8-2014

A Combined Approach to the Inference of Conditional Factor Models

Yan LI
Temple University

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Yuewu XU
Fordham University

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

A Combined Approach to the Inference of Conditional Factor Models

Yan Li, Liangjun Su and Yuewu Xu

August 2014

Paper No. 10-2014

A Combined Approach to the Inference of Conditional Factor Models*

YAN LI[a], LIANGJUN SU[b] AND YUEWU XU[c]

[a]Department of Finance, Temple University, Philadelphia, PA 19122

[b]School of Economics, Singapore Management University, Singapore, 178903

[c]School of Business, Fordham University, New York, NY 10019

May 3, 2014

Abstract

This paper develops a new methodology for estimating and testing conditional factor models in finance. We propose a two-stage procedure that naturally unifies the two existing approaches in the finance literature—the parametric approach and the nonparametric approach. Our combined approach possesses important advantages over both methods. Using our two-stage combined estimator, we derive new test statistics for investigating key hypotheses in the context of conditional factor models. Our tests can be performed on a single asset or jointly across multiple assets. We further propose a novel test to directly check whether the parametric model used in our first stage is correctly specified. Simulations indicate that our estimates and tests perform well in finite samples. In our empirical analysis, we use our new method to examine the performance of the conditional CAPM, which has generated controversial results in the recent asset-pricing literature.

JEL Classification: C51, C52, G12

Keywords: Conditional Factor Models, Specification Tests, Semiparametric Method, Nonparametric Method, Conditional CAPM.

*Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; Email: ljsu@smu.edu.sg; Tel: (65)6828 0386.

1 Introduction

Since the advance of the Capital Asset Pricing Model (Sharpe 1964; Lintner 1965; Mossin 1966) and the Arbitrage Pricing Model (Ross 1977), factor models have become one of the most important tools in modern finance. The existing research literature on both theoretical studies and empirical analysis is vast (see Campbell, Lo, and MacKinlay 1996 or Cochrane 2005 for excellent overviews). Although earlier studies tend to focus on the unconditional models (e.g., Fama and MacBeth 1973 for CAPM, Fama and French 1993 for a three-factor model), many recent studies have focused on conditional factor models (e.g., Cochrane 1996; Jagannathan and Wang 1996; Lettau and Ludvigson 2001; Santos and Veronesi 2006; Lustig and Nieuwerburgh 2005; among many others), because conditional models are theoretically more appealing (e.g., Jensen 1968; Dybvig and Ross 1985).

Under a conditional factor model, parameters (i.e., the factor loadings) change with the investor's information set, which is unobservable to econometricians. Therefore, it is empirically challenging to estimate and test a conditional factor model. To tackle this issue, two approaches have emerged in existing studies. The first approach is the traditional method that specifies factor loadings to be parametric functions of the state variables, which approximate the investor's information set. It has been used in Shanken (1990), Ferson and Harvey (1999), among many others. This parametric approach has the advantage of clear economic modeling of the conditioning information but also has the disadvantage that the model is very likely to be misspecified. To avoid the potential misspecification associated with the traditional approach, nonparametric methods have been proposed for the inference of conditional factor models. An incomplete list of studies relying on nonparametric approaches in the test of conditional models includes French, Schwert, and Stambaugh (1987), Bansal, Hsieh, and Viswanathan (1993), Bansal and Viswanathan (1993), Wang (2003), Lewellen and Nagel (2006), Nagel and Singleton (2011), Ang and Kristensen (2010), Li and Yang (2011), and Roussanov (2014). However, compared with the traditional parametric approach, nonparametric approaches have the disadvantage of being less efficient as evidenced by their slow rate of convergence.

The goal of this paper is to develop a new methodology that unifies the above two approaches to the inference of conditional factor models. Our approach consists of two stages. In the first stage, we follow the traditional approach by specifying factor loadings as a parametric function of state variables. In the second stage, we use nonparametric methods to fit the residuals obtained from the first-stage estimation. Because the parametric models in the traditional approach are usually carefully built using clear economic intuitions, our two-stage procedure allows the parametric estimation from the traditional approach to take care of a substantial portion of the problem first, and then it allows the nonparametric approach to take care of any remaining tasks.

To our knowledge, this is the first paper to combine the parametric and nonparametric approaches in a unified way for the inference of factor models, although ideas similar to ours have appeared in other finance contexts. For example, in their study of state-price densities, Ait-Sahalia and Lo (1998) use a two-stage procedure by first transforming the option prices into the Black-Scholes implied volatilities, and then they use kernel regression to fit nonparametrically the implied volatility curve/surface. The use of the Black-Scholes option-pricing formula in Ait-Sahalia and Lo (1998) is very similar to the first-stage parametric estimation in our procedure. Similar to the Ait-Sahalia and Lo method which does not assume the Black-Scholes option pricing formula to hold perfectly, our method does not assume that the parametric models used in the first stage is correctly specified.

Our paper contributes to the literature in several important ways. First, our new methodology joins two distinct approaches in the literature on the inference of conditional models in finance. Because we use a parametric model from the traditional approach in the first step to get the first-stage estimator, the method retains the desirable feature of traditional methods of more explicit modeling of the investor's information set. Since the state variables used in the traditional approach are usually selected carefully based on economic reasoning, it is reasonable to expect that after the first-stage estimation in our approach, we have already obtained a substantial part for the unknown parameter values. The second-stage estimation then uses a nonparametric technique to take care of any potential model misspecification in the first-stage estimation, thereby avoiding the pitfalls of model misspecification, an important issue stressed in Ghysels (1998), Harvey (2001), and Brandt and Chapman (2008), among others.

Second, we establish the asymptotic distributions for our two-stage estimators under both general conditions and the correct specification of the parametric model in the first stage. These results illustrate that our new two-stage estimators have some crucial advantages with respect to both parametric and nonparametric methods. It is well known that the parametric estimator is inconsistent when the parametric model is incorrectly specified; our new combined estimator, on the other hand, remains consistent even if the first-stage parametric model is misspecified. Our two-stage estimator is also superior to the nonparametric estimator of Cai, Fan, and Yao (2000): if the first-stage parametric model is correctly specified, then our new two-stage estimator converges at the faster parametric rate; even if the first-stage parametric model is incorrectly specified, our two-stage estimator still possesses important potential gains in achieving a smaller asymptotic bias than the nonparametric estimator.

Third, we propose new tests to investigate important hypotheses in the context of conditional factor models, such as tests on conditional alphas and conditional betas. These tests can be performed for a single asset or jointly across multiple assets. We derive the asymptotic distribution of these tests and propose a bootstrap procedure which can be implemented in small samples.

Fourth, because our combined approach naturally nests the traditional parametric approach, we can directly test whether or not the parametric model used in the first stage is correctly specified. This has not been done by existing parametric or nonparametric methods. The test for the correct specification of the parametric part is important because the parametric estimation of conditional models is widely used in the literature, and it is well known that all the p-values/critical values from the traditional parametric approach implicitly assume that the parametric model is correctly specified. If the model is in fact incorrectly specified, then all the p-values (or the critical values) are likely wrong, which could result in misleading conclusions.

Finally, in light of the conflicting results generated from existing parametric and nonparametric methods, we apply our new method to re-evaluate the performance of the conditional CAPM. The conditioning variables examined in our study are from several recent influential studies (Jagannathan and Wang 1996; Lettau and Ludvigson 2001; Santos and Veronesi 2006). In addition to examining the performance of the conditional CAPM in pricing portfolios sorted by book-to-market ratios, we also formally test whether conditional CAPM betas are time varying and whether the first-stage parametric form is correctly specified.

The rest of this paper is organized as follows. Section 2 develops our new methodology that combines the existing two approaches. Section 3 describes the simulation studies, and Section 4 provides empirical analysis. Section 5 concludes the paper. All proofs are relegated to the technical appendix.

Notation. For natural numbers a and b, we use I_a to denote an $a \times a$ identity matrix, $\mathbf{1}_{a\times b}$ is an $a \times b$ matrix of ones, and $\mathbf{0}_{a\times b}$ is an $a \times b$ matrix of zeros. Let $\otimes$ and $\odot$ denote the Kronecker and Hadamard products, respectively. For a matrix A, $A^\top$ denotes its transpose and $\|A\|$ its Euclidean norm ($\|A\| = \{\mathrm{tr}(AA^\top)\}^{1/2}$). We use $\xrightarrow{d}$ to denote convergence in distribution.

2 Statistical Methodology

In this section, we develop our new approach to estimate and test conditional models.

Let $R_t = (R_{1,t}, \cdots, R_{N,t})^\top$ denote a vector of asset returns on N assets. The traditional unconditional k-factor model attempts to explain the returns $R_{j,t}$ of asset j by a linear model of factors $F_t = (F_{1,t}, \cdots, F_{k,t})^\top$:

$$R_{j,t} = \beta_j^\top F_t + \zeta_{j,t}, \tag{2.1}$$

where β_j is a $k \times 1$ vector of factor loadings that is usually assumed to be constant over time, and $\zeta_{j,t}$ is the usual error term. The conditional factor model extends the above unconditional factor model (2.1) by allowing the vector of factor loadings β_j to depend on the information set $\mathcal{F}_{t-1}$ available to the investor at time $t-1$: $\beta_j = \beta_j(\mathcal{F}_{t-1})$, where $\beta_j(\mathcal{F}_{t-1})$ is a function of the investors' information set $\mathcal{F}_{t-1}$ at time $t-1$, and hence can be time-varying. To make the inference feasible, we assume that $\beta_j(\mathcal{F}_{t-1})$ depends on $\mathcal{F}_{t-1}$ only through a finite d-dimensional vector U_t, i.e., $\beta_j(\mathcal{F}_{t-1}) = \beta_j(U_t)$. As a result, the conditional factor model that we will study in this paper is

$$R_{j,t} = \beta_j\left(U_t\right)^\top F_t + \zeta_{j,t}, \tag{2.2}$$

where the coefficient $\beta_j(U_t)$ depends on the state variable U_t through a vector of unknown functions $\beta_j(\cdot)$ and we allow the error term $\zeta_{j,t}$ to exhibit heteroskedasticity, serial correlations, and cross-sectional dependence across different asset j. The model in (2.2) will be regarded as the true model in our subsequent analysis. Throughout this paper, we assume that the conditioning variable U_t is observable, but this assumption can be relaxed as in Mishra, Su, and Ullah (2010).

2.1 A Semiparametric Approach to the Conditional Factor Model

In this section, we propose a two-stage procedure to estimate the conditional factor model in (2.2). In the first stage of our procedure, we assume that a parametric model for the factor loadings is available to us:

$$\beta_j\left(U_t\right) = g(U_t, \gamma_j), \tag{2.3}$$

where γ_j is an unknown parameter and g is a $k \times 1$ vector of functions whose functional forms are assumed to be known. Different asset-pricing models can provide different U_t; for example, Lettau and Ludvigson (2001) propose U_t as the consumption-to-wealth ratio, Santos and Veronesi (2006) propose U_t as the labor-to-income ratio, and all of these models can be used in the estimation of (2.3). In the extreme case in which no U_t is available, one can always use the null (empty) model or the constant model in the first stage, and in this case our combined approach will reduce to a pure nonparametric one.

Once the above parametric specification for β_j has been chosen, we can estimate γ_j in the model $R_{j,t} = g(U_t, \gamma_j)^\top F_t + \zeta_{j,t}$ by several ways such as the nonlinear least squares (NLS) and quasi-maximum likelihood methods. In this paper we focus on the NLS estimator $\hat{\gamma}_j$ of γ_j, which gives us our first-stage parametric estimator $g(u, \hat{\gamma}_j)$ for $\beta_j(u)$. After the first-stage estimation is done, we obtain the residuals as

$$\hat{\zeta}_{j,t} = R_{j,t} - g(U_t, \hat{\gamma}_j)^\top F_t. \tag{2.4}$$

In the second stage, we explore the remaining information in the residuals $\hat{\zeta}_{j,t}$ by estimating nonparametrically the following model:

$$\hat{\zeta}_{j,t} = m_j(U_t)^\top F_t + \eta_{j,t}, \tag{2.5}$$

where m_j is a $k \times 1$ vector of functions whose functional forms are unknown. Although there are many ways to estimate m_j in the nonparametric literature, we consider the local linear estimator in this paper due to its simple structure of asymptotic bias and its automatic boundary bias correction mechanism (see Fan and Gijbels 1996). In the finance literature, the local linear estimator has become popular (e.g., Ait-Sahalia and Duarte 2003; Nagel and Singleton 2011).

Once we obtain the nonparametric estimator $\hat{m}_j(\cdot)$ of $m_j(\cdot)$, our two-stage estimator for $\beta_j(\cdot)$ can be obtained additively as follows:

$$\hat{\beta}_j(u) = g(u, \hat{\gamma}_j) + \hat{m}_j(u) \text{ for any } u, \tag{2.6}$$

due to the linear structure of the original conditional model in F_t. In some sense, one can regard $\hat{m}_j(u)$ as a nonparametric correction term added to the first-stage parametric estimator $g(u, \hat{\gamma}_j)$. Ideally, this term is not needed if the the first-stage parametric model is correctly specified (which we do not assume), otherwise it plays an important role.

Because our new estimator of $\hat{\beta}_j(u)$ is the sum of the first-stage parametric estimator $g(u, \hat{\gamma}_j)$ and a second-stage nonparametric estimator $\hat{m}_j(u)$, we also call our two-stage estimator as a semiparametric estimator. It is worth mentioning that the idea of semiparametric combined estimation has appeared in the econometric and statistics literature. See Glad (1998), Fan and Ullah (1999), Mishra, Su, and Ullah (2010), Long, Su, and Ullah (2011) and Su, Murtazashvili, and Ullah (2013) for different approaches in the standard conditional mean and variance models. Our additive semiparametric approach is closely related to the work of Martins-Filho, Mishra, and Ullah (2008), who demonstrated that for the estimation of unknown conditional mean function, the local polynomial estimator, the multiplicatively combined estimator of Glad (1998), and the additively combined estimator like that in (2.6) can all be regarded as the minimization of a suitably defined Cressie-Read discrepancy. Nevertheless, to the best of our knowledge, our paper is the first to consider combined estimation for the conditional factor models in finance.

To see the motivation for our approach, note that (2.4) and (2.2) imply that

$$\begin{aligned} \hat{\zeta}_{j,t} &= R_{j,t} - g(U_t, \hat{\gamma}_j)^\top F_t \\ &= \zeta_{j,t} + \left[\beta_j(U_t) - g(U_t, \gamma_j^0)\right]^\top F_t - \left[g(U_t, \hat{\gamma}_j) - g(U_t, \gamma_j^0)\right]^\top F_t \\ &= \zeta_{j,t} + m_j(U_t)^\top F_t - \delta_j(U_t)^\top F_t, \end{aligned} \tag{2.7}$$

where $\delta_j(u) = g(u, \hat{\gamma}_j) - g(u, \gamma_j^0)$ represents the estimation error from the first-stage parametric regression and

$$m_j(u) = \beta_j(u) - g(u, \gamma_j^0) \tag{2.8}$$

represents the remaining information about $R_{j,t}$ that cannot be explained by the fitted parametric conditional factor model in the first stage.

The last expression provides intuition for our two-stage semiparametric approach. If the first-stage parametric model $g(\cdot,\cdot)$ is correctly specified, then we expect that no useful information should be retained in the residual $\hat{\zeta}_{j,t}$ because the nonparametric function $m_j(u)$ is now identically zero. The second-stage nonparametric estimate of such a zero function can be accurate enough so that adding the second-stage nonparametric correction term $\hat{m}_j(u)$ to $g(u, \hat{\gamma}_j)$ as in (2.6) will not affect the latter's asymptotic distributional property. On the other hand, if the first-stage parametric model $g(\cdot,\cdot)$ is misspecified, then much information about the returns will be carried on to the residual $\hat{\zeta}_{j,t}$ through the nonparametric correction term $m_j(\cdot)$, and the second-stage nonparametric regression will pick up extra useful information about the returns. In this case, it is easy to show that the parametric estimator $g(u, \hat{\gamma}_j)$ is inconsistent for $\beta_j(u)$, whereas our semiparametric estimator $\hat{\beta}_j(u)$ remains consistent.

Now, we discuss how to use the local linear method to obtain the nonparametric estimator for m_j. Let $m_{ij}(\cdot)$ denote the ith element of $m_j(\cdot)$ for $i = 1, \cdots, k$ and $j = 1, \cdots, N$. Assume that $m_{ij}(\cdot)$ has a second-order partial derivatives. For any given u and U_t in the neighborhood of u, it follows from a first-order Taylor expansion that $m_{ij}(U_t) \approx m_{ij}(u) + \dot{m}_{ij}(u)^\top (U_t - u)$, where $\dot{m}_{ij}(u) = \partial m_{ij}(u)/\partial u$. To estimate $\{m_{ij}(u)\}_i$ and $\{\dot{m}_{ij}(u)\}_i$, we choose $\{a_{ij}\}_i$ and $\{b_{ij}\}_i$ to minimize

$$\sum_{t=1}^{T}\left[\hat{\zeta}_{j,t} - \sum_{i=1}^{k}\left\{a_{ij} + b_{ij}^\top ((U_t - u)/h)\right\} F_{i,t}\right]^2 K_h(U_t - u), \tag{2.9}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a product kernel function defined by $K(u) = \Pi_{i=1}^d k(u_i)$, $k(\cdot)$ is a symmetric probability density function (PDF) on the real line, and $h = h(T)$ is a bandwidth that typically shrinks to 0 as the sample size T goes to infinity. Let $\hat{a}_{ij}(u)$ and $\hat{b}_{ij}(u)$ denote the solution to the above minimization problem. Then, the local linear regression estimator for $m_{ij}(u)$ is given by $\hat{m}_{ij}(u) = \hat{a}_{ij}(u)$ for $i = 1, \cdots, k$, and $\hat{m}_j(u) = (\hat{m}_{1j}(u), \cdots, \hat{m}_{kj}(u))^\top$ is the local linear estimator of $m_j(u)$. To obtain the expression for $\hat{m}_j(u)$, let $\mathbf{X}_u$ denote an $T \times k(1+d)$ matrix with $X_t(u) = (F_t^\top, F_t^\top \otimes ((U_t - u)/h)^\top)$ as its tth row, where $\otimes$ denotes the Kronecker product. Let $\hat{\zeta}_j = (\hat{\zeta}_{j,1}, \cdots, \hat{\zeta}_{j,T})^\top$ and $\mathbf{K}_u$ =diag$\{K_{1u}, K_{2u}, \cdots, K_{Tu}\}$, where $K_{tu} = K_h(U_t - u)$ for $t = 1, 2, \cdots, T$. Then it is easy to verify that

$$\hat{m}_j(u) = \mathbf{s}\left(\mathbf{X}_u^\top \mathbf{K}_u \mathbf{X}_u\right)^{-1} \mathbf{X}_u^\top \mathbf{K}_u \hat{\zeta}_j, \tag{2.10}$$

where $\mathbf{s} = (I_k, \mathbf{0}_{k\times kd})$, I_k is an $k \times k$ identity matrix, and $\mathbf{0}_{k\times kd}$ is a $k \times kd$ matrix of zeros. Note that for notational simplicity we use the same bandwidth h for all conditioning variables in U_t. In practice, it is advisable to choose h as a vector of bandwidths if the spreads of different elements of U_t are quite different from each other. Then we can use $\mathbf{h} = (h_1, \cdots, h_d)$ to denote the bandwidth, and $K_h(U_t - u)$ changes to $K_{\mathbf{h}}(U_t - u) = \Pi_{i=1}^d k\left((U_{it} - u_i)/h_i\right)/h_i$. The results in the following analysis can easily be adjusted to accommodate this case with a little bit more complicated notation.

Intuitively, $m_j(u)$ specifies how $\hat{\zeta}_{j,t}$, which contains the remaining information in $R_{j,t}$ after the first-stage estimation, varies with the factor F_t when the state variable $U_t = u$. For any asset j, we can obtain the first-stage parametric estimator $g(u, \hat{\gamma}_j)$ and the second-stage nonparametric estimator $\hat{m_j}(u)$. Our combined semiparametric estimator for asset j is

$$\hat{\beta}_j(u) = g(u, \hat{\gamma}_j) + \hat{m}_j(u).$$

We will study the asymptotic properties of $\hat{\beta}_j(u)$ in the next two subsections.

2.2 Asymptotic Distribution of $\hat{\beta}_j(u)$ Under General Conditions

In this section, we show that our semiparametric estimator has asymptotic normal distribution under general nonparametric assumptions. Notably, compared with the pure nonparametric estimator, our semiparametric estimator has potential gains in the reduction of asymptotic bias. Furthermore, we will demonstrate in the next section that our estimator can converge at the faster parametric $\sqrt{T}$-rate when the parametric part is correctly specified.

Some notations are needed to state our result. Let $p_U(\cdot)$ denote the PDF of U_t, and set $\mu_{r_1,r_2} = \int_{\mathbb{R}} v^{r_1} k(v)^{r_2} \, dv$ for $r_1, r_2 = 0, 1, 2$. The following result shows that our semiparametric estimator $\hat{\beta}_j(u)$ has asymptotic normal distribution.

Theorem 1 *Assume that Assumptions A1-A4 in the Appendix hold. Then, for each interior point u in the support of U_t, we have*

$$\sqrt{Th^d}\left[\hat{\beta}_j(u) - \beta_j(u) - h^2 B_j(u)\right] \xrightarrow{d} N(0, \Sigma_{jj}(u)), \; j = 1, \cdots, N. \tag{2.11}$$

Here, $B_j(u) = \frac{1}{2}\mu_{2,1} M_j(u)$, $M_j(u) = \left(\sum_{s=1}^d m_{1j,ss}(u), \cdots, \sum_{s=1}^d m_{kj,ss}(u)\right)^\top$ *with* $m_{ij,s_1 s_2} \equiv \frac{\partial^2 m_{ij}(u)}{\partial u_{s_1} \partial u_{s_2}}$ *for* $i, j = 1, \cdots, k$, *and* $s_1, s_2 = 1, \cdots, d$; $\Sigma_{ij}(u) = \mu_{0,2}^d \Omega(u)^{-1} \Omega_{ij}^*(u) \Omega(u)^{-1} / p_U(u)$, $\Omega(u) = \mathbb{E}(F_t F_t^\top | U_t = u)$, $\Omega_{ij}^*(u) = \mathbb{E}[F_t F_t^\top \sigma_{ij}(U_t, F_t) | U_t = u]$, *and* $\sigma_{ij}(u, f) = \mathbb{E}[\zeta_{i,t}\zeta_{j,t} | U_t = u, F_t = f]$.

Remark 1. The above result does not need the assumption that the first-stage parametric model is correctly specified. It can be used to construct confidence bands for the parameters in the conditional factor models. To this goal, one can estimate $p_U(u)$, $\Omega(u)$, and $\Omega^*_{ij}(u)$ respectively by $\hat{p}_U(u) = \frac{1}{T}\sum_{t=1}^{T} K_h(U_t - u)$, $\hat{\Omega}(u) = \frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top K_h(U_t - u)/\hat{p}_U(u)$, and $\hat{\Omega}^*_{ij}(u) = \frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top \tilde{\zeta}_{i,t}\tilde{\zeta}_{j,t} K_h(U_t - u)/\hat{p}_U(u)$, where $\tilde{\zeta}_{j,t}$ is the residual obtained from our semiparametric fit: $\tilde{\zeta}_{j,t} = R_{j,t} - \hat{\beta}_j(U_t)^\top F_t$. It is standard to justify that the above estimators are consistent.

Remark 2. Many applications in finance involve multiple assets or portfolios. Theorem 1 can be extended to cover the situation easily. Suppose that we have N assets and our interest is not to estimate the coefficients for a single asset j (which can be done using Theorem 1), but to estimate the collection of estimators $\hat{\beta}_j(u)$, $j = 1, \cdots, N$. Let $\boldsymbol{\beta}(u)$ collect the coefficients from N different assets, and let $\hat{\boldsymbol{\beta}}(u)$ denote the semiparametric estimator for $\boldsymbol{\beta}(u)$: $\boldsymbol{\beta}(u) = (\beta_1(u)^\intercal, ..., \beta_N(u)^\intercal)^\intercal$, and $\hat{\boldsymbol{\beta}}(u) = (\hat{\beta}_1(u)^\intercal, ..., \hat{\beta}_N(u)^\intercal)^\intercal$. Following the proof of Theorem 1 and the Cramér-Wold device, we can readily show that $\sqrt{Th^d}\left(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u) - h^2\mathbf{B}(u)\right) \xrightarrow{d} N(0, \boldsymbol{\Sigma}(u))$, where

$$\mathbf{B}(u) = \begin{pmatrix} B_1(u) \\ \vdots \\ B_N(u) \end{pmatrix} \text{ and } \boldsymbol{\Sigma}(u) = \begin{pmatrix} \Sigma_{11}(u) & \cdots & \Sigma_{1N}(u) \\ \vdots & \ddots & \vdots \\ \Sigma_{N1}(u) & \cdots & \Sigma_{NN}(u) \end{pmatrix}.$$

Note that because we allow cross-sectional dependence among $\zeta_{j,t}$'s, the variance-covariance matrix $\boldsymbol{\Sigma}(u)$ is not block-diagonal. On the other hand, as expected, the serial dependence in $\zeta_{j,t}$ does not contribute to the asymptotic distribution of $\hat{\boldsymbol{\beta}}(u)$ after local smoothing. To implement the above asymptotic distribution in practice, the quantities involved in the asymptotic distributions for $\hat{\boldsymbol{\beta}}$ can be estimated consistently by their sample analogs, very similar to the one-asset case that we have discussed before.

Remark 3. It is very helpful to compare our two-stage estimator to one-step local linear estimator of Cai, Fan, and Yao (2000). It is well known that the pure one-step local linear estimator $\tilde{\beta}_j(u)$ of $\beta_j(u)$ in (2.2) has the asymptotic distribution of

$$\sqrt{Th^d}\left[\tilde{\beta}_j(u) - \beta_j(u) - h^2\bar{B}_j(u)\right] \xrightarrow{d} N(0, \Sigma_{jj}(u)), \tag{2.12}$$

where $\bar{B}_j(u)$ is analogously defined as $B_j(u)$ but with $M_j(u)$ being replaced by

$$\bar{M}_j(u) = \left[\sum_{s=1}^{d} \beta_{1j,ss}(u), \cdots, \sum_{s=1}^{d} \beta_{kj,ss}(u)\right]^\top \quad \text{with } \beta_{ij,s_1 s_2}(u) = \frac{\partial^2 \beta_{ij}(u)}{\partial u_{s_1}\partial u_{s_2}},$$

and $\beta_{ij}(u)$ denotes the i-th element in $\beta_j(u)$. Apparently, comparing the results in Theorem 1 with that in (2.12), we find that our two-step semiparametric estimator $\hat{\beta}_j(u)$ shares the same asymptotic variance as the one-step nonparametric estimator $\tilde{\beta}_j(u)$, but they have different asymptotic biases. If the first-stage parametric model in our combined procedure can reasonably capture the curvature of the unknown function $\beta_j(\cdot)$ at the point of interest u, then one can demonstrate that our two-step semiparametric estimator $\hat{\beta}_j(u)$ has smaller asymptotic bias than $\tilde{\beta}_j(u)$. To see the last point clearly, we focus on the case where $d = 1$ and compare the asymptotic biases of the first element $\hat{\beta}_{1j}(u)$ of $\hat{\beta}_j(u)$ with that of the first element $\tilde{\beta}_{1j}(u)$ of $\tilde{\beta}_j(u)$. The leading asymptotic bias term of our two-stage estimator $\hat{\beta}_{1j}(u)$ can be written as $B_{1j}(u) = \frac{1}{2}\mu_{2,1}\frac{\partial^2 m_{1j}(u)}{\partial u^2}$ and that of Cai, Fan, and Yao's (2000) one-stage estimator $\tilde{\beta}_{1j}(u)$ is given by $\bar{B}_{1j}(u) = \frac{1}{2}\mu_{2,1}\frac{\partial^2 \beta_{1j}(u)}{\partial u^2}$. It follows that our two-stage estimator will achieve bias reduction in comparison with the one-stage estimator if one can fit the first-stage parametric model such that

$$\left|\frac{\partial^2 m_{1j}(u)}{\partial u^2}\right| < \left|\frac{\partial^2 \beta_{1j}(u)}{\partial u^2}\right|. \tag{2.13}$$

In other words, if the parametric function $g_1(\cdot, \gamma_j^0)$ can capture some of the shape features of $\beta_{1j}(\cdot)$ at u, $m_{1j}(\cdot)$ will be less rough than $\beta_{1j}(\cdot)$ at u so that (2.13) can be satisfied and we achieve bias reduction. The bias reduction condition in (2.13) is analogous to that in Glad (1998) and crucially depends on whether the first-stage parametric model is reasonably good or not. In the special case where $\partial^2\beta_{1j}(u)/\partial u^2 = \partial^2 g_1(u, \gamma_j^0)/\partial u$, $B_{1j}(u) = 0$, and our semiparametric estimator is asymptotically unbiased up to the order $O(h^2)$ and thus more efficient than the one-step local linear estimator. If the first-stage parametric model $g_1(\cdot, \gamma_j)$ is correctly specified, i.e., $\beta_{1j}(U_t) = g_1(U_t, \gamma_j^0)$ a.s. for some γ_j^0, then such a derivative condition is automatically satisfied and we demonstrate below that our two-stage estimator can achieve the parametric $\sqrt{T}$-rate of consistency under certain conditions whereas the one-stage estimator can only achieve the usual nonparametric rate of consistency.

Remark 4. As a referee kindly points out, several nonparametric and semiparametric methods have been recently proposed to estimate functional/time-varying coefficient models. For example, Su, Chen, and Ullah (2009) allow both continuous and discrete variables in the vector U_t of state variables. But they focus on the one-step local linear estimation as in Cai, Fan, and Yao (2000). Gao, Gu, and Hernandez-Verme (2012) consider a special semiparametric varying coefficient model

$$Y_t = \beta(U_t)^\top F_t + \lambda(U_t)\, t + \varsigma_t$$

where $\beta(U_t)$ is a $k \times 1$ vector of unknown smooth functions of U_t, $\lambda(U_t)$ is a scalar function, and ς_t is the error term. Apparently, the above model is a special functional coefficient model when one factor is replaced by the deterministic time trend t and is applicable to the case when Y_t

may exhibit non-stationary feature. They also consider the one-stage local linear estimation but their estimators of $\beta(u)$ and $\lambda(u)$ have different convergence rates because of the appearance of the time trend. Sun and Wu (2005) consider a semiparametric time-varying coefficient model for longitudinal data:

$$Y_j(t) = \beta(t)^\top F_j^{\mathrm{I}}(t) + \lambda(t;\theta)^\top F_j^{\mathrm{II}}(t) + \varsigma_j(t),$$

where $j = 1, ..., N$ denote the individuals, t denotes time, the functional coefficient $\beta(t)$ is a vector of unspecified smooth functions of t, $\lambda(t;\theta)$ is a vector of smooth functions of t known up to the finite dimension parameter θ, $\varsigma_j(t)$ is the error term, and the observations of $Y_j(t)$ are taken at time points $t_{j1} < t_{j2} < ... < t_{jT_j}$ with T_j denoting the total number of observations on the jth object. Note that the regressors $F_j^{\mathrm{I}}(t)$ and $F_j^{\mathrm{II}}(t)$ are not common factors. The above model represents a functional coefficient unbalanced panel data model (without individual fixed effects) and is more complicated than the model considered in this paper. Sun and Wu (2005) propose two ways to estimate the above model and argue that one way is more efficient than the other. In addition, Borak and Weron (2008) consider the model:

$$Y_{j,t} = \beta(U_{j,t})^\top F_t + \varsigma_{j,t} = \beta_0(U_{j,t}) + \beta_1(U_{j,t}) F_{1,t} + ... + \beta_k(U_{j,t}) F_{k,t} + \varsigma_{j,t},$$

where $j = 1, ..., J$, $t = 1, ..., T$, $F_t = (F_{1,t}, ..., F_{k,t})^\top$ is a $k \times 1$ vector of *unobserved* common factors, $U_{j,t}$ is observed, $\beta(\cdot) = (\beta_0(\cdot), \beta_1(\cdot), ..., \beta_k(\cdot))^\top$ is $(k+1)$-vector of unknown smooth functions, and $\varsigma_{j,t}$ is the error term. Borak and Weron term the above model as semiparametric dynamic factor model (DSFM) and apply it to model the electricity forward curve dynamics. Park, Mammen, Härdle, and Borak (2009) propose an iterative algorithm to fit the model and study the asymptotic properties of the resulting estimators, but they do not have any asymptotic distributional results.

2.3 Asymptotic Distribution of the Estimator $\hat{\beta}_j(u)$ under the Correct Specification of the Parametric Part

Having established the asymptotic distribution for our semiparametric estimator under general conditions, we proceed to study the asymptotic properties of $\hat{\beta}_j(u)$ when the parametric model in the first stage is actually correctly specified. We show that when the parametric part of the model is correctly specified, our two-stage estimator converges to the true β_j at the faster parametric $\sqrt{T}$-rate when holding the bandwidth h constant. Note that the assumption that the parametric part is correctly specified is equivalent to

$$g(U_t, \gamma_j^0) = \beta_j(U_t) \text{ almost surely (a.s.) for some } \gamma_j^0. \tag{2.14}$$

Recall that Theorem 1 implies that when $h \to 0$ as $T \to \infty$, the first-stage parametric estimator does not contribute to the asymptotic variance of our semiparametric estimator $\hat{\beta}_j(u)$, and the serial dependence among $\zeta_{j,t}$ does not play a role either. Nevertheless, this is not the case when h is held fixed. The following theorem indicates that when h is held fixed, both the first-stage estimation and the serial dependence in the process $\{\zeta_{j,t}\}$ play important roles in the asymptotic distribution of our semiparametric estimator $\hat{\beta}_j(u)$.

Theorem 2 *Suppose Assumptions A1-A3 and A5 in the Appendix hold. Suppose (2.14) holds. If h is held fixed as $T \to \infty$, then we have*

$$\sqrt{T}\left[\hat{\beta}_j(u) - \beta_j(u)\right] \xrightarrow{d} N\left(0, \bar{\Sigma}_{jj}(u)\right), \ j = 1, \cdots, N, \tag{2.15}$$

where $\bar{\Sigma}_{ij}(u) = \mathbb{E}[\bar{\zeta}_{i,1}\bar{\zeta}_{j,1}^{\top}] + \sum_{t=2}^{\infty} \mathbb{E}[\bar{\zeta}_{i,1}\bar{\zeta}_{j,t}^{\top} + \bar{\zeta}_{i,t}\bar{\zeta}_{j,1}^{\top}]$ *denotes the long-run covariance of the process* $\{(\bar{\zeta}_{i,t}, \bar{\zeta}_{j,t}), t \geq 1\}$ *with* $\bar{\zeta}_{j,t} = [\mathbf{s}\bar{S}(u)^{-1}\bar{U}_t(u)F_t + \bar{\Lambda}_j(u)A_{j,t}(\gamma_j^0)]\zeta_{j,t}$. *Here,* $\bar{S}(u) = \mathbb{E}[S_T(u)]$; $A_{j,t}(\gamma_j^0)$, $S_T(u)$, *and* $\bar{U}_t(u)$ *are respectively defined in (A.2), (B.2) and (B.9) in the Appendix;* $\bar{\Lambda}_j(u) = D_{\gamma}g(u, \gamma_j^0) - \mathbf{s}\bar{S}(u)^{-1}\mathbb{E}[\bar{U}_t F_t F_t^{\top} D_{\gamma}g(U_t, \gamma_j^0)]$ *with* $D_{\gamma}g(u, \gamma_j) \equiv \partial g(u, \gamma_j)/\partial \gamma_j^{\top}$.

Remark 5. Theorem 2 indicates that under the correct specification of the parametric conditional model, $\hat{\beta}_j(u)$ is asymptotically unbiased up to $O\left(T^{-1/2}\right)$. The asymptotic variance of $\hat{\beta}_j(u)$ has a quite complicated formula because it is affected by the first-stage parametric estimation through the term $A_{j,t}(\gamma_j^0)$ in the definition of $\bar{\zeta}_{j,t}$ and the serial dependence among $\{\zeta_{j,t}\}$. In general, we cannot simplify this formula as we keep h fixed. But as we show in the appendix, $\bar{\Lambda}_j(u) = o(1)$ provided $h = o(1)$, so that the contribution from the first-stage parametric estimation will be asymptotically negligible by permitting $h \to 0$.

Remark 6. To compare our estimator with the parametric estimator $g(u, \hat{\gamma}_j)$ of $\beta_j(u)$, we consider two situations: $h \to 0$ as $T \to \infty$, and h being fixed. In the case where $h \to 0$ as $T \to \infty$, Theorem 1 indicates that when the parametric component is correctly specified, our estimator is usually less efficient than the parametric one since our estimator has a slower convergence rate in this case, as expected. In the case where h is kept fixed, Theorem 2 indicates that our estimator converges at the parametric $\sqrt{T}$-rate. In this sense, we say that our estimator is as good as the parametric estimator in terms of convergence rates when h is kept fixed, which is consistent with Glad (1998) even though she did not explicitly point this fact out. In contrast, Fan and Ullah (1999) consider a combined estimator of the regression mean in the cross section framework with independent and identically distributed (i.i.d.) observations. Their combined estimator is a linear combination of a parametric estimator and a nonparametric estimator with the weights automatically determined by the data. The parametric rate of convergence of their estimator in case of correct parametric specification can be achieved by letting the bandwidth approach zero.

In the multiple-asset case, a similar result holds. Following the proof of Theorem 1 and the Cramér-Wold device, we can also show that $\sqrt{T}\left(\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}(u)\right)\xrightarrow{d}N\left(0,\bar{\boldsymbol{\Sigma}}(u)\right)$, where $\bar{\boldsymbol{\Sigma}}(u)$ is defined analogously as $\boldsymbol{\Sigma}(u)$ with typical block $\Sigma_{ij}(u)$ being replaced by $\bar{\Sigma}_{ij}(u)$.

2.4 Tests for the Constancy of Coefficients and the Correct Specification of the Parametric Part

In this section, we develop tests for testing important hypotheses in empirical finance studies.

2.4.1 Hypotheses and test statistics

To state the hypothesis testing problem, we first split up the set of factors in F_t into two components: $F_t^{\mathrm{I}}=(F_{1,t},...,F_{k_1,t})^{\top}$ and $F_t^{\mathrm{II}}=(F_{k_1+1,t},...,F_{k,t})^{\top}$. As before, to keep notation compact, we focus on the test for a single generic asset j and make remarks on the more general case. Correspondingly, the coefficients $\beta_j(u)$ are also partitioned into two components: $\beta_j^{\mathrm{I}}(u)=(\beta_{1j}(u),\cdots,\beta_{k_1,j}(u))^{\top}$ and $\beta_j^{\mathrm{II}}(u)=(\beta_{k_1+1,j}(u),\cdots,\beta_{k,j}(u))^{\top}$, and the original factor model (2.2) can be written as

$$R_{j,t}=\beta_j^{\mathrm{I}}(u)^{\top}F_t^{\mathrm{I}}+\beta_j^{\mathrm{II}}(U_t)^{\top}F_t^{\mathrm{II}}+\zeta_{j,t}. \tag{2.16}$$

The first hypothesis of interest to us is to test for the constancy of the first set of coefficients $\beta_j^{\mathrm{I}}(u)$ for F_t^{I} while allowing the second set of coefficients $\beta_j^{\mathrm{II}}(u)$ for F_t^{II} to depend on the set of exogenous regressors U_t. Formally, the general form of the null and the alternative hypotheses can be formulated as

$$\begin{aligned}&\mathbb{H}_0^{(1j)}:\beta_j^{\mathrm{I}}(U_t)=\theta_j^{\mathrm{I}}\text{ a.s. for some parameter }\theta_j^{\mathrm{I}}\in\mathbb{R}^{k_1},\\&\mathbb{H}_1^{(1j)}:\text{ negation of }\mathbb{H}_0^{(1j)}.\end{aligned} \tag{2.17}$$

In other words, under the null hypothesis $\mathbb{H}_0^{(1j)}$, k_1 of the k coefficients in $\beta_j(U_t)$ are constant over U_t, whereas under the alternative at least one of the k_1 coefficients in $\beta_j^{\mathrm{I}}(U_t)$ is not constant.

It should be noted that the above formulation covers two interesting hypotheses in the context of conditional factor models in finance—the case of constant alphas:

$$\mathbb{H}_0^{(alpha)}:\beta_j^{\mathrm{I}}(U_t)=\theta_j^{\mathrm{I}}\in\mathbb{R}\text{ with }k_1=1,\ F_t^{\mathrm{I}}=1,$$

and the case of constant betas:

$$\mathbb{H}_0^{(beta)} : \beta_j^{\mathrm{I}}(U_t) = \theta_j^{\mathrm{I}} \in \mathbb{R}^{k-1} \text{ with } k_1 = k-1,\ F_t^{\mathrm{I}} = f_t,$$

where we recall that f_t denotes the set of non-constant factors in F_t. If we are interested in whether the conditional alphas are equal to zero, we can simply test $\mathbb{H}_0^{(alpha)}$ by setting $\theta_j^{\mathrm{I}} = 0$. The hypothesis of $\mathbb{H}_0^{(beta)}$ can be used to test whether the conditional betas are constant.

To test $\mathbb{H}_0^{(1j)}$, we propose a Wald-type test statistic that has the advantage of requiring only a one-time consistent estimation of the unrestricted model. Our proposed test statistic is

$$W_{1jT} = h^{d/2} \sum_{t=1}^{T} \left\| \hat{\beta}_j^{\mathrm{I}}(U_t) - \hat{\theta}_j^{\mathrm{I}} \right\|^2, \tag{2.18}$$

where $\hat{\beta}_j^{\mathrm{I}}(U_t)$ is the first k_1 element of $\hat{\beta}_j(u)$, $\hat{\theta}_j^{\mathrm{I}} = T^{-1}\sum_{t=1}^{T} \hat{\beta}_j^{\mathrm{I}}(U_t)$ is an estimator of θ_j^{I} under $\mathbb{H}_0^{(1j)}$, and $\|\cdot\|$ denotes the Frobenius norm. Clearly, if θ_j^{I} is known, as in the cases of testing conditional alphas and betas, then one can replace $\hat{\theta}_j^{\mathrm{I}}$ by the known value of θ_j^{I}, which will not affect the asymptotics developed below.

Another important hypothesis in finance concerns whether the parametric part is correctly specified. This hypothesis is important because the usual t-stat derived from fitting a parametric model implicitly assumed that the parametric model is correctly specified. If in fact the parametric model is misspecified, then all the reported statistics such as the t-stat will be wrong. Hence, it is very desirable to test for whether the parametric part is correctly specified, given that the parametric estimation of conditional factor models is widely used in the literature. One important feature of our approach is that it naturally nests the parametric model in the first stage into a nonparametric setup, and hence it can provide a direct check for the correct specification of the parametric model. This has not been done for the existing methods that have been proposed in the literature.

Recall that in our approach $m_j(u) = \beta_j(u) - g(u, \gamma_j^0)$, and hence the assertion that the parametric part is correctly specified is equivalent to the assertion that the $m_j(u)$ is zero. Specifically, the null hypothesis of correct specification of the parametric part can be formulated as

$$\mathbb{H}_0^{(2j)} : m_j(U_t) = \mathbf{0}_{k\times 1} \text{ a.s.} \tag{2.19}$$

The alternative hypothesis $\mathbb{H}_1^{(2j)}$ is the negation of $\mathbb{H}_0^{(2j)}$. To test the above hypothesis of correct specification of the parametric part, we propose the following test statistic

$$W_{2jT} = h^{d/2} \sum_{t=1}^{T} \left\| \hat{m}_j (U_t) - \overline{\hat{m}}_j \right\|^2, \tag{2.20}$$

where $\overline{\hat{m}}_j = T^{-1} \sum_{t=1}^{T} \hat{m}_j (U_t)$ serves as an estimator of $\mathbf{0}_{k\times 1}$ under $\mathbb{H}_0^{(2j)}$. Alternatively, one can consider $\bar{W}_{2jT} = h^{d/2} \sum_{t=1}^{T} \|\hat{m}_j (U_t)\|^2$, which is asymptotically equivalent to W_{2jT} under $\mathbb{H}_0^{(2j)}$.

In the next subsection, we develop the asymptotic distribution for our test statistics W_{1jT} and W_{2jT} under general conditions.

2.4.2 Asymptotic distributions of the test statistics

It can be shown that under mild regularity conditions, $\hat{\theta}_j^{\mathrm{I}}$ converges to θ_j^{I} at the $\sqrt{T}$-rate. In this section, we show that our test statistic for testing the constancy of the coefficients follow an asymptotic normal distribution.

Theorem 3 (Test for Constancy of Coefficients) *Suppose that Assumptions A1-A4 and A6-A8 in the Appendix hold. Suppose that $Th^{d/2+4} \to 0$ as $T \to \infty$. Then,*

$$W_{1jT} - B_{\mathbf{s}_1,jT} \xrightarrow{d} N(0, \Theta_{\mathbf{s}_1,jj}) \text{ under } \mathbb{H}_0^{(1j)}. \tag{2.21}$$

where $B_{\mathbf{s}_1,jT} = \frac{h^{d/2}}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \|\varphi_{\mathbf{s}_1} (\xi_{j,t}, \xi_{j,s})\|^2$, $\varphi_{\mathbf{s}_1} (\xi_{j,t}, \xi_{j,s}) = \mathbf{s}_1 \mathbf{s} \bar{S} (U_t)^{-1} \bar{U}_s (U_t) F_s \zeta_{j,s}$;

$$\Theta_{\mathbf{s}_1,ij} = \lim_{h\to 0} 2h^d \mathbb{E}_t \mathbb{E}_s [\bar{\varphi}_{\mathbf{s}_1} (\xi_{i,t}, \xi_{i,s}) \bar{\varphi}_{\mathbf{s}_1} (\xi_{j,t}, \xi_{j,s})] = \bar{\kappa} \int \left\| \mathbf{s}_1 \Omega (u)^{-1} \Omega_{ij}^{*} (u) \Omega (u)^{-1} \mathbf{s}_1^{\top} \right\|^2 du,$$

$\bar{\kappa} = \int [\int k(u) k(u+v) du]^2 dv$, $\bar{\varphi}_{\mathbf{s}_1} (\xi_{j,t}, \xi_{j,s}) = \int \varphi_{\mathbf{s}_1} (\xi, \xi_{j,t})^{\top} \varphi_{\mathbf{s}_1} (\xi, \xi_{j,s}) dP_{\xi_j} (\xi)$, $\xi_{j,t} = \left(U_t^{\top}, F_t^{\top}, \zeta_{j,t}\right)^{\top}$, *$P_{\xi_j}$ denotes the CDF of $\xi_{j,t}$, and $\mathbb{E}_t$ denotes expectation with respect to variables indexed by time t only.*

Remark 7. If $d \le 3$ as in most applications, the above theorem also holds if we replace $B_{\mathbf{s}_1,jT}$ by its nonstochastic version: $B_{\mathbf{s}_1,j} = h^{-d/2} \mu_{0,2}^d \mathrm{trace}(\mathbf{s}_1 \int \Omega (u)^{-1} \Omega_{jj}^{*} (u) \Omega (u)^{-1} du \mathbf{s}_1^{\intercal})$, where $\mu_{0,2} = 27/(32\sqrt{\pi}) \simeq 0.4760$ for the standard normal kernel. As before, the extra condition on the bandwidth in the above theorem ensures that the bias term from the nonparametric regression does not contribute to the asymptotic distribution of W_{1jT}. If we do not assume that the first-stage parametric model is correctly specified, then clearly undersmoothing is needed here. Nevertheless, if we assume that the first-stage parametric model is correctly specified, then $m_j (U_t) = 0$ a.s., and this extra condition is not required any more.

To implement the test in finance applications, we need to consistently estimate $B_{\mathbf{s}_1,jT}$ and $\Theta_{\mathbf{s}_1,jj}$. This can be done using

$$\hat{B}_{\mathbf{s}_1,jT} \equiv \frac{h^{d/2}}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \|\hat{\varphi}_{\mathbf{s}_1,ts}\|^2 \text{ and } \hat{\Theta}_{\mathbf{s}_1,jj} = \frac{2h^d}{T(T-1)} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \left[\frac{1}{T} \sum_{r=1}^{T} \hat{\varphi}_{\mathbf{s}_1,jrt}^{\top} \hat{\varphi}_{\mathbf{s}_1,jrs} \right]^2,$$

where $\hat{\varphi}_{\mathbf{s}_1,jts} = \mathbf{s}_1 \mathbf{s} S_T(U_t)^{-1} \bar{U}_s(U_t) F_s \tilde{\zeta}_{j,s}$. Equivalently, Theorem 3 can be stated more conveniently as follows:

$$\hat{W}_{\mathbf{s}_1,jT} \equiv \frac{W_{1jT} - \hat{B}_{\mathbf{s}_1,jT}}{\sqrt{\hat{\Theta}_{\mathbf{s}_1,jj}}} \xrightarrow{d} N(0,1) \text{ under } \mathbb{H}_0^{(1j)}. \tag{2.22}$$

This is due to the fact that $\hat{B}_{\mathbf{s}_1,jT} - B_{\mathbf{s}_1,jT} = o_P(1)$ and $\hat{\Theta}_{\mathbf{s}_1,jj} - \Theta_{\mathbf{s}_1,jj} = o_P(1)$, which can be justified easily. In applications, one can compare $\hat{W}_{\mathbf{s}_1,jT}$ with the one-sided critical value z_α, the upper α percentile from the $N(0,1)$ distribution, and reject the null at the asymptotic nominal level α if $\hat{W}_{\mathbf{s}_1,jT} > z_\alpha$.

The next theorem establishes the asymptotic distribution of our statistic for testing the correct specification of the parametric part.

Theorem 4 (Test for Correct Specification of Parametric Part) *Suppose that Assumptions A1-A4 and A6-A8 in the Appendix hold. Then,*

$$W_{2jT} - B_{I_k,jT} \xrightarrow{d} N(0, \Theta_{I_k,jj}) \tag{2.23}$$

under the null hypothesis $\mathbb{H}_0^{(2j)}$, *where* $B_{I_k,jT}$ *and* $\Theta_{I_k,jj}$ *are analogously defined as* $B_{\mathbf{s}_1,jT}$ *and* $\Theta_{\mathbf{s}_1,jj}$, *with the selection matrix* $\mathbf{s}_1$ *being replaced by the* $k \times k$ *identity matrix* I_k.

Remark 8. Note that the above theorem does not require that $Th^{d/2+4} \to 0$ as $T \to \infty$ as in Theorem 3. The reason is that under $\mathbb{H}_0^{(2)}$, $m_j(U_t) = 0$ a.s., so that the asymptotic bias from the second-stage nonparametric estimation vanishes automatically. Following the remark after Theorem 3, a feasible version of $W_{2,jT}$ is given by

$$\hat{W}_{I_k,jT} \equiv \frac{W_{2,jT} - \hat{B}_{I_k,jT}}{\sqrt{\hat{\Theta}_{I_k,jj}}}. \tag{2.24}$$

which is asymptotically distributed as $N(0,1)$ under $\mathbb{H}_0^{(2j)}$. Here, the definitions of $\hat{B}_{I_k,jT}$ and $\hat{\Theta}_{I_k,jj}$ follow from those of $\hat{B}_{\mathbf{s}_1,jT}$ and $\hat{\Theta}_{\mathbf{s}_1,jj}$, with $\mathbf{s}_1$ being replaced by I_k.

2.4.3 Joint tests for multiple assets

In many applications, it is of interest to test whether (2.17) or (2.19) holds for multiple assets. In this case, the null hypothesis will be either

$$\mathbb{H}_0^{(1)} : \beta_j^{\mathrm{I}}(U_t) = \theta_j^{\mathrm{I}} \text{ a.s. for } j = 1, 2, \cdots, N, \tag{2.25}$$

or

$$\mathbb{H}_0^{(2)} : m_j(U_t) = \mathbf{0}_{k\times 1} \text{ a.s. for } j = 1, 2, \cdots, N. \tag{2.26}$$

Natural test statistics for $\mathbb{H}_0^{(1)}$ and $\mathbb{H}_0^{(2)}$ would, respectively, be

$$\hat{W}_{\mathbf{s}_1,T} \equiv \frac{\sum_{j=1}^N W_{1,jT} - \sum_{j=1}^N \hat{B}_{\mathbf{s}_1,jT}}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \hat{\Theta}_{\mathbf{s}_1,ij}}} \tag{2.27}$$

and

$$\hat{W}_{I_k,T} \equiv \frac{\sum_{j=1}^N W_{2,jT} - \sum_{j=1}^N \hat{B}_{I_k,j}}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \hat{\Theta}_{I_k,ij}}}, \tag{2.28}$$

where $\hat{\Theta}_{\mathbf{s}_1,ij} = \frac{2h^d}{T(T-1)} \sum_{t=1}^T \sum_{s\neq t}^T (\frac{1}{T}\sum_{r=1}^T \hat{\varphi}_{\mathbf{s}_1,irt}^\top \hat{\varphi}_{\mathbf{s}_1,irs})(\frac{1}{T}\sum_{r=1}^T \hat{\varphi}_{\mathbf{s}_1,jrt}^\top \hat{\varphi}_{\mathbf{s}_1,jrs})$ and $\hat{\Theta}_{I_k,ij}$ is analogously defined. Following the proof of Theorems 3 and 4, we can readily show that under hypothesis $\mathbb{H}_0^{(1)}$, $\hat{W}_{\mathbf{s}_1,T} \xrightarrow{d} N(0,1)$ as $T \to \infty$, and under $\mathbb{H}_0^{(2)}$, $\hat{W}_{I_k,T} \xrightarrow{d} N(0,1)$ as $T \to \infty$.

2.4.4 A Bootstrap Version of Our Test

It is well known that many nonparametric tests based on their asymptotic normal null distributions may perform poorly in finite samples. We follow the literature (e.g., Hansen 2000; Su and Ullah 2013; Su, Murtazashvili, and Ullah 2013 (SMU hereafter)) and recommend using the *fixed-design wild bootstrap* method to obtain the bootstrap p-value for our test statistics.

Below we focus on the case of testing $\mathbb{H}_0^{(1j)}$: $\beta_j^{\mathrm{I}}(U_t) = \theta_j^{\mathrm{I}}$ a.s. using statistic $\hat{W}_{\mathbf{s}_1,jT}$ as the case for other test statistics is similar. The method can be described as follows:

1. First, obtain the semiparametric estimate $\hat{\beta}_j(U_t) = (\hat{\beta}_j^{\mathrm{I}}(U_t)^\top, \hat{\beta}_j^{\mathrm{II}}(U_t)^\top)^\top$ by using the bandwidth h^* and kernel function K and calculate the unrestricted residuals $\tilde{\zeta}_{j,t} = R_{j,t} - \hat{\beta}_j(U_t)^\top F_t$.
2. Generate the wild bootstrap residuals $\{\hat{u}_{j,t}\}$, $t = 1, \cdots, T$ from the centered fitted residuals $\hat{u}_{j,t} = \tilde{\zeta}_{j,t} - \bar{\tilde{\zeta}}_j$ with $\bar{\tilde{\zeta}}_j = \frac{1}{T}\sum_{t=1}^T \tilde{\zeta}_{j,t}$.
3. Define the bootstrap sample $R_{j,t}^* = [\hat{\theta}_j^{\mathrm{I}}]^\top F_t^{\mathrm{I}} + \hat{\beta}_j^{\mathrm{II}}(U_t)^\top F_t^{\mathrm{II}} + u_{j,t}^*$ with $u_{j,t}^* = \hat{u}_{j,t} \cdot \epsilon_{j,t}$, where

$\{\epsilon_{j,t}\}_{t=1}^{T}$ is a sequence of i.i.d. random variables with zero mean and unit variance that are independent of the data. If θ_j^{I} is known, then $\hat{\theta}_j^{\mathrm{I}} = \theta_j^{\mathrm{I}}$; otherwise, set $\hat{\theta}_j^{\mathrm{I}} \equiv \frac{1}{T}\sum_{t=1}^{T} \hat{\beta}_j^{\mathrm{I}}(U_t)$.

4. Calculate the bootstrap test statistic $\hat{W}^{*}_{\mathbf{s}_1,jT}$ in the same way as $\hat{W}_{\mathbf{s}_1,jT}$ using the bootstrap sample $\{R^{*}_{j,t}, F_t, U_t\}$ and the same bandwidth h and kernel function K as used to obtain $\hat{W}_{\mathbf{s}_1,jT}$.

5. Repeat the above steps 1-4 B times to obtain B bootstrap test statistics and label them as $\{\hat{W}^{*(b)}_{\mathbf{s}_1,jT}\}_{b=1}^{B}$. The bootstrap p-value for $\hat{W}_{\mathbf{s}_1,jT}$ is defined as $p^* = B^{-1}\sum_{b=1}^{B} 1(\hat{W}^{*(b)}_{\mathbf{s}_1,jT} \geq \hat{W}_{\mathbf{s}_1,jT})$. We reject the null hypothesis $\mathbb{H}_0^{(1j)}$ if p^* is smaller than the prescribed level of significance.

Remark 9. The above algorithm is similar to that in SMU who consider testing the correct specification of functional coefficient based on local linear GMM estimation in the cross section setting. We obtain the unrestricted residuals in Step 1 and center them to ensure zero sample mean in Step 2. The centering is commonly used but not required for the asymptotic theory because the way $\epsilon_{j,t}$ is generated in Step 3 can ensure the bootstrap error term $u^{*}_{j,t}$ to have zero mean conditional on the data. Note that in Step 3 we impose the null hypothesis $\mathbb{H}_0^{(1j)}$: $\beta_j^{\mathrm{I}}(U_t) = \theta_j^{\mathrm{I}}$ a.s. whereas SMU impose the null hypothesis: $\mathbb{H}_0^{(j)}$: $\beta_j(U_t) = \theta_j$ a.s. The latter is typically stronger than the tested one (unless $k_1 = k$) but can facilitate the justification of the asymptotic validity of the bootstrap procedure. As SMU remark, either way is fine and has both pros and cons. The way we generate the bootstrap resample requires that we should use oversmoothing bandwidth h^* to obtain $\hat{\beta}_j(U_t)$ used in the construction of the bootstrap observations $R^{*}_{j,t}$; see Härdle and Marron (1991) for the explanation. In the simulation and application below, we generate $\{\epsilon_{j,t}\}_{t=1}^{T}$ as an i.i.d. sequence from the standard normal distributions. After the bootstrap sample is generated, then one recalculate the bootstrap statistics and p-values as stated in Steps 4 and 5.

Remark 10. Following SMU and Härdle and Marron (1991), we can justify the asymptotic validity of the bootstrap test. Intuitively, because we impose $\mathbb{H}_0^{(1j)}$ in Step 3, the bootstrap test statistics $\{\hat{W}^{*(b)}_{\mathbf{s}_1,jT}\}_{b=1}^{B}$ have the asymptotic distribution $N(0,1)$ *no matter whether the original sample is generated under this null hypothesis or not.* Note that the original test statistic $\hat{W}_{\mathbf{s}_1,jT}$ is asymptotically $N(0,1)$ under $\mathbb{H}_0^{(1j)}$ and our bootstrap statistic has the same asymptotic distribution. This ensures the correct asymptotic size of our bootstrap test. Further, one can follow SMU and show that the original test statistic $\hat{W}_{\mathbf{s}_1,jT}$ diverges to infinity at the rate $Th^{d/2}$ under the alternative whereas the bootstrap test statistics $\{\hat{W}^{*(b)}_{\mathbf{s}_1,jT}\}_{b=1}^{B}$ remain asymptotically distributed as $N(0,1)$. This ensures the consistency of our bootstrap test.

As a referee kindly points out, it is possible to consider other resampling schemes. For example, it is well known that subsampling also works in a variety of hypothesis testing problems and it does not need to impose the null hypothesis to generate the resampling data; see Politis, Romano, and Wolf (1999). Nevertheless, subsampling typically does not work as well as a bootstrap method in terms of asymptotic power when the latter works. To see this, let n_T be a sequence of positive integers such that $n_T \to \infty$ and $n_T/T \to 0$ as $T \to \infty$. Let $\hat{W}^{**}_{\mathbf{s}_1,jn_T}$ be the subsampling analogue of $\hat{W}_{\mathbf{s}_1,jT}$ based on a subsample of $\{U_t, F_t, R_{j,t}\}$ with n_T observations. Under the null hypothesis, both $\hat{W}_{\mathbf{s}_1,jT}$ and $\hat{W}^{*}_{\mathbf{s}_1,jn_T}$ are asymptotically distributed as $N(0,1)$, which ensures the correct asymptotic size for the subsampling-based test. Under the alternative, $\hat{W}_{\mathbf{s}_1,jT}$ diverges to infinity at a speed faster than $\hat{W}^{**}_{\mathbf{s}_1,jn_T}$, which ensures the asymptotic power of the subsampling-based test but at the same time indicates such an asymptotic power is lower than that of the bootstrap-based test because $\hat{W}^{*}_{\mathbf{s}_1,jn_T}$ remains asymptotically distributed as $N(0,1)$ even if the alternative hypothesis holds. Other bootstrap methods like moving block bootstrap or stationary bootstrap take into account the weak dependence structure in the data and may also work for our testing problem. For example, Hwang and Shin (2012) have recently justified the asymptotic validity of the stationary bootstrap applied to kernel estimators of densities and derivatives. We conjecture that we can also generate the bootstrap observations $\{U_t^*, F_t^*\}$ via the stationary bootstrap first and then obtain $\{R^*_{j,t}\}$ as in Step 3 by replacing $\{U_t, F_t\}$ by $\{U_t^*, F_t^*\}$, and then justify the asymptotic validity of such a bootstrap method. We leave the formal study of such a bootstrap procedure in future research.

2.5 Choice of Bandwidth and Kernel Function

In this section, we discuss the choice of the bandwidth and the kernel function for our methods. It is well known that the choice of bandwidth parameter plays a critical role in many kernel-based nonparametric inferences. It is desirable to have a reliable bandwidth selection procedure that is data-driven and yet easily implementable.

For the estimation of $\beta_j(u)$, several approaches are possible. One approach is to apply a "plug-in" method to obtain an estimate of h as described in Fan and Gijbels (1996) (Ch. 4.2 for the single regressor case). Without assuming the correct specification of the parametric part, we can consider choosing h to minimize the asymptotic mean integrated squared error (AMISE) of our estimator $\hat{\beta}_j(u)$. As long as the second derivatives of the $g(u, \gamma_j^0)$ and $\beta_j(u)$ with respect to u do not fully match each other, the resulting "optimal" bandwidth h_T^* converges to zero at the rate $T^{-1/(d+4)}$. However, as Mishra, Su, and Ullah (2010) remarked, such an approach can not be easily implemented in the case of combined estimation for two reasons. First, since h_T^* depends on several unknown quantities that need to be estimated by some pilot bandwidth, the performance of our estimate $\hat{\beta}_j(u)$ will be contingent upon the choice of such a pilot bandwidth and the estimates of these unknown quantities. Second, the AMISE can not be minimized in the case of correct parametric specification because Theorem 2 implies that the optimal bandwidth should now be a fixed, finite constant.

In this paper, we consider two data-driven ways to choose the bandwidth to obtain the estimate $\hat{\beta}_j(u)$. One is to use the leave-one-out least squares cross validation (LSCV) to obtain choices of data-driven bandwidth, and the other is to adopt the bias-corrected AIC (AIC_c) of Hurvich, Simonoff, and Tsai (1997). To allow different variations of conditioning variables in $U_t = (U_{1,t}, ..., U_{d,t})^\top$, we choose $\mathbf{h} = (h_1, ..., h_d)$ to minimize the following LSCV criterion function

$$\text{CV}_j(\mathbf{h}) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\zeta}_{j,t} - \hat{m}_{j,-t}^{(\mathbf{h})}(U_t)^\top F_t\right)^2 w(U_t), \tag{2.29}$$

where $\hat{\zeta}_{j,t}$ is the residual from the first-stage parametric regression, $\hat{m}_{j,-t}^{(\mathbf{h})}$ is the second-step local linear estimator obtained using all observations except the one at time t, and $w(\cdot)$ is a nonnegative weight function, e.g., $w(U_t) = \Pi_{i=1}^{d}\mathbf{1}\left\{\left\|U_{i,t} - \bar{U}_i\right\| \le 2s_i\right\}$, with $\bar{U}_i$ and s_i being the sample mean and standard deviation of $U_{i,t}$, respectively.

To consider the AIC_c criterion, we need to find the effective number of parameters. Let e_i denote a $k \times 1$ vector with one in the ith position and zeros elsewhere. Let $s_i(u) \equiv e_i^\top \mathbf{s}\left(\mathbf{X}_u^\top \mathbf{K}_u \mathbf{X}_u\right)^{-1}\mathbf{X}_u^\top \mathbf{K}_u$, a $1\times T$ vector. Then the local linear estimate $\hat{m}_{ij}(u)$ of $m_{ij}(u)$ is given by $\hat{m}_{ij}(u) = s_i(u)\hat{\zeta}_j$. Let $S_i \equiv \left(s_i(U_1)^\top, ..., s_i(U_T)^\top\right)^\top$ and $\hat{m}_{ij} \equiv (\hat{m}_{ij}(U_1), ..., \hat{m}_{ij}(U_T))^\top$. Then $\hat{m}_{ij} = S_i\hat{\zeta}_j$ and

$$\hat{\sigma}^2 \equiv \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\zeta}_{j,t} - \hat{m}_j(U_t)^\top F_t\right]^2 = \frac{1}{T}\left\|\hat{\zeta}_j - \sum_{i=1}^{k}\left(S_i\hat{\zeta}_j\right)\odot \mathbf{F}_i\right\|^2 = \frac{1}{T}\left\|(I_T - H)\hat{\zeta}_j\right\|^2$$

where $\odot$ denotes the Hadamard product, $\mathbf{F}_i = (F_{i,1}, ..., F_{i,T})^\intercal$ for $i = 1, ..., k$, and $H = \sum_{i=1}^{k} S_i \odot (\mathbf{F}_i \mathbf{1}_{1\times T})$ is the $T \times T$ "hat matrix". Note that we have suppressed the dependence of S_i, H, and $\hat{\sigma}^2$ on the bandwidth. Analogous to the case of linear regression models, the effective number of parameters in our model is given by $\text{tr}(H)$. Then the AIC_c of Hurvich, Simonoff, and Tsai (1997) is defined as follows:

$$\text{AIC}_c = \log\left(\hat{\sigma}^2\right) + 1 + \frac{2\{\text{tr}(H) + 1\}}{T - \text{tr}(H) - 2}.$$

One chooses the bandwidth by minimizing the above AIC_c criterion.

Hart and Vieu (1990) claimed that the usual leave-one-out LSCV is robust to moderate amount of dependence in the data but some improvement can be obtained by considering leave-$(2\ell + 1)$-out LSCV with $\ell \ge 1$ when the data are sufficiently highly dependent. Yao and Tong (1998) argued that this is true only for regressions with fixed design and thus it does not apply to our setup with random covariates. Huang and Shen (2004) compared the finite sample performance of AIC, AIC$_c$, BIC, and the modified cross-validation (MCV) of Cai, Fan, and Yao (2000) in their spline polynomial estimation of functional coefficient models for nonlinear time series, and found that the AIC and AIC$_c$ behave similarly and both outperform the BIC and MCV. So in this paper, we focus on the comparison of AIC$_c$ with the LSCV method.

For the hypothesis testing part, to construct the test statistics, we will consider the sensitivity of our test for different choices of bandwidth by setting $\mathbf{h} = c s_{U_t} T^{-1/(d/2+3)}$ for different values of c (say, $c = 1, 1.5, 2$), where s_{U_t} stacks the sample standard deviations of elements of U_t. In our simulation and empirical work, we choose the Gaussian density as the kernel function: $k(u) = \exp\left(-u^2/2\right)/\sqrt{2\pi}$.

3 Monte Carlo Simulations

In this section we conduct a small set of Monte Carlo simulations to illustrate the finite sample performance of our semiparametric estimators and tests.

3.1 Evaluation of the Semiparametric Estimates

To study the finite performance of our semiparametric estimator, we simulate 500 random samples with sample size $T = 400$ according to the following data generating process (DGP):

$$R_t = \beta(U_t) R_{m,t} + e_t,$$

where $\beta(U_t) = 1 + U_t + U_t^2 + U_t^3$, $U_t = 0.9U_{t-1} + \varepsilon_t$, $R_{m,t} \sim N(0.017, 0.08^2)$, $e_t \sim N(0, 0.1^2)$, $\varepsilon_t \sim N(0, 0.5^2)$ and is truncated between $[-1, 1]$.

For each random sample, we obtain three different estimator for $\beta(u)$: the parametric estimator $g(u, \hat{\gamma})$, the one-step nonparametric estimator $\tilde{\beta}(u)$, and our semiparametric estimator $\hat{\beta}(u)$. To obtain $g(u, \hat{\gamma})$ and $\hat{\beta}(u)$, we use three parametric specifications: (1) the cubic specification with $g(U_t, \gamma) = \gamma_0 + \gamma_1 U_t + \gamma_2 U_t^2 + \gamma_3 U_t^3$, (2) the quadratic specification with $g(U_t, \gamma) = \gamma_0 + \gamma_1 U_t + \gamma_2 U_t^2$, and (3) the linear specification with $g(U_t, \gamma) = \gamma_0 + \gamma_1 U_t$. We consider both the LSCV and AIC$_c$ methods to choose the bandwidth to compute $\hat{\beta}(u)$ and $\tilde{\beta}(u)$. As mentioned, we use the Gaussian kernel throughout. The average time for the estimation per replication is 3.1 seconds for LSCV and 14.3 seconds for AIC$_c$ on our Intel(R) Core (TM) i7-2820QM CPU.

To evaluate the finite sample performance of different estimators, we calculate both the mean absolute deviation (MAD) and mean squared error (MSE) for each estimate evaluated at all T data points. In the case of the semiparametric estimator $\hat{\beta}(u)$, we have:

$$MAD = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{\beta}(U_t) - \beta(U_t)\right| \text{ and } MSE = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\beta}(U_t) - \beta(U_t)\right]^2.$$

The MAD and MSE measures are defined analogously for $g(u,\hat{\gamma})$ and $\tilde{\beta}(u)$.

Table 1 provides the results where the MADs and MSEs are averages over 500 replications for each estimator. When the parametric form is correctly specified (Panel A), the parametric estimator $g(u,\hat{\gamma})$ produces the smallest MAD and MSE among the three estimators. However, when the parametric form is misspecified (Panels B and C), the parametric estimator produces the largest MAD and MSE. The semiparametric estimator $\hat{\beta}(u)$ produces much smaller MAD and MSE than the one-step nonparametric estimator $\tilde{\beta}(u)$ when the parametric form is correctly specified (Panel A). When the first-stage parametric form is linear (Panel B), the semiparametric estimator and the one-step nonparametric estimator are quite close to each other and they produce almost the same MAD and MSE. When the first-stage parametric form is quadratic (Panel C), the semiparametric estimator produces smaller MAD and MSE than the one-step nonparametric estimator. While both the linear and quadratic specifications are misspecified, the quadratic specification captures some of the shape features of $\beta(U_t)$ and therefore the semiparametric estimator achieves bias reduction in comparison to the one-step nonparametric estimator. This is consistent with our discussion in Remark 3.

Theorem 1 provides the asymptotic standard errors for our semiparametric estimator $\hat{\beta}(u)$. In Table 2, we use simulations to examine the accuracy of the asymptotic standard errors. We choose five points on the support of U for the above DGP: $u = -0.2$, -0.1, 0, 0.1, and 0.2. We then estimate $\hat{\beta}(u)$ using the bandwidth chosen by LSCV and AIC$_c$ methods, respectively. $E[\hat{\beta}(u)]$ is the average estimated $\hat{\beta}(u)$ over 500 random samples, $std[\hat{\beta}(u)]$ is the average estimated standard errors of $\hat{\beta}(u)$ over 500 random samples, where the standard errors are computed using Theorem 1 , and $sstd[\hat{\beta}(u)]$ is the simulated standard deviation of estimated $\hat{\beta}(u)$ that is obtained over 500 random samples. We observe that the theoretical standard errors $std[\hat{\beta}(u)]$ are quite close to the simulated standard errors $sstd[\hat{\beta}(u)]$.

3.2 Size and Power of the Tests

In this section, we study the size and power for our new tests. To match our empirical studies later, we generate artificial data based on cases in which the conditional CAPM holds or fails.

More specifically, we consider the following data-generating process:

$$R_t = \alpha\left(U_t\right) + \beta\left(U_t\right) R_{m,t} + e_t, e_t \sim N(0, 0.1^2),$$

where $R_{m,t} \sim N(0.017, 0.08^2)$, and the state variable U_t is generated according to an AR(1) process: $U_t = 0.9U_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0, 0.5^2)$ and is truncated between $[-1, 1]$. We consider the following four specifications for the evolutions of $\alpha\left(U_t\right)$ and $\beta\left(U_t\right)$:

$$\alpha\left(U_t\right) = 0, \qquad \beta\left(U_t\right) = 1 + U_t, \tag{DGP1}$$

$$\alpha\left(U_t\right) = 0, \qquad \beta\left(U_t\right) = 1 + U_t + U_t^2 + U_t^3, \tag{DGP2}$$

$$\alpha\left(U_t\right) = 0.3 + 0.8U_t, \qquad \beta\left(U_t\right) = 1, \tag{DGP3}$$

$$\alpha\left(U_t\right) = 0.3 + 0.8U_t + 0.4U_t^2 + 0.4U_t^3, \quad \beta\left(U_t\right) = 1. \tag{DGP4}$$

When testing the conditional CAPM, a key test is on the conditional alpha $\alpha\left(U_t\right)$. If the conditional CAPM holds, then $\alpha\left(U_t\right) = 0$ for all time t. DGP1 and DGP2 simulate data when the conditional CAPM holds, and they are designed to examine the size of the constancy test on $\alpha\left(U_t\right)$. Another important test in empirical asset pricing is the constancy test on the conditional betas $\beta\left(U_t\right)$. DGP3 and DGP4 simulate the data when betas are constant, and are designed to examine the size of the constancy test on $\beta\left(U_t\right)$. Because $\alpha\left(U_t\right)$ varies with state variables in DGP3 and DGP4 rather than being zero, DGP3 and DGP4 also serve to examine the power of the constancy test on $\alpha\left(U_t\right)$. Because $\beta\left(U_t\right)$ varies with the state variable in DGP1 and DGP2 rather than being constant, DGP1 and DGP2 also serve to examine the power of the constancy test on $\beta\left(U_t\right)$. An important contribution of our semiparametric method is the test on $m\left(U_t\right)$, i.e., the test on whether the first-stage parametric form is correctly specified. To examine the size and the power of the specification test on $m\left(U_t\right)$, in DGP1 and DGP3, $\alpha\left(U_t\right)$ and $\beta\left(U_t\right)$ are linear functions of state variables U_t, while in DGP2 and DGP4 either $\alpha\left(U_t\right)$ or $\beta\left(U_t\right)$ is a nonlinear function of U_t.

For each of DGP1-4, we generate 500 random samples with sample size $T = 100$ and $T = 400$, respectively. For each random sample, we obtain the semiparametric estimators for $\alpha\left(U_t\right)$ and $\beta\left(U_t\right)$ by conducting a two-stage estimation. In the first-stage regression, we consider two parametric forms for $\alpha\left(U_t\right)$ and $\beta\left(U_t\right)$: (i) a linear parametric specification with

$$g^{\alpha}(U_t, \gamma) = a_0 + a_1 U_t \text{ and } g^{\beta}\left(U_t, \gamma\right) = b_0 + b_1 U_t,$$

and (ii) a quadratic parametric specification with

$$g^{\alpha}(U_t, \gamma) = a_0 + a_1 U_t + a_2 U_t^2 \text{ and } g^{\beta}(U_t, \gamma) = b_0 + b_1 U_t + b_2 U_t^2.$$

In the second-stage regression, we first get the residuals from the first-stage regression: $\hat{\zeta}_t = R_t - g^{\alpha}(U_t, \hat{\gamma}) - g^{\beta}(U_t, \hat{\gamma}) R_{m,t}$ where $\hat{\gamma}$ is the first-stage parametric estimator of γ. Then we fit $\hat{\zeta}_t$ as a nonparametric function of U_t: $\hat{\zeta}_t = m^{\alpha}(U_t) + m^{\beta}(U_t) R_{m,t} + \eta_t$. The estimators $\hat{m}^{\alpha}(U_t)$ and $\hat{m}^{\beta}(U_t)$ can be obtained by minimizing (2.9), and they are the second-stage nonparametric estimators for $\alpha(U_t)$ and $\beta(U_t)$, respectively. Finally, our semiparametric estimators for $\alpha(U_t)$ and $\beta(U_t)$ are given by $\hat{\alpha}(U_t) = g^{\alpha}(U_t, \hat{\gamma}) + \hat{m}^{\alpha}(U_t)$ and $\hat{\beta}(U_t) = g^{\beta}(U_t, \hat{\gamma}) + \hat{m}^{\beta}(U_t)$, respectively.

For each random sample, we construct three test statistics to examine three hypotheses: (1) $\mathbb{H}_0^{(1)}$: $\alpha(U_t) = 0$, (2) $\mathbb{H}_0^{(2)}$: $\beta(U_t) = 1$, and (3) $\mathbb{H}_0^{(3)}$: $m^{\alpha}(U_t) = m^{\beta}(U_t) = 0$. The test statistics for $\mathbb{H}_0^{(1)}$ and $\mathbb{H}_0^{(2)}$ are provided in equation (2.18), and the test statistic for $\mathbb{H}_0^{(3)}$ is provided in equation (2.20). To construct the relevant test statistic, we choose $h = c s_U T^{-1/(d/2+3)}$, where s_U is the sample standard deviation of U_t. We try three different values of c to check the sensitivity of our test to the choice of bandwidth: $c = 1, 1.5, 2$. Because we do not want to assume that the first-stage parametric model is correctly specified, our tests require some sort of undersmoothing. The p-values of the test statistics are obtained using the procedure described in Section 2.4.4 with $B = 200$ bootstrap resamples. When generating bootstrap data under the null, we use the parameters estimated using $h^* = s_U T^{-1/(d+6)}$. To compare the performance of our semiparametric test with the one-step nonparametric test, we also construct test statistics for $\mathbb{H}_0^{(1)}$ and $\mathbb{H}_0^{(2)}$ according to equation (2.18) by replacing $\hat{\beta}_j^{\mathrm{I}}(U_t)$ with the one-step nonparametric estimator.

For each DGP, we calculate the rejection frequency of different test statistics across 500 random samples. Table 3 provides the simulation results when the sample size $T = 100$. Our semiparametric estimator exhibits very good sizes regardless of whether the first-stage regression adopts a linear parametric form (Panel A), or the first-stage regression adopts a quadratic form (Panel B). When the first-stage regression adopts a linear parametric form, our semiparametric test has similar power to that of the one-step nonparametric test (Panel C). However, when the first-stage regression adopts a quadratic parametric form, our semiparametric test has higher power than that of the one-step nonparametric test. Table 4 provides the simulation results when the sample size T increases to 400. We observe that the sizes of all three tests generally improve and their powers increase fast with the majority of powers close to 1.

4 Empirical Applications

The performance of the conditional CAPM has attracted enormous research efforts in recent asset-pricing studies. Depending on the methods used (parametric or nonparametric methods), the literature has offered controversial results. In our empirical studies, we use our new method to examine the performance of the conditional CAPM in the presence of three influential state variables, which have been emphasized in the recent asset-pricing literature.

The state variables examined in our study are: the consumption–wealth ratio of Lettau and Ludvigson (2001) (*cay*), the labor income-consumption ratio of Santos and Veronesi (2006) (*yc*), and the corporate bond spread as in Jagannathan and Wang (1996) (*def*). The *cay* data are obtained from Martin Lettau's Website. Following Santos and Veronesi (2006), we obtain *yc* as the labor income component of *cay*. The *def* series is calculated as the yield difference between Baa- and Aaa-rated bonds, obtained from the Federal Reserve Bank of St. Louis. The data on these state variables run from 1952.Q1 to 2012.Q2. Because the portfolios sorted by book-to-market (B/M) ratios have presented arguably the greatest empirical challenge to the unconditional CAPM (Fama and French 1993), we use B/M portfolios as our test portfolios. More specifically, from the 25 size-B/M portfolios obtained from Kenneth French's Website, we form three B/M portfolios. G is the average of the five portfolios in the lowest B/M quintile, V is the average of the five portfolios in the highest B/M quintile, and V-G is their the difference. We compound monthly portfolio returns to obtain quarterly returns which run from 1952.Q2 to 2012.Q3.

The conditional CAPM states that

$$R_{j,t} = \alpha_j\left(U_t\right) + \beta_j\left(U_t\right) R_{m,t} + e_{j,t}. \tag{4.1}$$

Here, $R_{j,t}$ is the excess return of portfolio j at time t, $R_{m,t}$ is the market excess return at time t, $\alpha_j\left(U_t\right)$ and $\beta_j\left(U_t\right)$ are portfolio j's conditional alpha and beta at time $t-1$, respectively. The state variable U_t summarizes the information set at time $t-1$. In our context, we consider three choices of U_t : cay_{t-1}, yc_{t-1}, and def_{t-1}. Our semiparametric method estimates (4.1) in two stages. Similar to our simulation analysis in Section 3.2, we use both a linear and a quadratic specification in the first-stage regression to obtain the parametric estimators.

To evaluate the performance of the conditional CAPM, we conduct three hypothesis tests. First, we examine whether or not, when conditioning on U_t, the conditional CAPM can price a single portfolio as well as multiple portfolios. If the conditional CAPM is able to price a portfolio j, then the conditional alpha (i.e., conditional pricing error) associated with portfolio j should be equal to zero at all time t. This amounts to testing the null hypothesis: $\mathbb{H}_0^{(1)} : \alpha_j\left(U_t\right) = 0$ a.s. for $j = 1, \cdots, N$. If the conditional CAPM is able to price all N portfolios jointly, the conditional pricing errors associated with any portfolio j should be zero at all time t, which means that $\mathbb{H}_0^{(1)}$ should hold across all N assets. The test statistics for a single portfolio j and across all N portfolios can be obtained from (2.22) and (2.27), respectively. Second, an important question in finance is whether betas are indeed time varying (e.g., Bollerslev, Engle, and Wooldridge 1988; Ferson and Harvey 1991; Ferson and Korajczyk 1995). We investigate whether the conditional CAPM betas are time varying by examining $\mathbb{H}_0^{(2)}$: $\beta_j(U_t) = \bar{\beta}_j$, where $\bar{\beta}_j$ is the unconditional CAPM beta for portfolio j. Similar to the test on the conditional alpha, the relevant test statistics can be obtained from (2.22) and (2.27). Finally, we conduct a model specification test on the first-step parametric form. If the first-stage parametric estimators are correctly specified, then the second-stage nonparametric estimators $m_j^{\alpha}(U_t)$ and $m_j^{\beta}(U_t)$ are not needed. That is, $\mathbb{H}_0^{(3)} : m_j^{\alpha}(U_t) = m_j^{\beta}(U_t) = 0$ a.s. We test $\mathbb{H}_0^{(3)}$ both for a single portfolio j and across all N portfolios by utilizing the test statistics in (2.24) and (2.28), respectively. The p-value for each test statistic is obtained based on 200 bootstraps using the procedure described in Section 2.4.4. We use $h = cs_U T^{-1/(d/2+3)} (c = 1, 1.5, 2)$ to construct the test statistics. To compare our semiparametric test results with those using a one-step nonparametric test, we also construct test statistics for testing $\mathbb{H}_0^{(1)}$ and $\mathbb{H}_0^{(2)}$.

Table 5 provides the bootstrap p-values of the tests on the conditional alpha. Panel A uses the linear first-stage parametric specification, and Panel B uses the quadratic first-stage specification. The results show that the conditional CAPM is strongly rejected for V and V-G when conditioning on either of the three state variables. The conditional CAPM is also strongly rejected for G when conditioning on *cay* and *yc*. The joint test testing that the conditional CAPM holds for all three portfolios yields a p-value of virtually 0 in all cases, indicating that the model is strongly rejected for pricing the three B/M portfolios simultaneously.

Figures 1 plots the quarterly conditional alphas for the three portfolios with respect to the state variable *yc*, together with their corresponding two-standard-deviation confidence bands. These conditional alphas are obtained from our two-stage estimation with first-stage specification being linear, and the confidence bands are based on the standard errors in Theorem 1. Consistent with our formal test in Panel A of Table 5, the conditional alphas of V and V-G largely stay above zero while the conditional alphas of G largely stay below zero.

Overall, our empirical results show that when conditioning on *cay*, *yc*, and *def*, the B/M portfolios remain a serious challenge for the conditional CAPM. These results are consistent with those of the one-step nonparametric test provided in Panel C of Table 5, but run counter to the conclusions of several recent influential studies (e.g., Jagannathan and Wang 1996; Lettau and Ludvigson 2001; Santos and Veronesi 2006), who argue that conditioning dramatically improves the performance of both the simple and consumption CAPMs. As illustrated in Lewellen and Nagel (2006), by focusing on cross-sectional regressions rather than time-series intercept tests, these extant studies ignore important restrictions on the cross-sectional slopes. Lewellen, Nagel, and Shanken (2010) further argue that high cross-sectional R^2 or low cross-sectional pricing errors are low hurdles for claiming the success of a model. Our new test, on the other hand, looks at the time series of squared deviations of conditional pricing errors from zero, which renders a more powerful test. Our empirical findings on the conditional CAPM also complement those in Nagel and Singleton (2011) who find that the consumption-based models in the presence of the same set of state variables fail to capture the return variations of size and B/M portfolios.

Table 6 presents the empirical results for testing the null hypothesis that the conditional betas are equal to their unconditional counterparts. The semiparametric tests using a linear and quadratic first-stage specification are provided in Panels A and B, respectively. Both tests show that for *cay*, we cannot reject the null for all three B/M portfolios at all conventional significance levels; for *yc*, we reject the null for every portfolio at least at the 10% level; for *def*, we reject the null for G and V-G at the 5% level, but not for V. The nonparametric test in Panel C provides similar results.

Figures 2 plots the quarterly conditional betas of the three portfolios with respect to *yc*, together with their two-standard-deviation confidence bands. These conditional betas are obtained from our two-stage estimation with first-stage specification being linear, and the confidence bands are based on the standard errors in Theorem 1. Consistent with our formal test in Panel A of Table 6, the conditional betas of all three portfolios appear time-varying, suggesting that the risk of these portfolios varies with the business cycle proxied by *yc*.

Table 7 provides the test results for $\mathbb{H}_0^3 : m_j^\alpha(U_t) = m_j^\beta(U_t) = 0$. When the first-stage parametric form is linear (Panel A), for *cay* the null is not rejected for any portfolio at all conventional significance levels; for *yc*, the null is strongly rejected for every portfolio; for *def*, the null is rejected for G and V-G at the 5% level, but not for V. The joint test testing $\mathbb{H}_0^3$ across all three portfolios is strongly rejected for *yc* and *def*. When the first-stage parametric form is quadratic (Panel B), the results are quite similar to those in Panel A except that the null is no longer rejected for G when the state variable is *yc*. Our analysis therefore shows whether the second-step nonparametric estimator is needed varies greatly with state variables and with portfolios. For certain portfolios, it is crucial to model the nonlinear dynamics of the state variables which can go beyond the commonly used linear and quadratic forms.

5 Conclusions

This paper develops a new methodology for estimating and testing conditional factor models in modern finance. Our method naturally unifies two existing approaches in the literature—the traditional parametric approach and the nonparametric approach—and thereby retains the distinct advantages of both approaches. We propose new tests for investigating important issues in the context of conditional factor models, such as the tests on the conditional alphas and betas, and the test on the correct specification of the first-stage parametric model.

In light of the controversial results in the empirical literature, we apply our new method to examine the performance of the conditional CAPM in explaining return variations of portfolios sorted by book-to-market ratios. Our results run counter to the conclusions of several recent studies, who argue that conditioning dramatically improves the performance of the conditional CAPM. Due to the unique advantage of our combined procedure, we further show that, for some portfolios it is important to model the nonlinear functional forms of the conditional alphas and betas.

Appendix

A Technical Assumptions

Our theoretical analysis is based on the model

$$R_{j,t} = \beta_j (U_t)^\top F_t + \zeta_{j,t},\ j = 1, \cdots, N,\ t = 1, \cdots, T,$$

where $\beta_j(\cdot) = (\beta_{1j}(\cdot), \cdots, \beta_{kj}(\cdot))^\top$, $F_t = (F_{1,t}, \cdots, F_{k,t})^\top$, and $F_{1,t} \equiv 1$ in most applications. If $F_{1,t} \equiv 1$, then one can replace F_t in the following assumptions by $f_t = (F_{2,t}, \cdots, F_{k,t})^\top$ everywhere for $k \geq 2$. Let $\zeta_t = (\zeta_{1,t}, \cdots, \zeta_{N,t})^\top$.

Recall $p_U(\cdot)$ denotes the PDF of U_t. Let $p_{U,F}(\cdot,\cdot)$ and $p_{U|F}(\cdot|f)$ denote the joint density of (U_t, F_t) and the conditional PDF of U_t given $F_t = f$, respectively. Let $p_t(\cdot,\cdot|f,\ \tilde{f})$ be the conditional density of (U_1, U_t) given $(F_1, F_t) = (f, \tilde{f})$ for $t \geq 2$. Let $\mathcal{U}$ and $\mathcal{F}$ denote the support of $p_U(\cdot)$ and that of the PDF of F_t, respectively. Let $Q_{jT}(\gamma_j) \equiv T^{-1}\sum_{t=1}^{T}[R_{j,t} - g(U_t, \gamma_j)^\intercal F_t]^2$, $A_{jT}(\gamma_j) \equiv \partial^2 Q_{jT}(\gamma_j)/\partial\gamma_j\partial\gamma_j^\intercal$, $Q_j(\gamma_j) \equiv E[Q_{jT}(\gamma_j)]$, and $A_j \equiv E[A_{jT}(\gamma_j^0)]$. Recall $D_\gamma g(\bar{u}, \gamma_j) \equiv \partial g(\bar{u}, \gamma_j)/\partial\gamma_j^\intercal$. Let $D_{\gamma\gamma} g_l(\bar{u}, \gamma_j) = \partial^2 g_l(\bar{u}, \gamma_j)/\partial\gamma_j\partial\gamma_j^\intercal$ for $l = 1, ..., k$. Let $\zeta_{j,t}(\gamma_j) \equiv R_{j,t} - g(U_t, \gamma_j)^\intercal F_t$. Note that $\zeta_{j,t}(\gamma_j^0) = \zeta_{j,t}$ if (2.14) holds and the two objects may differ otherwise. We use C to denote a generic finite constant whose value may vary across lines.

The following assumptions are used in the establishment of the asymptotic distributions of our estimators and test statistics.

Assumption A1. (i) The process $\{(\zeta_t, U_t, F_t),\ t \geq 1\}$ is a strictly stationary α-mixing process with coefficients $\alpha(s)$ satisfying $\sum_{s=1}^{\infty} s^c [\alpha(s)]^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$ and $c > \delta/(2+\delta)$. $E(\zeta_t|U_t, F_t) = 0_{N\times 1}$ a.s.

(ii) $p_U(\cdot)$ is Lipschitz continuous of order 1 and $0 < p_U(u) \leq C < \infty$. $p_{U|F}(\bar{u}|\bar{f}) \leq C < \infty$ and $p_t(\bar{u}, \tilde{u}|\bar{f}, \tilde{f}) \leq C < \infty$ for all $t \geq 2$, $\bar{u}, \tilde{u}, \bar{f}$, and $\tilde{f}$. $\Omega(u)$ and $\Omega_{ij}^*(u)$ are positive definite for $i, j = 1, \cdots, N$. $\Omega(\cdot)$ and $\Omega_{ij}^*(\cdot)$ are continuous on $\mathcal{U}$ for $i, j = 1, \cdots, N$.

(iii) The second order partial derivatives of $\beta_j(\cdot)$ exist and are bounded and uniformly continuous on $\mathcal{U}$.

(iv) $E\|F_t\|^{2(2+\delta)} < \infty$ where δ is given in (i). $E[R_{j,1}^2 + R_{j,t}^2|(U_1, F_1) = (\bar{u}, \bar{f}), (U_t, F_t) = (\tilde{u}, \tilde{f})] \leq C < \infty$ for all $\bar{u}$ and $\tilde{u}$ in the neighborhood of u and $\bar{f},\ \tilde{f} \in \mathcal{F}$. There exists $\bar{\delta} > 2 + \delta$

such that $E[|R_{j,1}|^{\bar{\delta}}|\,(U_1, F_1) = (\tilde{u}, \tilde{f})] \leq C < \infty$ for all $\tilde{f} \in \mathcal{F}$ and all $\tilde{u}$ in the neighborhood of u. $\alpha(s) = O(s^{-\kappa})$, where $\kappa \geq (2+\delta)\,\bar{\delta}/\{2(\bar{\delta} - 2 - \delta)\}$.

(v) There exists a sequence of positive integers s_T such that $s_T \to \infty$, $s_T = o(T^{1/2}h^{d/2})$, and $T^{1/2}h^{-d/2}\alpha(s_T) \to 0$.

Assumption A2. (i) $Q_j(\gamma_j)$ is continuous on the compact parameter space Γ_j. γ_j^0 uniquely minimizes $Q_j(\gamma_j)$ over $\gamma_j \in \Gamma_j$ and is an interior point of Γ_j.

(ii) $\|g(\bar{u}, \gamma_j) - g(\bar{u}, \bar{\gamma}_j)\| \leq C_g(\bar{u})\,||\gamma_j - \bar{\gamma}_j||$ for some continuous function C_g, for all $\bar{u} \in \mathcal{U}$, and $\gamma_j, \bar{\gamma}_j \in \Gamma_j$, where $E\{[C_g(U_t) + C_g(U_t)^2]\,\|F_t\|^2\} \leq C < \infty$.

(iii) Both the Jacobian matrix $D_\gamma g(\bar{u}, \gamma_j)$ and the Hessian matrices $D_{\gamma\gamma} g_l(\bar{u}, \gamma_j)$, $l = 1, ..., k$, are uniformly continuous in γ_j in the neighborhood of γ_j^0 for all $\bar{u} \in \mathcal{U}$. $E[||D_{\gamma\gamma} g_l(U_t, \gamma_j^0)||^{1+\delta}$ $|F_{l,t}\zeta_{j,t}(\gamma_j^0)|^{1+\delta}] \leq C < \infty$ for $l = 1, ..., k$. A_j is positive definite. $E[||(D_\gamma g(U_t, \gamma_j^0))^\intercal F_t \zeta_{j,t}(\gamma_j^0)||^{2+\delta}]$ $\leq C < \infty$.

(iv) $g(\cdot, \gamma_j)$ has continuous second order partial derivatives with respect to its first argument for all γ_j in the neighborhood of γ_j^0.

Assumption A3. The kernel function $K(\cdot)$ is product kernel of $k(\cdot)$, which is a continuous, bounded, and symmetric PDF on the real line $\mathbb{R}$. Let $\mathbf{k}_s(z) \equiv z^s k(z)$ for $s = 0, 1, 2, 3, 4$. $\mathbf{k}_4(\cdot)$ is integrable on $\mathbb{R}$ with respect to Lebesgue measure, and $zk(z)$ is uniformly bounded on $\mathbb{R}$. For some $C_1 < \infty$ and $C_2 < \infty$, either $k(\cdot)$ is compactly supported such that $k(z) = 0$ for $\|z\| > C_1$, and $|\mathbf{k}_s(z) - \mathbf{k}_s(\tilde{z})| \leq C_2\,|z - \tilde{z}|$ for any z, $\tilde{z} \in \mathbb{R}$ and $s = 0, 1, 2, 3$; or $k(\cdot)$ is differentiable, $\|\partial \mathbf{k}_s(z)/\partial z\| \leq C_1$, and for some $\iota_0 > 1$, $|\partial \mathbf{k}_s(z)/\partial z| \leq C_1 |z|^{-\iota_0}$ for all $|z| > C_2$ and for $s = 0, 1, 2, 3$.

Assumption A4. As $T \to \infty$, $h \to 0$, $Th^d \to \infty$, $Th^{d+4} \to c \in [0, \infty)$.

Assumption A5. $\bar{S}(u) = E[S_T(u)]$ is positive definite, where $S_T(u)$ is defined in (B.2) below.

Assumption A6. The support $\mathcal{U}$ of $p_U(\cdot)$ is compact, and $p_U(\cdot)$ is uniformly bounded and bounded away from zero on $\mathcal{U}$. $\Omega(\bar{u})$ and $\Omega_{ij}^*(\bar{u})$ are positive definite for i, $j = 1, ..., N$ and for all $\bar{u} \in \mathcal{U}$.

Assumption A7. $E(\|F_t\zeta_{j,t}\|^{8(1+\delta)}) \leq C < \infty$ and $E(\|F_t\zeta_{j,t}\|^{8(1+\delta)}\,|U_t = \bar{u})$ is a continuous in $\bar{u}$. For each $1 < t_1 < ... < t_l$ $(l = 1, 2, 3)$, the joint density $p_{t_1, \cdots, t_l}(\cdot)$ of $(U_1, U_{t_1}, \cdots, U_{t_l})$ exists and satisfies the Lipschitz condition: $|p_{t_1, \cdots, t_l}(u^{(1)} + v^{(1)}, \cdots, u^{(l+1)} + v^{(l+1)}) - p_{t_1, \cdots, t_l}(u^{(1)}, \cdots,$ $u^{(l+1)})| \leq D_{t_1, ..., t_l}(\mathbf{u})\,\|\mathbf{v}\|$, where $\mathbf{u} = (u^{(1)\top}, \cdots, u^{(l+1)\top})^\top$, $\mathbf{v} = (v^{(1)\top}, \cdots, v^{(l+1)\top})^\top$, $\int D_{t_1, \cdots, t_l}$ $(\mathbf{u})\,\|\mathbf{u}\|^{2(1+\delta)}\,d\mathbf{u} < C$, and $\int D_{t_1, \cdots, t_l}(\mathbf{u})\,p_{t_1, \cdots, t_l}(\mathbf{u})\,d\mathbf{u} \leq C$.

Assumption A8. As $T \to \infty$, $Th^{d(1+3\delta)/(1+\delta)} \to 0$ and $\sum_{s=1}^{\infty} s^3 \alpha(s)^{\delta/(2+\delta)} < \infty$.

Assumption A1 is similar to Conditions A1-A2 in Cai, Fan, and Yao (2000). In particular, A1(i) is standard in the nonparametric regression for time series and it is satisfied by many well-known processes such as linear stationary ARMA processes, bilinear processes, nonlinear autoregressive processes, ARCH processes, and functional coefficient autoregressive processes; see Cai, Fan, and Yao (2000). Assumptions A2(i)-(iv) are about the parametric model used in the first-stage regression and can be easily verified for the commonly used models where $g(U_t, \gamma_j)$ is either linear or quadratic in U_t. A2(i) ensures unique identification of γ_j^0 and A2(ii) ensures that $\sup_{\gamma_j \in \Gamma_j} |Q_{jT}(\gamma_j) - Q_j(\gamma_j)| = o_P(1)$ by Theorem 21.11 in Davidson (1994) and the law of large number for stationary strong mixing processes (e.g., White 2001, Theorem 3.34 and Proposition 3.44). As a result, A2(i)-(ii) ensures the consistency of the NLS estimator $\hat{\gamma}_j$. In conjunction with A2(ii) and A1(i), A2(iii) ensures that $\sup_{\|\gamma_j-\gamma_j^0\|\leq\epsilon}\|A_{jT}(\gamma_j)-A_j\|=o_P(1)$ for any small $\epsilon>0$ and the $\sqrt{T}$-consistency of the NLS estimator $\hat{\gamma}_j$:

$$
\begin{aligned}
\sqrt{T}(\hat{\gamma}_j-\gamma_j^0) &= -\left[A_{jT}(\gamma_j^*)\right]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(D_\gamma g(U_t,\gamma_j^0))^\intercal F_t\zeta_{j,t}(\gamma_j^0) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}A_{j,t}(\gamma_j^0)\zeta_{j,t}(\gamma_j^0)+o_P(1)=O_P(1) \qquad \text{(A.1)}
\end{aligned}
$$

where γ_j^* is the mean value between $\hat{\gamma}_j$ and γ_j^0 and

$$A_{j,t}(\gamma_j^0)\equiv -A_j^{-1}(D_\gamma g(U_t,\gamma_j^0))^\intercal F_t. \qquad \text{(A.2)}$$

A2 (iv) is used to study the asymptotic distribution of our combined estimator. In addition, it is easy to verify that $A_j=E[(D_\gamma g(U_t,\gamma_j^0))^\top F_tF_t^\top D_\gamma g(U_t,\gamma_j^0)]-\sum_{l=1}^{k}E[D_{\gamma\gamma}g_l(U_t,\gamma_j^0)F_{l,t}\zeta_{j,t}(\gamma_j^0)]$, which simplifies to $A_j=E[(D_\gamma g(U_t,\gamma_j^0))^\top F_tF_t^\top D_\gamma g(U_t,\gamma_j^0)]$ if (2.14) holds.

Assumption A3 specifies conditions on the kernel function used in the second-stage estimation. It is used to obtain uniform consistency of our local linear estimator by applying the results of Masry (1996) and Hansen (2008) and is satisfied for the commonly used kernels, e.g., normal or Epanechnikov kernels. Assumption A4 imposes some basic conditions on the bandwidth, which are assumed to hold in all theorems but Theorem 2. Assumption A5 is required only in Theorem 2 where the bandwidth h is held fixed. It is automatically ensured by Assumptions A1(ii) and A3 if $h\to 0$ as $T\to\infty$. Assumption A6-A8 are standard in the literature and are used only in the proofs of Theorems 3-4. Note that A6 is required because we need to obtain estimates of $\beta_j(u)$ at each data point, and A7-A8 strengthen the conditions on the process $\{(\zeta_t,U_t,F_t),t\geq 1\}$ and the bandwidth to ensure some higher order term from the Hoeffding decomposition of our test statistics are asymptotically negligible and enable us to apply a version of central limit theorem for degenerate second order U-statistics for strong mixing processes. Note by choosing $\delta>0$ small enough in A1(i), the first condition in A8 is only slightly stronger than $Th^d\to 0$. But this would require that the mixing rate must decay sufficiently fast.

B Proof of Results in Section 2

Recall $\delta_j(u)=g(u,\hat{\gamma}_j)-g(u,\gamma_j^0)$ and $m_j(u)=\beta_j(u)-g(u,\gamma_j^0)$. It follows that

$$
\begin{aligned}
\hat{\beta}_j(u)-\beta_j(u) &= \hat{m}_j(u)-\left[\beta_j(u)-g(u,\gamma_j^0)\right]+\left[g(u,\hat{\gamma}_j)-g(u,\gamma_j^0)\right] \\
&= \hat{m}_j(u)-m_j(u)+\delta_j(u). \qquad \text{(B.1)}
\end{aligned}
$$

Proof of Theorem 1. For ease of notation, let

$$S_T(u) = \begin{pmatrix} S_{T,0}(u) & S_{T,1}(u) \\ S_{T,1}(u)^\top & S_{T,2}(u) \end{pmatrix} \text{ and } Q_{jT}(u) = Q_{jT,1}(u) + Q_{jT,2}(u) - Q_{jT,3}(u), \tag{B.2}$$

where

$$\begin{aligned}
S_{T,0}(u) &= T^{-1}\sum_{t=1}^{T} F_t F_t^\top K_{tu}, \\
S_{T,1}(u) &= T^{-1}\sum_{t=1}^{T} \left(F_t F_t^\top\right) \otimes \left(\left(U_t - u\right)/h\right)^\top K_{tu}, \\
S_{T,2}(u) &= T^{-1}\sum_{t=1}^{T} \left(F_t F_t^\top\right) \otimes \left(\left(\left(U_t - u\right)/h\right)\left(\left(U_t - u\right)/h\right)^\top\right) K_{tu}, \\
Q_{jT,1}(u) &= T^{-1}\sum_{t=1}^{T} \begin{pmatrix} F_t\zeta_{j,t} \\ \left(F_t\zeta_{j,t}\right) \otimes \left(\left(U_t - u\right)/h\right) \end{pmatrix} K_{tu}, \\
Q_{jT,2}(u) &= T^{-1}\sum_{t=1}^{T} \begin{pmatrix} F_t F_t^\top m_j(U_t) \\ \left(F_t F_t^\top m_j(U_t)\right) \otimes \left(\left(U_t - u\right)/h\right) \end{pmatrix} K_{tu}, \\
Q_{jT,3}(u) &= T^{-1}\sum_{t=1}^{T} \begin{pmatrix} F_t F_t^\top \delta_j(U_t) \\ \left(F_t F_t^\top \delta_j(U_t)\right) \otimes \left(\left(U_t - u\right)/h\right) \end{pmatrix} K_{tu}.
\end{aligned}$$

Then $\hat{m}_j(u) = \mathbf{s}S_T(u)^{-1} Q_T(u)$ by (2.10) and (2.7). Let $\theta_j(u) = (m_{1j}(u), \cdots, m_{kj}(u), h\dot{m}_{1j}(u)^\top, \cdots, h\dot{m}_{kj}(u)^\top)^\top$, where recall $m_{ij}(\cdot)$ denotes the ith element of $m_j(\cdot)$ and $\dot{m}_{ij}(u) = \partial m_{ij}(u)/\partial u$ for $i = 1, ..., k$. It follows that

$$\begin{aligned}
&\hat{m}_j(u) - m_j(u) \\
&= \mathbf{s}S_T(u)^{-1}\left[Q_{jT}(u) - S_T(u)\theta_j(u)\right] \\
&= \mathbf{s}S_T(u)^{-1} Q_{jT,1}(u) + \mathbf{s}S_T(u)^{-1}\left[Q_{jT,2}(u) - S_T(u)\theta_j(u)\right] - \mathbf{s}S_T(u)^{-1} Q_{jT,3}(u) \\
&\equiv \mathbb{V}_{jT}(u) + \mathbb{B}_{jT}(u) - \mathbb{R}_{jT}(u), \text{ say,}
\end{aligned} \tag{B.3}$$

where $\mathbb{V}_{jT}(u)$ and $\mathbb{B}_{jT}(u)$ are the usual asymptotic variance and bias terms, respectively, and $\mathbb{R}_{jT}(u)$ results from the first stage parametric estimation. Then by (B.1) we have

$$\hat{\beta}_j(u) - \beta_j(u) = \mathbb{V}_{jT}(u) + \mathbb{B}_{jT}(u) - \left[\mathbb{R}_{jT}(u) - \delta_j(u)\right] \tag{B.4}$$

We prove the theorem by showing that

$$\sqrt{Th^d}\mathbb{V}_{jT}(u) \xrightarrow{d} N\left(0, \mu_{0,2}^d \Omega(u)^{-1}\Omega_{jj}^*(u)\Omega(u)/p_U(u)\right), \tag{B.5}$$

$$\mathbb{B}_{jT}(u) = h^2 B_j(u) + o_P(h^2), \tag{B.6}$$

$$\mathbb{R}_{jT}(u) - \delta_j(u) = O_P(T^{-1/2}). \tag{B.7}$$

Define the $k(1+d) \times k(1+d)$ diagonal matrices $S(u)$ and $\Gamma_{ij}(u)$ by

$$S(u) = \Upsilon_1 \otimes \Omega(u) \text{ and } \Gamma_{ij}(u) = \Upsilon_2 \otimes \Omega_{ij}^*(u),$$

where $\Upsilon_1 = \text{diag}(1, \mu_{2,1}, \cdots, \mu_{2,1})$ and $\Upsilon_2 = \text{diag}(\mu_{0,2}^d, \mu_{2,2}, \cdots, \mu_{2,2})$ are both $(d+1) \times (d+1)$ diagonal matrices. Noting that $\mathbf{s}S(u)^{-1}\Gamma_{jj}(u)S(u)^{-1}\mathbf{s} = \mu_{0,2}^d\Omega(u)^{-1}\Omega_{jj}^*(u)\Omega(u)$, (B.5) holds by Lemmata A.1 and A.2 of Su, Chen, and Ullah (2009) as the former lemma implies that $S_T(u) = S(u)p_U(u) + o_P(1)$ and the latter implies that $\sqrt{Th^d}Q_{jT,1}(u) \xrightarrow{d} N(0,\ p_U(u)\Gamma_{jj}(u))$. [Su, Chen, and Ullah (2009) assume the kernel function has compact support, but this can be relaxed as in Hansen (2008).] Applying Lemma A.3 of Su, Chen, and Ullah (2009) with our $m_j(\cdot)$ and F_t in place of their $\alpha(\cdot)$ and X_i delivers (B.6). So we are left to show (B.7) only. Let $e_{jt} = F_t F_t^\top \delta_j(U_t)$. Noting that

$$\|\delta_j(u)\| = \left\|g(u, \hat{\gamma}_j) - g(u, \gamma_j^0)\right\| \leq C_g(u)\left\|\hat{\gamma}_j - \gamma_j^0\right\|, \tag{B.8}$$

we have $e_{jt}^\top e_{jt} = \delta_j(U_t)^\top F_t F_t^\top F_t F_t^\top \delta_j(U_t) \leq C_g(U_t)^2 \|F_t\|^4 \|\hat{\gamma}_j - \gamma_j^0\|^2$. Then by Minkowski inequality, straightforward calculations, Markov inequality, and (A.1), we have

$$\begin{aligned}
\|Q_{jT,3}(u)\| &\leq T^{-1}\sum_{t=1}^T \left\| \begin{pmatrix} e_{jt} \\ e_{jt} \otimes ((U_t - u)/h) \end{pmatrix} K_{tu} \right\| \\
&= T^{-1}\sum_{t=1}^T \left\{ e_{jt}^\top e_{jt}[1 + \|(U_t - u)/h\|^2] \right\}^{1/2} K_{tu} \\
&\leq \left\|\hat{\gamma}_j - \gamma_j^0\right\| \chi_T = O\left(\left\|\hat{\gamma}_j - \gamma_j^0\right\|\right) = O_P(T^{-1/2}),
\end{aligned}$$

where $\chi_T \equiv T^{-1}\sum_{t=1}^T C_g(U_t)\|F_t\|^2 \left\{[1 + \|(U_t - u)/h\|^2]\right\}^{1/2} K_{tu}$ satisfies $E(\chi_T) \leq C < \infty$ by Assumptions A2(ii) and A3. This, in conjunction with the fact that $S_T(u) = S(u)p_U(u) + o_P(1)$ and Assumptions A1(ii)-(iii), implies that $\|\mathbb{R}_{jT}(u)\| = O_P(T^{-1/2})$. Then (B.7) follows. ■

Proof of Theorem 2. The proof parallels that of Theorem 1. The major differences lie in two aspects: (a) we now hold h as fixed; (b) we rely on the fact that $m_j(u) \equiv 0$ under the correct specification of the first stage conditional factor model.

The decomposition in (B.3) continues to hold. Under (b), $\mathbb{B}_{jT}(u) \equiv 0$ so that our semiparametric estimator $\hat{\beta}_j(u)$ is asymptotically unbiased (up to order $T^{-1/2}$, which is the magnitude of $\mathbb{R}_{jT}(u) - \delta_j(u)$). Under (a), both $\mathbb{V}_{jT}(u)$ and $\mathbb{R}_{jT}(u) - \delta_j(u)$ contribute to the asymptotic variance of $\hat{\beta}_j(u)$. Noting that $\begin{pmatrix} A \\ A \otimes B \end{pmatrix} = \begin{pmatrix} I_k \\ I_k \otimes B \end{pmatrix} A$ for any $k \times 1$ vector A and $l \times 1$ vector B, we can write $Q_{jT,1}(u)$ and $Q_{jT,3}(u)$ respectively as

$$Q_{jT,1}(u) = T^{-1} \sum_{t=1}^{T} \bar{U}_t(u) F_t \zeta_{j,t} \text{ and } Q_{jT,3}(u) = T^{-1} \sum_{t=1}^{T} \bar{U}_t(u) F_t F_t^{\top} \delta_j(U_t),$$

where

$$\bar{U}_t(u) = K_{tu} \begin{pmatrix} I_k \\ I_k \otimes (U_t - u)/h \end{pmatrix}. \tag{B.9}$$

By (A.1) and Assumption A2(iii), $\delta_j(\bar{u}) = g(\bar{u}, \hat{\gamma}_j) - g(\bar{u}, \gamma_j^0) = D_\gamma g(\bar{u}, \gamma_j^0)(\hat{\gamma}_j - \gamma_j^0) + o_P\left(T^{-1/2}\right)$ uniformly in $\bar{u}$. Then

$$\begin{aligned}
&\sqrt{T}\left[\mathbb{V}_{jT}(u) - \mathbb{R}_{jT}(u) + \delta_j(u)\right] \\
&= \mathbf{s} S_T(u)^{-1} T^{-1/2} \sum_{t=1}^{T} \bar{U}_t(u) F_t \zeta_{j,t} - \mathbf{s} S_T(u)^{-1} T^{-1/2} \sum_{t=1}^{T} \bar{U}_t(u) F_t F_t^{\top} \delta_j(U_t) + \sqrt{T} \delta_j(u) \\
&= \mathbf{s} S_T(u)^{-1} T^{-1/2} \sum_{t=1}^{T} \bar{U}_t(u) F_t \zeta_{j,t} - \mathbf{s} S_T(u)^{-1} T^{-1} \sum_{t=1}^{T} \bar{U}_t(u) F_t F_t^{\top} D_\gamma g\left(U_t, \gamma_j^0\right) \sqrt{T}(\hat{\gamma}_j - \gamma_j^0) \\
&\qquad + D_\gamma g\left(u, \gamma_j^0\right) \sqrt{T}(\hat{\gamma}_j - \gamma_j^0) + o_P(1) \\
&= \mathbf{s} \bar{S}(u)^{-1} T^{-1/2} \sum_{t=1}^{T} \bar{U}_t(u) F_t \zeta_{j,t} + \bar{\Lambda}_j(u) \sqrt{T}(\hat{\gamma}_j - \gamma_j^0) + o_P(1) \\
&= T^{-1/2} \sum_{t=1}^{T} \left[\mathbf{s} \bar{S}(u)^{-1} \bar{U}_t(u) F_t + \Lambda_j(u) A_{j,t}\left(\gamma_j^0\right)\right] \zeta_{j,t} + o_P(1) \\
&\xrightarrow{d} N\left(0, \bar{\Sigma}_{jj}(u)\right)
\end{aligned}$$

where the third equality follows from the weak law of large number for strong mixing processes, $\bar{\Lambda}_j(u) = D_\gamma g(u, \gamma_j^0) - \Lambda_j(u)$, $\Lambda_j(u) = \mathbf{s} \bar{S}(u)^{-1} \mathbb{E}\left[\bar{U}_t(u) F_t F_t^{\top} D_\gamma g\left(U_t, \gamma^0\right)\right]$, $\bar{S}(u) = \mathbb{E}\left[S_T(u)\right]$, $A_{j,t}$ is defined in (A.2), and $\bar{\Sigma}_{jj}(u)$ is defined in the theorem. The last CLT result follows from Theorem 5.20 of White (2001) by Assumptions A1(i) and A2(ii) and by noting that $\left\|\bar{U}_t(u)\right\|$ is uniformly bounded in t under Assumption A3 and that $\bar{\Lambda}_j(u)$ and $\bar{S}(u)^{-1}$ are bounded too under Assumptions A2(iv) and the positive definiteness of $\bar{S}(u)$. ■

Remark. When h is held fixed, we cannot simplify the expression for $\Lambda_j(u)$. Nevertheless, if $h \to 0$ as $T \to \infty$, then we can show $\bar{S}(u) = S(u) p_U(u) + o(1)$ and $\Lambda_j(u) = D_\gamma g\left(u, \gamma_j^0\right) + o(1)$. In this case, $\bar{\Lambda}_j(u) = D_\gamma g(u, \gamma_j^0) - \Lambda_j(u) = o(1)$ so that the contribution from the first stage parametric estimation is asymptotically negligible.

Proof of Theorem 3. Decompose W_{1jT} as follows

$$\begin{aligned} W_{1jT} &= h^{d/2} \sum_{t=1}^{T} \left\| \left(\hat{\beta}_j^{\mathrm{I}} (U_t) - \beta_j^{\mathrm{I}} \right) - \left(\hat{\theta}_j^{\mathrm{I}} - \beta_j^{\mathrm{I}} \right) \right\|^2 \\ &= h^{d/2} \sum_{t=1}^{T} \left[\hat{\beta}_j^{\mathrm{I}} (U_t) - \beta_j^{\mathrm{I}} \right]^\top \left[\hat{\beta}_j^{\mathrm{I}} (U_t) - \beta_j^{\mathrm{I}} \right] - T h^{d/2} \left[\hat{\theta}_j^{\mathrm{I}} - \beta_j^{\mathrm{I}} \right]^\top \left[\hat{\theta}_j^{\mathrm{I}} - \beta_j^{\mathrm{I}} \right] \\ &\equiv W_{1jT,1} - W_{1jT,2}, \text{ say.} \end{aligned}$$

It suffices to show that under $\mathbb{H}_0^{(1j)}$, (i) $W_{1jT,1} - B_{\mathbf{s}_1,jT} \xrightarrow{d} N(0, \Theta_{\mathbf{s}_1,jj})$ and (ii) $W_{1jT,2} = o_P(1)$.

We first prove (i). Under $\mathbb{H}_0^{(1j)}$ we have by (B.4) that

$$\begin{aligned} W_{1jT,1} &= h^{d/2} \sum_{t=1}^{T} \{\mathbf{s}_1 \mathbb{V}_{jT}(U_t)\}^\top \{\mathbf{s}_1 \mathbb{V}_{jT}(U_t)\} + h^{d/2} \sum_{t=1}^{T} \{\mathbf{s}_1 \mathbb{B}_{jT}(U_t)\}^\top \{\mathbf{s}_1 \mathbb{B}_{jT}(U_t)\} \\ &+ h^{d/2} \sum_{t=1}^{T} \{\mathbf{s}_1 [\mathbb{R}_{jT}(U_t) - \delta_j(U_t)]\}^\top \{\mathbf{s}_1 [\mathbb{R}_{jT}(U_t) - \delta_j(U_t)]\} \\ &+ 2h^{d/2} \sum_{t=1}^{T} \{\mathbf{s}_1 \mathbb{V}_{jT}(U_t)\}^\top \mathbf{s}_1 \mathbb{B}_{jT}(U_t) - 2h^{d/2} \sum_{t=1}^{T} \{\mathbf{s}_1 \mathbb{V}_{jT}(U_t)\}^\top \{\mathbf{s}_1 [\mathbb{R}_{jT}(U_t) - \delta_j(U_t)]\} \\ &- 2h^{d/2} \sum_{t=1}^{T} \{\mathbf{s}_1 \mathbb{B}_{jT}(U_t)\}^\top \{\mathbf{s}_1 [\mathbb{R}_{jT}(U_t) - \delta_j(U_t)]\} \\ &\equiv W_{1jT,11} + W_{1jT,12} + W_{1jT,13} + 2W_{1jT,14} - 2W_{1jT,15} - 2W_{1jT,16}, \text{ say.} \end{aligned}$$

We prove (i) by showing that $W_{1jT,11} - B_{\mathbf{s}_1,jT} \xrightarrow{d} N(0, \Theta_{\mathbf{s}_1,jj})$ and $W_{1jT,1s} = o_P(1)$ under $\mathbb{H}_0^{(1j)}$ for $s = 2, \cdots, 6$.

First, noting that (B.6) also holds uniformly in u, we can show that $W_{1jT,12} = O_P\left(Th^{d/2+4}\right) = o_P(1)$, and by the proof of Theorem 1, $W_{1jT,13} = O_P\left(h^{d/2}\right) = o_P(1)$. It follows that $W_{1jT,16} = o_P(1)$ by Chebyshev inequality. Next, for $W_{1jT,14}$ we have

$$\begin{aligned} W_{1jT,14} &\simeq h^{d/2+2} \sum_{t=1}^{T} Q_{jT,1}(U_t)^\top \bar{S}_T(U_t)^{-1} \mathbf{s}^\top \mathbf{s}_1^\top \mathbf{s}_1 B_j(U_t) \\ &= h^{d/2+2} \sum_{s=1}^{T} \zeta_{j,s} F_s^\top \left[T^{-1} \sum_{t=1}^{T} \bar{U}_s(U_t)^\top \bar{S}_T(U_t)^{-1} \mathbf{s}^\top \mathbf{s}_1^\top \mathbf{s}_1 B_j(U_t) \right] \simeq \bar{W}_{1jT,14} \end{aligned}$$

where $\bar{W}_{1jT,14} = h^{d/2+2} \sum_{s=1}^{T} \zeta_{j,s} F_s^\top \Omega(U_s)^{-1} \mathbf{s}_1^\top \mathbf{s}_1 B_j(U_s)$, and the last results holds because uniformly in u

$$T^{-1}\sum_{t=1}^{T} K_h\left(u-U_t\right)\left(\mathbf{s}_1\mathbf{s}\bar{S}_T^{-1}\left(U_t\right)\begin{pmatrix} I_k \\ I_k\otimes\left(u-U_t\right)/h \end{pmatrix}\right)^{\top}\mathbf{s}_1^{\top}\mathbf{s}_1 B_j\left(U_t\right)$$
$$\simeq\left(\mathbf{s}S^{-1}\left(u\right)\begin{pmatrix} I_k \\ 0_{kd\times 1} \end{pmatrix}\right)^{\top}\mathbf{s}_1^{\top}\mathbf{s}_1 B_j\left(u\right)=\Omega\left(u\right)^{-1}\mathbf{s}_1^{\top}\mathbf{s}_1 B_j\left(u\right).$$

Let $\zeta_{j,s}^{*}=\zeta_{j,s}F_s^{\top}\Omega\left(U_s\right)^{-1}\mathbf{s}_1^{\top}\mathbf{s}_1 B_j\left(U_s\right).$ By Davydov inequality for strong mixing processes (e.g., Bosq 1996, p. 19), we have

$$E\left[\left(\bar{W}_{1jT,14}\right)^2\right]=h^{d+4}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\zeta_{j,s}^{*}\zeta_{j,t}^{*}\right]\leq CTh^{d+4}\sum_{\tau=1}^{\infty}\alpha\left(\tau\right)^{\delta/(2+\delta)}\left\{\mathbb{E}|\zeta_{j,1}^{*}|^{2+\delta}\right\}^{2/(2+\delta)}$$
$$=O\left(Th^{d+4}\right)=o\left(1\right).$$

It follows that $\bar{W}_{1jT,14}=o_P\left(1\right)$ by Chebyshev inequality. By analogous arguments we can show that $W_{1jT,15}=o_P\left(1\right).$

Now we analyze $W_{1jT,11}$. Noting that $S_T\left(u\right)-\bar{S}\left(u\right)=O_P\left(T^{-1/2}h^{-d/2}\sqrt{\log T}\right)$ uniformly in u by Masry (1996), we can show that

$$W_{1jT,11}=h^{d/2}\sum_{t=1}^{T}Q_{jT,1}\left(U_t\right)^{\top}S_T\left(U_t\right)^{-1}\mathbf{s}\mathbf{s}_1^{\top}\mathbf{s}_1\mathbf{s}S_T\left(U_t\right)^{-1}Q_{jT,1}\left(U_t\right)=\bar{W}_{1jT,11}+o_P\left(1\right),\quad\text{(B.10)}$$

where $\bar{W}_{1jT,11}=h^{d/2}\sum_{t=1}^{T}Q_{jT,1}\left(U_t\right)^{\top}\bar{S}\left(U_t\right)^{-1}\mathbf{s}^{\top}\mathbf{s}_1^{\top}\mathbf{s}_1\mathbf{s}\bar{S}\left(U_t\right)^{-1}Q_{jT,1}\left(U_t\right).$ Let

$$\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)=\mathbf{s}_1\mathbf{s}\bar{S}\left(U_t\right)^{-1}\bar{U}_s\left(U_t\right)F_s\zeta_{j,s}.\quad\text{(B.11)}$$

Then $\mathbf{s}_1\mathbf{s}\bar{S}\left(U_t\right)^{-1}Q_{jT,1}\left(U_t\right)=T^{-1}h^{d/2}\sum_{s=1}^{T}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)$ and

$$\begin{aligned}\bar{W}_{1jT,11}&=T^{-2}h^{d/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)^{\top}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,r}\right)\\&=T^{-2}h^{d/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)\right\|^2+T^{-2}h^{d/2}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\sum_{r\neq t,s}^{T}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)^{\top}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,r}\right)\\&\quad+2T^{-2}h^{d/2}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)^{\top}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,t}\right)\\&\equiv B_{1jT}+V_{1jT}+R_{1jT},\ \text{say.}\end{aligned}\quad\text{(B.12)}$$

Let $\varphi(\xi_{j,t},\xi_{j,s},\xi_{j,r})\equiv[\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,s}\right)^{\top}\varphi_{\mathbf{s}_1}\left(\xi_{j,t},\xi_{j,r}\right)+\varphi_{\mathbf{s}_1}\left(\xi_{j,s},\xi_{j,t}\right)^{\top}\varphi_{\mathbf{s}_1}\left(\xi_{j,s},\xi_{j,r}\right)+\varphi_{\mathbf{s}_1}\left(\xi_{j,r},\xi_{j,t}\right)^{\top}$

$\varphi_{\mathbf{s}_1}(\xi_{j,r}, \xi_{j,s})]/3$, which is symmetric in its arguments. Then

$$V_{1jT} = \frac{6h^{d/2}}{T^2} \sum_{1 \leq t < s < r \leq T} \varphi(\xi_{j,t}, \xi_{j,s}, \xi_{j,r}) = \frac{(T-1)(T-2)}{T} \bar{V}_{1jT},$$

where $\bar{V}_{1jT} \equiv \frac{6h^{d/2}}{T(T-1)(T-2)} \sum_{1 \leq t < s < r \leq T} \varphi(\xi_{j,t}, \xi_{j,s}, \xi_{j,r})$. Clearly, $\int \int \varphi(w, \xi_1, \xi_2) dP_{\xi_j}(\xi_1) \, dP_{\xi_j}(\xi_2) = 0$. Let $\varphi_2(w_1, w_2) = \int \varphi(w_1, w_2, \xi) dP_{\xi_j}(\xi) = \frac{1}{3} \int \varphi_{\mathbf{s}_1}(\xi, w_1)^\top \varphi_{\mathbf{s}_1}(\xi, w_2) \, dP_{\xi_j}(\xi)$. Let $\varphi_3(\xi_{j,t}, \xi_{j,s}, \xi_{j,r}) = \varphi(\xi_{j,t}, \xi_{j,s}, \xi_{j,r}) - \varphi_2(\xi_{j,t}, \xi_{j,s}) - \varphi_2(\xi_{j,t}, \xi_{j,r}) - \varphi_3(\xi_{j,s}, \xi_{j,r})$. By Hoeffding decomposition (c.f. Lee 1990, Chapter 1.6),

$$\bar{V}_{1jT} = 3H_{jT}^{(2)} + H_{jT}^{(3)},$$

where $H_{jT}^{(2)} \equiv \frac{2h^{d/2}}{T(T-1)} \sum_{1 \leq t < s \leq T} \varphi_2(\xi_{j,t}, \xi_{j,s})$ and $H_{jT}^{(3)} \equiv \frac{6h^{d/2}}{T(T-1)(T-2)} \sum_{1 \leq t < s < r \leq T} \varphi_3(\xi_{j,t}, \xi_{j,s}, \xi_{j,r})$. Noting that $\int \varphi_3(w_1, w_2, \xi) \, dP_{\xi_j}(\xi) = 0$ and that φ_3 is symmetric in its arguments by construction, we can apply Lemma A.2 of Gao (2007) (pp. 193-194) to obtain $\mathbb{E}[H_{jT}^{(3)}]^2 \leq Ch^d T^{-3} h^{-2d(1+2\delta)/(1+\delta)} = O\left(T^{-3} h^{-d(1+3\delta)/(1+\delta)}\right)$. Hence, $H_{jT}^{(3)} = O_P(T^{-3/2} h^{-d(1+3\delta)/(2(1+\delta))}) = o_P\left(T^{-1}\right)$ by Chebyshev inequality and Assumption A7. It follows that $V_{1jT} = \frac{T(T-2)}{T-1} \bar{V}_{1jT} = \{1 + o(1)\} \mathcal{V}_{1jT} + o_P(1)$, where

$$\mathcal{V}_{1jT} \equiv \frac{2h^{d/2}}{T} \sum_{1 \leq t \leq s \leq T} 3\varphi_2(\xi_{j,t}, \xi_{j,s}) = \frac{2h^{d/2}}{T} \sum_{1 \leq t < s \leq T} \int \varphi_{\mathbf{s}_1}(\xi, \xi_{j,t})^\top \varphi_{\mathbf{s}_1}(\xi, \xi_{j,s}) \, dP_{\xi_j}(\xi).$$

As $\mathcal{V}_{1jT}$ is a second order degenerate U-statistic, it is straightforward but tedious to verify that all the conditions of Theorem A.1 of Gao (2007) (p. 198) are satisfied, implying that a central limit theorem applies to $\mathcal{V}_{1jT}$: $\mathcal{V}_{1jT} \xrightarrow{d} N(0, \Theta_{\mathbf{s}_1, jj})$, where the asymptotic variance of $\mathcal{V}_{1jT}$ is given by $\Theta_{\mathbf{s}_1, jj} \equiv \lim_{T \to \infty} \Theta_{\mathbf{s}_1, jjT}$ and $\Theta_{\mathbf{s}_1, jjT} \equiv 2h^d E_t E_s \left[\int \varphi_{\mathbf{s}_1}(\xi, \xi_{j,t})^\top \varphi_{\mathbf{s}_1}(\xi, \xi_{j,s}) \, dP_{\xi_j}(\xi)\right]^2$, where E_t denotes expectation with respect to variables indexed by time t only. [A careful examination of the proof in the theorem indicates that the geometric strong mixing rate in Gao (2007) can be relaxed to our arithmetic rate.] For R_{1jT}, we can apply Lemma A.2 of Gao (2007) to obtain $\mathbb{E}(R_{1jT}^2) \leq CT^{-2} h^{-d(2+3\delta)/(1+\delta)} = o(1)$ and Assumptions A4 and A7. So $R_{1jT} = o_P(1)$ by Chebyshev inequality. It follows that $W_{1jT,11} - B_{1jT} \xrightarrow{d} N(0, \Theta_{\mathbf{s}_1, jjT})$.

Now we show (ii). Under $\mathbb{H}_0^{(1j)}$, we have by (B.4) and (B.2) that

$$\begin{aligned}
\sqrt{T}\left(\hat{\theta}_j^{\mathrm{I}} - \beta_j^{\mathrm{I}}\right) &= T^{-1/2} \sum_{t=1}^{T} \mathbf{s}_1 \mathbf{s} S_T(U_t)^{-1} Q_{jT,1}(U_t) \\
&\quad + T^{-1/2} \sum_{t=1}^{T} \mathbf{s}_1 \mathbf{s} S_T(U_t)^{-1} \left[Q_{jT,2}(U_t) - S_T(U_t)\, \theta_j(U_t)\right] \\
&\quad - T^{-1/2} \sum_{t=1}^{T} \mathbf{s}_1 \left[\mathbf{s} S_T(U_t)^{-1} Q_{jT,3}(U_t) - \delta_j(U_t)\right] \equiv A_{1jT} + A_{2jT} + A_{3jT}, \text{ say.}
\end{aligned}$$

It follows that $W_{1jT,2} = h^{d/2} (A_{1jT} + A_{2jT} + A_{3jT})^\top (A_{1jT} + A_{2jT} + A_{3jT}) \leq 3h^{d/2}(A_{1jT}^\top A_{1jT} + A_{2jT}^\top A_{2jT} + A_{3jT}^\top A_{3jT})$. Under our assumptions (note that $h \to 0$ here), it is easy to show that $A_{1jT}^\top A_{1jT} = O_P(1)$, $A_{2jT}^\top A_{2jT} = O_P(Th^4)$, and $A_{3jT}^\top A_{3jT} = O_P(1)$. It follows that $W_{1jT,2} = O_P\left(h^{d/2} + Th^{d/2+4}\right) = o_P(1)$. ■

Proof of Theorems 4. The proof is analogous to that of Theorem 3. So we only outline the difference. Decompose W_{2jT} as follows $W_{2jT} = h^{d/2} \sum_{t=1}^T \|\hat{m}_j(U_t)\|^2 - Th^{d/2}\overline{\hat{m}}_j^2 \equiv W_{2jT,1} - W_{2jT,2}$, say. It is easy to show that $W_{2jT,2} = o_P(1)$. For $W_{2jT,1}$, by (B.3) and arguments analogous to those used in the analysis of $W_{1jT,1}$, we can show that

$$W_{2jT,1} = h^{d/2} \sum_{t=1}^T \|\mathbb{V}_{jT}(U_t) + \mathbb{B}_{jT}(U_t) + \mathbb{R}_{jT}(U_t)\|^2 = W_{2jT,11} + o_P(1) \text{ under } \mathbb{H}_0^{(2j)},$$

where $W_{2jT,11} = h^{d/2} \sum_{t=1}^T \|\mathbb{V}_{jT}(U_t)\|^2$. The major difference is that $\mathbb{B}_{jT}(U_t) = 0$ a.s. under $\mathbb{H}_0^{(2j)}$ so that we do not need the condition $Th^{d/2+4} = o(1)$ as in the proof of Theorem 3.

Furthermore, $W_{2jT,11} = \bar{W}_{2jT,11} + o_P(1)$, where $\bar{W}_{2jT,11} = h^{d/2} \sum_{t=1}^T Q_{jT,1}(U_t)^\top \bar{S}(U_t)^{-1} \mathbf{s}^\top \mathbf{s}\bar{S}(U_t)^{-1} Q_{jT,1}(U_t)$. So the asymptotic bias and variance of $\bar{W}_{2jT,11}$ are determined as those of $\bar{W}_{1jT,11}$ with $\mathbf{s}_1$ being replaced by I_k everywhere. Consequently $W_{2jT,11} - B_{I_k,jT} \xrightarrow{d} N(0, \Theta_{I_k,jj})$. This completes the proof. ■

ACKNOWLEDGMENTS

The authors sincerely thank Rong Chen, the associate editor, and two anonymous referees for their many insightful comments and suggestions that lead to a substantial improvement of the presentation. We also thank Hendrik Bessembinder, Frank Diebold, Yongmiao Hong, Jingzhi Huang, David Ng, David Reeb, Yu Ren, Oleg Rytchkov, and the participants in the SMU-ESSEC Symposium on Empirical Finance & Financial Econometrics and 2012 China International Conference in Finance for their helpful comments and suggestions. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

References

Ait-Sahalia, Y. and Duarte, J. (2003), "Nonparametric Option Pricing under Shape Restrictions," *Journal of Econometrics* 116, 9-47.

Ait-Sahalia, Y. and Lo, A. W. (1998), "Nonparametric Estimation of State-price Densities Implicit in Financial Asset Prices," *Journal of Finance* 53, 499-547.

Ang, A. and Kristensen, D. (2010), "Testing Conditional Factor Models," *Journal of Financial Economics* 106, 132-156.

Bansal, R., Hsieh, D. A., and Viswanathan, S. (1993), "A New Approach to International Arbitrage Pricing," *Journal of Finance* 48, 1719-1747.

Bansal, R. and Viswanathan, S. (1993), "No Arbitrage and Arbitrage Pricing: A New Approach," *Journal of Finance* 48, 1231-1262.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988), "A Capital Asset Pricing Model with Time-varying Covariances," *Journal of Political Economy* 96, 116-131.

Borak, S. and Weron, R. (2008), "A Semiparametric Factor Model for Electricity Forward Curve Dynamics," *Journal of Energy Markets* 1, 3-16.

Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Springer-Verlag, Berlin.

Brandt, M. W. and Chapman, D. A. (2008), "Composition of Wealth, Conditioning Information, and the Cross-section of Stock Returns," *Working Paper*, Duke University.

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-coeffcient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association* 95, 941-956.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1996), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

Cochrane, J. H. (1996), "A Cross-sectional Test of an Investment-based Asset Pricing Model," *Journal of Political Economy* 104, 572-621.

Cochrane, J. H. (2005), *Asset Pricing*, Princeton University Press, New Jersey.

Davidson, J. (1994), *Stochastic Limit Theory,* Oxford University Press, New York.

Dybvig, P. H. and Ross, S. A. (1985), "Differential Information and Performance Measurement Using a Security Market Line," *Journal of Finance* 40, 383-399.

Fama, E. F. and French, K. R. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.

Fama, E. F. and MacBeth, J. D. (1973), "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy* 81, 607-636.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall/CRC, New York.

Fan, Y. and Ullah, A. (1999), "Asymptotic Normality of a Combined Regression Estimator," *Journal of Multivariate Analysis* 71, 191-240.

Ferson, W. E. and Harvey, C. R. (1991), "The Variation of Economic Risk Premiums," *Journal of Political Economy* 99, 385-415.

Ferson, W. E. and Harvey, C. R. (1999), "Conditioning Variables and the Cross Section of Stock Returns," *Journal of Finance* 54, 1325-1360.

Ferson, W. E. and Korajczyk, R. A. (1995), "Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?" *Journal of Business* 68, 309-349.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19, 3-29.

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman and Hall/CRC, New York.

Gao, Q., Gu, J. and Hernandez-Verme, P. (2012), "A Semiparametric Time Trend Varying Coefficients Model: with an Application to Evaluate Credit Rationing in U.S. Credit Market," *Annals of Economics and Finance* 13, 189-210.

Ghysels, E. (1998), "On Stable Factor Structures in the Pricing of Risk: Do Time-varying Betas Help or Hurt?" *Journal of Finance* 53, 549-573.

Glad, I. K. (1998), "Parametrically Guided Non-parametric Regression," *Scandinavian Journal of Statistics* 25, 649-668.

Hansen, B. E. (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics* 97, 93-115.

Hansen, L. P. (2008), "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory* 24, 726-748.

Härdle, W. and Marron, J. S. (1991), "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *Annals of Statistics* 19, 778-796.

Hart, J. D. and Vieu, P. (1990), "Data-driven Bandwidth Choice for Density Estimation Based on Dependent Data," *Annals of Statistics* 18, 873-890.

Harvey, C. R. (2001), "The Specification of Conditional Expectations," *Journal of Empirical Finance* 8, 573-637.

Huang, J. Z. and Shen, H. (2004), "Functional Coefficient Regression Models for Non-linear Time Series: a Polynomial Spline Approach," *Scandinavian Journal of Statistics* 31, 515-534.

Hurvich, C. M., Simonoff, J. S., and Tsai, C. L. (1998), "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of the Royal Statistical Society Series B* 60, 271-293.

Hwang, E. and Shin, D. W. (2012), "Stationary Bootstrap for Kernel Density Estimators under Weak Dependence," *Computational Statistics and Data Analysis* 56, 1581-1593.

Jagannathan, R. and Wang, Z. (1996), "The Conditional CAPM and the Cross-section of Expected Returns," *Journal of Finance* 51, 3-53.

Jensen, M. C. (1968), "The Performance of Mutual Funds in the Period 1945-1964," *Journal of Finance* 23, 389-416.

Lee, A. J. (1990), *U-statistics: theory and practice*, Marcel Dekker, New York.

Lettau, M. and Ludvigson, S. (2001), "Resurrecting the (C)CAPM: A Cross-sectional Test When Risk Premia Are Time-varying," *Journal of Political Economy* 109, 1238-1287.

Lewellen, J. and Nagel, S. (2006), "The Conditional CAPM Does Not Explain Asset-pricing Anomalies," *Journal of Financial Economics* 82, 289-314.

Lewellen, J., Nagel, S., and Shanken, J. (2010), "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics* 96, 175-194.

Li, Y. and Yang, L. (2011), "Testing Conditional Factor Models: A Nonparametric Approach," *Journal of Empirical Finance* 18, 972-992.

Lintner, J. (1965), "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics* 47, 13-37.

Long, X., Su, L., and Ullah, A. (2011), "Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model," *Journal of Business & Economic Statistics* 29, 109-125.

Lustig, H. N. and Nieuwerburgh, S. G. V. (2005), "Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective," *Journal of Finance* 60, 1167-1219.

Martins-Filho, C., Mishra, S., and Ullah, A. (2008), "A Class of Improved Parametrically Guided Nonparametric Regression Estimators," *Econometric Reviews* 24, 542-573.

Masry, E. (1996), "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency Rates," *Journal of Time Series Analysis* 17, 571-599.

Mishra, S., Su, L., and Ullah, A. (2010), "Semiparametric Estimator of Time Series Conditional Variance," *Journal of Business & Economic Statistics* 28, 256-274.

Mossin, J. (1966), "Equilibrium in a Capital Asset Market," *Econometrica* 34, 768-783.

Nagel, S. and Singleton, K. J. (2011), "Estimation and Evaluation of Conditional Asset Pricing Models," *Journal of Finance* 66, 873-909.

Park, B. U., Mammen, E., Härdle, W., and Borak, S. (2009), "Time Series Modeling with Semiparametric Factor Dynamics," *Journal of the American Statistical Association* 104, 284-298.

Politis, D., Romano, J., and Wolf, M. (1999), *Subsampling,* New York: Springer-Verlag.

Ross, S. A. (1977), "The Capital Asset Pricing Model (CAPM), Short-sale Restrictions and Related Issues," *Journal of Finance* 32, 177-183.

Roussanov, N. (2014), "Composition of Wealth, Conditioning Information, and the Cross-section of Stock Returns," *Journal of Financial Economics* 111, 352-380.

Santos, T. and Veronesi, P. (2006), "Labor Income and Predictable Stock Returns," *Review of Financial Studies* 19, 1-44.

Shanken, J. (1990), "Intertemporal Asset Pricing: An Empirical Investigation," *Journal of Econometrics* 45, 99-120.

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

Su, L., Chen, Y., and Ullah, A. (2009), "Functional Coefficient Estimation with Both Categorical and Continuous Data," *Advances in Econometrics* 23, 131-167.

Su, L., Murtazashvili, I., and Ullah, A. (2013), "Local Linear GMM Estimation of Functional Coefficient IV Models with an Application to Estimating the Rate of Return to Schooling," *Journal of Business & Economic Statistics* 31, 184-207.

Su, L. and Ullah, A. (2013), "A Nonparametric Goodness-of-fit-based Test for Conditional Heteroskedasticity," *Econometric Theory* 29, 187-212.

Sun, Y. and Wu, H. (2005), "Semiparametric Time-varying Coefficients Regression Model for Longitudinal Data," *Scandinavian Journal of Statistics* 32, 21-47.

Wang, K. Q. (2003), "Asset Pricing with Conditioning Information: A New Test," *Journal of Finance* 58, 161-196.

White, H. L. (2001), *Asymptotic Theory for Econometricians*, Academic Press, New York.

Yao, Q. and Tong, H. (1998), "Cross-validatory Bandwidth Selection for Regression Estimation Based on Dependent Data," *Journal of Statistical Planning and Inference* 68, 387-715.

Table 1: Mean absolute deviation (MAD) and mean squared error (MSE) comparisons

Panel A: Cubic parametric specification				
	LSCV		AIC$_c$	
	MAD	MSE	MAD	MSE
Parametric	0.088	0.016	0.088	0.016
Nonparametric	0.195	0.082	0.164	0.080
Semiparametric	0.114	0.036	0.096	0.022
Panel B: Linear parametric specification				
	LSCV		AIC$_c$	
	MAD	MSE	MAD	MSE
Parametric	1.662	7.030	1.662	7.030
Nonparametric	0.195	0.082	0.164	0.080
Semiparametric	0.195	0.082	0.164	0.080
Panel C: Quadratic parametric specification				
	LSCV		AIC$_c$	
	MAD	MSE	MAD	MSE
Parametric	1.288	3.939	1.288	3.939
Nonparametric	0.195	0.082	0.164	0.080
Semiparametric	0.191	0.078	0.162	0.077

Table 2: Simulation on standard errors

		LSCV			AIC$_c$		
U	$\beta(u)$	$E[\hat{\beta}(u)]$	$std[\hat{\beta}(u)]$	$sstd[\hat{\beta}(u)]$	$E[\hat{\beta}(u)]$	$std[\hat{\beta}(u)]$	$sstd[\hat{\beta}(u)]$
-0.2	0.832	0.852	0.103	0.110	0.843	0.117	0.121
-0.1	0.909	0.948	0.102	0.103	0.930	0.115	0.116
0	1.000	1.059	0.101	0.106	1.032	0.114	0.120
0.1	1.111	1.189	0.102	0.116	1.154	0.115	0.128
0.2	1.248	1.348	0.104	0.130	1.305	0.117	0.143

Table 3: Size and power with sample size T=100

Panel A: Semiparametric test using linear first-stage specification										
		1%			5%			10%		
		$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$
DGP1	$cc = 1$	0.008	0.866	0.000	0.046	0.946	0.038	0.100	0.976	0.080
	$cc = 1.5$	0.012	0.936	0.002	0.046	0.978	0.042	0.090	0.990	0.086
	$cc = 2$	0.006	0.956	0.006	0.052	0.990	0.050	0.092	0.994	0.088
DGP2	$cc = 1$	0.016	0.982	0.798	0.054	0.996	0.914	0.106	0.998	0.940
	$cc = 1.5$	0.012	0.992	0.822	0.056	0.998	0.938	0.094	0.998	0.956
	$cc = 2$	0.006	0.994	0.822	0.054	0.998	0.936	0.100	0.998	0.958
DGP3	$cc = 1$	1.000	0.016	0.000	1.000	0.068	0.042	1.000	0.136	0.078
	$cc = 1.5$	1.000	0.018	0.002	1.000	0.066	0.050	1.000	0.126	0.078
	$cc = 2$	1.000	0.018	0.002	1.000	0.068	0.050	1.000	0.116	0.078
DGP4	$cc = 1$	1.000	0.022	0.720	1.000	0.062	0.844	1.000	0.122	0.890
	$cc = 1.5$	1.000	0.012	0.654	1.000	0.038	0.820	1.000	0.078	0.874
	$cc = 2$	1.000	0.010	0.560	1.000	0.032	0.742	1.000	0.078	0.834
Panel B: Semiparametric test using quadratic first-stage specification										
		1%			5%			10%		
		$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$
DGP1	$cc = 1$	0.008	0.998	0.008	0.052	1.000	0.064	0.094	1.000	0.116
	$cc = 1.5$	0.010	1.000	0.008	0.058	1.000	0.068	0.104	1.000	0.118
	$cc = 2$	0.010	1.000	0.014	0.064	1.000	0.072	0.106	1.000	0.122
DGP2	$cc = 1$	0.008	1.000	0.946	0.050	1.000	0.992	0.092	1.000	0.994
	$cc = 1.5$	0.008	1.000	0.966	0.058	1.000	0.992	0.102	1.000	0.996
	$cc = 2$	0.010	1.000	0.970	0.064	1.000	0.996	0.118	1.000	0.998
DGP3	$cc = 1$	1.000	0.022	0.008	1.000	0.084	0.070	1.000	0.136	0.116
	$cc = 1.5$	1.000	0.020	0.008	1.000	0.078	0.070	1.000	0.136	0.118
	$cc = 2$	1.000	0.020	0.014	1.000	0.082	0.072	1.000	0.136	0.122
DGP4	$cc = 1$	1.000	0.004	0.842	1.000	0.052	0.928	1.000	0.112	0.958
	$cc = 1.5$	1.000	0.010	0.830	1.000	0.046	0.914	1.000	0.102	0.940
	$cc = 2$	1.000	0.006	0.768	1.000	0.052	0.892	1.000	0.126	0.930
Panel C: One-step nonparametric test										
		1%			5%			10%		
		$H_0^{(1)}$	$H_0^{(2)}$		$H_0^{(1)}$	$H_0^{(2)}$		$H_0^{(1)}$	$H_0^{(2)}$	
DGP1	$cc = 1$	0.010	0.846		0.048	0.936		0.090	0.958	
	$cc = 1.5$	0.004	0.896		0.038	0.962		0.082	0.980	
	$cc = 2$	0.002	0.908		0.028	0.980		0.086	0.990	
DGP2	$cc = 1$	0.015	0.980		0.055	0.990		0.088	0.998	
	$cc = 1.5$	0.008	0.992		0.042	0.998		0.080	0.998	
	$cc = 2$	0.006	0.994		0.032	0.998		0.090	0.998	
DGP3	$cc = 1$	1.000	0.026		1.000	0.088		1.000	0.148	
	$cc = 1.5$	1.000	0.016		1.000	0.074		1.000	0.136	
	$cc = 2$	1.000	0.016		1.000	0.064		1.000	0.128	
DGP4	$cc = 1$	1.000	0.020		1.000	0.074		1.000	0.138	
	$cc = 1.5$	1.000	0.018		1.000	0.044		1.000	0.070	
	$cc = 2$	1.000	0.010		1.000	0.032		1.000	0.068	

Table 4: Size and power with sample size T=400

Panel A: Semiparametric test using linear first-stage specification

		1%			5%			10%		
		$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(3)}$	$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(3)}$	$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(3)}$
	$cc = 1$	0.014	1.000	0.012	0.048	1.000	0.052	0.084	1.000	0.094
DGP1	$cc = 1.5$	0.016	1.000	0.010	0.056	1.000	0.048	0.090	1.000	0.096
	$cc = 2$	0.014	1.000	0.008	0.056	1.000	0.050	0.092	1.000	0.092
	$cc = 1$	0.010	1.000	1.000	0.046	1.000	1.000	0.084	1.000	1.000
DGP2	$cc = 1.5$	0.018	1.000	1.000	0.054	1.000	1.000	0.092	1.000	1.000
	$cc = 2$	0.016	1.000	1.000	0.052	1.000	1.000	0.082	1.000	1.000
	$cc = 1$	1.000	0.016	0.012	1.000	0.056	0.052	1.000	0.106	0.088
DGP3	$cc = 1.5$	1.000	0.016	0.010	1.000	0.054	0.048	1.000	0.110	0.096
	$cc = 2$	1.000	0.014	0.008	1.000	0.058	0.050	1.000	0.100	0.092
	$cc = 1$	1.000	0.008	0.982	1.000	0.048	0.996	1.000	0.100	1.000
DGP4	$cc = 1.5$	1.000	0.004	0.970	1.000	0.042	0.994	1.000	0.072	0.996
	$cc = 2$	1.000	0.000	0.906	1.000	0.030	0.974	1.000	0.070	0.996

Panel B: Semiparametric test using quadratic first-stage specification

		1%			5%			10%		
		$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(3)}$	$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(3)}$	$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(3)}$
	$cc = 1$	0.014	1.000	0.012	0.050	1.000	0.052	0.086	1.000	0.086
DGP1	$cc = 1.5$	0.018	1.000	0.010	0.052	1.000	0.040	0.090	1.000	0.094
	$cc = 2$	0.016	1.000	0.004	0.060	1.000	0.046	0.092	1.000	0.100
	$cc = 1$	0.010	1.000	1.000	0.046	1.000	1.000	0.084	1.000	1.000
DGP2	$cc = 1.5$	0.012	1.000	1.000	0.040	1.000	1.000	0.086	1.000	1.000
	$cc = 2$	0.024	1.000	1.000	0.052	1.000	1.000	0.084	1.000	1.000
	$cc = 1$	1.000	0.014	0.010	1.000	0.054	0.058	1.000	0.102	0.094
DGP3	$cc = 1.5$	1.000	0.016	0.010	1.000	0.050	0.046	1.000	0.102	0.102
	$cc = 2$	1.000	0.014	0.006	1.000	0.050	0.056	1.000	0.102	0.110
	$cc = 1$	1.000	0.006	0.978	1.000	0.058	0.996	1.000	0.112	1.000
DGP4	$cc = 1.5$	1.000	0.010	0.966	1.000	0.062	0.996	1.000	0.108	0.998
	$cc = 2$	1.000	0.014	0.934	1.000	0.070	0.994	1.000	0.126	0.996

Panel C: One-step nonparametric test

		1%		5%		10%	
		$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$	$\mathbb{H}_0^{(1)}$	$\mathbb{H}_0^{(2)}$
	$cc = 1$	0.008	1.000	0.056	1.000	0.116	1.000
DGP1	$cc = 1.5$	0.014	1.000	0.064	1.000	0.114	1.000
	$cc = 2$	0.014	1.000	0.072	1.000	0.116	1.000
	$cc = 1$	0.006	1.000	0.060	1.000	0.114	1.000
DGP2	$cc = 1.5$	0.016	1.000	0.060	1.000	0.118	1.000
	$cc = 2$	0.018	1.000	0.066	1.000	0.108	1.000
	$cc = 1$	1.000	0.014	1.000	0.062	1.000	0.106
DGP3	$cc = 1.5$	1.000	0.016	1.000	0.056	1.000	0.102
	$cc = 2$	1.000	0.016	1.000	0.062	1.000	0.108
	$cc = 1$	1.000	0.014	1.000	0.046	1.000	0.112
DGP4	$cc = 1.5$	1.000	0.018	1.000	0.044	1.000	0.084
	$cc = 2$	1.000	0.008	1.000	0.030	1.000	0.056

Table 5: Bootstrap p-values for tests on conditional alpha

		V	G	V-G	Joint test
Panel A. Semiparametric test with linear first-stage form					
	$cc = 1$	0.010	0.005	0.000	0.000
cay	$cc = 1.5$	0.000	0.005	0.000	0.000
	$cc = 2$	0.000	0.005	0.000	0.000
	$cc = 1$	0.000	0.000	0.000	0.000
yc	$cc = 1.5$	0.005	0.020	0.000	0.000
	$cc = 2$	0.005	0.025	0.000	0.000
	$cc = 1$	0.010	0.115	0.005	0.000
def	$cc = 1.5$	0.005	0.075	0.000	0.000
	$cc = 2$	0.000	0.045	0.000	0.000
Panel B. Semiparametric test with quadratic first-stage form					
	$cc = 1$	0.005	0.005	0.000	0.000
cay	$cc = 1.5$	0.015	0.010	0.000	0.000
	$cc = 2$	0.015	0.000	0.000	0.000
	$cc = 1$	0.005	0.005	0.000	0.000
yc	$cc = 1.5$	0.000	0.000	0.000	0.000
	$cc = 2$	0.000	0.000	0.000	0.000
	$cc = 1$	0.000	0.145	0.000	0.000
def	$cc = 1.5$	0.005	0.145	0.000	0.000
	$cc = 2$	0.030	0.195	0.005	0.005
Panel C. One-step nonparametric test					
	$cc = 1$	0.005	0.010	0.000	0.000
cay	$cc = 1.5$	0.005	0.010	0.000	0.000
	$cc = 2$	0.005	0.005	0.000	0.000
	$cc = 1$	0.000	0.000	0.000	0.000
yc	$cc = 1.5$	0.000	0.005	0.000	0.000
	$cc = 2$	0.000	0.015	0.000	0.000
	$cc = 1$	0.000	0.155	0.000	0.020
def	$cc = 1.5$	0.000	0.115	0.000	0.010
	$cc = 2$	0.000	0.080	0.000	0.000

Table 6: Bootstrap p-values for tests on conditional beta

		V	G	V-G	Joint test
Panel A. Semiparametric test with linear first-stage form					
	$cc = 1$	0.370	0.940	0.240	0.375
cay	$cc = 1.5$	0.440	0.915	0.340	0.480
	$cc = 2$	0.515	0.890	0.405	0.565
	$cc = 1$	0.010	0.000	0.000	0.000
yc	$cc = 1.5$	0.050	0.000	0.000	0.000
	$cc = 2$	0.095	0.005	0.000	0.000
	$cc = 1$	0.535	0.020	0.020	0.015
def	$cc = 1.5$	0.525	0.010	0.015	0.025
	$cc = 2$	0.525	0.010	0.040	0.055
Panel B. Semiparametric test with quadratic first-stage form					
	$cc = 1$	0.285	0.895	0.305	0.385
cay	$cc = 1.5$	0.380	0.900	0.335	0.425
	$cc = 2$	0.425	0.925	0.280	0.385
	$cc = 1$	0.025	0.005	0.000	0.000
yc	$cc = 1.5$	0.015	0.000	0.000	0.000
	$cc = 2$	0.020	0.000	0.005	0.000
	$cc = 1$	0.570	0.045	0.005	0.025
def	$cc = 1.5$	0.510	0.035	0.020	0.035
	$cc = 2$	0.495	0.050	0.005	0.025
Panel C. One-step nonparametric test					
	$cc = 1$	0.455	0.915	0.325	0.500
cay	$cc = 1.5$	0.455	0.915	0.325	0.500
	$cc = 2$	0.560	0.925	0.385	0.685
	$cc = 1$	0.010	0.000	0.000	0.000
yc	$cc = 1.5$	0.045	0.005	0.000	0.000
	$cc = 2$	0.070	0.005	0.000	0.005
	$cc = 1$	0.465	0.035	0.010	0.160
def	$cc = 1.5$	0.450	0.010	0.015	0.110
	$cc = 2$	0.460	0.015	0.035	0.115

Table 7: Model specification test

		V	G	V-G	Joint test
Panel A. Semiparametric test with linear first-stage form					
cay	$cc = 1$	0.410	0.950	0.575	0.660
	$cc = 1.5$	0.405	0.885	0.575	0.615
	$cc = 2$	0.465	0.880	0.635	0.735
yc	$cc = 1$	0.005	0.000	0.000	0.000
	$cc = 1.5$	0.015	0.000	0.000	0.000
	$cc = 2$	0.020	0.000	0.000	0.000
def	$cc = 1$	0.375	0.000	0.005	0.010
	$cc = 1.5$	0.315	0.000	0.005	0.005
	$cc = 2$	0.360	0.000	0.010	0.005
Panel B. Semiparametric test with quadratic first-stage form					
cay	$cc = 1$	0.335	0.875	0.380	0.450
	$cc = 1.5$	0.300	0.825	0.395	0.460
	$cc = 2$	0.335	0.845	0.395	0.445
yc	$cc = 1$	0.020	0.220	0.035	0.010
	$cc = 1.5$	0.030	0.245	0.035	0.000
	$cc = 2$	0.030	0.200	0.025	0.005
def	$cc = 1$	0.375	0.010	0.005	0.010
	$cc = 1.5$	0.365	0.025	0.005	0.010
	$cc = 2$	0.425	0.020	0.005	0.020

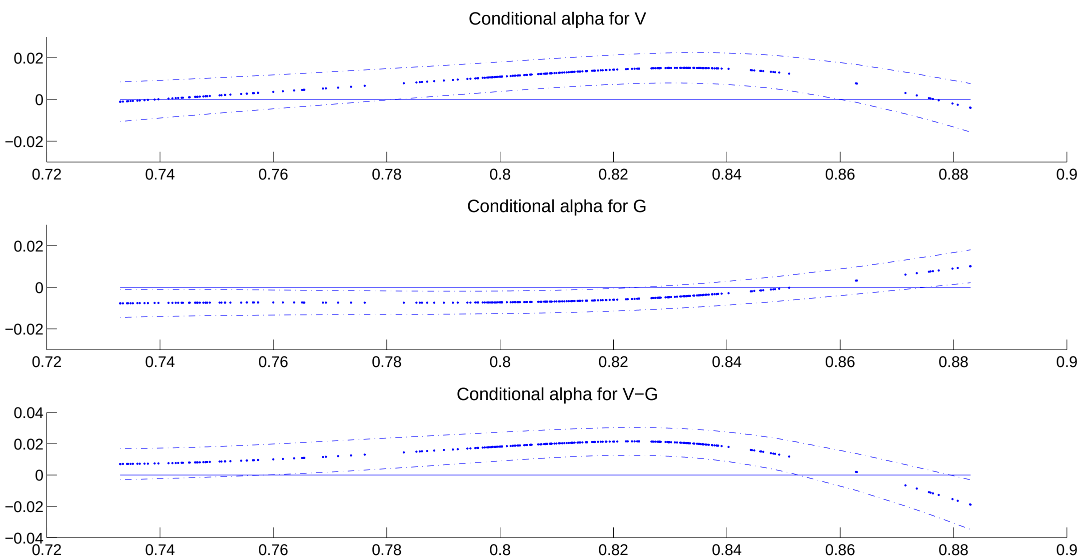

Figure 1: Plots of conditional alphas for V, G, and V-G when conditioning on labor income-consumption ratio. The conditional alphas are estimated using our two-stage estimation method where the first-stage specification is linear. The 95% pointwise confidence bands are computed based on Theorem 1. The solid line corresponds to the value zero.

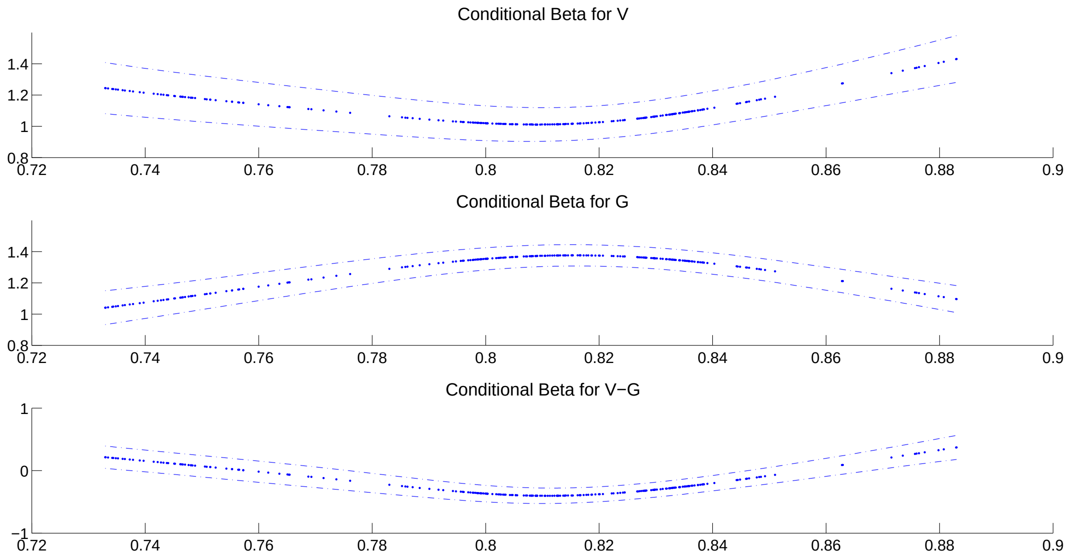

Figure 2: Plots of conditional betas for V, G, and V-G when conditioning on labor income-consumption ratio. The conditional betas are estimated using our two-stage estimation method where the first-stage specification is linear. The 95% pointwise confidence bands are computed based on Theorem 1.