) in Hamilton (1994, p.419) for parametric IV models. To ensure that the equations in (2.9) have a unique solution, the dimension of $Q(\cdot)$ must satisfy that $d_q \geq L + 1$ since the number of parameters in (2.9) is $L + 1$. Therefore, solving the above equations leads to the NPGEE estimate of $m(z_0)$, denoted by $\hat{m}(z_0)$, and the NPGEE estimate of $m'(z_0)$, denoted by $\hat{m}'(z_0)$; that is,

$$\begin{pmatrix} \hat{m}(z_0) \\ \hat{m}'(z_0) \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (S_T^\top S_T)^{-1} S_T^\top L_T, \tag{2.10}$$

where with $Q_t^* = \begin{pmatrix} 1 \\ Z_t - z_0 \end{pmatrix}$,

$$S_T = \frac{1}{T} \sum_{t=1}^{T} Q_t \, Q_t^{*\top} \, r_{p,t+1} K_h(Z_t - z_0) \, r_{i,t+1} \quad \text{and} \quad L_T = \frac{1}{T} \sum_{t=1}^{T} Q_t \, K_h(Z_t - z_0) \, r_{i,t+1}.$$

When $d_q = L + 1$ and S_T is nonsingular, $\begin{pmatrix} \hat{m}(z_0) \\ \hat{m}'(z_0) \end{pmatrix}$ becomes $S_T^{-1} L_T$. Clearly, (2.10) provides a formula for computational implementation, which can be carried out by any standard statistical package.

I now turn to the choice of $Q(Z_t)$ in (2.9). Motivated by the estimation equations in Cai (2003) and following a similar idea in Cai and Li (2008), I choose Q_t as

$$Q_t = Q_t^*; \tag{2.11}$$

see Remark 5 later for more discussion. Then, $\begin{pmatrix} \hat{m}(z_0) \\ \hat{m}'(z_0) \end{pmatrix}$ becomes $S_T^{-1} L_T$. Finally, notice that the method proposed in Cai (2003) can be regarded as a special case of the aforementioned NPGEE estimation procedure.

2.3 Distribution Theory

In this subsection, I discuss the large sample theory for the proposed estimator based on the nonparametric generalized estimation equations. Let $e_t = e_{i,t+1} = m_{t+1}\, r_{i,t+1} = [1 - m(Z_t) r_{p,t+1}]\, r_{i,t+1}$, which is called the pricing error in the finance literature.

2.3.1 Assumption

Assumption A:

A1. $\{Z_t, r_{i,t+1}, r_{p,t+1}, e_t\}$ is a strictly stationary α-mixing process with the mixing coefficient satisfying $\alpha(t) = O(t^{-\tau})$, where $\tau = (2+\delta)(1+\delta)/\delta$, for some $\delta > 0$. Also, assume that $E(r_{p,t+1}) < \infty$, $E(r_{i,t+1}) < \infty$, and $E(r_{i,t+1}^2 r_{p,t+1}^2) < \infty$.

A2. **(i)** Assume that for each t and s, and $\sup_{z_1,z_2} |E(e_t\, e_s | Z_s = z_1, Z_t = z_2)| < \infty$.

(ii) Define $M(z) = E(r_{p,t+1} r_{i,t+1} | Z_t = z)$ and $\sigma_0^2(z) = E(e_t^2 | Z_t = z)$. Assume that $m(\cdot)$ and $M(\cdot)$ are twice differentiable, and $\sigma_0^2(\cdot)$ is continuous. Furthermore, assume that $\sigma_0^2(z)$ and $M(z)$ are positive for all z.

(iii) $\sigma_0^2(z)$ satisfy Lipschitz conditions. There exists some $\delta > 0$, there exists some $\delta > 0$ such that $E\{|e_t|^{2+\delta} | Z_t = z\}$ is continuous at z_0.

(iv) Assume that for all τ, $f_\tau(\cdot, \cdot)$ exists and satisfies the Lipschitz condition, where $f_\tau(\cdot, \cdot)$ is the joint probability density function of Z_1 and Z_τ. Also, assume that the marginal density function $f(z)$ of Z_t is continuous.

A3. The kernel $K(\cdot)$ is symmetric, bounded and compactly supported.

A4. $h \to 0$ and $Th^L \to \infty$ as $T \to \infty$.

A5. $Th^{L[1+2/(1+\delta)]} \to \infty$.

Remark 2: (Discussion of conditions). A similar discussion of the foregoing assumptions has been given by Cai (2003) and Cai and Li (2008). Assumption A1 requires that observations are stationary, which is a standard assumption in the literature. α-mixing condition is one of the weakest mixing conditions for weakly dependent stochastic processes. Many stationary time series or Markov chains, including many financial time series fulfilling certain (mild) conditions are α-mixing with exponentially decaying coefficients; see Cai (2002), Carrasco and Chen (2002) and Chen and Tang (2005) for additional examples. Assumption A1 also gives some standard moment conditions. Assumption A2 includes some smoothness conditions on functionals involved. The requirement in A3 that $K(\cdot)$ be compactly supported is imposed for the sake of brevity of proofs and can be removed at the cost of lengthier arguments. In particular, the Gaussian kernel is allowed. Assumption A4 is a standard condition for a nonparametric kernel smoothing. Finally, notice that A5 is not restrictive; e.g., if one considers the optimal bandwidth such that $h_{opt} = O(T^{-1/(L+4)})$ (see Remark 4 later), then A5 is satisfied when $\delta > L/2 - 1$. Therefore, the conditions imposed here are quite mild and standard.

2.3.2 Large Sample Theory

Before I derive the asymptotic distribution of NPGEE estimate, I list some notations. To this effect, define $\mu_2(K) = \int u\, u^\top K(u)du$ and $\nu_0(K) = \int K^2(u)du$. Set $H = \text{diag}\{1, h^2 I_L\}$, where I_L is an $L \times L$ identity matrix. Finally, define $S(z) = M(z)\text{diag}\{1, \mu_2(K)\}$ and $S^*(z) = \text{diag}\{\nu_0(k),\, h^2\mu_2(K^2)\}\sigma_0^2(z)$. The asymptotic normality of the NPGEE estimator is established in Theorem 2.1 with detailed proof given in Section 2.6.

Theorem 2.1. *Under Assumptions A(1) - A(5), for any grid point z_0, then,*

$$\sqrt{Th^L}\left[H\left\{\begin{pmatrix}\hat{m}(z_0)\\ \hat{m}'(z_0)\end{pmatrix}-\begin{pmatrix}m(z_0)\\ m'(z_0)\end{pmatrix}\right\}-B(z_0)\right]\to N(0,\,\Delta_m(z_0)),\quad(2.12)$$

where the asymptotic bias term is $B(z_0) = h^2/2\begin{pmatrix}tr(\mu_2(K)m''(z_0))\\ 0\end{pmatrix}$ *and the asymptotic variance is* $\Delta_m(z) = f(z)^{-1}S^{-1}(z)S^*(z)S^{-1}(z)$*, Particularly,*

$$\sqrt{Th^L}\left[\hat{m}(z_0)-m(z_0)-\frac{h^2}{2}tr(\mu_2(K)m''(z_0))\right]\to N(0,\,\sigma_m^2(z_0)),\qquad(2.13)$$

where $\sigma_m^2(z_0) = \nu_0(K)\,\sigma_0^2(z_0)f^{-1}(z_0)M^{-2}(z_0)$.

Remark 3: (Consistent estimate of asymptotic variance). The first consequence of Theorem 2.1 is to provide an easy way to obtain a consistent estimator for the asymptotic variance $\sigma_m^2(z)$. After estimating the nonparametric pricing kernel, one can obtain the estimated pricing error as $\hat{e}_t = [1-\hat{m}(Z_t)r_{p,t+1}]\,r_{i,t+1}$. Then, any nonparametric kernel smoothing method, say the local linear technique, can be applied to obtaining a consistent estimate for $\sigma_0^2(z)$, $f(z)$, and $M(z)$, and one can apply some existing optimal bandwidth selectors, like plugging in, cross-validation, generalized cross-validation, nonparametric the Akaike information criterion, and others. Therefore, a consistent estimate for $\sigma_m^2(z)$ is $\hat{\sigma}_m^2(z) = \nu_0(K)\,\hat{\sigma}_0^2(z)\hat{f}^{-1}(z)\hat{M}^{-2}(z)$. Thus, a 95% pointwise confidence interval with bias ignored can be constructed as

$$\hat{m}(z)\pm 1.96\times\frac{\hat{\sigma}_m(z)}{\sqrt{Th^L}},\qquad(2.14)$$

which will be used in computing the confidence interval for the real example presented in Section 2.5.

Remark 4: (A rule of thumb for bandwidth selection). It is well known that the bandwidth plays an essential role in the trade-off between reducing bias and variance. To the best of our knowledge, almost nothing has been done about selecting the bandwidth in the context of nonparametric estimation equations method. In many applications, one would like to have a quick idea on how large the amount of smoothing should be. A rule of thumb is very appealing for such a case. Such a rule is meant to be somewhat crude, but possesses simplicity and requires little programming effort that other methods can not compete with. Toward this end, one can see easily from Theorem 2.1 that the weighted integrated asymptotic mean squared error (AMSE) is given by

$$\text{AMSE} = \int \left[\text{Var} + (\text{Bias})^2\right]^2 f(z_0)dz_0 = \frac{C_1}{Th^L} + \frac{h^4}{4}C_2,$$

where $C_1 = \int \sigma_m^2(z_0)f(z_0)dz_0 = E\left[\sigma_m^2(Z_t)\right]$ and $C_2 = \int \left[tr(\mu_2(K)m''(z_0))\right] f(z_0)dz_0 = E\left[tr(\mu_2(K)m''(Z_t))\right]$. By minimizing AMSE with respect to h, one obtains the optimal theoretical bandwidth

$$h_{opt} = \left(\frac{L\,C_1}{C_2}\right)^{1/(L+4)} T^{-1/(L+4)} \equiv C_3 T^{-1/(L+4)}. \tag{2.15}$$

With the above choice of h_{opt}, it is easily seen that the optimal AMSE has the order of $O(T^{-4/(L+4)}$. Clearly, the formulation in (2.15) provides an easy way to find a data-driven fashion bandwidth selection method, say a plugging in method. Toward this end, one needs to estimate C_3 consistently, which can be done as follows. First, take a pilot bandwidth h_0 which is much smaller than $T^{-1/(L+4)}$, say $h_\sigma = 0.1 \times T^{-1/(L+4)}$ or smaller. Using this pilot bandwidth, one can estimate $\sigma_m^2(z_0)$, so that one obtains $\hat{C}_1$ using the average. To estimate $m''(z_0)$ consistently and easily, one can use a simple way to do so. That is to fit a multivariate polynomial of certain order L_m (say $L_m = \log(T)$ or larger) globally to $m(z)$, leading to a parametric fit. Other global parametric approaches, including series and spline methods, can be used too. Then, the generalized method of moment (GMM) of Hansen (1982) can be used for estimating the parameters. The choice of a global fit results in a derivative function $\hat{m}''(z)$ which is a multivariate polynomial of order $L_m - 2$. Thus, $\hat{C}_2$ is obtained by average. Hence, one has $\hat{C}_3$ and $\hat{h}_{opt} = \hat{C}_3 T^{-1/(L+4)}$.

Remark 5: (Choice of instruments). After establishing the asymptotic property of the estimator, I now turn to the choice of $Q(Z_t)$. At this moment, I assume that $Q(Z_t) = \begin{pmatrix} Q_0(Z_t) \\ Q_0(Z_t)(Z_t - z_0) \end{pmatrix}$, where $Q_0(Z_t)$ is an unknown scale function. By following the same proofs used in the proof of Theorem 2.1, one can show that the asymptotic normality in (2.1) holds true for this situation with the asymptotic variance

$$\triangle_{m,0}(z_0) = f^{-1}(z_0) S_1^{-1}(z_0) S_1^*(z_0) S_1^{-1}(z_0),$$

where $S_1(z) = Q_0(z)S(z)$ and $S_1^*(z) = Q_0^2(z)S^*(z)$. It is clearly that the asymptotic variance $\triangle_{m,0}(z) = \triangle_m(z)$, which is not related to the choice of $Q_0(\cdot)$. Hence, I assume $Q(\cdot)$ has the form given in (2.11).

2.4 Model Extension

Theoretically, for a valid SDF, equation (2.1) is supposed to hold all the assets in the market. In reality, asset pricing models are at best approximations. This implies no stochastic discount factor proxies can price portfolios perfectly in general. Therefore, it is important to conduct a measure of pricing errors produced by SDFs so that we are able to compare and evaluate SDFs. For this purpose, Hansen and Jagannathan (1997) introduced the Hansen-Jagannathan distance method (HJ-distance) which is a measure that is widely used for diagnosis and estimation of asset pricing models. This method gained tremendous popularity in the empirical asset pricing literature by many researchers. The measure is in the quadratic form of the pricing errors weighted by the inverse of the second moment matrix of returns

$$HJ = \sqrt{E[e_t]^\top E(r_{t+1}^\top r_{t+1})^{-1} E[e_t]}, \tag{2.16}$$

where e_t is the pricing error.

To have some intuition idea, one may provide a geometric interpretation of HJ-distance in terms of the minimum-variance frontiers of the test assets. Actually, the HJ-distance is a special form of generalized method of moments (GMM) of Hansen (1982). Thus, the estimation procedure can be conducted in the framework of GMM.

Nagel and Singleton (2008) attempted to provide an improved understanding of the HJ-distance by focusing on the conditional version of HJ distance. It has a similar econometric interpretation comparing to the unconditional one. However, it measures the pricing error on the condition of current information set

$$HJ_c = \sqrt{E[e_t|\Omega_t]^\top E(r_{t+1}^\top r_{t+1}|\Omega_t)^{-1} E[e_t|\Omega_t]}. \tag{2.17}$$

The conditional HJ-distance has more advantage than the unconditional one. In the case that the two different SDFs may generate the same unconditional HJ-distance statistically, the conditional measure makes it possible to discriminate them.

In this chapter, I attempt to provide an improved understanding of the HJ-distance by focusing on the case of conditional pricing models and combining local linear technique. The conditional HJ-distance will serve as an extension of the model. Still, I assume $m(\cdot)$ is twice continuously differentiable. By the Taylor expansion, one has the same locally weighted orthogonality condition as (2.9).

Define

$$A_T(\beta(z)) = \frac{1}{T}\sum_{t=1}^{T} Q_t\{1 - [m(z) + \nabla m(z)^\top (Z_t - z)]r_{p,t+1}\}r_{i,t+1}K_h(Z_t - z), \quad (2.18)$$

$$\varepsilon_t = Q_t(1 - m(Z_t)r_{p,t+1})r_{i,t+1}K_h(Z_t - z), \quad (2.19)$$

and

$$\beta(z) = \begin{pmatrix} m(z) \\ \nabla m(z) \end{pmatrix} = \begin{pmatrix} a(z) \\ b(z) \end{pmatrix}.$$

For local estimation purpose, we may need some additional assumption for the model.

2.4.1 Distribution Theorem

During the initial estimation, we might choose different Λ_0 as the weighting matrix. Λ_T is a consistent estimate of certain finite positive definite matrix Λ_0. Different choice of weighting matrix Λ_T would result in different asymptotic property in the estimation of $\beta(z)$.

Assumption B:

B1. For all $\beta(z) \in \theta(z)$, $E[\|\varepsilon_t(\beta(z))\|^2|Z_t = z]$, Λ_0 is the weighting matrix. Let Λ_T be a finite positive definite matrix for all T, as is $\Lambda_0 = plim_{T\to\infty}\Lambda_T$. $E[Q_tQ_t^\top r_{p,t+1}r_{i,t+1}f(z)|Z_t = z]$ and $\text{Var}(Q_tQ_t^\top r_{p,t+1}r_{i,t+1}f(z)|Z_t = z)$ are finite and continuous at z.

B2. $E[\varepsilon_t(\beta(z))]$ is finite and twice differentiable in the vector $\beta(z)$ for all $\beta(z)$ in some compact set $\theta(z)$.

Firstly, in Theorem (2.2), we would like to show the asymptotic distribution of $\beta(z)$ under different weighting matrix Λ_T.

If we estimate $\beta(z)$ by minimizing the square conditional HJ-distance, the estimation can be conducted in the framework of local nonparametric GMM.

The proposed local estimator is

$$\hat{\beta}(z) = arginf_{\beta(z)\in\Theta(z)} A_T(\beta(z))^\top \Lambda_T A_T(\beta(z)), \tag{2.20}$$

By taking derivative with respect to $\beta(z)$ and solving for $\beta(z)$, one obtain

$$\hat{\beta}(z) = (S_T^\top \Lambda_T S_T)^{-1} S_T^\top \Lambda_T L_T. \tag{2.21}$$

Under some assumptions, the distribution of the sample HJ-distance estimator is presented in the following theorem.

Theorem 2.2. *Under Assumptions A(1) - A(5), B(1) - B(2), Λ_T is consistent matrix of $\Lambda_0(z)$ for any grid point z, we have,*

$$\sqrt{Th^L}\,\left[H\,(\hat{\beta}(z) - \beta(z)) - B_1(z)\right] \;\rightarrow\; N(0,\, \Delta\Delta_m(z)), \tag{2.22}$$

where the asymptotic bias term is

$$B_1(z) = (S^\top(z)\Lambda_0 S(z))^{-1} S^\top(z)\Lambda_0 M(z) B(z)$$

and the asymptotic variance is

$$\Delta\Delta_m(z) = [S(z)^\top \Lambda_0(z) S(z)]^{-1} S(z)^\top \Lambda_0(z) S(z)^* \Lambda_0(z) S(z) [S(z)^\top \Lambda_0(z) S(z)]^{-1}.$$

For practical purpose, in the standard two step GMM procedure. We would like to choose the weighting matrix Λ_T to be the inverse of the sample variance of

ε_t, where $\Lambda_0 = (\text{Var}(\varepsilon_t))^{-1}$, $\Lambda_T \xrightarrow{p} \Lambda_0$.

In the first step to get an consistent estimate of $\text{Var}(\varepsilon_t)$, one needs an initial estimate of $\beta(z)$. An initial estimate of $\beta(z)$ is obtained by minimizing $A_T(\beta(z))^\top A_T(\beta(z))$ with identity weighting matrix. By taking derivative with respect to $\beta(z)$ and solving for $\beta(z)$, one obtain

$$\hat{\beta}^0(z) = (S_T^\top S_T)^{-1} S_T^\top L_T, \tag{2.23}$$

which coincides with (2.10). $\hat{\beta}^0(z)$ is then used to produce a weighting matrix as

$$\hat{\Lambda}_T = [(Th^L) A_T(\hat{\beta}^0(z)) A_T^\top(\hat{\beta}^0(z))]^{-1}.$$

Hence, the final estimate is given by

$$\hat{\beta}^1(z) = (S_T^\top \hat{\Lambda}_T S_T)^{-1} S_T^\top \hat{\Lambda}_T \, L_T. \tag{2.24}$$

The following theorem gives the asymptotic property of $\hat{\beta}^1(z)$. It is interesting that, if the weighting matrix Λ_T is chosen to be consistent estimate of $\text{Var}^{-1}(\varepsilon_t)$, the asymptomatic distribution is the same as Theorem 2.1; see (2.12).

Theorem 2.3. *Under Assumptions A(1) - A(5), B(1) - B(2), for any grid point z, we have,*

$$\sqrt{Th^L}\ \left[H\,(\hat{\beta}^1(z) - \beta(z)) - B_2(z)\right] \ \rightarrow\ N(0,\, \Delta_m(z)),$$

where the asymptotic bias is $B_2(z) = [S^\top(z)(S^*)^{-1}S(z)]^{-1}S^\top(z)(S^*)^{-1}M(z)B(z) = B(z)$ and the asymptotic variance is $\Delta_m(z) = f(z)^{-1}[S(z)(S^*(z))^{-1}S(z)]^{-1}$, Particularly,

$$\sqrt{Th^L}\ \left[\hat{m}(z) - m(z) - \frac{h^2}{2}\, tr(\mu_2(K)m''(z_0))\right] \ \rightarrow\ N(0,\, \sigma_m^2(z_0)),$$

where $\sigma_m^2(z_0) = \nu_0(K)\,\sigma_0^2(z_0) f^{-1}(z_0) M^{-2}(z_0)$.

Remark 6: (Discussion of iteration estimation). This estimation process can be iterated until $\hat{\beta}^j(z) \approx \hat{\beta}^{j+1}(z)$, though the estimate based on a single iteration $\beta^1(z)$ has the same asymptotic distribution as that based on an arbitrarily large number of iterations. Iterating offers the practical advantage that the final estimates are invariant with respect to the scale of the data and to the initial weighting matrix.

2.5 Empirical Examples

In this section, I use three simulated examples and two real examples to illustrate the proposed model and its nonparametric estimation procedure. Among the simulated examples, the first two examples are for one-dimensional case and the last one is for two-dimensional setting. Notice that the Gaussian kernel is used.

2.5.1 Simulated Examples

Example 1: For simplicity of implementation, I first choose only one covariate Z_t following an autoregressive (AR) model as

$$Z_t = 0.2\, Z_{t-1} + \epsilon_{t,1}, \tag{2.25}$$

where $\epsilon_{t,1}$ is standard normally distributed. To illustrate the proposed methods, I consider simulated examples under mean-variance efficient condition as in Wang (2002) for the portfolio. The conditional mean of $r_{p,t+1}$ takes the form $r_{p,t+1} = g(Z_t) + 0.05\, \epsilon_{t,2}$, where $g(Z_t) = 0.1 + 0.1\, Z_t^2$ and $\epsilon_{t,2}$ is standard normally distributed. In order to generate $m(Z_t)$ to satisfy mean-variance efficiency in (1.14), I choose

$$m(Z_t) = E(r_{p,t+1}|Z_t)/E(r_{p,t+1}^2|Z_t) = g(Z_t)/[(0.05)^2 + g(Z_t)^2]. \tag{2.26}$$

Then $r_{i,t+1}$ is determined by (2.1) as $r_{i,t+1} = e_t/\,[1 - m(Z_t) r_{p,t+1}]$, where

$$e_t = 0.05\, e_{t-1} + v_t, \tag{2.27}$$

and v_t is also standard normal. Next I choose three sample sizes: $T = 300$, 500, and 1000. The performance of the proposed nonparametric estimators is evaluated by the mean absolute deviation error (MADE), defined as

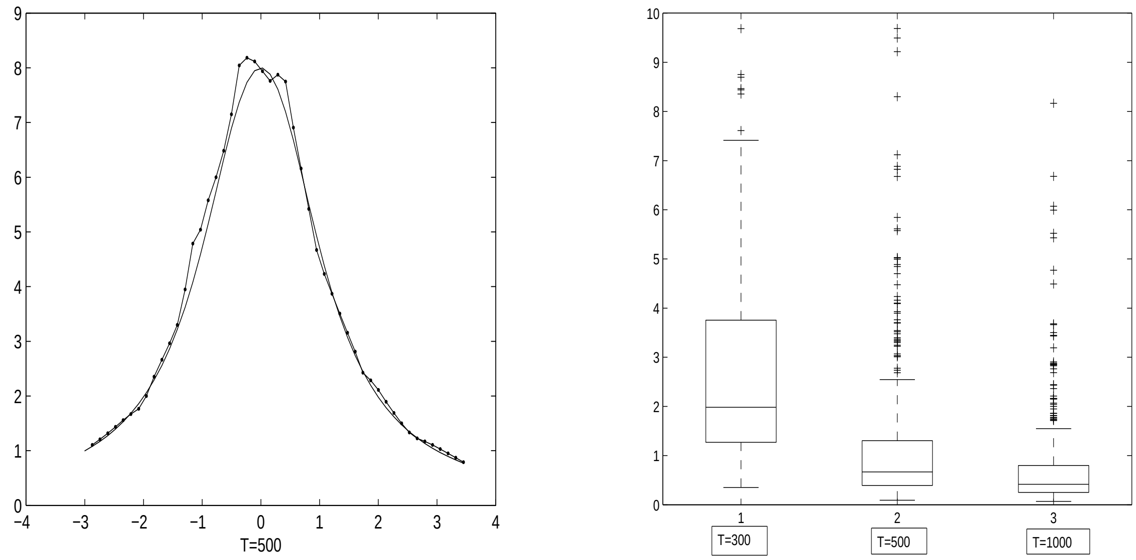
$$\mathcal{E}_m = \frac{1}{n_0}\sum_{k=1}^{n_0} |\hat{m}(z_k) - m(z_k)|,$$

where $\{z_k\}_{k=1}^{n_0}$ are grid points. For each sample size, I compute the mean absolute deviation errors and the experiment is repeated 500 times.

The 500 $\mathcal{E}_m$ values are computed and plotted in the form of box-plots in Figure 2.1(b), from which, one can see clearly that the median of the MADE 500 values is decreasing when the sample size gets larger. This implies that the estimation becomes stable when sample size becomes larger. This supports the asymptotic theory that the proposed estimator is consistent. Furthermore, I choose a typical sample to show how close the nonparametric estimate $\hat{m}(z)$ is to its true curve. The typical sample is selected in such a way that its MADE value is equal to the median in the 500 MADE values. The true curve (solid line) of $m(z)$ defined in (2.26) is plotted in Figure 2.1(a) together with its nonparametric estimated curve (dotted line) for sample size $T = 500$ based on the typical sample. One can observe that the nonparametric estimate for $m(\cdot)$ is very close to its true curve, and $\hat{m}(\cdot)$ performs fairly well.

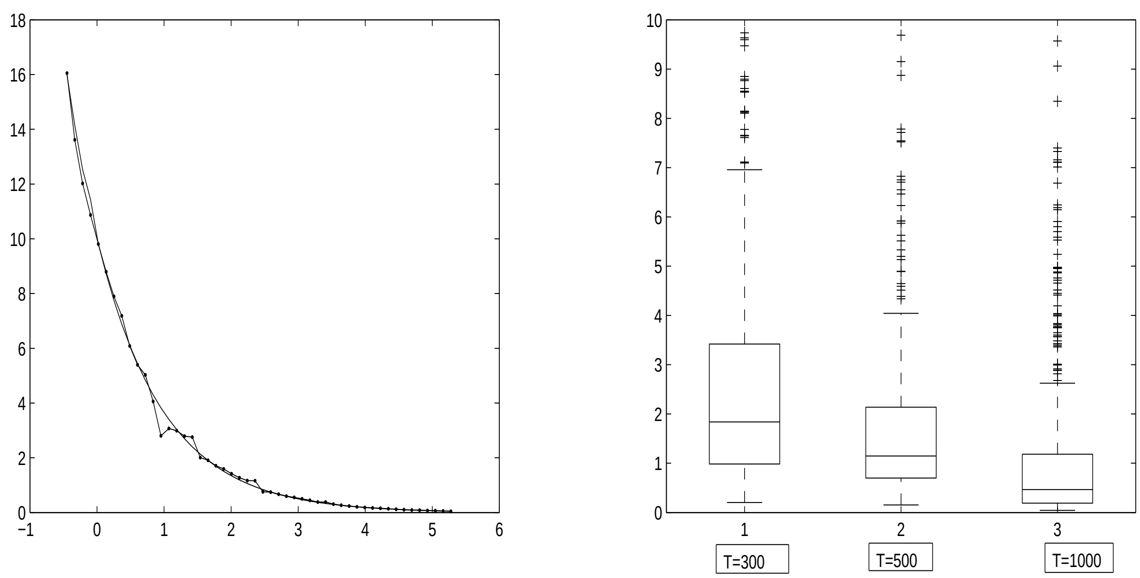
Example 2: I now consider a new model without imposing the mean-variance efficient condition. Here I assume that only the general orthogonal condition (2.1) is satisfied. Z_t is generated in as in (2.25), $r_{i,t+1} = e_t/\{1 - m(Z_t)r_{p,t+1}\}$, where e_t is the same as in Example 1 (see (2.27)) and $m(Z_t)$ is given by

$$m(Z_t) = \frac{0.1\ \exp(Z_t) + Z_t}{0.01\ \exp(2Z_t) + 0.1\ \exp(Z_t)\, Z_t + 0.0025 Z_t^2},$$



(a) (b)

Figure 2.1: *(a) The true curve of* $m(z)$ *and its nonparametric estimate for sample size* $T = 500$. *(b) The boxplots of* $MADE500$ *for three sample sizes.*



(a) (b)

Figure 2.2: *(a) The true curve of* $m(z)$ *and its nonparametric estimate for sample size* $T = 500$. *(b) The boxplots of* $MADE500$ *for three sample sizes.*

and $r_{p,t+1} = 0.1\exp(Z_t) + 0.05\,Z_t\,\epsilon_{t,3}$, where $\epsilon_{t,3}$ is standard normally distributed. Clearly, equation (2.1) or (2.2) is satisfied but $m(Z_t) \neq E(r_{p,t+1}|Z_t)/E(r_{p,t+1}^2|Z_t)$, so that the mean-variance efficient condition (1.14) is not satisfied. Similar to Example 1, I compute the 500 MADE values and the nonparametric estimate for a typical sample. Figure 2.2(a) shows the true curve of $m\cdot)$ and its nonparametric estimate based on the typical sample when $T = 500$ and boxplots for 500 MADE values are reported in Figure 2.2(b) for three sample sizes $T = 300$, 500 and 1000. Obviously, the same conclusion similar to that in Example 1 can be made.

Example 3: In the foregoing examples, I only consider the case where Z_t is a scalar. To gain a further insight, I consider the multivariate situation. The scenario is similar to the one dimensional case. I first generate the model under mean-variance efficient condition. That is, Z_t is generated from the following two AR models:

$$Z_{1t} = 2 + 0.5\,Z_{1t-1} + \eta_{t,1}, \quad Z_{2t} = 2 + 0.3\,Z_{2t-1} + \eta_{t,2}, \quad \text{and} \quad \eta_{t,2} = 0.1\,\eta_{t,1} + 0.1\,u_t, \tag{2.28}$$

where $\eta_{t,1}$ and u_t are standard normally distributed. Clearly, Z_{1t} and Z_{2t} is correlated. Similar to Example 1, $r_{p,t+1} = g(Z_t) + 0.05\,\epsilon_{t,2}$, where $g(Z_t) = 1 + 0.1\,Z_{1t}^2 + 0.1\,Z_{2t}^2$ and $\epsilon_{t,2}$ is the same as in Example 1.

By equation (1.14),

$$m(Z_t) = E(r_{p,t+1}|Z_t)/E(r_{p,t+1}^2|Z_t) = g(Z_t)/[0.0025 + g(Z_t)^2].$$

Then, $r_{i,t+1}$ is determined by $r_{i,t+1} = e_t/\,[1 - m(Z_t)r_{p,t+1}]$, where e_t is the same as in Example 1 (see (2.27)).

I still choose three sample sizes: $T = 300$, 500, and 1000. For each sample size, I replicate the design 500 times, and the boxplots of the MADE 500 are presented

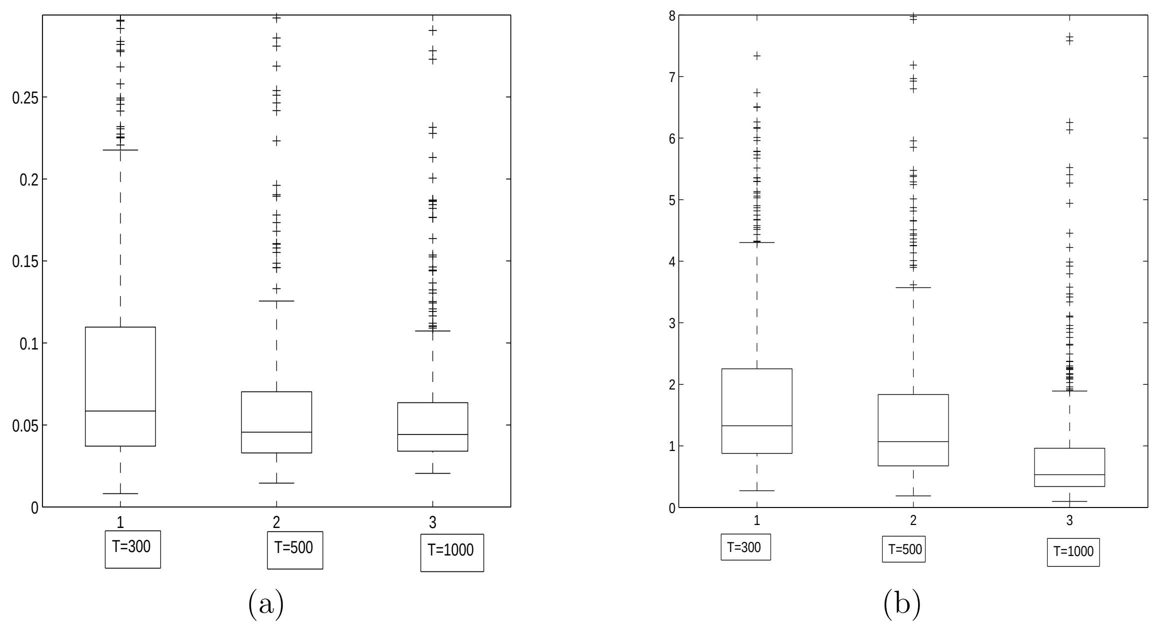


Figure 2.3: *The boxplots of $MADE500$ for three sample sizes $T = 300$, $T = 500$, $T = 1000$. (a) under mean-covariance efficiency; (b) without mean-covariance efficiency.*

in Figure 2.3(a). The two dimensional mean absolute deviation error is defined as

$$\mathcal{E}_m = \frac{1}{n_0^2} \sum_{k_1=1,k_2=1}^{n_0} |\hat{m}(z_{1k_1}, z_{2k_2}) - m(z_{1k_1}, z_{2k_2})|$$

where $\{(z_{1k_1}, z_{2k_2})\}$ are grid points. I choose the kernel function to be the product kernel as $K(u, v) = K(u)K(v)$, where $K(\cdot)$ is the standard normal density function. From Figure 2.3(a), it is evident that when sample size increases, the performance of nonparametric estimator becomes better.

Finally, I consider an example with two covariates without imposing the mean-covariance efficiency. The model is generated as follows. $r_{i,t+1}$ is generated by the orthogonal condition (2.1) as $r_{i,t+1} = e_t / [1 - m(Z_{1t}, Z_{2t}) r_{p,t+1}]$, where e_t is the same as in Example 1 (see (2.27)), $r_{p,t+1} = 0.1 \exp(Z_{1t}) + \epsilon_{t,4}$, and

$$m(Z_{1t}, Z_{2t}) = \frac{0.1 \exp(Z_{1t}) + Z_{2t}}{0.01 \exp(2\, Z_{1t}) + 0.1 \exp(Z_{1t})\, Z_{2t} + 0.0025}.$$

Here, $\epsilon_{t,4} \sim N(0, 1)$ and (Z_{1t}, Z_{2t}) is generated based on (2.28). I again choose three sample sizes: $T = 300$, 500, and 1000 and replicate the experiment 500 times for each sample size. The boxplots of MADE 500 are reported in Figure 2.3(b). It can be seen obviously from Figure 2.3(b) that the same conclusion similar to the foregoing case.

2.5.2 A Real Example

Example 4: I now apply the proposed nonparametric method to estimate $m(\cdot)$ for a real example. The data are monthly excess returns from January 31, 1966 to December 29, 2006, which are downloaded from CRSP. For the benchmark portfolio, I use NYSE value-weighted (including dividend) as $r_{p,t+1}$ and the value-weighted NYSE size decile 1 (SZ1) is used as asset $r_{i,t+1}$. The covariates are chosen to be the logarithm of dividend-price ratio (DPR), the logarithm of default premium (DEF), the logarithm of the one month treasury bill rate (RTB), and the excess return on NYSE equally weighted index (EWR). DPR is the dividend yield (in percent) on the NYSE value-weighted index, DEF is the difference between Baa-rated corporate bond yield and Aaa-rated bond yield, and RTB is the 1-month T-bill yield; see Wang (2002) for details. The choice of kernel is the same as in the simulation study and the bandwidth is selected based on the rule of thumb described in Remark 4.

Similar to the simulation, I now begin with the one dimensional estimation. Each time I use only one variable as the covariate; that is $m(z)$ is estimated as a univariate function. I also obtain the 95% confidence intervals for the estimates of $m(z)$ with the bias ignored; see (2.14) in Remark 3. The result is presented in Figure 2.4(a) for DPR, 2.4(b) for DEF, 2.4(c) for RTB and 2.4(d) for EWR. The dashed curves represent the 95% confidence interval and the solid line is the nonparametric estimator of $m(\cdot)$.

From these graphs, it is obvious that the estimates are noisy in some cases. Thus it is very hard to give a clear conclusion for these patterns, but these graphs do suggest the nonlinearity of $m(\cdot)$. Since I do not assume any functional form of $m(\cdot)$, presence of the nonparametric estimation method is advantageous if $m(\cdot)$ is nonlinear.

Remark 7: In Wang (2002), he gives an estimate of conditional betas using non-parametric way where $b(Z_t) = \frac{E(r_{p,t+1}|Z_t)}{E(r^2_{p,t+1}|Z_t)}$. $E(r_{p,t+1}|Z_t)$ and $E(r^2_{p,t+1}|Z_t)$ is estimated by Nadaraya-Watson regression. Also, Wang presents plots of conditional betas by focusing on "one-dimensional snap shots". While betas are estimated as multivariate functions, he plots the univariate functions by conditioning on one variable and keeping all the other conditioning variables at their means. Plots of conditional betas serve as an interesting way to illustrate nonlinearity in the time-varying be-

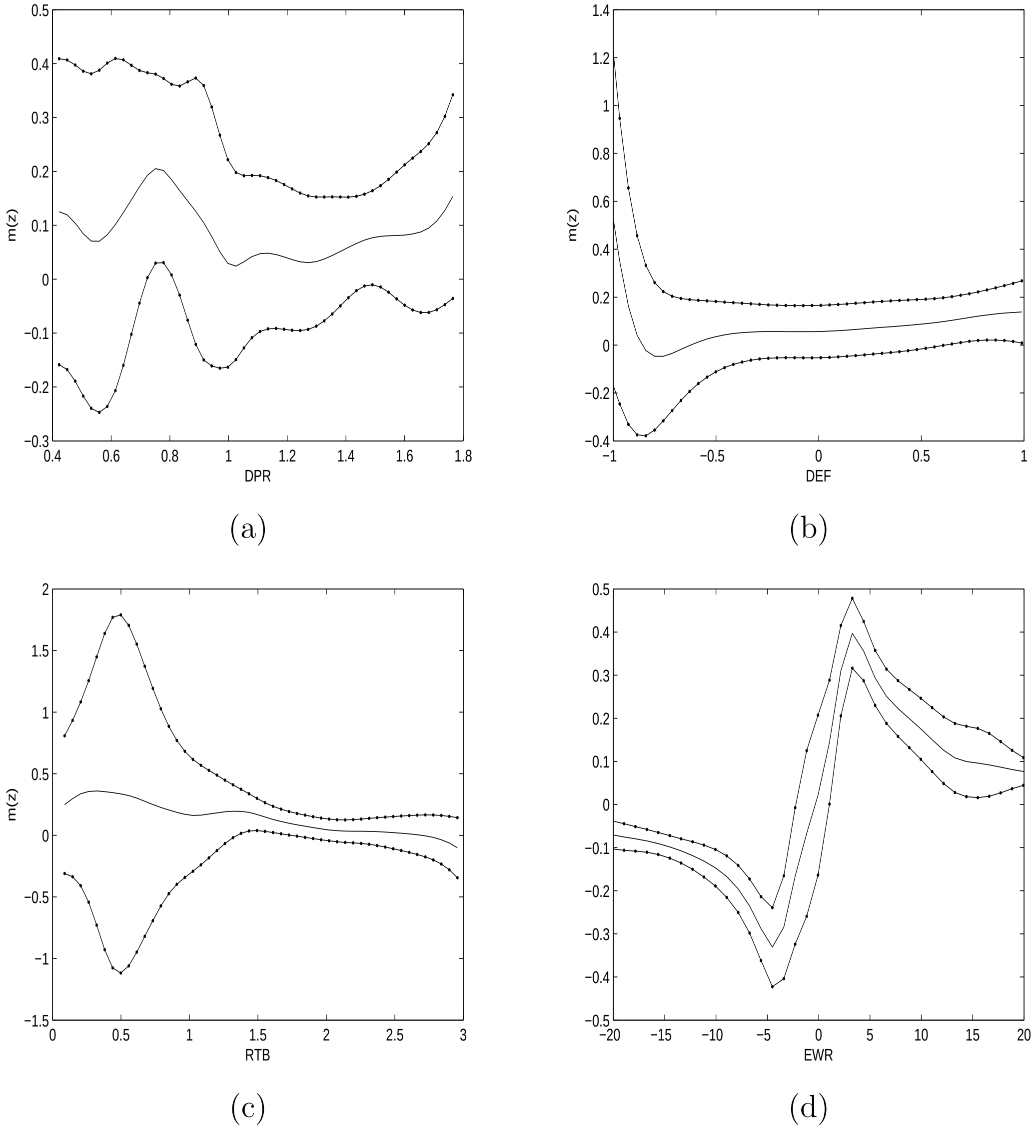
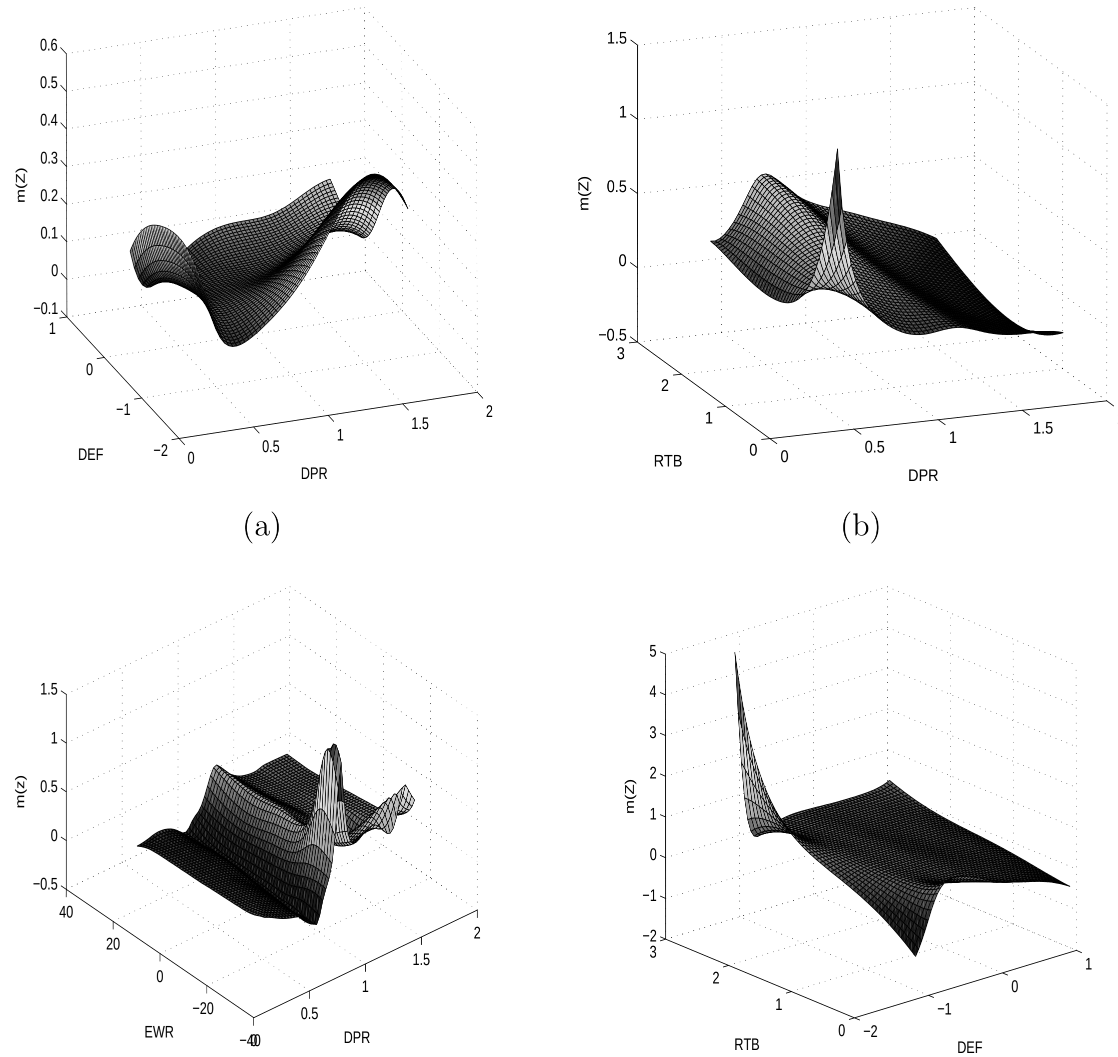


Figure 2.4: The one dimensional nonparametric estimate of $m(\cdot)$. (a) Z_t is DPR; (b) Z_t is DEF; (c) Z_t is RTB; (d) Z_t is EWR.

tas. Moreover, under mean-variance efficiency, $m(Z_t)$ reduces to $b(Z_t) = \frac{E(r_{p,t+1}|Z_t)}{E(r_{p,t+1}^2|Z_t)}$. Therefore, $m(Z_t) = b(Z_t)$ is equivalent to the mean-variance efficiency. While under general orthogonal condition (2.5), when $m(Z_t)$ is a scalar function, has expression as $m(Z_t) = \frac{E(r_{i,t+1}|Z_t)}{E(r_{i,t+1}r_{p,t+1}|Z_t)}$.

Finally, in this example I apply the proposed method to estimate $m(Z_t)$ un-



(a) (b) (c) (d)

Figure 2.5: The two dimensional nonparametric estimate of $m(\cdot)$. (a) $Z_t =$(DPR, DEF); (b) $Z_t =$(DPR, RTB); (c) $Z_t =$(DPR, EWR); (d) $Z_t =$(DEF, RTB).

der the multidimensional setting. To simplify computation, each time I take two variables as covariates. The kernel is the product kernel mentioned in simulation examples. The results are summarized in Figure 2.5 which shows the surfaces of $m(\cdot)$ versus bivariate covariant.

One can observe an unstable and nonlinear curved surfaces from Figure 2.5.

This is similar to the finding in Ghysels (1998). In Figures 2.5(a) and 2.5(c), $m(\cdot)$ is volatile and changes rapidly versus the covariant. But in Figure 2.5(d), $m(\cdot)$ as a function of DEF is less volatile and is modestly stable. Therefore, if I use a specific model to measure time-varying structure for different APT-type models, the result would be sophisticated and fragile. That is the reason why one needs a more flexible model in the nonlinear APT models.

2.6 Proof of Theorems

To prove the theorem, I need the following four lemmas, which are stated below without a proof. For the detailed proofs, the reader is referred to the book by Hall and Hyde (1980) for Lemma 2.1, papers by Volkonskii and Rozanov (1959) for Lemma 2.2, Shao and Yu (1996) for Lemma 2.3, papers by Hjort and Pollard (1996) for Lemma 2.4. Finally, I use the same notation introduced in Section 2. Throughout this Appendix, C denotes a generic positive constant, which may take different values at different places.

Lemma 2.1. *Davydov's Lemma: Suppose that two random variables X and Y that are $\mathbb{F}_{-\infty}^{t}$ and $\mathbb{F}_{t+\tau}^{\infty}$, respectively, and that $||X||_p < \infty$ and $||Y||_q < \infty$, where $||X||_p = \{E|X|^p\}^{1/p}$, $p, q \geq 1$, and $1/p + 1/q < 1$. Then,*

$$\sup_t |Cov(X, Y)| \leq 8\alpha^{1/r}(\tau)\{E|X|^p\}^{1/p}\{E|Y|^q\}^{1/q},$$

where $r = (1 - 1/p - 1/q)^{-1}$ and $\alpha(\cdot)$ is the mixing coefficient.

Lemma 2.2. *Let $V_1, \ldots, V_{L_1}$ be α mixing stationary random variables that are $\mathbb{F}_{i_1}^{j_1}, \cdots, \mathbb{F}_{i_{L_1}}^{j_{L_1}}$-measurable, respectively with $1 \leq i_1 < j_1 < \cdots < j_{L_1}, i_{l+1} - j_l \geq \tau$,* *and* $|V_l| \leq 1$ *for* $l = 1, \ldots L_1$. *Then,*

$$\left| E\left(\prod_{l=1}^{L_1} V_l\right) - \prod_{l=1}^{L_1} E(V_l) \right| \leq 16(L_1 - 1)\alpha(\tau),$$

where $\alpha(\cdot)$ *is the mixing coefficient.*

Lemma 2.3. *Let* V_t *be an* α*-mixing process with* $E(V_t) = 0$ *and* $\| V_t \|_r < \infty$ *for* $2 < p < r \leq \infty$. *Define* $S_n = \sum_{t=1}^n V_t$ *and assume that* $\alpha(\tau) = O(\tau^{-\theta})$ *for some* $\theta > pr/(2(r-p))$. *Then,*

$$E|S_n|^p \leq K n^{p/2} \max_{t \leq n} \| V_t \|_r^p,$$

where K *is a finite positive constant.*

Lemma 2.4. *Convexity Lemma: Let* $\{\lambda_n(\theta : \theta \in \Theta)\}$ *be a sequence of random convex functions defined on a convex, open subset* Θ *of* $\Re^d$. *Suppose* $\lambda(d)$ *is a real-valued function on* Θ *for which* $\Lambda_n(\theta) \to \lambda(\theta)$ *in probability, for each* θ *in* Θ. *Then for each compact subset* C_0 *of* Θ,

$$\sup_{\theta \in C_0} |\lambda_n(\theta) - \lambda(\theta)| \xrightarrow{p} 0$$

Moreover, the function $\lambda(.)$ *is necessarily convex on* Θ.

Recall and define some notations as follows:

$$e_t = (1 - m(Z_t) r_{p,t+1}) r_{i,t+1}, \quad S_T = Q_t Q_t^\top K_h(Z_t - z_0) r_{i,t+1} r_{p,t+1},$$

$$\varepsilon_t = Q_t(1 - m(Z_t) r_{p,t+1}) r_{i,t+1} K_h(Z_t - z), \qquad G_T = \frac{1}{T}\sum_{t=1}^T \varepsilon_t,$$

$$B_T = \frac{1}{T}\sum_{t=1}^T \frac{1}{2} Q_t r_{i,t+1} r_{p,t+1} (Z_t - z_0)^\top m''(z_0)(Z_t - z_0) K_h(Z_t - z_0),$$

$$
\begin{aligned}
R_t &= m(Z_t) - m(z_0) - m'(z_0)^\top (Z_t - z_0) - \frac{1}{2}(Z_t - z_0)^\top m''(z_0)(Z_t - z_0), \\
R_T^* &= \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t - z_0) Q_t r_{i,t+1} r_{p,t+1} \left[m(Z_t) - m(z_0) - m'(z_0)^\top (Z_t - z_0) \right. \\
&\quad \left. - \frac{1}{2}(Z_t - z_0)^\top m''(z_0)(Z_t - z_0) \right],
\end{aligned}
$$

$M(z) = E(r_{p,t+1}, r_{i,t+1} | Z_t = z)$, $S(z) = M(z) diag\{1, \mu_2(K)\}$ and

$$
S^*(z) = diag\{\nu_0(K), h^2 \mu_2(K^2)\}\sigma_0^2(z).
$$

Finally, set $\tilde{S}_T = H^{-1} S_T$ and $H = \text{diag}\{1, h^2 I_L\}$.

Proposition 2.1. *Under Assumption A1 - A5, we have*

(i) $\tilde{S}_T = f(z_0) S\{1 + o_p(1)\}$.

(ii) $B_T = f(z_0) M(z_0) B(z_0) + o_p(h^2)$.

(iii) $R_T^* = o_p(h^2)$.

Proof: By the stationary assumption and A1 - A5, we have,

$$
\begin{aligned}
E(\tilde{S}_T) &= E\left(\frac{1}{T}\sum_{t=1}^{T} H^{-1} Q_t Q_t^{*\top} K_h(Z_t - z_0) r_{i,t+1} r_{p,t+1} \right) \\
&= E(H^{-1} Q_t Q_t^{*\top} K_h(Z_t - z_0) r_{i,t+1} r_{p,t+1}) \\
&= E(E(H^{-1} Q_t Q_t^{*\top} K_h(Z_t - z_0) r_{i,t+1} r_{p,t+1} \,|Z_t)) \\
&= \int \begin{pmatrix} M(z_0 + hu) & M(z_0 + hu)u^\top \\ M(z_0 + hu)u & M(z_0 + hu)uu^\top \end{pmatrix} K(u) f(z_0 + hu) du \\
&\to f(z_0) S(z_0)
\end{aligned}
$$

and

$$
\begin{aligned}
& Th^L \operatorname{Var}\left(\frac{1}{T}\sum_{t=1}^{T} r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)\right) \\
= \; & h^L \operatorname{Var}(r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)) \\
& + \frac{2h^L}{T}\sum_{t=1}^{T-1}(T-t)\operatorname{Cov}(r_{i,2}r_{p,2}K_h(Z_1 - z_0), r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)) \\
\equiv \; & I_1 + I_2.
\end{aligned}
$$

By Assumption A1 and A2,

$$
\operatorname{Var}(r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)) = O(h^{-L}),
$$

which implies that

$$
I_1 = O(1).
$$

Next we prove that $I_2 \to 0$. To this end we split I_2 into two parts as $I_2 = I_3 + I_4$, where $I_3 = 2h^L/T\sum_{t=1}^{d_T}(\cdots)$ and $I_4 = 2h^L/T\sum_{t>d_T}(\cdots)$. Let $d_T \to \infty$ be a sequence of integers such that $d_T h^L \to 0$. Firstly we show that $I_3 \to 0$. By conditional on Z_1, Z_t, and using assumption A_2, we obtain

$$
\operatorname{Cov}(r_{i,2}r_{p,2}K_h(Z_1 - z_0), r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)) = O(1)
$$

It follows that $I_3 \leq d_T h^L \to 0$. We now consider the contribution of I_4. For an α-mixing process, we use Davydov's inequality (see, e.g., Lemma 2.1).

$$
\begin{aligned}
& |\operatorname{Cov}(r_{i,2}r_{p,2}K_h(Z_1 - z_0), r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0))| \\
\leq \; & C[\alpha(t)]^{\frac{\delta}{2+\delta}} \parallel r_{i,2}r_{p,2}K_h(Z_1 - z_0) \parallel_{2+\delta} \parallel r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0) \parallel_{2+\delta}
\end{aligned}
$$

By Assumption A2, we have

$$\begin{aligned} & E|r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)|^{2+\delta} \\ = \;& h^{-L(1+\delta)} f(z_0)|E(r_{i,t+1}r_{p,t+1}|z_0)^{2+\delta}| \int K^{2+\delta}(u)du + o(h^{-L(1+\delta)}) \\ \leq \;& O(h^{-L(1+\delta)}) \end{aligned}$$

Thus,

$$|\mathrm{Cov}(r_{i,2}r_{p,2}K_h(Z_1 - z_0), r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0))| = O(\alpha^{\delta/(2+\delta)}(t)h^{-2L(1+\delta)/(2+\delta)}),$$

and

$$|I_4| = C\frac{h^L}{T}\sum_{t>d_T}(T-t)\alpha^{\delta/(2+\delta)}(t)h^{-2L(1+\delta)/(2+\delta)} \leq C\sum_{t>d_T}\alpha^{\delta/2+\delta}(t)h^{-\delta L/2+\delta}$$

By Assumption A1, and choosing $d_T^{2+\delta}h^L = O(1)$, we have

$$I_4 = C\sum_{t>d_T}\alpha^{\delta/2+\delta}(t)h^{-L\delta/2+\delta} = o(h^{-L\delta/2+\delta}d_T^{-\delta}) = o(1)$$

d_T satisfies the requirement that $d_T h^L \to 0$. Note that, in Assumption A4, we assume $h \to 0$ and $Th^L \to \infty$ as $T \to \infty$,

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)\right) = o(1).$$

Using similar arguments, we can show that

$$\frac{1}{T}\sum_{t=1}^{T} r_{i,t+1}r_{p,t+1}K_h(Z_t - z_0)(Z_t - z_0)/h = o_p(1), \tag{2.29}$$

$$\frac{1}{T}\sum_{t=1}^{T} r_{i,t+1}r_{p,t+1}K_h(Z_t-z_0)(Z_t-z_0)(Z_t-z_0)^\top/h^2 = f(z_0)M(z_0)\mu_2(K)+o_p(1), \tag{2.30}$$

By (2.29) and (2.30), we can obtain immediately that

$$\tilde{S}_T = f(z_0)S\{1+o_p(1)\}.$$

Hence, we have proved (i).

Next, we show (ii). Note that by stationary assumption and A2,

$$\begin{aligned}
& E(B_T) \\
= \; & \frac{h^2}{2}E\left(Q_tK_h(Z_t-z_0)r_{i,t+1}r_{p,t+1}\left(\frac{Z_t-z_0}{h}\right)^\top m''(z_0)\left(\frac{Z_t-z_0}{h}\right)\right) \\
= \; & \frac{h^2}{2}E\left(E[K_h(Z_t-z_0)r_{i,t+1}r_{p,t+1}Q_t\left(\frac{Z_t-z_0}{h}\right)^\top m''(z_0)\left(\frac{Z_t-z_0}{h}\right)|Z_t]\right) \\
= \; & \frac{h^2}{2}\int\begin{pmatrix} M(z_0+hu)u^\top m''(z_0)u \\ h\,M(z_0+hu)uu^\top m''(z_0)u\end{pmatrix}K(u)f(z_0+hu)du \\
\to \; & f(z_0)M(z_0)\,B(z_0).
\end{aligned}$$

where $\int u^\top uK(u) = tr(\mu_2(K))$. By the same token, we can show that the variance of $h^{-2}B_T$ converges to 0. Hence, we proved (ii). Finally, we have

$$\begin{aligned}
h^{-2}E(R_T^*) &= h^{-2}E\left[K_h(Z_t-z_0)r_{i,t+1}r_{p,t+1}R_t\,Q_t\right] \\
&= h^{-2}E\left[K_h(Z_t-z_0)M(Z_t)R_tQ_t\right] \\
&= h^{-2}\int M(z_0+hu)K(u)f(z_0+hu)R(z_0+hu)\begin{pmatrix}1 \\ hu\end{pmatrix}du,
\end{aligned}$$

where, by assumption A2,

$$R(z) = m(z) - m(z_0) - m'(z_0)hu - \frac{1}{2}(z - z_0)^\top m''(z_0)(z - z_0),$$

so that $R(z_0 + hu) = o(h^2)$. Then,

$$E[h^{-2}R_T^*] = o(1).$$

Similarly, we can show that $\mathrm{Var}[h^{-2}R_T^*] = o(1)$. This proves the proposition.

Proposition 2.2. *Under assumption A1-A5, then,*

$$Th^L\, Var(G_T) \;\rightarrow\; f(z_0)S^*. \tag{2.31}$$

Proof: By the orthogonal condition in (2.1), we know that $E(G_T) = 0$ and

$$\begin{aligned} Th^L\,\mathrm{Var}(G_T) &= \frac{h^L}{T} Var\left(\sum_{t=1}^{T} Q_t e_t K_h(Z_t - z_0)\right) \\ &= h^L\,\mathrm{Var}\left(Q_t e_t K_h(Z_t - z_0)\right) \\ &\quad + \frac{2h^L}{T}\sum_{t=1}^{T-1}(T-t)\mathrm{Cov}(Q_1 e_1 K_h(Z_1 - z_0), Q_t e_t K_h(Z_t - z_0)) \\ &\equiv I_5 + I_6. \end{aligned}$$

By Assumption A2, similar to the proof of Proposition 2.1, we have

$$I_5 \;\rightarrow\; f(z_0)S^*.$$

The same, we split I_6 into two parts as $I_6 = I_7 + I_8$, where $I_7 = 2h^L/T\sum_{t=1}^{d_T}(\cdots) \leq d_T h^L \rightarrow 0$ and $I_8 = 2h^L/T\sum_{t>d_T}(\cdots)$. Take v_1, v_2 as 0,1, by Davydov's inequality to obtain,

$$
\begin{aligned}
&|\mathrm{Cov}(e_1 K_h(Z_1 - z_0)(Z_1 - z_0)^{v_1}, e_t K_h(Z_t - z_0)(Z_t - z_0)^{v_2})| \\
\le\ & C[\alpha(t)]^{\frac{\delta}{2+\delta}} \parallel e_1 K_h(Z_1 - z_0)(Z_1 - z_0)^{v_1} \parallel_{2+\delta} \parallel e_t K_h(Z_t - z_0)(Z_t - z_0)^{v_2} \parallel_{2+\delta}
\end{aligned}
$$

$$E|e_1 K_h(Z_1 - z_0)(Z_1 - z_0)^{v_1}|^{2+\delta} \le O(h^{-L(1+\delta)}).$$

Thus, by assumption $A.1$ and choosing $d_T^{2+\delta} h^L = O(1)$

$$I_8 = o(d_T^{-\delta} h^{-L\delta/2+\delta}) = o(1)$$

and that

$$Th^L \,\mathrm{Var}(G_T) \ \rightarrow \ f(z_0) S^*.$$

This proves the proposition.

Proof of Theorem 2.1: Recall that

$$H \left\{ \begin{pmatrix} \hat{m}(z_0) \\ \hat{m}'(z_0) \end{pmatrix} - \begin{pmatrix} m(z_0) \\ m'(z_0) \end{pmatrix} \right\} - \tilde{S}_T^{-1} B_T - \tilde{S}_T^{-1} R_T^* = \tilde{S}_T^{-1} G_T$$

It follows from Propositions 2.1 and 2.2 that

$$H \left\{ \begin{pmatrix} \hat{m}(z_0) \\ \hat{m}'(z_0) \end{pmatrix} - \begin{pmatrix} m(z_0) \\ m'(z_0) \end{pmatrix} \right\} - B(z_0) + o_p(h^2) = f^{-1}(z_0) S^{-1} G_T \{1 + o_p(1)\}. \tag{2.32}$$

To prove Theorem 2.1, it suffices to establish the asymptotic normality of $\sqrt{Th^L} G_T$. Now we use the Wold-Craḿer device, so that we consider linear combination with an unit vector $d^\top G_T$. It is easy to show by a simple algebra that

$$\sqrt{Th^L} d^\top G_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t,$$

where $w_t = \sqrt{h^L} d^\top \{Q_t r_{i,t+1} K_h(Z_t - z_0)(1 - m(Z_t) r_{p,t+1})\}$. Now the problem reduces to proving the asymptotic normality of $\sum_{t=1}^T w_t/\sqrt{T}$. By Proposition 2.2, one can show that

$$\text{Var}(w_t) = f(z_0) d^\top S^* d (1+o(1)) \equiv \theta^2(z_0)(1+o(1)), \quad \text{and} \quad \sum_{t=2}^{T} |\text{Cov}(w_1, w_t)| = o(1).$$

Therefore,

$$\text{Var}\left(\sqrt{Th^L}\, d^\top G_T\right) = \theta^2(z_0)(1 + o(1)). \tag{2.33}$$

We employ so-called small-block and large-block method. For this setting, we partition the set $\{1, 2, \cdots, T\}$ into $2q_T + 1$ subsets with large-blocks of size r_T and small blocks of size s_T. Let $T/(r_T + s_T)$ be the number of blocks. Let the random variables η_j and ϵ_j be the sum over the jth large block, the jth small block, and ξ be the sum over the residual block. That is,

$$\eta_j = \sum_{t=j(r_T+s_T)+1}^{(j+1)(r_T+s_T)+r_T} w_t, \quad \text{and} \quad \epsilon_j = \sum_{t=j(r_T+s_T)+r_T+1}^{(j+1)(r_T+s_T)} w_t.$$

Then,

$$\sqrt{Th^L}\, d^\top G_T = \frac{1}{\sqrt{T}} \left\{ \sum_{j=0}^{q_T-1} \eta_j + \sum_{j=0}^{q_T-1} \epsilon_j + \xi \right\} \equiv \frac{1}{\sqrt{T}} \{Q_{T,1} + Q_{T,2} + Q_{T,3}\}.$$

We will show that as $T \to \infty$,

$$\frac{1}{T} E[Q_{T,2}]^2 \;\to\; 0, \quad \frac{1}{T} E[Q_{T,3}]^2 \;\to\; 0, \tag{2.34}$$

$$\left| E\left[\exp(itQ_{T,1})\right] - \prod_{j=0}^{q_T-1} E\left[\exp(it\eta_j)\right] \right| \;\to\; 0, \tag{2.35}$$

$$\frac{1}{T}\sum_{j=0}^{q_T-1} E(\eta_j^2) \;\to\; \theta^2(z_0),, \tag{2.36}$$

and for every $\epsilon^* > 0$,

$$\frac{1}{T}\sum_{j=0}^{q_T-1} E\left[\eta_j^2 I\{|\eta_j| \geq \epsilon^*\,\theta(z_0)\sqrt{T}\}\right] \;\to\; 0. \tag{2.37}$$

Clearly, these four statements imply that the sums over small and residual blocks $Q_{T,2}/\sqrt{T}$, $Q_{T,3}/\sqrt{T}$ are asymptotically negligible in probability, and $\{\eta_j\}$ in $Q_{T,1}$ are asymptotically independent. Also, (2.36) and (2.37) are standard Lindeberg-Fellow conditions for asymptotic normality of $Q_{T,1}/\sqrt{T}$. To show the asymptotical normality of $d^\top G_T$, it suffices to establish the four statements stated in (2.34) - (2.37). First, we choose the block sizes,

$$r_T = \left\lfloor (Th^L)^{1/2} \right\rfloor, \qquad s_T = \left\lfloor (Th^L)^{1/2}/\log T \right\rfloor,$$

where $\tau = (2+\delta)(1+\delta)/\delta$. It can be easily shown that

$$s_T/r_T \;\to\; 0, \quad r_T/T \;\to\; 0, \quad q_T\,\alpha(s_T) \;\to\; 0. \tag{2.38}$$

Now we establish (2.34) and (2.36). Clearly,

$$E[Q_{T,2}^2] = \sum_{j=0}^{q_T-1} \mathrm{Var}(\epsilon_j) + 2\sum_{0\leq k<j<q_T-1} \mathrm{Cov}(\epsilon_k, \epsilon_j) \equiv J_1 + J_2.$$

By stationarity and (2.33),

$$J_1 = q_T\,\mathrm{Var}(\epsilon_1) = q_T\,\mathrm{Var}\left(\sum_{t=1}^{s_T} w_t\right) = q_T\,s_T[\theta^2(z_0) + o(1)],$$

and

$$|J_2| \le 2 \sum_{j_1=1}^{T-r_T} \sum_{j_2=j_1+r_T}^{T} |\mathrm{Cov}(w_{j_1}, w_{j_2})| \le 2T \sum_{j=r_T+1}^{T} |\mathrm{Cov}(w_1, w_j)| = o(T).$$

Hence, by (2.38),

$$q_T s_T = o(T), \quad \text{so that} \quad E(Q_{T,2})^2 = q_T s_T \theta^2(z_0) + o(T) = o(T).$$

It follows from the stationarity, (2.38) and Proposition 2.2 that

$$\mathrm{Var}(Q_{T,3}) = \mathrm{Var}\left(\sum_{t=1}^{T-q_T(s_T+r_T)} w_t \right) = O(T - q_t(r_T + s_T)) = o(T).$$

From Lemma 2.2, we then proceed as follows:

$$\left| E\left[\exp\left(it \sum_{j=0}^{q_T-1} Q_{T,1} \right) \right] - \prod_{j=0}^{q_T-1} E\left[\exp(it\eta_j) \right] \right| \le 16\, q_T \alpha(s_T) \ \to\ 0.$$

This proves (2.35). It remains to show that

$$\frac{1}{T} \sum_{j=0}^{q_T-1} E\left[\eta_j^2 I\{|\eta_j| \ge \epsilon\theta(z_0)\sqrt{T}\} \right] \ \to\ 0.$$

Now, it follows from Lemma 2.3 that

$$\begin{aligned} & E\left[\eta_j^2 I\{|\eta_j| \ge \epsilon\theta(z_0)\sqrt{T}\} \right] \\ \le\ & CT^{-\delta/2} E(|\eta_j|^{2+\delta}) \le CT^{-\delta/2} r_T^{1+\delta/2} \{E|w_t|^{2(1+\delta)}\}^{(2+\delta)/2(1+\delta)} \end{aligned}$$

One can easily show that

$$E\left(|w_1|^{2+2\delta} \right) \le c\, h^{-L\delta}.$$

By plugging into the above the right-hand side, we obtain

$$\begin{aligned} & \frac{1}{T}\sum_{j=0}^{q_T-1} E\left[\eta_j^2 I\{|\eta_j| \geq \epsilon^*\theta(z_0)\sqrt{T}\}\right] \\ = \;& O(r_T^{\delta/2}T^{-\delta/2}h^{-L(2+\delta)\delta/2(1+\delta)}) \\ = \;& O(T^{-\delta/4}h^{-L[1+2/(1+\delta)]\delta/4}) \to 0 \end{aligned}$$

by Assumption A5. This proves the theorem.

Proof of Theorem 2.2: Solving for $\hat{\beta}(z)$ in (2.21), we have

$$\begin{aligned} & H(\hat{\beta}_T(z) - \beta(z)) \\ = \;& H[(S_T^\top \Lambda_T S_T)^{-1}(S_T^\top \Lambda_T L_T) - \beta(z)] \\ = \;& H(S_T^\top \Lambda_T S_T)^{-1} S_T^\top \Lambda_T (L_T - S_T\beta(z)) \\ = \;& H(S_T^\top \Lambda_T S_T)^{-1} S_T^\top \Lambda_T A_T \\ = \;& [\tilde{S}_T^\top \Lambda_T \tilde{S}_T]^{-1}\tilde{S}_T^\top(\Lambda_T)A_T. \end{aligned}$$

Then,

$$H(\hat{\beta}_T - \beta_T) - (\tilde{S}_T(\Lambda_T)\tilde{S}_T)^{-1}\tilde{S}_T\Lambda_T B_T - (\tilde{S}_T(\Lambda_T)\tilde{S}_T)^{-1}\tilde{S}_T\Lambda_T R_T^* = (\tilde{S}_T(\Lambda_T)\tilde{S}_T)^{-1}\tilde{S}_T\Lambda_T G_T.$$

Since in the general nonparametric GMM framework, we may choose different weighing matrix Λ_T. Λ_T is consistent estimate of some positive definite matrix Λ_0, where $\lim \Lambda_T \xrightarrow{p} \Lambda_0(z)$. By Propositions 2.1 and 2.2

$$\begin{aligned} & H[\hat{\beta}_T(z) - \beta_T(z)] - (S(z)^\top \Lambda_0(z) S(z))^{-1} S^\top \Lambda_0(z) M(z) B(z) + o_p(h^2) \\ = \;& f^{-1}[S(z)^\top \Lambda_0(z) S(z)]^{-1} S^\top \Lambda_0(z) G_T\{1 + o_p(1)\} \end{aligned}$$

By applying the proof of asymptotic normality of $\sqrt{Th^L}G_T$ in the proof of Theorem 2.1, Theorem 2.2 is proved.

In Proposition 2.3, we take the weighting matrix $\Lambda_0 = Var^{-1}(\varepsilon_t)$.

Proposition 2.3. *Under Assumption A1 - A5, B1-B2, we have*

(i) $Th^L(\hat{A}_T(z) - A_T(z))(\hat{A}_T(z) - A_T(z))^\top = o_p(1)$

(ii) $Th^L A_T(\hat{A}_T - A_T)^\top = o_p(1)$.

(iii) $Th^L \hat{A}_T(z)\hat{A}_T^\top(z) \xrightarrow{p} \Lambda_0^{-1}(z)$, *where* $\Lambda_0 = Var^{-1}(\varepsilon_t)$

(iv) $\sup_{\beta\in\Theta} |\Lambda_T(z) - \Lambda_0(z)| \xrightarrow{p} 0$.

Proof: Firstly, we prove proposition 2.3(i). Recall that

$$\hat{A}_T = \frac{1}{T}\sum_{t=1}^{T} Q_t\{1 - [\hat{m}(z) + \hat{\nabla} m(z)^\top (Z_t - z)]r_{p,t+1}\}r_{i,t+1}K_h(Z_t - z)$$

By decomposition we have that $\hat{A}_T = R_T^* + G_T + B_T$

$$
\begin{aligned}
& Th^L(\hat{A}_T(z) - A_T(z))(\hat{A}_T(z) - A_T(z))^\top \\
= \; & \frac{1}{Th^L}\sum_{t=1}^{T} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top \begin{pmatrix} \hat{m}(z) - m(z) \\ \nabla\hat{m}(z) - \nabla m(z) \end{pmatrix} \begin{pmatrix} 1 & \\ & h^4 I \end{pmatrix} \\
& \begin{pmatrix} \hat{m}(z) - m(z) \\ \nabla\hat{m}(z) - \nabla m(z) \end{pmatrix}^\top \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top K^2(\frac{Z_t - z}{h}) r_{p,t+1}^2 r_{i,t+1}^2 \\
+ \; & \frac{1}{Th^L}\sum_{s \neq t} r_{p,t+1} r_{i,t+1} r_{p,s+1} r_{i,s+1} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top \begin{pmatrix} \hat{m}(z) - m(z) \\ \nabla\hat{m}(z) - \nabla m(z) \end{pmatrix} \\
& \begin{pmatrix} \hat{m}(z) - m(z) \\ \nabla\hat{m}(z) - \nabla m(z) \end{pmatrix}^\top \begin{pmatrix} 1 & \\ & h^4 I \end{pmatrix} \begin{pmatrix} 1 \\ (Z_s - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_s - z)/h \end{pmatrix}^\top \\
& K(\frac{Z_t - z}{h}) K(\frac{Z_s - z}{h}) \\
\equiv \; & G_1(z) + G_2(z).
\end{aligned}
$$

We have proved in Theorem 2.1 that

$$
H\left\{ \begin{pmatrix} \hat{m}(z_0) \\ \hat{m}'(z_0) \end{pmatrix} - \begin{pmatrix} m(z_0) \\ m'(z_0) \end{pmatrix} \right\} - \tilde{S}_T^{-1} B_T - \tilde{S}_T^{-1} R_T^* = \tilde{S}_T^{-1} G_T
$$

Since $\hat{\beta}(z)$ is a kernel estimate, by Masry (1996) and standard conditions, we can easily show the uniform consistency,

$$
\sup_z |\tilde{S}_T - f(z)S(z)| \xrightarrow{p} 0 \quad \text{and} \quad \sup_z |B_T - f(z)M(z)B(z)| = o_p(h^2).
$$

Then

$$H\left\{\begin{pmatrix}\hat{m}(z)\\ \hat{m}'(z)\end{pmatrix}-\begin{pmatrix}m(z)\\ m'(z)\end{pmatrix}\right\}-B(z)+o_p(h^2)=f^{-1}(z)S^{-1}G_T\{1+o_p(1)\}. \quad (2.39)$$

Regarding on the first term of $(\hat{A}_T(z)-A_T(z))(\hat{A}_T(z)-A_T(z))^\top$,

$$\begin{aligned}&G_1(z)\\ =\ &\frac{1}{Th^L}\sum_t\Bigg(\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}^\top(B(z)+f^{-1}(z)S^{-1}(z)G_T)\\ &(B(z)^\top+f^{-1}(z)S^{-1}(z)^\top G_T^\top)K^2(\frac{Z_t-z}{h})r_{p,t+1}^2r_{i,t+1}^2\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}^\top\Bigg)\\ &+o_p(1).\end{aligned}$$

By Proposition 2.1 and 2.2. $Th^LVar(G_T)\to f(z)S^*$, $B_T=f(z)M(z)B(z)+o_p(h^2)$, $B(z)=O(h^2)$, $R_T^*=o_p(h^2)$ we can obtain that

$$\begin{aligned}&E\Bigg(\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}^\top\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}\begin{pmatrix}1\\ (Z_t-z)/h\end{pmatrix}^\top\\ &(f^{-2}(z)S^{-1}(z)G_TG_T^\top S^{-1}(z)^\top)K^2(\frac{Z_t-z}{h})r_{p,t+1}^2r_{i,t+1}^2\Bigg)\\ &=O(1/(Th^L))+O(h^4)=o(1)\end{aligned}$$

By law of large number,

$$G_1(z)=o_p(1).$$

Similar to the proof of Proposition 2.1, by Assumption A1 that under mixing condition and Davydov's inequality,

$$E(\Big(\begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top (f^{-1}(z) G_T S^{-1}(z) K(\frac{Z_t - z}{h}) r_{p,t+1} r_{i,t+1})$$
$$(f^{-1}(z) G_T S^{-1}(z)^\top) K(\frac{Z_s - z}{h}) r_{p,t+1} r_{i,s+1} \begin{pmatrix} 1 \\ (Z_s - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_s - z)/h \end{pmatrix}^\top \Big)$$
$$= O(\alpha^{\delta/(2+\delta)}(t) h^{2\delta L/(2+\delta)})$$

Hence $G_2(z) = o_p(1)$, Proposition 2.3 (i) is proved.

Moreover, by Proposition 2.1

$$\begin{aligned} & Th^L(\hat{A}_T - A_T) A_T^\top \\ = \;& Th^L(\hat{A}_T - A_T)(G_T + B_T + R_T^*) \\ = \;& \frac{1}{Th^L} \sum_{t=1}^{T} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top (B(z)^\top + f^{-1}(z) S^{-1}(z)^\top G_T^\top) \\ & (G_T + f(z) M(z) B(z)) + o_p(1) \end{aligned}$$

Also

$$E(\begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix} \begin{pmatrix} 1 \\ (Z_t - z)/h \end{pmatrix}^\top (B(z)^\top + f^{-1}(z) S^{-1}(z)^\top G_T^\top)(G_T + f(z) M(z) B(z)))$$
$$= O(1/Th^L) + O(h^4) = o(1)$$

Using similar argument in Proposition 2.3(i), we have $Th^L(\hat{A}_T - A_T) A_T^\top = o_p(1)$.

Next, we show (iii). Note that

$$Var(\varepsilon_t) = f(z) S^*$$

$$Th^L A_T(z) A_T(z)^\top \xrightarrow{p} f(z) S^*.$$

Thus $Th^L A_T(z) A_T(z)^\top \xrightarrow{p} \Lambda_0^{-1}(z)$ where $\Lambda_0 = Var^{-1}(\varepsilon_t)$.

By decomposition, it is easy to see that

$$\begin{aligned} & \hat{A}_T(z)\hat{A}_T(z) \\ = \;& (\hat{A}_T(z) - A_T(z))(\hat{A}_T(z) - A_T(z))^\top + 2(\hat{A}_T - A_T)A_T^\top + A_T A_T^\top \end{aligned} \tag{2.40}$$

Thus, by Proposition 2.3 (i) and (ii), we have that

$$Th^L \hat{A}_T(z)\hat{A}_T^\top(z) \xrightarrow{p} \Lambda_0^{-1}(z)$$

Proposition 2.3 (iii) is proved.

We have already proved the point wise convergence Λ_T. Since the limiting process of our target function $A_T A_T^\top$ is a quadratic form, using the convexity lemma 2.4 given in Pollard (1991), see lemma 2.4 one can easily establish the uniform convergence. According to Proposition 2.3, $\Lambda(z)$ is a real-valued function on Θ for which $\Lambda_T \to \Lambda(z)$ in probability, for each β in Θ. Then for each compact subset C of Θ,

$$\sup_{\beta \in \Theta} |\Lambda_T(z) - \Lambda(z)| \xrightarrow{p} 0.$$

Hence, Proposition 2.3 is proved.

. Proof of Theorem 2.3 It follows from Proposition 2.3,

$$\begin{aligned} & H[\hat{\beta}_T(z) - \beta_T(z)] - (S(z)\Lambda(z)S(z))^{-1} S^\top \Lambda(z) M(z) B(z) + o_p(h^2) \\ = \;& f^{-1}[S(z)^\top \Lambda(z) S(z)]^{-1} S^\top \Lambda(z) G_T\{1 + o_p(1)\}. \end{aligned}$$

By combining with Proposition 2.3 and multiple by $\sqrt{Th^L}$, Theorem 2.3 is proved.

CHAPTER 3: TEST OF MISSPECIFICATION OF PRICING KERNEL

In Chapter 2, I consider a general nonlinear pricing kernel model and propose a new nonparametric estimation procedure by combining local polynomial estimation technique and generalized estimation equations, termed as *nonparametric generalized estimation equation* (NPGEE). The nonparametric method may provide a useful insight for future parametric fitting. Parametric models for time-varying betas can be most efficient if the underlying betas are correctly specified. However, a misspecification may cause serious bias and model constraints may distort the betas in local area if the underlying betas are not correctly specified. Hence, to test whether the SDF model has a linear structure or whether some parametric form is correctly specified becomes essentially important in practice. In this chapter, I propose a consistent nonparametric testing procedure to test whether the model is correctly specified under a U-statistic framework. I adopt general GMM of Hansen (1982) method to estimate the parameter inside parametric form of SDF under the null hypothesis. Under fairly general stationarity, continuity, and the moment condition that the expectations of the pricing errors delivered by SDF equal to zero, the estimate inside SDF is consistent. My test combines the methodology of the conditional moment test and nonparametric techniques, it may help to avoid the model misidentification in pricing kernel and may enhance the efficiency of the parametric model with prior information. Further, the limiting distributions under both the null and alternative hypotheses are derived.

The rest of the chapter is organized as follows. In Section 3.1, I use GMM to estimate functional form involved in the pricing kernel. I consider testing misspecification of the pricing error and propose a nonparametric consistent test statistic and establish its limiting distributions under both the null and alternative hypotheses. Also in this section, I compare the nonparametric and parametric pricing kernel model and their testing inference. I show that there is no power if I plug in the nonparametric estimator into the pricing kernel. Results based on the Monte Carlo simulation study and real examples are reported in Section 3.4 to illustrate the finite sample performance.

3.1 Parametric Models

In this section, I concentrate on a parametric approach to asset pricing models. I follow Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), Ghysels (1998), and Wang (2003) closely and start with a stochastic discount factor model; see Campbell, Lo and MacKinlay (1997) and Cochrane (2001) for details about theory and methods and references of recent studies on the SDF approach. A very simplified version of the SDF framework is universal and admits a basic pricing representation

$$E[m_{t+1}\, r_{i,t+1} \,|\, \Omega_t] = 0,$$

where Ω_t denotes the information set at time t, m_{t+1} is the SDF or the marginal rate of substitution (MRS) or the pricing kernel, and $r_{i,t+1}$ is the excess return on the i-th asset or portfolio. In empirical finance, different models impose different constraints on the SDF. Particularly, the SDF is usually assumed to be a linear function of factors in various applications. Further, when the SDF is fully parameterized such as linear form, the general method of moments (GMM) of Hansen (1982) can be used to estimate parameters; see Campbell, Lo and MacKinlay (1997) and Cochrane (2001) for details.

Recently, some more flexible SDF models have been studied by several authors. For example, Bansal, Hsieh and Viswanathan (1993) and Bansal and Viswanathan

(1993) were the pioneers to propose nonlinear APT models in empirical asset pricing by assuming that the SDF or MRS is a nonlinear function of a few state variables. Ghysels (1998) further advocated these models. When the exact form of the nonlinear pricing kernel is unknown, Bansal and Viswanathan (1993) suggested using the polynomial expansion to approximate it. As pointed by Wang (2003), although this approach is very intuitive and general, one of the shortcomings is that it is difficult to obtain the distribution theory and the effective assessment of finite sample performance. To overcome this difficulty, instead of considering the nonlinear pricing kernel, Ghysels (1998) focused on the nonlinear parametric model and used a set of moment conditions suitable for GMM estimation of parameters involved. Wang (2003) studied the nonparametric conditional CAPM and gave an explicit expression for the nonparametric form of conditional CAPM for the excess return; that is, $m_{t+1} = 1 - b(Z_t)\, r_{p,t+1}$, where Z_t is a $L \times 1$ vector of conditioning variables from Ω_t, $b(z) = E(r_{p,t+1}|Z_t = z)/E(r_{p,t+1}^2|Z_t = z)$ is an unknown function, and $r_{p,t+1}$ is the return on the market portfolio in excess of the riskless rate. Since the functional form of $b(\cdot)$ is unknown, Wang (2003) suggested estimating $b(\cdot)$ by using the Nadaraya-Watson method to two regression functions $E(r_{p,t+1}|Z_t = z)$ and $E(r_{p,t+1}^2|Z_t = z)$. Also, he conducted a simple nonparametric test about the pricing error. Further, Wang (2003) extended this setting to multifactor models by allowing $b(\cdot)$ to change over time; that is, $b(Z_t) = b(t)$. Finally, Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), and Ghysels (1998) did not assume that m_{t+1} is a linear function of $r_{p,t+1}$ and instead they considered a parametric model by using the polynomial expansion.

In reality, people would be interested in identifying the parametric form of the pricing kernel and to test the mean-variance efficiency. My test statistic U_T introduced in Section 3.2 is essentially based on a leave-one-out Nadaraya-Watson estimator formulated from $E(\xi_t E(\xi_t|Z_t) f(Z_t))$ and I show in Section 3.2 that to test

$E(\xi_t|Z_t) = 0$ is equivalent to testing $E(\xi_t E(\xi_t|Z_t)f(Z_t)) = 0$. Furthermore, there is no power if I plug in the nonparametric GEE estimator into my test statistic. The reason is obviously that nonparametric GEE method can always give us a consistent estimate which satisfies the orthogonal condition

$$E[\{1 - m(Z_t)\, r_{p,t+1}\}\; r_{i,t+1}\,|\,\Omega_t] = 0. \tag{3.1}$$

Especially, when $m(.)$ is a scale function, the alternative expression is

$$m(Z_t) = \frac{E(r_{i,t+1}|\Omega_t)}{E(r_{i,t+1}r_{p,t+1}|\Omega_t)}.$$

In other words, the null hypothesis $H_0 : E[(1 - m(Z_t)r_{p,t+1})r_{i,t+1}|Z_t] = 0$ can not be rejected since such $m(Z_t)$ satisfying H_0 always exists and its NPGEE can always give an consistent estimate of $m(Z_t)$.

Indeed, H_0 can be regarded as a parametric moment (orthogonal) condition. Hence, my interest is to estimate function form of $m(Z_t, \theta)$ and thus I can test misspecification of $m(Z_t, \theta)$ and the pricing error $E(\xi_{i,t+1}|\Omega_t) = E((1-m(Z_t, \theta)r_{p,t+1})r_{i,t+1}|\Omega_t) = 0$, where $m(\cdot, \theta)$ is an assumed known function of Z_t and Z_t is a $L \times 1$ vector of conditioning variables from Ω_t.

3.2 Estimation and Statistical Inference

Let θ denote an unknown $(a \times 1)$ vector of parameters and Z_t be a $L \times 1$ vector of conditional variables from Ω_t. Then, (3.1) becomes

$$E[\{1 - m(Z_t, \theta)r_{p,t+1}\}\; r_{i,t+1}\,|\,Z_t] = 0 \tag{3.2}$$

which implies

$$E[\{1 - m(Z_t, \theta)r_{p,t+1}\}\; r_{i,t+1}] = 0$$

its sample version is

$$\frac{1}{T}\sum_{t=1}^{T}\{1 - m(Z_t,\theta)r_{p,t+1}\}\ r_{i,t+1} = 0. \tag{3.3}$$

Let $\hat{e}^p_{i,t+1}$ be the estimated pricing error in parametric model; that is, $\hat{\xi}_t = \hat{e}^p_{i,t+1} = \hat{m}_{t+1}\, r_{i,t+1}$, where $\hat{m}_{t+1} = 1 - m(Z_t,\hat{\theta})r_{p,t+1}$. I use GMM method of Hensen (1982) to estimate the unknown parameter θ and the formulation of this estimation problem is as follows. Under one factor model, I have (L×1) vector-valued function

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta, Z_t, r_{p,t+1}, r_{i,t+1}),$$

where

$$g_t(\theta, y_t) \equiv g_t(\theta, Z_t, r_{p,t+1}, r_{i,t+1}) = Z_t(1 - m(Z_t,\theta)r_{p,t+1})r_{i,t+1}$$

Under H_0, the true value of θ is characterized by

$$E[g_t(\theta, y_t)] = 0.$$

To choose θ so as to make the sample moment $g_t(\theta, y_t)$ as close as possible to the population moment of zero, the GMM estimator $\hat{\theta}$ is

$$\hat{\theta} = \text{argmin}\ g_T^{\top}(\theta)W^{-1}g_T(\theta), \tag{3.4}$$

where W is a weighting matrices. The optimal weighting matrix is the asymptotic variance of $\sqrt{T}g_T(\cdot)$. If the vector process $g_t(\theta, y_t)$ is serially uncorrelated, then the matrix $W = \lim_{T\to\infty} E(g_T(\theta)g_T(\theta)^{\top})$ can be consistently estimated by

$$W_T = (1/T)\sum_t [g_t(\theta, y_t)][g_t(\theta, y_t)]^{\top}$$

Of course, W should be estimated by Newey-West (1987) when $g_t(\theta, y_t)$ is serially correlated

$$W_T \equiv \sum_{\upsilon=-\infty}^{\infty} \tau_\upsilon, \quad \tau_\upsilon = E(g_t(\theta, y_t) g_t^\top(\theta, y_{t-\upsilon})), \tag{3.5}$$

and

$$\hat{W}_T = \tau_{0,T} + \sum_{\upsilon=1}^{q} \{1 - [\upsilon/(q+1)]\}(\hat{\tau}_{\upsilon,T} + \hat{\tau}'_{\upsilon,T}), \tag{3.6}$$

where

$$\hat{\tau}_{\upsilon,T} = (1/T) \sum_{t=\upsilon+1}^{T} [g_t(\hat{\theta}, y_t)][g_t(\hat{\theta}, y_{t-\upsilon})]. \tag{3.7}$$

Clearly, it turns out that

$$W_T \,{}^{p} \longrightarrow W.$$

Indeed, Hansen (1982) showed that under some regularity conditions $(\hat{\theta} - \theta) = o_p(1)$, $\hat{\theta}$ is asymptotically normally distributed and $\sqrt{T} g_T(\theta, Z_t, r_{p,t+1}, r_{i,t+1}) \stackrel{d}{\longrightarrow} N(0, W)$.

To test $E(\xi_t | \, \Omega_t) = E((1 - m(\theta, Z_t) r_{p,t+1}) r_{i,t+1} \,|\, \Omega_t) = 0$, Wang (2002, 2003) considered a simple test as follows. First, to run a multiple regression

$$\hat{\xi}_t = \hat{e}^p_{i,t+1} = I_t^T \, \delta_i + v_{i,t+1},$$

where I_t is a $q \times 1$ $(q \geq k)$ vector of observed variables from Ω_t,[1] and then test if all the regression coefficients are zero; that is, $H_0 : \delta_1 = \cdots = \delta_q = 0$ by using a F-test. Also, Wang (2002) discussed two alternative test procedures. Indeed, the above model can be viewed as a liner approximation of $E[\xi_t \,|\, I_t]$. To examine the magnitude of pricing errors, Ghysels (1998) considered the mean square error (MSE) as a criterion to test if the conditional CAPM or APT model is misspecified relative to the unconditional one.

[1]Wang (2003) took I_t to be Z_t in his empirical analysis.

To check the misspecification of the model, I am going to construct test based on U-Statistics technique. This test combines the methodology of the conditional moment test and nonparametric techniques and it is similar to the test in Zheng (1996). In this section, with the help of central limit theory and U-statistics theories, I am able to derive the asymptotic normality under the null and alternative hypothesis. I am interested in testing the null hypothesis

$$H_0 : E(\xi_t | \, I_t) = 0 \quad \text{versus} \quad H_a : E(\xi_t | \, I_t) \neq 0, \tag{3.8}$$

where I_t is a $q \times 1$ ($q \geq k$) vector of observed variables from Ω_t. Similar to Wang (2003), we take I_t to be Z_t. It is clear that $E(\xi_t | \, Z_t) = 0$ if and only if $\left[E(\xi_t | \, Z_t)\right]^2 f(Z_t) = 0$ if and only if $E\left[\xi_t E(\xi_t | \, Z_t) f(Z_t)\right] = 0$. Interestingly, the testing problem on conditional moment becomes unconditional. Obviously, the test statistic could be postulated as

$$\frac{1}{T} \sum_{t=1}^{T} \xi_t E(\xi_t | \, Z_t) f(Z_t), \tag{3.9}$$

if ξ_t and $E(\xi_t | \, Z_t) f(Z_t)$ would be known. Since $E(\xi_t | \, Z_t) f(Z_t)$ is unknown, its leave-one-out Nadaraya-Watson estimator can be formulated as

$$E(\xi_t \widehat{| \, Z_t) f}(Z_t) = \frac{1}{T(T-1)} \sum_{s \neq t}^{T} \xi_t \, K_h(Z_s - Z_t). \tag{3.10}$$

Plugging (3.10) into (3.9) and replacing ξ_t by its estimate $\hat{\xi}_t$, I obtain the test statistic, which is indeed similar to that in Zheng (1996) for classical regression settings, denoted by U_T, which is a second order U-statistics

$$U_T = \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t) \hat{\xi}_t \hat{\xi}_s \tag{3.11}$$

as shown later.

3.3 Distribution Theory

In this subsection, I discuss the large sample theory for the proposed test statistic.

3.3.1 Assumptions

Assumption C:

C1. **(i)** For each t and s, $E(r_{p,t+1}) < \infty$, $E(r_{i,t+1}) < \infty$, $E(r_{i,t+1}^2 r_{p,t+1}^2) < \infty$, $E(\xi_t|Z_s, Z_t) < \infty$, and $E(\xi_t^2|Z_t) < \infty$.

(ii) $m(.,\theta)$ and $\frac{\partial m}{\partial \theta}(\cdot,\theta)$ are Borel measurable for each θ, $m(z,\theta)$ and $\frac{\partial m}{\partial \theta}(\cdot,\theta)$ are continuous for each $z \subseteq R^L$, and $E(\xi_t)$ and $E(\frac{\partial m}{\partial \theta})^2$ exist and are finite for all θ.

(iii) The parameter space Θ is a compact and convex subset of R^L, and $E[\xi^2(\theta)]$ takes a unique minimum at $\theta_0 \in \Theta$. Under H_0, θ_0 is an interior point of Θ, $E[\sup_{\theta\in\Theta} m^2(z,\theta)] < \infty$, $E[\sup_{\theta\in\Theta} \parallel \frac{\partial m(z,\theta)}{\partial \theta} . \frac{\partial m(z,\theta)}{\partial \theta'} \parallel] < \infty$, and the matrix $E[\frac{\partial \xi(\theta_0,z)}{\partial \theta} . \frac{\partial \xi(\theta_0,z)}{\partial \theta'}]$ is nonsingular.

(iv) $A(z) = E(\epsilon_t|Z_t = z)$ and $\sigma_0^2(z) = E(\epsilon_t^2|Z_t = z)$ satisfy Lipschitz conditions. There exists some $\delta > 0$ such that $E\{|\epsilon_t|^{2+\delta}|Z_t = z\}$ is continuous at z.

(v) Let $f_\tau(.,.)$ be the joint probability density function of Z_1 and Z_τ. Then for all τ, $f_\tau(.,.)$ exists and satisfies a Lipschitz condition.

C2. The process $\{Z_t, r_{i,t+1}, r_{p,t+1}, \xi_t\}$ is strictly stationary and absolutely regular with mixing coefficient $\beta(k)$ satisfying $\sum_{k=1}^{\infty} k^2\beta(k) < \infty$.

C3. For some δ, $Th^{\frac{1+3\delta}{1+\delta}L} \to \infty$ and $Th^{4L} \to \infty$ as $T \to \infty$

C4. $\kappa_j = E[\xi_0|\xi_{-j}, \xi_{-j-1}...] - E[\xi_0|\xi_{-j-1}, \xi_{-j-2}...]$ for $j \geq 0$, $\sum_{j=0}^{\infty} E[\kappa_j^2]^{1/2}$ is finite.

C5. The kernel $K(\cdot)$ is symmetric, bounded and compactly supported.

3.3.2 Large Sample Theory

Due to the lack of the asymptotic result for a U-statistic under α-mixing context, I consider only β-mixing given in Assumption C2. As aforementioned, the test statistic U_T given in (3.11) is a U-statistic. First, I decompose ξ_t to be

$$\hat{\xi}_t = \epsilon_t + \mu_t - b_t,$$

where

$$\epsilon_t = (1 - m(Z_t) r_{p,t+1}) r_{i,t+1},$$

$$\mu_t = (m(Z_t) - m(Z_t, \theta)) r_{p,t+1} r_{i,t+1},$$

and

$$b_t = (m(Z_t, \hat{\theta}) - m(Z_t, \theta)) r_{p,t+1} r_{i,t+1}.$$

Hence, the test statistic U_T is decomposed as

$$\begin{aligned} U_T &= \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t) \hat{\xi}_s \hat{\xi}_t \\ &= \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t) [\epsilon_t \epsilon_s - 2\epsilon_t b_s + b_s b_t] \\ &+ \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t) [\mu_t \mu_s - 2\mu_t b_s + 2\epsilon_s \mu_t] \\ &\equiv (L_{T1} - 2L_{T2} + L_{T3}) + (L_{T4} - 2L_{T5} + 2L_{T6}) \\ &\equiv U_{T1} + U_{T2}. \end{aligned} \tag{3.12}$$

It is showed in the proof of Theorem 3.1 that under H_0, U_{T2} is a higher order by comparing with U_{T1} and U_{T1} is asymptotically unbiased and it is a degenerate U-statistics. To prove the asymptotic result under the null hypothesis, I use martingale central limit theorem which is similar to Hjellvik etc (1998), and the proof is presented in the Appendix.

Theorem 3.1. *If Assumptions C1 - C5 are satisfied, then under the null hypothesis, one has*

$$T\, h^{L/2} U_{T1} \xrightarrow{d} N(0, \Sigma_A),$$

where $\Sigma_A = 2\int[\mu_2(K)\sigma_0^4(z)]f^2(z)dz$ *with* $\sigma_0^2(z) = E(\epsilon_t^2|Z_t = z)$ *and*

$$\hat{\Sigma}_A = 2/(T^2h^L)\sum_{s\neq t} K^2(\frac{Z_s - Z_t}{h})\hat{\xi}_t^2\hat{\xi}_s^2$$

is a consistent estimator of Σ_A.

Next, we now discuss the asymptotic properties of the statistic U_T under the alternative. By Proposition 2.3, it is easy to show that $\sqrt{T}U_{T1} = o_p(1)$. Hence, the dominating term in U_T under H_a is still U_{T2}.

Theorem 3.2. *If Assumptions C1 - C5 are satisfied, under the alternative hypothesis,* U_{T2} *is asymptotically normally distributed*

$$[U_{T2} - E(U_{T2})]/\sqrt{Var(U_{T2})}\ {}^{d}\longrightarrow N(0,\ 1)$$

and

$$Var(U_{T2}) = \frac{4}{T}\{\,Var(h_T^{(1)}(V_1)) + 2\sum_{t=1}^{T-1} Cov\left(h_T^{(1)}(V_1), h_T^{(1)}(V_{t+1})\right)\} + o(T^{-1}) \rightarrow \Sigma_B,$$

where

$$V_t = [Z_t, \xi_t], \quad h_T(V_s, V_t) = K_h(Z_s - Z_t)[\mu_t\mu_s + \epsilon_t\mu_s + \epsilon_s\mu_t - \mu_t b_s - \mu_s b_t],$$

and

$$\gamma_T = E^{\otimes} h_T(V_s, V_t) \ , \ \ h_T^{(1)} = E(h_T(v, V_t)) - \gamma_T.$$

Moreover, since $\sqrt{T}\, U_{T_1} = o_p(1)$, it is easy to establish the asymptotic normality for the test statistic U_T

$$\sqrt{T}(U_T - E(U_T)) \ \stackrel{d}{\longrightarrow} \ N(0, \ \Sigma_B).$$

Now I am going to investigate the power of the test under local misidentification which departures from the null hypothesis under the rate $\delta_T = T^{-1/2} h^{-L/4}$. It is assumed that the local alternative is given by

$$\begin{aligned} H_0 \ &: \ E[(1 - m(Z_t, \theta) r_{p,t+1}) r_{i,t+1} | Z_t = z] = 0, \\ H_a \ &: \ E[(1 - m(Z_t, \theta) r_{p,t+1}) r_{i,t+1} | Z_t = z] = \delta_T \, q(z), \end{aligned} \tag{3.13}$$

where $m(\cdot)$ is a known function and $q(\cdot)$ is bounded and continuously differentiable.

Theorem 3.3. *If Assumptions C1 - C5, are satisfied and $\delta_T = T^{-1/2} h^{-L/4}$, under the local alternative hypothesis (3.13), one has*

$$T h^{L/2}(U_T) \ \stackrel{d}{\longrightarrow} \ N(\int q^2(z) f^2(z) dz, \ \Sigma_A),$$

where Σ_A is given in Theorem 3.1.

3.3.3 Comparing Nonparametric with Parametric Pricing Kernel

It is interesting to compare the nonparametric with parametric pricing kernel model. Similarly, we assume that

$$\begin{aligned} H_0 &: E[(1 - m(Z_t) r_{p,t+1}) r_{i,t+1} | Z_t = z] = 0, \\ H_a &: E[(1 - m(Z_t) r_{p,t+1}) r_{i,t+1} | Z_t = z] = q(z), \end{aligned} \tag{3.14}$$

where where $m(\cdot)$ is unspecified function and $q(\cdot)$ is bounded and continuously differentiable. If we use the estimated NPGEE pricing error to construct test statistic $U_T^N = \frac{1}{T}\sum_{t\neq s} K(Z_t - Z_s)\hat{e}_t\hat{e}_s$, where $\hat{e}_t = (1 - \hat{m}(Z_t)r_{p,t+1})r_{i,t+1}$, the following theorem shows that, the test will have no power against the alternative hypotheses.

Theorem 3.4. *If Assumptions A1 - A5 and C1-C5 are satisfied, under the alternative hypothesis (3.14), U_T^N is asymptotically normally distributed*

$$Th^{L/2}(U_T^N) \stackrel{d}{\longrightarrow} N(0,\ \Sigma_C),$$

where $\Sigma_C = 2\int [\mu_2(K)\sigma_a^4] f^2(z)dz$ with $\sigma_a^2 = E(\epsilon_{a,t}^2 | Z_t = z)$ and $\epsilon_{a,t} = \{1 - [m(Z_t) + q(z)/(r_{p,t+1}r_{i,t+1})]r_{p,t+1}\}r_{i,t+1}$, and $\hat{\Sigma}_C = 2/(T^2h^L)\sum_{s\neq t} K^2(Z_s - Z_t)\hat{e}_t^2\hat{e}_s^2$ is consistent estimator of Σ_C.

Remark 8: In Theorem 3.4, we have an alternative hypothesis that conditional expectation of the pricing error $E[e_t|Z_t] = E[(1 - m(Z_t)r_{p,t+1})r_{i,t+1}|Z_t]$ is $q(Z_t)$. Under H_a, we can rewrite this as $\epsilon_{a,t} = \{1 - [m(Z_t) + q(z)/(r_{p,t+1}r_{i,t+1})]r_{p,t+1}\}$, $E[\epsilon_{a,t}|Z_t] = 0$. Hence under NPGEE estimation framework, the orthogonal condition setting mentioned in (2.2) turns out to be $E(1 - [m(Z_t) + q(z)/(r_{p,t+1}r_{i,t+1})]r_{p,t+1}|Z_t) = 0$. The pseudo pricing kernel becomes $m^*(Z_t) = m(Z_t) + E(\frac{q(Z_t)}{r_{p,t+1}r_{i,t+1}}|Z_t)$, meanwhile in the working orthogonality condition (2.2), $\hat{a}$ becomes the estimator of $m^*(z)$ and $\hat{b}$ becomes the estimator of $\nabla m^*(z)$. It is easy to see that $\epsilon_{a,t}$ is the pseudo pricing error where $E(\epsilon_{a,t}|z) = 0$. It is difficult to separate our true pricing kernel $m(z)$ from $q(z)$ since without imposing any constrain or specific form of real pricing kernel $m(.)$, we can not estimate $m(z)$ separately using NPGEE. Moreover in the testing procedure, we test misspecification of the pricing kernel by testing whether pricing error $E(e_t|z_t)$ is zero. However as we have discussed that by using nonparametric estimation technique the pseudo pricing error from orthogonal condition (2.2) has conditional mean zero $E(\epsilon_{a,t}|z) = 0$. It is naturally that in case we use NPGEE estimate $\hat{e}_t$ and plug into the test statistic 3.11, it will not show any power and can not detect the deviance between $m(z)$ and $m^*(z)$.

3.4 Empirical Examples

To investigate how well the proposed test works, I need to calculate the size and power function of the test. In both simulations and real data applications, the bandwidth h_i is chosen to be $\sigma_i\,((L+2)T/4)^{-1/(L+4)}$ suggested by Zhang (2004), where σ_i is the standard deviation for the i^{th} variable, and L is the dimension. I choose the kernel function to be the product kernel as $K(u, v) = K(u)K(v)$, where $K(\cdot)$ is the standard normal density function.

3.4.1 Simulated Examples

To illustrate the methods proposed, I consider two simulated examples, one is testing for linearity and the other is testing for mean-variance efficient condition.

Example 3.1 (Testing for Linearity): For simplicity of implementation, I choose only two covariates $Z_t = (Z_{1t},\, Z_{2t})^T$ and Z_t follows a vector autoregressive (VAR) model as

$$Z_t = \mu + \Phi\, Z_{t-1} + \eta_t, \tag{3.15}$$

T	α	Rejection rate
T=400	$\alpha = 0.10$	0.089
	$\alpha = 0.05$	0.046
	$\alpha = 0.01$	0.02
T=800	$\alpha = 0.10$	0.092
	$\alpha = 0.05$	0.053
	$\alpha = 0.01$	0.016
T=1200	$\alpha = 0.10$	0.097
	$\alpha = 0.05$	0.049
	$\alpha = 0.01$	0.012

Table 3.1: Simulation results with three sample sizes $T = 400$, 800, 1200 for three significance levels $\alpha = 0.10$, 0.05, and 0.01.

where

$$\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Phi = \begin{pmatrix} 0.4 & 0.3 \\ 0.5 & 0.35 \end{pmatrix}$$

and η_t is standard normally distributed.

$$r_{p,t+1} = 1 + Z_{1t}^2 + Z_{2t}^2 + \epsilon_t \equiv g(Z_t) + \epsilon_t,$$

where ϵ_t is standard normally distributed and $m(Z_t)$ is taken to be $m(Z_t) = \theta^\top Z_t$. First, I consider the test size. Toward the end, under H_0, $r_{i,t+1}$ is generated by

$$[1 - m(Z_t) r_{p,t+1}]\, r_{i,t+1} = e_t,$$

where $e_t = \rho_1 e_{t-1} + v_t$, and v_t is also standard normally distributed and $\rho_1 = 0.05$. Hansen's (1982) GMM is used for estimating θ, and the pricing error is

$$\hat{e}_t = [1 - (\hat{\theta}^\top Z_t) r_{p,t+1}] r_{i,t+1}.$$

Define

$$Y_T = \frac{Th^{L/2}U_T}{\sqrt{\hat{\sigma}_T}} = \frac{\sum_{s\neq t} K(\frac{Z_s - Z_t}{h})\hat{e}_s\hat{e}_t}{\{\sum_{s\neq t} 2K^2(\frac{Z_s-Z_t}{h})\hat{e}_s^2\hat{e}_t^2\}^{1/2}}. \tag{3.16}$$

Then, Theorem 3.1 implies that the asymptotic distribution of Y_T is $N(0,1)$ under H_0. For each sample size of 400, 800, and 1200, the simulation is replicated 1000 times. For each level of significance α, based on the asymptotic distribution of the test statistic, the rejection rate is computed. The simulation results are reported in Table 3.1, from which it is easy to see that the test size performs fairly well.

Next I consider the test power. To this effect, I will evaluate the power under a sequence of alternative models:

$$H_a : E[\{1 - m_0(Z_t, \theta)r_{p,t+1}\}r_{i,t+1}|Z_t] = \delta\, Z_{1t}^2$$

parameterized by δ, where Z_t is generated based on equation (3.15). Here, $m_0(Z_t, \theta) = \theta' Z_t$ is a linear form, where $\theta = (1, 3)^T$ and $r_{i,t+1}$ is generated by

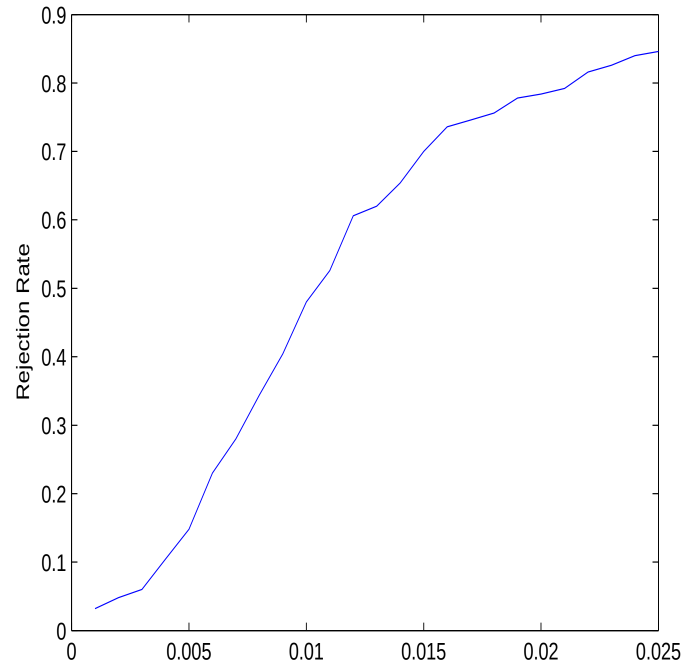
$$r_{it+1} = [\epsilon_t + \delta(Z_{1t}^2)] / [1 - m_0(Z_t, \theta)r_{p,t+1}],$$

where $r_{p,t+1} = Z_{2t} + u_t$, ϵ_t and u_t are standard normally distributed. This sequence of models ranges from the null model ($\delta = 0$) to the models departure from the null model. $\delta(1 + Z_{1t}^2)$ measures the degree of departure from the null hypothesis. The test statistics is defined in (3.16), the rejection frequencies are computed based on 1000 simulations.

Note that when δ is equal to zero, the setting becomes to the null hypotheses. In that case, the power is the significant level. On the other hand, when δ increases, it is expected that the rejection rate of the null hypothesis should become larger. The power functions under significance level $\alpha = 0.05$ are reported in Figure (3.1). The vertical line represents the rejection rate and the horizontal line represents δ

ranging from 0 to 0.025. The result is inline with the theory.

Figure 3.1: Power function $\alpha = 0.05$



Example 3.2 (Testing for Mean-Variance Efficiency:) To further illustrate the testing procedure, I consider testing mean-variance efficient condition as in Wang (2002) for the portfolio. The conditional mean of $r_{p,t+1}$ takes the form

$$E(r_{p,t+1}\,|\,Z_t) = \, 1 + Z_{1t}^2 + Z_{2t}^2 \equiv g(Z_t),$$

and

$$r_{p,t+1} = g(Z_t) + \epsilon_t,$$

where the residual ϵ_t is standard normally distributed. In order to generate $m(Z_t)$ to satisfy the mean-variance efficient condition as in Wang (2002) for the portfolio, I consider

$$E(r_{i,t+1}|\Omega_t) = E(r_{p,t+1}|\Omega_t)\frac{\text{Cov}(r_{i,t+1},\ r_{p,t+1}|\Omega_t)}{\text{Var}(r_{p,t+1}|\Omega_t)},$$

which implies

$$E(r_{i,t+1}|\Omega_t) = E(r_{p,t+1}|\Omega_t)\frac{E(r_{i,t+1}r_{p,t+1}\,|\Omega_t)}{E(r_{p,t+1}^2|\Omega_t)} = m(Z_t)\,E(r_{i,t+1}r_{p,t+1}\,|\Omega_t), \qquad (3.17)$$

where

$$m(Z_t) = E(r_{p,t+1}|Z_t)/E(r_{p,t+1}^2|Z_t) = g(Z_t)/[1+g(Z_t)^2]. \qquad (3.18)$$

Then, $r_{i,t+1}$ is determined by (3.17). In this experiment, under H_0, $r_{i,t+1}$ is generated by

$$[1 - m(Z_t)r_{p,t+1}]\,r_{i,t+1} = e_t, \qquad (3.19)$$

where $e_t = \rho_1 e_{t-1} + v_t$, v_t is also standard normally distributed, and $\rho_1 = 0.05$. $m(\cdot)$ and $r_{i,t+1}$ are generated by (3.17) and (3.19), respectively.

From equation (3.18), it is easy to see $m(\cdot)$ is nonlinear and does not have specific form. Therefore, the standard GMM method is not suitable for the estimation. First, the proposed NPGEE method can be used to estimate the unknown form of $m(\cdot)$. Thus, by plugging the NPGEE estimator $\hat{m}(z_t)$ into (3.19), $\hat{e}_t$ can be obtained and so is the test statistic. For each of sample sizes $T = 400$, 800, and 1200, the simulation is repeated 1000 times. For each level of significance α, the rejection rate is computed and is displayed in Table 3.2. One can see easily from Table 3.2 that the rejection rate is closer to its nominal rate when the sample size is increasing. This implies that the result is consistent with the asymptotic theory.

3.4.2 A Real Example

Similar to the simulated examples, I am now applying the methods to test the linearity of the pricing kernel and the mean-variance efficiency condition. The data are monthly excess returns from January 31, 1966 to December 29, 2006. For the benchmark portfolio, I use NYSE value-weighted (including dividend) as $r_{p,t+1}$ and

T	α	Rejection rate
T=400	$\alpha = 0.10$	0.079
	$\alpha = 0.05$	0.035
	$\alpha = 0.01$	0.004
T=800	$\alpha = 0.10$	0.08
	$\alpha = 0.05$	0.033
	$\alpha = 0.01$	0.007
T=1200	$\alpha = 0.10$	0.094
	$\alpha = 0.05$	0.045
	$\alpha = 0.01$	0.01

Table 3.2: Simulation results with three sample sizes $T = 400$, 800, 1200 for three significance levels $\alpha = 0.10$, 0.05, and 0.01.

Table 3.3: Testing of the linearity

Forecasting Variables	z-statistic	p-value
DPR,DEF	5.4224	< 0.001
DPR,RTB	3.3892	< 0.001
DEF,RTB	10.2596	< 0.001
EWR,DPR	48.8062	< 0.001

the value-weighted NYSE size decile 1 (SZ1) is used as asset $r_{i,t+1}$. The covariates are chosen to be the logarithm of dividend-price ratio (DPR), the logarithm of default premium (DEF), the logarithm of the one month treasury bill rate (RTB), and the excess return on NYSE equally weighted index (EWR).

First, I consider testing whether the pricing kernel has a linear form $m(Z_t) = \theta^\top Z_t$. To do so, I use the GMM method to obtain the estimated coefficient $\hat{\theta}$ and pricing error $\hat{\xi}$ and then compute the test statistic. Each time I take two variables as covariates, and only take one risky portfolio as our test asset $r_{i,t+1}$. The results are given in Table (3.3). All tests on linearity of the pricing kernel are rejected. This coincides with unstable and nonlinear curved surfaces shown in Figure 2.5 and similar to the findings in Ghysels (1998) and Wang (2002, 2003) that the pricing kernel is nonlinear.

Table 3.4: Testing the conditional mean-variance efficiency of the portfolio.

Forecasting Variables	z-statistic	p-value
DPR,DEF	-0.6375	0.2619
DPR,RTB	-0.0175	0.4930
DEF,RTB	0.0711	0.5283
EWR,DPR	16.4524	< 0.001

Next, I would like to test the conditional mean-variance efficiency. As discussed in Section (2.1), when $m(\cdot)$ is a scale function if benchmark portfolio is conditionally mean-variance efficient, $m(Z_t) = \frac{E(r_{p,t+1}|Z_t)}{E(r_{p,t+1}^2|Z_t)}$. A natural way to test the condition is to use local constant method to get the consistent estimate of $\hat{b}(Z_t) = \frac{\widehat{E}(r_{p,t+1}|Z_t)}{\widehat{E}(r_{p,t+1}^2|Z_t)}$, where

$$\widehat{E}(r_{p,t+1}|z) = \frac{\sum_t^T K_h(Z_t - z) r_{p,t+1}}{\sum_t^T K_h(Z_t - z)} \quad \text{and} \quad \widehat{E}(r_{p,t+1}^2|z) = \frac{\sum_t^T K_h(Z_t - z) r_{p,t+1}^2}{\sum_t^T K_h(Z_t - z)},$$

and $\hat{\varepsilon}_t = (1 - \hat{b}(Z_t) r_{p,t+1}) r_{i,t+1}$ to check whether the orthogonal condition (3.1) is satisfied. Table (3.4) presents the testing results based on nonparametric local constant estimate. When only two covariates DPR and EWR are used as the forecasting variables, the null hypothesis is strongly rejected, but for other cases, one would accept that the NYSE market proxies satisfy the conditional orthogonality, i.e., conditionally mean-variance efficiency.

Remark 10 (Comparing with the results in Wang (2002)): Wang takes the vector of conditioning variables as x_t = (DPR, DEF, RTB, EWR)' and the instrument vector is $z_t = (1 \ \ x_t')'$ for constructing the tests. The benchmark portfolio $r_{p,t+1}$ is selected as NYSE value-weighted portfolio. He tests the mean-variance efficiency by choosing the risky assets as five NYSE size portfolios, $r_{i,t+1}$ is a vector. Since in our test, we only use one test asset, $r_{i,t+1}$ is scalar, intuitively we will have smaller chance to reject the null hypothesis of mean-variance efficiency.

Wang gives three different test statistics from the chi-squared distribution and in each sample time interval case, UM-tests produce much higher p-value than the R2-tests. The difference of the p-value range from 10.8% to 52.0%. However, Wang claims that R1-tests producing the strongest rejections when he increases the number of test assets. During 1971-1995, both R1 and UM test failed to reject mean-variance efficiency hypotheses, while during 1947-1970, both R1 and UM test provide strong rejections.

3.5 Proof of Theorems

To prove the theorems, the following lemmas are needed and listed below without their proof. Indeed, Lemma 3.1 is Lemma A in Hjellvik (1999), Lemma 3.3 is Lemma 1 in Yoshihara (1976), and Lemma 3.4 is Lemma A.0 in Fan and Li (1999).

Lemma 3.1. *Let $\psi(.,.)$ be a symmetric Borel function defined on $R^p \times R^p$. Assume that for any fixed $x \in R^p, E\{\psi(\varsigma_1, x)\} = 0$. Then,*

$$E\{\sum_{1\le i<j\le n} \psi(\varsigma_i, \varsigma_j)\}^2 \le cn^2\{M^{\frac{1}{1+\delta}} \sum_{k=1}^{n} k\beta^{\frac{\delta}{1+\delta}}(k), \max_{i>1} E[\psi(\varsigma_1, \varsigma_i)]^2\},$$

where $\delta > 0$ is a constant,

$$M = \max_{1<i\le n} \max\{E|\psi(\varsigma_1, \varsigma_i)|^{2(1+\delta)}, \int |\psi(\varsigma_1, \varsigma_i)|^{2(1+\delta)} dP(\varsigma_1) dP(\varsigma_i)\}$$

and $c > 0$ is a constant independent of n and the function ψ.

Lemma 3.2. *Let $\psi(.,.,.)$ be a symmetric Borel function defined on $R^p \times R^p \times R^p$. Assume that for any fixed $x, y \in R^p$, $E\{\psi(\varsigma_1, x, y)\} = 0$. Then,*

$$E\{\sum_{1\le i<j<k\le n} \psi(\varsigma_i, \varsigma_j, \varsigma_k)\}^2 \le cn^2\{M^{\frac{1}{1+\delta}} \sum_{k=1}^{n} k^2\beta^{\frac{\delta}{1+\delta}}(k), \max_{i>1} E[\psi(\varsigma_1, \varsigma_i, \varsigma_j)]^2\},$$

where $\delta > 0$ is a constant,

$$M = \max_{1<i\leq n} \max\{E|\psi(\varsigma_1,\varsigma_i,\varsigma_j)|^{2(1+\delta)}, \int |\psi(\varsigma_1,\varsigma_i,\varsigma_j)|^{2(1+\delta)} dP(\varsigma_1)dP(\varsigma_i,\varsigma_j),$$
$$\int |\psi(\varsigma_1,\varsigma_i,\varsigma_j)|^{2(1+\delta)} dP(\varsigma_1,\varsigma_i)dP(\varsigma_j), \int |\psi(\varsigma_1,\varsigma_i,\varsigma_j)|^{2(1+\delta)} dP(\varsigma_1)dP(\varsigma_i)dP(\varsigma_j)\}$$

and $c > 0$ is a constant independent of n and the function ψ.

Lemma 3.3. *If $(V_i)_{i\in Z}$ is an absolutely regular process with mixing coefficients $\beta(j)$ and, if for some $\delta > 0$ and $1 \leq j < k$,*

$$M_n = \max\{E|h_n(V_{t_1},\ldots,V_{t_k})|^{1+\delta}, E^{j\otimes}|h_n(V_{t_1},\ldots,V_{t_k})|^{1+\delta}\} < \infty,$$

then,

$$|Eh_n(V_{t_1},\ldots,V_{t_k}) - E^{j\otimes}|h_n(V_{t_1},\ldots,V_{t_k})| \leq 4M_n^{\frac{1}{1+\delta}}\beta^{\frac{\delta}{1+\delta}}(t_{j+1}-t_j),$$

where

$$E^{j\otimes}h_n(V_{t_1},...,Vt_k) = \int h_n dP^{(V_{t_1},...,Vt_k)} \otimes dP^{(V_{t_{j+1}},...,Vt_k)}.$$

Lemma 3.4. *Let $\varepsilon_1,\ldots,\varepsilon_n$ be a random vectors taking values in R^p satisfying an absolute regularity condition with coefficient β_m. Let $h(y,z)$ be a Borel measurable function, $\delta > 0$, $M = \max\{E[|h(\eta,\zeta)|^{1+\delta}], \int\int |h(y,z)|^{1+\delta}Q(dy)R(dz)\}$ exists, where Q and R are probability distributions of η and ζ, respectively. Further, let $g(y) = E[h(y,\zeta)]$. Then,*

$$E|E\{h(\eta,\zeta)\} - g(\eta)| \leq 4M^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)}.$$

Now, I embark on the proofs of the theorems.

Proof of Theorem 3.1: Under H_0, U_{T2} will vanish. The dominating term in U_{T1}

is L_{T1} which can be expresses as a partial sum of martingale differences. I will first show that L_{T2} and L_{T3} are asymptotically of order $o_p(T^{-1}h^{-k/2})$, and L_{T1} is asymptotically normally distributed, which are given in the following proposition.

Proposition 3.1. *Assume that Assumptions C1 - C6 hold, under H_0, one has*

(i) $Th^{L/2}L_{T_1} \xrightarrow{d} N(0, \Sigma_A)$;

(ii) $L_{T_2} = o_p((Th^{L/2})^{-1})$;

(iii) $L_{T_3} = o_p((Th^{L/2})^{-1})$;

(iv) $\hat{\Sigma} = \Sigma + o_p(1)$.

Proof of Proposition 3.1(i):

It follows from (3.11) that

$$\begin{aligned} L_{T1} &= \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t) e_s e_t \\ &= \frac{2}{T(T-1)} \sum_{1 \leq s < t \leq T} K_h(Z_s - Z_t) e_s e_t \\ &\equiv \frac{2}{T(T-1)} L_{T1}^* \end{aligned} \tag{3.20}$$

Define

$$L_{T1}^* = \sum_{t=2}^{T} J_t, \quad J_t = \sum_{s=1}^{t-1} \phi^{(1)}(V_s, V_t), \quad \phi^{(1)} = K_h(Z_s - Z_t)\epsilon_s \epsilon_t,$$

$$\sigma_T^2 = \sum_{1 \leq s < t \leq T} \sigma_{st}^2, \quad \text{and} \quad \sigma_{st}^2 = E(\phi^{(1)2}).$$

Since (V_t) is an absolutely regular process, by Lemma 3.3, it is easy to calculate that

$$\sigma_{st}^2 = h^{-L} \int [\mu_2(K)\sigma_0^4(z)] f^2(z) dz + o(\beta^{\frac{\delta}{1+\delta}}(t-s)h^{-L}), \quad \text{and} \quad \sigma_T^2 = O(T^2 h^{-L}).$$

Under the null hypothesis, L^*_{T1} is a second order degenerate U-Statistic and J_t is a martingale difference. To show the asymptotic normality of the dominate term L^*_{T1}, martingale central limit theorem can be applied; see Shiryaev (1995, Theorem 4 in VII.8). Then, it suffices to prove the following conditions:

$$\text{Var}(L^*_{T_1}) \simeq \sum_{t=2}^{T} E(J_t^2) \sim \sigma_T^2 \tag{3.21}$$

and

$$\sum_{t=2}^{T} J_t^2 \xrightarrow{p} \sigma_T^2, \tag{3.22}$$

and Lindeberg condition:

$$\frac{\sum_{t=2}^{T} E(J_t^2 I_{\{J_t > \sigma_T \epsilon\}} | \mathcal{F}_{t-1})}{\sigma_T^2} \xrightarrow{p} 0 \tag{3.23}$$

and

$$T^2 h^L \text{Var}(L_{T1}) \xrightarrow{p} \Sigma_A. \tag{3.24}$$

Under H_0, L_{T1} is a degenerate U-Statistics. Therefore, $E(J_t) = 0$. By the definition of L^*_{T1},

$$\begin{aligned} \text{Var}(L^*_{T1}) &= \sum_{t=2}^{T} \text{Var}(J_t) + \sum_{t \neq s} cov(J_t, J_s) \\ &= E(\sum_{t=2}^{T} J_t^2) + \sum_{t \neq s} E(\sum_{i=1}^{t-1} \sum_{j=1}^{s-1} \phi_{it}^{(1)} \phi_{js}^{(1)}). \end{aligned}$$

It follows by using the assumption $\sum_{k=1}^{T}\{k^2\beta^{\frac{1}{1+\delta}}(k)\}<\infty$ that

$$\begin{aligned}
&\sum_{t\neq s}E(\sum_{i=1}^{t-1}\sum_{j=1}^{s-1}\phi_{it}^{(1)}\phi_{js}^{(1)})\\
=\ & 2\sum_{2<s<t\leq T}\sum_{i=1}^{t-1}\sum_{j=1}^{s-1}E(\phi_{it}^{(1)}\phi_{js}^{(1)})\\
\leq\ & C\sum_{2<s<t\leq T}t\beta^{\delta/(1+\delta)}(t)s\beta^{\delta/(1+\delta)}(s)\\
\leq\ & \sum_{t}\beta^{\delta/(1+\delta)}(t)\sum_{s}\beta^{\delta/(1+\delta)}(s)=O(1)=o(\sigma_T^2).
\end{aligned}$$

Thus,

$$\mathrm{Var}(L_{T1})\approx E(\sum_{t=2}^{T}J_t^2)=\sum_{1\leq s<t\leq T}\sigma_{st}+2\sum_{1\leq s<t<k\leq T}\phi_{sk}^{(1)}\phi_{tk}^{(1)}.$$

By Lemma 3.3, one can show that

$$\sum_{1\leq s<t<k<T}E(\phi_{sk}^{(1)}\phi_{tk}^{(1)})=o(\sigma_T^2),$$

and

$$E(\phi_{sk}^{(1)}\phi_{tk}^{(1)})\leq 4M_1^{\frac{1}{1+\delta}}\{\beta(t-s)\}^{\frac{\delta}{1+\delta}},$$

where $\delta>0$ is a constant, and

$$M_1=max_{1\leq s<t\leq T}\{E|\phi_{1s}^{(1)}\phi_{st}^{(1)}|^{1+\delta},E^{j\otimes}|\phi_{1s}^{(1)}\phi_{st}^{(1)}|^{1+\delta}\quad j=1,2\}=O(h^{-2\delta L}).$$

Next I will show that $\sum_{t=2}^{T}J_t^2/\sigma_T^2\ \xrightarrow{p}\ 1$. By a simple calculation, one can easily show that

$$
\begin{aligned}
E(\sum_{t=1}^{T} J_t^2 - \sigma_T^2)^2 &= E\{[\sum_{t=1}^{T} J_t^2 - E(\sum_{t=1}^{T} J_t^2)] + [E(\sum_{t=1}^{T} J_t^2) - \sigma_T^2]\}^2 \\
&\leq 2E[\sum_{t=1}^{T} J_t^2 - E(\sum_{t=1}^{T} J_t^2)]^2 + 2E[(\sum_{t=1}^{T} J_t^2) - \sigma_T^2]^2 \\
&\leq 2E\{\sum_{1\leq s<t\leq T} (\phi_{st}^{(1)2} - \sigma_{st}^2)\}^2 + 8E\{\sum_{1\leq s<k<t\leq T} \phi_{st}^{(1)}\phi_{tk}^{(1)}\}^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{1\leq s<k<t<T} E(\phi_{sk}^{(1)}\phi_{tk}^{(1)}) &\leq C\sum_{s=1}^{T-2}\sum_{t=s+1}^{T-1}\sum_{t=k+1}^{T} M_1^{\frac{1}{1+\delta}}\{\beta^{\frac{1}{1+\delta}}(t-s)\} \\
&\leq CM_1^{\frac{1}{1+\delta}}\sum_{s=1}^{T-2}\sum_{t=s+1}^{T-1}\{(t-s)\beta^{\frac{1}{1+\delta}}(t-s)\} \\
&\leq CTM_1^{\frac{1}{1+\delta}}\{\sum_{k=1}^{T} k\beta^{\frac{1}{1+\delta}}(k)\} = O(Th^{-2\delta L}) = o(\sigma_T^2).
\end{aligned}
$$

To build a symmetric third order U-statistic kernel, I project the three dimensional U-statistic onto two dimensional space and define

$$
q_{stk} = 1/3(\phi_{st}^{(1)}\phi_{kt}^{(1)} + \phi_{sk}^{(1)}\phi_{kt}^{(1)} + \phi_{st}^{(1)}\phi_{ks}^{(1)}), \qquad \text{and} \quad q_{sk} = 1/3\int \phi_{st}^{(1)}\phi_{kt}^{(1)}dP(V_t).
$$

It is easy to see that $E\{\sum_{1\leq s<k<t\leq T}(\phi_{kt}^{(1)}\phi_{st}^{(1)})\}^2 = E\{\sum_{1\leq s<k<t\leq T} q_{stk}\}^2$. By applying Lemma 3.1

$$
\begin{aligned}
& E\{\sum_{1\leq s<k<t\leq T} q_{stk}\}^2 \\
\leq\ & 2E\{\sum_{1\leq s<k<t\leq T}(q_{stk} - q_{st} - q_{sk} - q_{tk})\}^2 + 8T^2E\{\sum_{1\leq s<t\leq T} q_{st}\}^2 \\
\leq\ & \{c_1T^3(M_2^{1/1+\delta} + M_3) + c_2T^4(M_5^{1/1+\delta} + M_4), \} = o(\sigma_T^4),
\end{aligned}
$$

where

$$M_2 = max_{1<i<t} max\{E[\phi_{it}^{(1)2} - E(\phi_{it}^{(1)2})]^{2(1+\delta)}, \int [\phi_{it}^{(1)2} - E(\phi_{it}^{(1)2})]^{2(1+\delta)} dP(V_1)dP(V_i)\},$$

$$M_3 = max_{s<t<k} E(q_{stk})^2, \qquad M_4 = max_{s<t} E(q_{st})^2,$$

and

$$M_5 = max_{1<s<t<T} max\{E|\int \phi_{1s}\phi_{1t} dP(V_1)|^{2(1+\delta)}, \int |\phi_{1s}\phi_{1t} dP(V_1)|^{2(1+\delta)} dP(V_s, V_t)\}.$$

By a simple calculation, one has

$$M_2^{1/1+\delta} = O(h^{-2L}), \qquad M_3 = O(1), \qquad M_4 = O(1), \quad \text{and} \quad M_5^{1/1+\delta} = O(h^{-2L+\frac{L}{1+\delta}}).$$

Again by Lemma 3.3, one can show that

$$E\{\sum_{1\le s<t\le T} \phi_{st}^{(1)2} - \sigma_{st}^2\}^2 = O(T^2 h^{-2L}) = o(\sigma_T^2).$$

Therefore,

$$\frac{E(\sum_{t=1}^T J_t^2 - \sigma_T^2)^2}{\sigma_T^4 \epsilon} \ \rightarrow \ 0,$$

which, together with Chebychev's inequality implies that

$$\frac{\sum_{t=1}^T J_t^2}{\sigma_T^2} \xrightarrow{p} \ 1.$$

Similar to the previous proofs, one can show that

$$\frac{\sum_{t=2}^T E(J_t^4)}{\sigma_T^4} \rightarrow 0,$$

and

$$\sum_{t=2}^{T} E(J_t^4 I_{\{J_t>\sigma_T\epsilon\}}|\mathcal{F}_{t-1}) \leq \sum_{t=2}^{T} E(J_t^4|\mathcal{F}_{t-1}).$$

It is clear that by Chebychev's inequality, Lindeberg condition is satisfied

$$\frac{\sum_{t=2}^{T} E(J_t^4 I_{\{J_t>\sigma_T\}}|\mathcal{F}_{t-1})}{\sigma_T^4} \xrightarrow{p} 0.$$

Next, I need to prove that

$$T^2 h^L Var(L_{T1}) \xrightarrow{p} \Sigma_A, \quad \text{and} \quad \Sigma_A = \int \mu_2(K)\sigma_0^4(z) f^2(z)dz.$$

It follows from (3.21), (3.22) and (3.23) that

$$T^2 h^L \text{Var}(L_{T1}) = T^{-2} h^L \text{Var}(L_{T1}^*) = T^{-2} h^L \sum_{s\neq t} E(\phi_{st}^{(1)2}) + o_p(1) = \Sigma_A + o_p(1)$$

It follows by martingale central limit theorem that $Th^{L/2} L_{T1} \xrightarrow{d} N(0, \Sigma_A)$.

Proof of Proposition 3.1(ii): It follows by (3.11) and Lemma 3.1 that

$$\begin{aligned}
E(L_{T2})^2 &= \frac{1}{T^2(T-1)^2} E(\sum_{s\neq t} K_h(Z_t - Z_s)\epsilon_t b_s)^2 \\
&= \frac{4}{T^2(T-1)^2} E(\sum_{1\leq s<t\leq T} K_h(Z_t - Z_s)\epsilon_t m_0'(Z_s, \widetilde{\theta})(\hat{\theta} - \theta) r_{p,s+1} r_{i,s+1})^2 \\
&\leq \frac{C}{T^2(T-1)^2} T^2 M_6^{\frac{1}{1+\delta}} \sum_{k=1}^{T} k\beta^{\frac{\delta}{1+\delta}}(k),
\end{aligned}$$

where

$$m_0(Z_t, \theta) - m_0(Z_t, \hat{\theta}) \sim \frac{\partial m_0}{\partial \theta}(\cdot, \widetilde{\theta})(\hat{\theta} - \theta), \quad \text{and} \quad \widetilde{\theta} \text{ is between } \theta \text{ and } \hat{\theta}.$$

Define

$$\phi^{(2)}(V_s, V_t) = K_h(Z_t - Z_s)\epsilon_t m_0'(Z_s, \widetilde{\theta})(\hat{\theta} - \theta) r_{p,s+1} r_{i,s+1},$$

and

$$M_6 = \max_{1<t\leq T} \max\{E|\phi^{(2)}(V_1, V_t)|^{2(1+\delta)}, \int |\phi^{(2)}(V_1, V_t)|^{2(1+\delta)} dP(V_1) dP(V_t)\}.$$

Hansen proved the consistency and asymptotic normality of GMM estimator of θ. W is weighting matrix which is positive semi-definite. $\hat{W}_T \xrightarrow{p} W$, $\theta_0 \in \Theta$, $g(V_t, \theta)$ is continuous at each θ with probability one. $E[\sup_{\theta in \Theta} ||g(V_t, \theta||] < \infty$. Under global identification condition $WE[g(V_t, \theta)] = 0$ only for $\theta = \theta_0$. Hansen showed that

$$\hat{\theta} \xrightarrow{p} \theta_0$$

Moreover, define $G = E[\nabla_\theta g(V_t, \theta_0)]$, $\Omega = E[g(V_t, \theta_0) g(V_t, \theta_0)^\top]$. Hansen also showed the asymptotic normality

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0,\, (G^\top \Omega G)^{-1}),$$

One can obtain from Holder's inequality and Hansen (1982) that

$$\begin{aligned}
& E|K_h(Z_s - Z_t)\epsilon_s m_0'(Z_t, \widetilde{\theta})(\hat{\theta} - \theta) r_{p,t+1} r_{i,t+1}|^{2(1+\delta)} \\
\leq\ & \{E|K_h(Z_s - Z_t)\epsilon_s m_0'(Z_t, \widetilde{\theta}) r_{p,t+1} r_{i,t+1}|^{2(1+\delta)p}\}^{1/p} \{E(\hat{\theta} - \theta)^{2(1+\delta)q}\}^{1/q} \\
\leq\ & o(h^{-2L(2+2\delta)}) o(T^{-(1+\delta)}).
\end{aligned}$$

Thus, by Assumption C3, one has

$$E(Th^{L/2} L_{T2})^2 = o(1).$$

Similarly, by Chebychev's inequality, it is easy to see that $L_{T2} = o_p(1)$.

Proof of Proposition 3.1(iii): Again by Holder's inequality and Hansen (1982), one has

$$E(L_{T3})^2 = 1/T^2 E(K_h(Z_t - Z_s)b_t b_s)^2 \leq T^{-2} E^{1/2}(K_h(Z_t - Z_s))^2 E^{1/2}(b_t b_s)^2) \leq o(\frac{1}{T^3 h^{L/2}}).$$

Thus $Th^{L/2} L_{T3} = o_p(1)$.

Proof of Proposition 3.1(iv): Recall that

$$\hat{\Sigma}_A = \frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s) \hat{e}_t^2 \hat{e}_s^2.$$

Using similar arguments, one can show that under H_0,

$$\begin{aligned} \hat{\Sigma}_A &= \frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s)[\epsilon_t^2 \epsilon_s^2 \\ &\quad + \epsilon_t^2(\mu_s - b_s)^2 + 2\epsilon_t^2 \epsilon_s(\mu_s - b_s) + (\mu_s - b_s)^2(\mu_t - b_t)^2 \\ &+ 2(\mu_t - b_t)^2 \epsilon_s(\mu_s - b_s) + 2\epsilon_t(\mu_t - b_t)\epsilon_s^2 + 2\epsilon_t(\mu_t - b_t)(\mu_s - b_s)^2 \\ &\quad + 4\epsilon_t \epsilon_s(\mu_t - b_t)(\mu_s - b_s)] \\ &= \frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s)\epsilon_t^2 \epsilon_s^2 + o_p(1). \end{aligned}$$

It remains to show that

$$\frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s)\epsilon_t^2 \epsilon_s^2 = \Sigma_A + o_p(1).$$

By Lemma 3.4,

$$\frac{4}{T^4}\{\sum_{s \neq t}[E(h^{-L} K^2(Z_t - Z_s)\epsilon_t^2 \epsilon_s^2) - \Sigma_A]\}^2 \leq \frac{C}{T^4} T^2 M_7 \sum_{k=1}^{T} k \beta^{\frac{\delta}{1+\delta}}(k), \qquad (3.25)$$

where

$$M_7 = max_{1<i\leq n}\max\{E|\phi^{(3)}(V_1,V_s)|^{2(1+\delta)}, \int |\phi^{(3)}(V_1,V_s)|^{2(1+\delta)}dP(V_1)dP(V_s)\}.$$

It follows by Lemma 3.3 that

$$\phi^{(3)}(V_t,V_s) = E(h^{-K}K^2(Z_t-Z_s)\epsilon_t^2\epsilon_s^2) - \Sigma_A \leq \frac{C}{T^2}M_7\sum_{k=1}^{T}k\beta^{\frac{\delta}{1+\delta}}(k) = o(1),$$

which, together with (3.25) implies that

$$E\{\frac{2}{T^2h^K}\sum_{s\neq t}K^2(Z_t-Z_s)\epsilon_t^2\epsilon_s^2 - \Sigma_A\}^2 = o(1),$$

Therefore, $\hat{\Sigma}_A \xrightarrow{p} \Sigma_A$. Thus Proposition 3.1 and Theorem 3.1 is proved.

Proof of Theorem 3.2: Recall that

$$U_{T2} = \frac{1}{T(T-1)}\sum_{s\neq t}K_h(Z_s-Z_t)[\mu_t\mu_s+\epsilon_t\mu_s+\epsilon_s\mu_t-\mu_t b_s-\mu_s b_t].$$

To establish a corresponding result in the non-degenerate case, I use Theorem 2 in Dette and Spreckelsen (2004). To this end, I apply the Hoeffding decomposition method. Now, define

$$U_{T2}^* = \gamma_T + 2\,H_T^{(1)} + H_T^{(2)}, \qquad V_t = \{Z_t,\epsilon_t,\mu_t\}, \qquad \text{and} \quad \gamma_T = E^{\otimes}h_T(V_s,V_t),$$

where $E^{\otimes}$ denotes the expectation with respect to the measure $P^{V_s}\otimes P^{V_t}$,

$$H_T^{(1)} = \frac{1}{T}\sum_{t=1}^{T}h_T^{(1)}(V_t), \text{ and } H_T^{(2)} = \frac{1}{T(T-1)}\sum_{s\neq t}h_T^{(2)}(V_s,V_t)$$

with

$$h_T(V_s, V_t) = K_h(Z_s - Z_t)[\mu_t\mu_s + \epsilon_t\mu_s + \epsilon_s\mu_t - \mu_t b_s - \mu_s b_t],$$

and

$h_T^{(1)} = E(h_T(v, V_t)) - \gamma_T$ and $h_T^{(2)}(v, w) = h_T(v, w) - E(h_T(v, V_t)) - E(h_T(V_s, w)) + \gamma_T$.

To prove the theorem, I need to check whether the conditions in Dette and Spreckelsen (2004) are satisfied which are stated in the following proposition.

Proposition 3.2. *Under Assumptions C1 - C5, one has*

(i) $\gamma_T = \int\{[m(z) - m_0(z, \theta)]E(r_{p,t+1}r_{i,t+1}|Z_t = z)f(z)\}^2 dz + o_p(1);$

(ii) $N_T = O(h^{-L(1+\delta)});$

(iii) $E|h_T^{(1)}(V_t)|^q = O(1), \quad$ *and* $\quad E|h_T^{(2)}(v, 2)|^2 = O(h^{-L});$

(iv) $Var(h_T^{(1)}(V_1)) = O(1).$

Proof: By Lemma 3.4, one can evaluate the conditional expectation of a β-mixing process by an independent process which has the same marginal distribution. By observing Assumptions C1-C5, it follows by a straightforward application of Lemma A.1 in Yoshihara (1976) that

$$h_T^{(1)}(\nu) = \mu E(\mu|z)f(z) + \epsilon E(\mu|z)f(z) - \gamma_T + o_p(1).$$

Similarly, one can show that

$$\gamma_T = \int\{[m(z) - m_0(z, \theta)]E(r_{p,t+1}r_{i,t+1}|Z_t = z)f(z)\}^2 dz + o_p(1).$$

Therefore, Proposition 3.2(i) holds. For Proposition 3.2(ii), on has

$$\begin{aligned} N_T = \quad & \max\{\sup_{s\neq t, t\neq m, m\neq n} E|h_T(V_s,V_t)h_T(V_m,V_n)|^{1+\delta}, \\ & \sup_{s\neq t, t\neq m, m\neq n} E^{j\otimes}|h_T(V_s,V_t)h_T(V_m,V_n)|^{1+\delta}, \text{j=1,2,3}\} = O(h^{-L(1+\delta)}). \end{aligned}$$

Note that the order of $E^{j\otimes}|h_T(Z_s,Z_t)h_T(Z_m,Z_n)|^{1+\delta}$ is the same as

$$\begin{aligned} & E|K_h(Z_s-Z_t)K_h(Z_m-Z_t)[\mu_t\mu_s+\epsilon_t\mu_s+\epsilon_s\mu_t-\mu_t b_s-\mu_s b_t] \\ & [\mu_m\mu_n+\epsilon_m\mu_n+\epsilon_m\mu_n-\mu_m b_n-\mu_m b_n]|^{1+\delta}. \end{aligned}$$

By an application of Holder's inequality with $1/\eta+1/\zeta=1$, and taking $\eta<2/(1+\delta)$, one has

$$\begin{aligned} & E|h_T(Z_s,Z_t)h_T(Z_m,Z_n)|^{1+\delta} \\ = \quad & E|K_h(Z_s-Z_t)K_h(Z_m-Z_t)[\mu_t\mu_s+\epsilon_t\mu_s+\epsilon_s\mu_t-\mu_t b_s-\mu_s b_t] \\ & [\mu_m\mu_n+\epsilon_m\mu_n+\epsilon_n\mu_m-\mu_m b_n-\mu_n b_m]|^{1+\delta} \\ \leq \quad & \{E|K_h(Z_s-Z_t)K_h(Z_m-Z_n)|^{\eta(1+\delta)}\}^{\frac{1}{\eta}}\{E|(\mu_t\mu_s+\epsilon_t\mu_s+\epsilon_s\mu_t-\mu_t b_s-\mu_s b_t) \\ & (\mu_m\mu_n+\epsilon_m\mu_n+\epsilon_n\mu_m-\mu_m b_n-\mu_n b_m|^{\zeta(1+\delta)}\}^{\frac{1}{\zeta}} \\ = \quad & O(h^{(-2L\eta(1+\delta)+2L)/\eta})*O(1)=O(h^{-L(1+\delta)}). \end{aligned}$$

For other cases, using the same techniques, one can obtain

$$E^{j\otimes}|h_T(Z_i,Z_j)h_T(Z_k,Z_l)|^{1+\delta}=O(h^{-L(1+\delta)}),\ \ j=1,2,3.$$

This proves Proposition 3.2(ii). To show Proposition 3.2(iii), using C_τ-inequality, one has

$$E|h_T^{(1)}(V_t)|^q \leq E[\mu E(\mu|z)f(z)]^q+\gamma_T^q+o(1)=O(1).$$

Similarly,

$$E|h_T^{(2)}(v,w)|^2 \leq C_1 E|h_T(v,w)|^2 + C_2 E|h_T^{(1)}(v)|^2 + C_3\gamma_T^2, \tag{3.26}$$

and

$$\begin{aligned} & E|h_T(V_s,V_t)|^2 = E\{K_h(Z_s-Z_t)[\mu_t\mu_s+\epsilon_t\mu_s+\epsilon_s\mu_t-\mu_t b_s-\mu_s b_t]\}^2 \\ = \ & O(h^{-L}). \end{aligned} \tag{3.27}$$

Thus, it follows by (3.26) and (3.27) that $E|h_T^{(2)}(v,w)|^2 = O(h^{-L})$. Finally,

$$\begin{aligned} \mathrm{Var}(h_T^{(1)}(V_1)) &= \int\{[m(z)-m_0(z,\theta)]^2 E^2(r_{p,t+1}r_{i,t+1}|Z_t=z)\}^2 f^3(z) \\ &+ \{E(\epsilon^2|z)[m(z)-m_0(z,\theta)]^2 E^2(r_{p,t+1}r_{i,t+1}|Z_t=z)\}f^3(z)dz - \gamma_T^2 \\ &= O(1), \end{aligned}$$

and

$$\mathrm{Cov}(h_T^{(1)}(V_1), h_T^{(1)}(V_{t+1})) = E[E^2(u|z_1)E^2(u|z_{t+1})f(z_1)f(z_{t+1})] - \gamma_T^2 + o_p(1).$$

The proof of Proposition 3.2 is complete. Therefore, the conditions in Dette and Spreckelsen (2004) are satisfied. Thus, Theorem 3.2 is proved.

Proof of Theorem 3.3: Firstly, it is easy to see that

$$U_T = U_{T_1} + M_{T_1} + M_{T_2},$$

where

$$M_{T_1} = \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t)[\mu_t(b_s + \epsilon_s) + \mu_s(b_t + \epsilon_t)]$$

and

$$M_{T_2} = \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t)\mu_t\mu_s.$$

Proposition 3.3. *Under local alternative, if* $E(\mu_t|Z_t) = \delta_T q(Z_t)$ *and* $\delta_T = T^{-1/2}h^{-L/4}$, *one has*

$$Th^{L/2}M_{T_1} = o_p(1), \quad \textit{and} \quad Th^{L/2}M_{T_2} \xrightarrow{p} E[q^2(z)f(z)].$$

Proof of Proposition 3.3: M_{T_1} is a non-degenerate U-statistic. Similarly to proof of Proposition 3.2, one can deal with conditional probability of $P(V_s|V_t)$ by treating V_s and V_t as independent using Lemma 3.4.

$$h_{T2}^{(1)}(v) = E(K_h(Z_s - Z_t)[\mu_t(b_s + \epsilon_s) + \mu_s(b_t + \epsilon_t)|V_s = v) = \epsilon\delta_T q(z)f(z) + o(1).$$

Hence when $\delta_T = T^{-1/2}h^{-L/4}$,

$$\begin{aligned}\mathrm{Var}(M_{T_1}) &= \frac{4}{T}\{\mathrm{Var}(h_{T2}^{(1)}(V_1)) + 2\sum_{t=1}^{T-1} cov(h_{T2}^{(1)}(V_1), h_{T2}^{(1)}(V_{t+1}))\} + o(T^{-1}) \\ &= O(T^{-1}\delta_T^2) = O(T^{-2}h^{-k}).\end{aligned}$$

Thus,

$$Th^{L/2}M_{T_1} = o_p(1).$$

Since M_{T2} is a non-degenerate U-statistic with

$$h_{T3}^{(1)}(v) = E(K_h(Z_s - Z_t)\mu_s\mu_t|V_s = v) = \mu E(\mu|z)f(z),$$

it is clear that

$$E(M_{T_2}) = \int E^2(\mu|z)f^2(z)dz = \int \delta_T^2 q^2(z)^2(z)f^2(z)dz. \tag{3.28}$$

and

$$\begin{aligned}\mathrm{Var}(M_{T_2}) &= \frac{4}{T}\{\mathrm{Var}(h_{T3}^{(1)}(V_1)) + 2\sum_{t=1}^{T-1} cov(h_{T3}^{(1)}(V_1), h_{T3}^{(1)}(V_{t+1}))\} + o(T^{-1}) \\ &= O(T^{-3}h^{-L}).\end{aligned}$$

Thus,

$$\mathrm{Var}(Th^{L/2}M_{T_2}) = o(1),$$

which, together with (3.28), implies that

$$Th^{L/2}M_{T_2} \xrightarrow{p} \int \delta_T^2 q^2(z)(z)f^2(z)dz.$$

By Proposition 3.3, it is easy to show that under local alternative hypothesis,

$$Th^{k/2}(U_T - E(U_T)) \xrightarrow{d} N(\int \delta_T^2 q(z)^2 f^2(z)dz,\ \Sigma_A).$$

This proves Theorem 3.3.

Proof of Theorem 3.4: First, decompose $\hat{e}_t$ to be

$$\hat{e}_t = \epsilon_{a,t} + n_t$$

where

$$\epsilon_{a,t} = \{1 - [m(Z_t) + q(z)/(r_{p,t+1}r_{i,t+1})]r_{p,t+1}\}r_{i,t+1}$$

and

$$n_t = \hat{m}(Z_t) - [m(Z_t) + q(z)/(r_{p,t+1}r_{i,t+1})].$$

Then,

$$\begin{aligned} U_T^N &= \frac{1}{T(T-1)}\sum_{s\neq t} K_h(Z_s - Z_t)\hat{e}_s\hat{e}_t \\ &= \frac{1}{T(T-1)}\sum_{s\neq t} K_h(Z_s - Z_t)[\epsilon_{a,t}\epsilon_{a,s} + (\epsilon_{a,t}n_s + \epsilon_{a,s}n_t) + n_s n_t] \\ &= N_{T1} + N_{T2} + N_{T3}. \end{aligned} \quad (3.29)$$

Suppose $m(Z_t)$ inside the pricing kennel satisfies H_a, there exists $m^N(Z_t) = m(Z_t) + q(Z_t)/(r_{p,t+1}r_{i,t+1})$ such that $E(\epsilon_{a,t}|Z_t) = 0$. By comparing with the parametric model, $\epsilon_{a,t}$ and ϵ_t have the same property. In the following Proposition, it shows that the asymptotic distributions of N_{T1} and L^*_{T1} are same.

Proposition 3.4. *Under the alternative hypothesis setting given in 3.14, if Assumptions A1 - A5 are satisfied, one has*

(i) $Th^{L/2}N_{T1} \xrightarrow{d} N(0, \Sigma_c)$,

(ii) $N_{T2} = o_p((Th^{L/2})^{-1})$,

(iii) $N_{T3} = o_p((Th^{L/2})^{-1})$,

(iv) $\hat{\Sigma}_c = \Sigma_c + o_p(1)$.

Proof of Proposition 3.4: Recall that by decomposition given in (3.29), after adding the deviance part of the model back to the pricing kernel, the conditional expectation of the pricing error is zero; that is $E(\epsilon_{a,t}|Z_t) = 0$ so that N_{T1} is a degenerate U-statistic. The process $V_t^N = (Z_t, r_{p,t+1}, r_{i,t+1}, n_t, \epsilon_{a,t})$ is absolutely regular with mixing coefficient satisfying $\sum_{k=1}^T k^2\beta(k) < \infty$. The dominating term N_{T1} is asymptotically normally distributed by the central limit theorem for a degenerate

U-statistic. The nonparametric GEE method actually gives the estimate of $m^N(\cdot)$ inside the pricing kernel $m_{t+1}^N = 1 - m^N(Z_t) r_{i,t+1}$.

Note that the rest proof of Proposition 3.4 is the same as that for Proposition 3.1. Therefore, details are omitted. Define

$$N_{T1} = \sum_{t=2}^{T} J_{a,t}, \quad J_{a,t} = \sum_{s=1}^{t-1} \phi_a^{(1)}(V_s, V_t), \quad \phi_a^{(1)} = K_h(Z_s - Z_t)\epsilon_s \epsilon_t,$$

$$\sigma_{a,T}^2 = \sum_{1 \le s < t \le T} \sigma_{a,st}^2, \quad \text{and} \quad \sigma_{a,st}^2 = E(\phi_a^{(1)2}).$$

Similar to the proof of Proposition 3.1, one can show that

$$\text{Var}(N_{T1}) \simeq \sum_{t=2}^{T} E(J_{a,t}^2) \; \sim \; \sigma_{a,T}^2,$$

and

$$\sum_{t=2}^{T} J_{a,t}^2 \xrightarrow{p} \sigma_{a,T}^2$$

and Lindeberg condition:

$$\frac{\sum_{t=2}^{T} E(J_{a,t}^2 I_{\{J_{a,t} > \sigma_{a,T}\epsilon\}} | \mathcal{F}_{t-1})}{\sigma_{a,T}^2} \xrightarrow{p} 0,$$

and

$$T^2 h^L Var(N_{T1}) \xrightarrow{p} \Sigma_c.$$

The β-mixing coefficient is always larger than the α-mixing one. Then, if a process is β-mixing, it is α-mixing. Therefore, Assumptions A1-A5 in Theorem 2.12 are still valid. By Theorem 2.12, one can show that $n_t = O_p(h^2)$. It follows by the proof of

Proposition 3.1 and Theorem 3.1 that

$$\begin{aligned} E(N_{T2})^2 &= \frac{1}{T^2(T-1)^2} E(\sum_{s \neq t} K_h(Z_t - Z_s)\epsilon_{a,t} n_s)^2 \\ &= \frac{4}{T^2(T-1)^2} E(\sum_{1 \leq s < t \leq T} K_h(Z_t - Z_s)\epsilon_{a,t} n_s)^2 = O(h^2/(T^2 h^L)), \end{aligned}$$

and

$$E(Th^{L/2} N_{T2})^2 = o(1).$$

By Chebychev's inequality, it is easy to see that $L_{T2} = o_p(1)$ and

$$\begin{aligned} E(N_{T3})^2 &= T^{-2} E(K_h(Z_t - Z_s) n_t n_s)^2 \leq T^{-2} E^{1/2}(K_h(Z_t - Z_s))^2 E^{1/2}(n_t n_s)^2) \\ &\leq o(T^{-2} h^{-L/2-2}). \end{aligned}$$

Hence, $Th^{L/2} N_{T3} = o_p(1)$. Since nonparametric estimation can always give an consistent estimator, term n_t, which is the difference between the nonparametric estimator and $m(Z_t) + q(Z_t)/(r_{p,t+1} r_{i,t+1})$ has small order and it therefore vanishes. Also,

$$\begin{aligned} \hat{\Sigma}_c &= \frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s)[\epsilon_{a,t}^2 \epsilon_{a,s}^2 + \epsilon_{a,t}^2 n_s^2 + 2\epsilon_{a,t}^2 \epsilon_{a,s} n_s + n_s^2 n_t^2 \\ &+ 2n_t \epsilon_{a,t} \epsilon_{a,s}^2 + 2n_s^2 \epsilon_{a,t} n_t + 4 n_t n_s \epsilon_{a,t} \epsilon_{a,s}] \\ &= \frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s) \epsilon_t^2 \epsilon_s^2 + o_p(1). \end{aligned}$$

Similarly, one can show that

$$\frac{2}{T^2 h^L} \sum_{s \neq t} K^2(Z_t - Z_s) \epsilon_{a,t}^2 \epsilon_{a,s}^2 = \Sigma_c + o_p(1).$$

Thus, Proposition 3.4 is proved.

CHAPTER 4: CONCLUSION

In this dissertation, my main goal is to study the nonparametric pricing kernel models. To estimate nonparametric pricing kernel function, I propose a nonparametric estimation procedure, term as nonparametric generalized estimation equations (NPGEE) which combines the local linear fitting and the generalized estimation equations. I establish the asymptotic properties of the resulting estimator. In order to test whether pricing kernel model has some specific parametric form, I propose a consistent nonparametric testing procedure under U-statistic framework.

There are still some related works which can be done in this research area. First, some other possible models could be considered in a similar context, such as a semi-parametric SDF model. Also, the SDF m_{t+1} is unnecessary to be specified as $m_{t+1} = 1 - m(Z_t)r_{p,t+1}$. For example, one can consider a more general setting like $m_{t+1} = m_1(Z_t, r_{p,t+1})$ with the unknown form of SDF. Alternatively, $m_{t+1} = m_1(m(Z_t), r_{p,t+1})$ with known $m_1(\cdot,\cdot)$ but unknown $m(\cdot)$. In addition, unit root issue may exist for financial variables in the SDF model. Furthermore, my studies in this dissertation strongly support using a conditional SDF. But the question arises is how to choose an appropriate model and related variables in a real application. All of these problems could be considered as future work in this area.

REFERENCES

[1] Ang, A. and J. Liu (2004). How to discount cashflows with time-varying expected return, *The Journal of Finance*, 59, 2745-2783.

[2] Bansal, R., D.A. Hsieh and S. Viswanathan (1993). A new approach to international arbitrage pricing, *The Journal of Finance*, 48, 1719-1747.

[3] Bansal, R. and S. Viswanathan (1993). No arbitrage and arbitrage pricing: A new approach, *The Journal of Finance*, 47, 1231-1262.

[4] Cai, Z. (2003). Nonparametric estimation equations for time series data, *Statistics and Probability Letters*, 62, 379-390.

[5] Cai, Z. and Q. Li (2008). Nonparametric estimation of varying coefficient dynamic panel models, *Econometric Theory*, 24, 1321-1342.

[6] Carrasco, M. and X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory*, 18, 17-39.

[7] Chen, S.X. and C.Y. Tang (2005). Nonparametric inference of value at risk for dependent financial returns, *Journal of Financial Econometrics*, 3, 227-255.

[8] Cochrane, J.H. (1996). A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy*, 104, 572-621.

[9] Dette, H. and Spreckelsen, I. (2004). Some comments on specification tests in nonparametric absolutely regular processes. *Journal of Time Series Analysis*, 25, 159C172.

[10] Fama, E. and K.R. French (1992). The cross-section of expected stock returns, *The Journal of Finance*, 47, 427-466.

[11] Fama, E. and K.R. French (1993). Common risk factors in the returns on bonds and stocks, *Journal of Financial Economics*, 33, 3-56.

[12] Fama, E. and K.R. French (1995). Size and book-to-market factors in earning and returns, *The Journal of Finance*, 50, 131-155.

[13] Ferson, W.E. (1989). Changes in expected security returns, risk and the level of interest rates, *The Journal of Finance*, 44, 1191-1214.

[14] Ferson, W.E. and C.R. Harvey (1991). The variation of economic risk premiums, *Journal of Political Economy* 99, 385-415.

[15] Ferson, W.E. and C.R. Harvey (1993). The risk and predictability of international equity returns, *Journal of Financial Studies*, 6, 527-566.

[16] Ferson, W.E. and C.R. Harvey (1998). Fundamental determinants of national equity market returns: A perspective on conditional asset pricing, *Journal of Banking and Finance*, 21, 1625-1665.

[17] Ferson, W.E. and C.R. Harvey (1999). Conditional variables and the cross section of stock return, *The Journal of Finance*, 54, 1325-1360.

[18] Ferson, W.E. and R.A. Korajczyk (1995). Do arbitrage pricing models explain the predictability of stock returns?, *Journal of Business*, 68, 309-349.

[19] Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time varying betas help or hurt?, *The Journal of Finance*, 53, 549-573.

[20] Gourieroux, C. and A. Monfort (2007). Econometric specification of stochastic discount factor models, *Journal of Econometrics*, 136, 509-530.

[21] Hamilton, J.D. (1994). *Time Series Analysis*, Princeton: Princeton University Precess, 409-434.

[22] Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, 64, 413-430.

[23] Hansen, L.P. (1982). Large sample properties of the generalized method of moments estimators, *Econometrica*, 50, 1029-1054.

[24] Harvey, C.R. (1989). Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics*, 24, 289-317.

[25] Harvey, C.R. (1991). The word price of covariance risk, *Journal of Finance*, 46, 111-157.

[26] Hjellvik, V., Yao, Q. Linearity Testing using Local Polynomial Approximation, *Journal of Statistical Planning and Inference*, 68, 295-321.

[27] Jagannathan, R. and Z. Wang (1996), The conditional CAPM and the cross-section of expected returns, *The Journal of Finance*, 51, 3-53.

[28] Jagannathan, R. and Z. Wang (2002). Empirical evaluation of asset pricing models: A comparison of the SDF and beta methods, *The Journal of Finance*, 57, 2337-2367.

[29] Lewbel, A. (2007). A local generalized method of moments estimator, *Economics Letters*, 94, 124-128.

[30] Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolio and capital budgets, *Review of Economics and Statistics*, 47, 13-37.

[31] Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates, *J.Time Series Analysis*, 17, 571-599.

[32] Reyes, M.G. (1999). Size, time-varying beta, and conditional heteroscedasticity in UK stock return, *Review of Financial Economics*, 8, 1-10.

[33] Shao, Q. and H. Yu (1996). Weak convergence for weighted empirical processes of dependent sequences, *The Annals of Probability*, 24, 2098-2127.

[34] Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, 19, 425-442.

[35] Shiryaev, A, N. (1995). *Probability*, Springer; 2nd edition, 309-353.

[36] Volkonskii, V.A. and Yu.A. Rozanov (1959). Some limit theorems for random functions, I, *Theory of Probability and Its Applications*, 4, 178-197.

[37] Wang, K.Q. (2002). Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio, *Journal of Empirical Finance*, 9, 133-169.

[38] Wang, K.Q. (2003). Asset pricing with conditioning information: A new test, *The Journal of Finance*, 58, 161-196.