cemmap
centre for microdata methods and practice

Semiparametric dynamic portfolio choice with multiple conditioning variables

Jia Chen
Degui Li
Oliver Linton
Zudi Lu

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP07/15

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Semiparametric Dynamic Portfolio Choice with Multiple Conditioning Variables

Jia Chen* Degui Li† Oliver Linton‡ Zudi Lu§

January 30, 2015

Abstract

Dynamic portfolio choice has been a central and essential objective for institutional investors in active asset management. In this paper, we study the dynamic portfolio choice depending on multiple conditioning variables, where the number of the conditioning variables can be either fixed or diverging to infinity at certain polynomial rate in comparison with the sample size. We propose a novel data-driven method to estimate the nonparametric optimal portfolio choice, motivated by the model averaging marginal regression approach suggested by Li, Linton and Lu (2014). Specifically, in order to avoid curse of dimensionality associated with the problem and to make it practically implementable, we first estimate the optimal portfolio choice by maximising the conditional utility function for each individual conditioning variable, and then construct the dynamic optimal portfolio choice through the weighted average of the marginal optimal portfolio across all the conditioning variables. Under some mild regularity conditions, we have established the large sample properties for the developed portfolio choice procedure. Both simulation studies and empirical application well demonstrate the performance of the proposed methodology with finite sample and real data.

JEL Subject Classifications: C13, C14, C32.

*Department of Economics and Related Studies, University of York, Heslington, YO10 5DD, UK. E-mail: jia.chen@york.ac.uk.

†Department of Mathematics, University of York, Heslington, YO10 5DD, UK. E-mail: degui.li@york.ac.uk.

‡Faculty of Economics, Cambridge University, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. E-mail: obl20@cam.ac.uk.

§School of Mathematical Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: Z.Lu@soton.ac.uk.

Keywords: Conditioning variables, kernel smoothing, model averaging, portfolio choice, utility function.

1 Introduction

Dynamic portfolio choice has been widely recognized as a central and essential objective for institutional investors in active asset management. In financial research, in fact, how to choose an optimal portfolio is a fundamental issue, see, Markowitz (1952), Merton (1969) and Fama (1970) for some early references, and Back (2010) and Brandt (2010) for some recent surveys. In practice, it is not uncommon that the dynamic portfolio choice depends on many conditioning (or forecasting) variables, which reflects the varying investment opportunities over time. Generally speaking, there are two ways to characterize the dependence of portfolio choice on the conditioning variables. One is to assume a parametric statistical model that relates the returns of risky assets to the conditioning variables and then solve for an investor's portfolio choice by using some traditional econometric approaches to estimate the conditional distribution of the returns. However, the assumed parametric models might be misspecified, which would lead to inconsistent or biased estimation of the optimal portfolio and invalid inference on it. The other way that can avoid the possible issue of model misspecification, is to use some nonparametric methods such as the kernel estimation method to characterize the dependence of the portfolio choice on conditioning variables. This latter way was first introduced by Brandt (1999), who also establishes the asymptotic properties for the estimated portfolio choice and provides an empirical application.

Although the nonparametric method allows for the financial data to "speak for themselves" and is robust to model misspecification, its performance is however often poor, such as very slow convergence rates in comparison with the sample size, owing to the "curse of dimensionality" widely identified in the literature (c.f., Fan and Yao, 2003), when the dimension of the conditioning variables is large (say, even only larger than three). This indicates that direct use of Brandt (1999)'s nonparametric method may be inappropriate when there are multiple conditioning variables. In this paper, our main objective is to address this issue associated with the nonparametric dynamic portfolio choice depending on multiple conditioning variables, where the number of the conditioning variables can be either fixed or diverging to infinity at certain polynomial rate in comparison with the sample size. We will propose a novel data-driven method to estimate the nonparametric optimal portfolio choice.

Specifically, in order to avoid curse of dimensionality associated with the problem and to make it practically implementable, we first consider the optimal portfolio choice which maximises the conditional utility function for a given individual conditioning variable, and then construct the dynamic optimal portfolio choice through the weighted average of the marginal optimal portfolio across all the conditioning variables. This method is partly motivated by the Model Averaging MArginal Regression (MAMAR) approach suggested in a recent paper by Li, Linton and Lu (2014), which shows that such a method performs well in estimating the conditional multivariate mean regression function and the out-of-sample prediction. Under some mild regularity conditions, we will establish the large sample properties to show the advantages in convergence for the developed portfolio choice procedure. Both simulation studies and empirical application will be carried out to demonstrate the performance of the proposed methodology with both finite sample and real data.

The structure of the paper is as follows. The proposed semiparametric dynamic portfolio choice with its methodology and estimation will be introduced in Section 2. The large sample theory for the estimators constructed in Section 2 will be presented in Section 3. The data-driven choice of the optimal weights for model averaging of the marginal optimal portfolios across all the conditioning variables is developed in Section 4. Numerical studies including both simulation and empirical application are reported in Section 5. To make the paper ease of reading, all assumptions and technical proofs are relegated to Appendices A and B, respectively.

2 Semiparametric dynamic portfolio choice: methodology and estimation

Suppose that there are n risky assets with an uncertain returns vector $R_t = (R_{1t}, \ldots, R_{nt})^\intercal$, where n is assumed to be fixed throughout this paper. Let $\mathcal{F}_t = (X_{1t}, \ldots, X_{Jt})^\intercal$, where J is the number of the conditioning or forecasting variables X_{jt}. In the present paper, we consider two cases: (i) $J \equiv J_0$ is a fixed positive integer, (ii) $J \equiv J_T$ is a positive integer which is diverging with T. The dynamic portfolio choice aims to choose the weights which maximise the conditional utility function defined by

$$\mathsf{E}\left[u(w^\intercal R_t)|\mathcal{F}_{t-1}\right] = \mathsf{E}\left[u(w^\intercal R_t)|(X_{1,t-1}, \ldots, X_{J,t-1})\right], \tag{2.1}$$

subject to $\mathbf{1}_n^\intercal w = \sum_{i=1}^n w_i = 1$, where $w = (w_1, \ldots, w_n)^\intercal$, $\mathbf{1}_n$ is an n-dimensional column vector with each element being one, $u(\cdot)$ is a concave utility function which measures the investor's utility of the wealth $w^\intercal R_t$ at time t. For simplicity, in this paper, we only focus on the case of single-period portfolio choice. Furthermore, we assume that the investors can borrow assets and sell them (shot selling), which indicates that some of the optimal weights may take negative values. The classic mean-variance paradigm considers the quadratic utility function $u(x) = -\frac{1}{2}(x-\beta)^2$; one may also work with the more general CRRA (Constant Relative Risk Aversion) utility function with risk aversion parameter γ:

$$u(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma}, & \gamma \neq 1, \\ \log x, & \gamma = 1. \end{cases}$$

More discussions on different classes of the utility function $u(\cdot)$ can be found in Chapter 1 of the book by Back (2010).

In order to solve the maximisation problem in (2.1), Brandt (1999) proposes a nonparametric conditional method of moments approach, which can be seen as an extension of the method of moments approach in Hansen and Singleton (1982). Taking the first-order derivative of $u(\cdot)$ in (2.1) with respect to w and considering the constraint of $\mathbf{1}_n^\intercal w = \sum_{i=1}^n w_i = 1$, we may obtain the dynamic portfolio choice by solving the following equation:

$$\mathsf{E}\left[(R_{it} - R_{nt})\dot{u}(w^\intercal R_t)|X_{1,t-1}, \ldots, X_{J,t-1}\right] = 0 \quad a.s. \tag{2.2}$$

for $i = 1, \ldots, n-1$, where $\dot{u}(\cdot)$ is the derivative of the utility function $u(\cdot)$. The last element in the optimal weights ($i = n$) can be determined by using the constraint of $\sum_{i=1}^n w_i = 1$. Brandt (1999) suggests a kernel-based smoothing method to estimate the solution to (2.2) which changes according to the conditioning variables. However, when J is large, the kernel-based nonparametric conditional method of moments approach would perform quite poorly due to the curse of dimensionality which has been discussed in Section 1. We next propose a novel dimension-reduction technique to address this problem.

We start with the consideration of the portfolio choice for each univariate conditioning variable in $\mathcal{F}_{t-1}$. Let $x = (x_1, \ldots, x_J)^\intercal$. For $j = 1, \ldots, J$, we define the conditional utility function as

$$\mathsf{E}\left[u(w^\intercal R_t)|X_{j,t-1} = x_j\right] \tag{2.3}$$

with the constraint $\mathbf{1}_n^\intercal w = \sum_{i=1}^n w_i = 1$. Then, we have the following first-order conditions at the optimum: for $i = 1, \ldots, n-1$,

$$\mathsf{E}\left[(R_{it} - R_{nt})\dot{u}(w_j^\intercal(x_j)R_t)|X_{j,t-1} = x_j\right] = 0 \quad a.s., \tag{2.4}$$

where:

$$w_j(x_j) = [w_{1j}(x_j), \ldots, w_{nj}(x_j)]^\intercal \quad \text{with} \quad w_{nj}(x_j) = 1 - \sum_{i=1}^{n-1} w_{ij}(x_j),$$

is the optimal portfolio choice that maximises the conditional utility function defined in (2.3). For any given j, this is essentially the problem posed and solved by Brandt (1999).

We next consider how to combine the portfolios selected above. Specifically, we shall consider a weighted average of the marginal portfolio choices $w_j(x_j)$ over $j = 1, \ldots, J$, and obtain the joint portfolio choice

$$w_a(x) = \sum_{j=1}^J a_j w_j(x_j), \quad \sum_{j=1}^J a_j = 1, \tag{2.5}$$

where the weights $a_j < 0$ can be allowed in our portfolio choice as we assume the existence of short selling. In Section 4 below, we will discuss how to choose the weights $a = (a_1, \ldots, a_J)^\intercal$ in the combination (2.5).

We now turn to the sample problem. Let $K(\cdot)$ be a kernel function and h be a bandwidth that converges to zero as T tends to infinity. Using the sample information, we may express the first-order conditions for the utility function as

$$\frac{1}{Th}\sum_{t=1}^{T}(R_{it}-R_{nt})\dot{u}(w^\intercal R_t)K\left(\frac{X_{j,t-1}-x_j}{h}\right)=0, \quad i=1,\ldots,n-1. \tag{2.6}$$

Denote $\widehat{w}_j(x_j) = [\widehat{w}_{1j}(x_j), \ldots, \widehat{w}_{nj}(x_j)]^\intercal$ as the solution to (2.6), where

$$\widehat{w}_{nj}(x_j) = 1 - \sum_{i=1}^{n-1}\widehat{w}_{ij}(x_j). \tag{2.7}$$

Then define the joint portfolio choice through the weighted average

$$\widehat{w}_a(x) = \sum_{j=1}^{J} a_j \widehat{w}_j(x_j), \quad \sum_{j=1}^{J} a_j = 1. \tag{2.8}$$

The asymptotic properties for $\widehat{w}_a(x)$ when the number of the conditioning variables is either fixed or divergent will be given in Section 3 below.

3 Large sample theory

We start with the case that $J = J_0$ is a fixed positive integer. Following (2.6) and (2.7), we next only study the asymptotic theory for $\widehat{w}_j^*(x_j) = [\widehat{w}_{1j}(x_j), \ldots, \widehat{w}_{n-1,j}(x_j)]^\intercal$, the estimate of $w_j^* = [w_{1j}(x_j), \ldots, w_{n-1,j}(x_j)]^\intercal$. Before stating the asymptotic theorems, we first introduce some notations. Let:

$$\mathbf{\Lambda}_j(x_j) = f_j(x_j)\mathsf{E}\left[R_t^*(R_t^*)^\intercal \ddot{u}\left(w_j^\intercal(x_j)R_t\right)|X_{j,t-1}=x_j\right]$$
$$Z_{jt}(x_j) = R_t^*\dot{u}\left(w_j^\intercal(X_{j,t-1})R_t\right)K\left(\frac{X_{j,t-1}-x_j}{h}\right)$$

for $j = 1, \ldots, J_0$ and $t = 1, \ldots, T$, where $R_t^* = (R_{1t} - R_{nt}, \ldots, R_{n-1,t} - R_{nt})^\intercal$, $\ddot{u}(\cdot)$ is the second-order derivative of $u(\cdot)$ and $f_j(\cdot)$ is the marginal density function of X_{jt}. Define

$$W_{jt}(x_j) = \mathbf{\Lambda}_j^{-1}(x_j)Z_{jt}(x_j) \quad \text{and} \quad W_t(x|a) = \sum_{j=1}^{J_0} a_j W_{jt}(x_j)$$

for $t = 1, \ldots, T$. Following the argument in the proof of Theorem 3.1 in Appendix B and letting

$$\widehat{w}_a^*(x) = \sum_{j=1}^{J} a_j \widehat{w}_j^*(x_j), \quad w_a^*(x) = \sum_{j=1}^{J} a_j w_j^*(x_j),$$

we may show that

$$\sqrt{Th}\left[\widehat{w}_a^*(x) - w_a^*(x)\right] = \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} W_t(x|a) + o_P(1) \tag{3.1}$$

for given $a = (a_1, \ldots, a_{J_0})^\intercal$. The asymptotic distribution theory for $\overline{w}_j(x_j)$ and $\overline{w}_a(x)$ is given in Theorem 3.1 below.

THEOREM 3.1. *Suppose that Assumptions 1–5 in Appendix A are satisfied and the number of the conditioning variables J is a fixed positive integer J_0.*

(i) For $j = 1, \ldots, J_0$, we have

$$\sqrt{Th}\left[\widehat{w}_j^*(x_j) - w_j^*(x_j)\right] \xrightarrow{d} \mathsf{N}\left(\mathbf{0},\ \mathbf{\Omega}_j(x_j)\right), \tag{3.2}$$

where $\mathbf{\Omega}_j(x_j) = \mathsf{E}\left[W_{jt}(x_j)W_{jt}^\intercal(x_j)\right] = \mathbf{\Lambda}_j^{-1}(x_j)\mathsf{E}\left[Z_{jt}(x_j)Z_{jt}^\intercal(x_j)\right]\mathbf{\Lambda}_j^{-1}(x_j)$.

(ii) For the estimated portfolio choice defined in (2.8) with a set of given weights, we have

$$\sqrt{Th}\left[\widehat{w}_a^*(x) - w_a^*(x)\right] \xrightarrow{d} \mathsf{N}\left(\mathbf{0},\ \mathbf{\Omega}(x|a)\right), \tag{3.3}$$

where $\mathbf{\Omega}(x|a) = \mathsf{E}\left[W_t(x|a)W_t^\intercal(x|a)\right]$.

Although there are multiple conditioning variables in the nonparametric dynamic portfolio choice, we can still achieve the root-Th convergence rates as shown in the above theorem, which means that we can successfully overcome the curse of dimensionality issue. The main reason is that, in the estimation methodology, we only apply the univariate kernel smoothing to estimate the optimal portfolio choice for each univariate conditioning variable and then obtain the final portfolio choice through weighted averages defined as in (2.8). In contrast, if we directly use the multivariate kernel smoothing as done in Brandt (1999), the convergence rate for the resulting estimation would be root-Th^{J_0}, slower than the rates in (3.2) and (3.3) when $J_0 > 1$.

In practice, it is not uncommon that the number of the potential conditioning variables is large, and so a more reasonable assumptions is that J is divergent, i.e., $J \equiv J_T \to \infty$. The following theorem gives the asymptotic distribution theory for this general case.

THEOREM 3.2. *Suppose that Assumptions 1–4 and 5′ in Appendix A are satisfied and the number of conditioning variables J is a positive integer J_T which is diverging with the sample size T. Then, (3.2) and (3.3) in Theorem 3.1 still hold.*

Theorem 3.2 above indicates that the root-Th convergence rates remain the same even when the number of the potential conditioning variables is diverging. The restriction on J_T is

$$\frac{T^{1-1/(2+\delta)}h}{J_T^{1/(2+\delta)}\log T} \to \infty,$$

which is given in Assumption 5′. Such a restriction means that J_T can possibly be larger than T, if we are only interested in $\overline{w}_a(\cdot)$ or $\widehat{w}_a(\cdot)$ for given a. However, some additional restrictions on J_T would be needed when we consider the choice of the optimal $a = (a_1, \ldots, a_{J_T})^\intercal$, see Section 4 below for details.

4 Data-driven optimal weight choice in model averaging

The performance of the dynamic portfolio choice defined in (2.8) relies on the choice of the weights $a_1, \ldots, a_J$. Let $w_{at} \equiv w_a(\mathcal{F}_t) = \sum_{j=1}^J a_j w_j(X_{jt})$ and define the objective function:

$$U(a) = \mathsf{E}\left[u(w_{a,t-1}^\intercal R_t)\right] = \mathsf{E}\left\{u\big[\sum_{j=1}^J a_j w_j^\intercal(X_{j,t-1})R_t\big]\right\}. \quad (4.1)$$

We may choose the optimal weights by maximising $U(a)$, i.e.,

$$a_0 = \arg\max_a U(a) = \arg\max_a \mathsf{E}\left\{u\big[\sum_{j=1}^J a_j w_j^\intercal(X_{j,t-1})R_t\big]\right\}. \quad (4.2)$$

subject to the constraint of $\sum_{j=1}^J a_j = 1$. This leads to the following first-order conditions:

$$\mathsf{E}\left[(R_{jt}^w - R_{Jt}^w)\dot{u}\big(\sum_{j=1}^J a_{j0}R_{jt}^*\big)\right] = 0 \ \text{ for } \ j = 1, \ldots, J-1, \quad (4.3)$$

and $a_{J0} = 1 - \sum_{j=1}^{J-1} a_{j0}$, where $R_{jt}^w = w_j^\intercal(X_{j,t-1})R_t$, $R_{Jt}^w = w_J^\intercal(X_{J,t-1})R_t$ and a_{j0} is the j-th element of a_0.

We propose a data-driven optimal weight choice in model averaging by using the sample information and replacing the unobservable $w_j(X_{j,t-1})$ by the estimated value $\widehat{w}_j(X_{j,t-1})$ which is constructed in (2.6), we may estimate $a_0 = (a_{10}, \ldots, a_{J0})^\intercal$ by $\widehat{a} = (\widehat{a}_1, \ldots, \widehat{a}_J)^\intercal$ which is the solution to

$$\frac{1}{T}\sum_{t=1}^T(\widehat{R}_{jt}^w - \widehat{R}_{Jt}^w)\dot{u}\big[\sum_{j=1}^J \widehat{a}_j\widehat{R}_{jt}\big] = 0 \ \text{ for } \ j = 1, \ldots, J-1, \quad (4.4)$$

and $\widehat{a}_J = 1 - \sum_{j=1}^{J-1} \widehat{a}_j$, where $\widehat{R}_{jt}^w = \widehat{w}_j^\intercal(X_{j,t-1})R_t$.

To facilitate the proof of the asymptotic theory for $\widehat{a}$, we need to establish the uniform consistency results for $\widehat{w}_j(x_j)$ over $x_j \in \mathcal{X}_j$ with $\mathcal{X}_j$ being the support of X_{jt} and $j = 1, \ldots, J$.

PROPOSITION 4.1. *Suppose that Assumptions 1–4 in Appendix A are satisfied.*

(i) If the number of the conditioning variables J is a fixed positive integer J_0 and

$$h \to 0, \quad \frac{T^{1-2/(2+\delta)}h}{\log T} \to \infty, \tag{4.5}$$

where $\delta > 0$ is specified in Assumption 3 in Appendix A, then

$$\max_{1\le j\le J_0} \sup_{x_j\in\mathcal{X}_j} \|\widehat{w}_j(x_j) - w_j(x_j)\| = O_P\left(h^2 + \sqrt{\log T/(Th)}\right), \tag{4.6}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix.

(ii) If the number of the conditioning variables J is a diverging positive integer J_T and

$$h \to 0, \quad \frac{T^{1-2/(2+\delta)}h}{J_T^{2/(2+\delta)}\log T} \to \infty, \tag{4.7}$$

then (4.6) still holds with J_0 replaced by J_T.

In fact, Proposition 4.1(i) can be included as a special case of Proposition 4.1(ii), and the above uniform consistency results can be seen as the extension of the uniform consistency results for the nonparametric kernel-based estimation in stationary time series (Hansen, 2008; Kristensen, 2009; Li, Lu and Linton, 2012) to the scenario of the nonparametric portfolio choice. By modifying the proof in Appendix B, we may further generalise (4.6) to the case when $\mathcal{X}_j$ is an expanding set.

We next study the asymptotic property for $\widehat{a}$. As $\widehat{a}_J = 1 - \sum_{j=1}^{J-1} \widehat{a}_j$, it suffices to consider $\widehat{a}^* \equiv (\widehat{a}_1, \ldots, \widehat{a}_{J-1})^\intercal$, the estimate of $a_0^* \equiv (a_{10}, \ldots, a_{J-1,0})^\intercal$. Define $\eta_t = \dot{u}\left[\sum_{j=1}^J a_{j0} w_j^\intercal(X_{j,t-1})R_t\right]$, $\eta_t^* = \ddot{u}\left[\sum_{j=1}^J a_{j0} w_j^\intercal(X_{j,t-1})R_t\right]$, and $R_t^*(w) = \left(R_{1t}^w, \ldots, R_{J-1,t}^w\right)^\intercal$, $V_t^* = R_t^*(w) - R_{Jt}^w \mathbf{1}_{J-1}$ and let

$$\mathbf{\Delta}_1 = \mathsf{E}\left[\eta_t^* V_t^* (V_t^*)^\intercal\right].$$

For $j = 1, \ldots, J$, define $\varepsilon_{jt} = \dot{u}\left(w_j^\intercal(X_{j,t-1})R_t\right) = \dot{u}\left(R_{jt}^w\right)$, and let

$$\varepsilon_t = (\varepsilon_{1t}a_{10}, \ldots, \varepsilon_{Jt}a_{J0})^\intercal, \quad Q_t = (Q_{1t}, \ldots, Q_{Jt})^\intercal$$

with $Q_{jt} = \left\{\mathsf{E}\left[\eta_s^* V_s^* R_s^\intercal \mathbf{W}\mathbf{\Lambda}_j^{-1}(X_{j,s-1}) | X_{j,s-1} = X_{j,t-1}\right]\right\} R_t^*$ and

$$\mathbf{W} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1 \\ -1 & \cdots & -1 \end{pmatrix}.$$

Define

$$\boldsymbol{\Delta}_2 = \mathsf{Cov}(V_t^* \eta_t + Q_t^\intercal \varepsilon_t, V_t^* \eta_t + Q_t^\intercal \varepsilon_t).$$

In the following theorem, we give the asymptotic distribution theory for $\widehat{a}$ when J is fixed.

THEOREM 4.2. *Suppose that the assumptions in Proposition 4.1(i) are satisfied and the matrix $\boldsymbol{\Delta}_1$ is non-singular. Then we have*

$$\sqrt{T}\left(\widehat{a}^* - a_0^*\right) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \boldsymbol{\Delta}_1^{-1}\boldsymbol{\Delta}_2\boldsymbol{\Delta}_1^{-1}\right). \tag{4.8}$$

We next deal with the case that $J = J_T$ is diverging with the sample size T. Let $\boldsymbol{\Delta}_T = \boldsymbol{\Delta}_1^{-1}\boldsymbol{\Delta}_2\boldsymbol{\Delta}_1^{-1}$ which indicates that the size of the matrix relies on T. As the number of the potential conditioning variables J_T tends to infinity, we cannot state the asymptotic normal distribution theory by the same way as in Theorem 4.2 above. As in Fan and Peng (2004), we let $\boldsymbol{\Psi}_T$ be a $J_* \times (J_T - 1)$ matrix with full row rank such that as $T \to \infty$, $\boldsymbol{\Psi}_T\boldsymbol{\Psi}_T^\intercal \to \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is non-negative definite $J_* \times J_*$ matrix with J_* being a fixed positive integer. The role of the matrix $\boldsymbol{\Psi}_T$ is to reduce the dimension from $(J_T - 1)$ to J_* in the derivation of the asymptotic normality, so it is only involved in the asymptotic analysis. If we are only interested in the asymptotic behavior for the first J_* components of $\widehat{a}$, we may choose $\boldsymbol{\Psi}_T = \left[I_{J_*},\ O_{J_* \times (J_T - J_*)}\right]$, where I_p is a $p \times p$ identity matrix and $O_{k\times j}$ is a $k \times j$ null matrix. We next state the asymptotic distribution theory for $\widehat{a}$ when J is diverging.

THEOREM 4.3. *Suppose that the assumptions in Proposition 4.1(ii) are satisfied, the matrix $\boldsymbol{\Delta}_1$ is non-singular and*

$$J_T\left(h^2 + \sqrt{\frac{\log T}{Th}}\right) \to 0. \tag{4.9}$$

Then we have

$$\sqrt{T}\boldsymbol{\Psi}_T\boldsymbol{\Delta}_T^{-1/2}\left(\widehat{a}^* - a_0^*\right) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \boldsymbol{\Psi}\right). \tag{4.10}$$

The above theorem is similar to some results in the existing literature such as Theorem 2(ii) in Fan and Peng (2004) and Theorem 4.3 in Li, Linton and Lu (2014). The condition (4.9) shows that the dimension J_T should not diverge too fast to infinity, and (4.10) indicates that the convergence rate is $\sqrt{T/J_T}$ due to the diverging number of the conditioning variables.

5 Numerical studies

In this section, we set the number of assets under consideration for investment to be $n = 5$. This value of n is chosen primarily for convenience of computation. Computation procedures for larger values of n are exactly the same.

EXAMPLE 5.1. The time series of gross returns R_t on the assets are generated via the conditioning variables by the following regression:

$$\log(R_t) = 0.06 + A * \log(X_t) + e_t, \tag{5.1}$$

where A is an $n \times J$ full-rank matrix so generated such that the elements of $1000 * A$ are random integers ranging between 1 and 30; e_t are i.i.d. random vectors distributed as $e_t \sim \mathsf{N}(\mathbf{0}, 0.001 * I_n)$ in which I_n is the $n \times n$ identity matrix; $\{\log(X_t)\}$ is a J-dimensional AR(1) process generated as

$$\log(X_t) = -0.01 + 0.9 * \log(X_{t-1}) + u_t \tag{5.2}$$

in which u_t are i.i.d. random vectors generated from $\mathsf{N}(\mathbf{0}, 0.002 * I_J)$. The variables in X_t, will be used as the conditioning variables, and the number of conditioning variables is set to satisfy $J = [0.5 * \sqrt{T}]$, where $[\cdot]$ denotes the operator that rounds a number to the nearest integer less than or equal to that number.

We use a CRRA utility function with $\gamma = 1$, 5, and 10. For each $j = 1, \ldots, J$, $t = 1, \ldots, T$, and the observed value, $X_{j,t-1}$, of the conditioning variable in the previous time period $t-1$, we calculate the j-th set of conditional optimal portfolio weights, $\widehat{w}_j(X_{j,t-1})$, by solving the nonparametric version of the conditioning equations, i.e., (2.6). Then by solving the equations in (4.4) with respect to a_j, we can obtain the joint optimal portfolio weights, $\widehat{w}_a(X_{t-1}) = \sum_{j=1}^{J} \widehat{a}_j \widehat{w}_j(X_{j,t-1})$, conditional on the values of all the conditioning variables in time period $t-1$, where $X_{t-1} = (X_{1,t-1}, \ldots, X_{J,t-1})^\intercal$. Note that in calculating the $\widehat{w}_j(X_{j,t-1})$ and $\widehat{a}_j$, we have imposed $\sum_{j=1}^{J} \widehat{w}_j(X_{j,t-1}) = 1$ and $\sum_{j=1}^{J} \widehat{a}_j = 1$ so that the budget constraint is satisfied.

We compare the single-period returns of portfolios constructed with weights calculated from the proposed semiparametric model averaging method (SMAM) and the unconditional parametric method (UPM) which solves for the weights that maximise the unconditional utility, i.e., $\frac{1}{T}\sum_{t=1}^{T} u(w^\intercal R_t)$, subject to $w^\intercal i = 1$. Table 5.1 reports the averages of the mean difference (MD) between returns on the SMAM and UPM constructed portfolios:

$$\mathrm{MD} = \frac{1}{T}\sum_{t=1}^{T} \left(R_t^s - R_t^u\right),$$

where $R_t^s = \widehat{w}_a^\intercal(X_t)R_t$ and $R_t^u = \widehat{w}_u^\intercal R_t$ with $\widehat{w}_u$ chosen by the UPM. Also reported in Table 5.1 are the averages of positive frequency (PF) of the NAM, i.e., the frequency at which the return on the SMAM constructed portfolio exceeds that of the UPM constructed portfolio. These results are based on 100 independent samples of $T = 100$, 300, or 500 observations. The numbers in parentheses are the respective standard errors.

It can be seen from Table 5.1 that in most time periods, the return on the portfolio chosen by the SMAM is larger than the return on the portfolio chosen by the UPM. This is especially so when the sample size is relatively small. For example, when $\gamma = 5$, the average gain in choosing portfolios by the SMAM than by the UPM is an additional 5.4% return when $T = 100$, and this reduces to 0.4% when $T = 500$. As γ measures the level of risk aversion of an investor with a higher value representing less willingness for risk taking, the portfolio returns generally decrease as γ increases. Hence, we see a decreasing trend in the MD values as γ increases in Table 5.1.

TABLE 5.1. Averages of MD between SMAM and UCM returns and PF of the SMAM for Example 5.1

[dir=NW]γT		$T = 100$	$T = 300$	$T = 500$
2*$\gamma = 1$	MD	0.0540(0.2695)	0.0993(0.2905)	0.1925(0.3019)
	PF	0.5358(0.0674)	0.5178(0.0374)	0.5158(0.0298)
2*$\gamma = 5$	MD	0.0543(0.0228)	0.0138(0.0108)	0.0038(0.0023)
	PF	0.7076(0.0404)	0.6000(0.0296)	0.5533(0.0259)
2*$\gamma = 10$	MD	0.0203(0.0110)	0.0081(0.0037)	0.0058(0.0027)
	PF	0.6797(0.0454)	0.6131(0.0309)	0.602(0.0299)

EXAMPLE 5.2. In this example, the gross returns R_t are generated from a stationary VAR:

$$\log(R_t) = 0.01 + B * \log(R_{t-1}) + e_t, \tag{5.3}$$

where e_t are generated in the same way as in Example 5.1, the AR coefficient matrix B is defined as the transpose of $0.01 * \text{magic}(n)$ in which magic(n) denotes the magic matrix of dimension $n = 5$.

The conditioning variables are taken as the lag-one and lag-two returns, i.e. $X_t = (R_{t-1}^\intercal, R_{t-2}^\intercal)^\intercal$. Hence, the number of conditioning variables is $J = 2n$. The results based on 100 independent samples of this example are given in Table 5.2. Similar findings can be obtained as those in Example 5.1.

TABLE 5.2. Averages of MD between SMAM and UPM returns and PF of the SMAM for Example 5.2

[dir=NW]γT		$T = 100$	$T = 300$	$T = 500$
2*$\gamma = 1$	MD	0.2838(0.8739)	0.1022(0.1125)	0.2158(0.0957)
	PF	0.5487(0.0590)	0.5373(0.0489)	0.5755(0.0403)
2*$\gamma = 5$	MD	0.0619(0.0270)	0.0178(0.0081)	0.0163(0.0061)
	PF	0.7229(0.0430)	0.6220(0.0305)	0.6138(0.0215)
2*$\gamma = 10$	MD	0.0226(0.0102)	0.0111(0.0052)	0.0108(0.0048)
	PF	0.6596(0.0475)	0.6295(0.0306)	0.6219(0.0253)

6 Conclusions and Extensions

We have solved a portfolio problem for each conditioning variable X_j and then combined the portfolio weights from each of those "experts". This is quite a common approach in the machine learning literature, see for example Györfi, Ottucsák, and Urbán (2011). We could instead seek to approximate the objective function $Q(w; x) = \mathsf{E}\left[u(w^\intercal R_t)|X_{t-1} = x\right]$ by using the MAMAR method to approximate this conditional expectation by a sum of one dimensional nonparametric regressions, i.e.,

$$\widetilde{Q}(w; x) = \sum_{j=1}^{J} \alpha_j \mathsf{E}\left[u(w^\intercal R_t)|X_{jt-1} = x_j\right]$$

for weights α_j, and then optimizing $\widetilde{Q}(w; x)$ with respect to w. This method is likely to give similar results except that it provides less diagnostic information, and it is perhaps harder to define a method for selecting α_j.

It is also possible to introduce constraints such as absence of short selling or position limits at each stage of our method at the cost of computational complexity.

A Assumptions

We next list the regularity conditions which are used to prove the asymptotic results. Some of these conditions might not be the weakest possible.

Assumption 1. (i) The utility function $u(\cdot)$ is concave and has continuous derivatives up to the second order.

(ii) The optimal weight functions $w_j(\cdot)$, $j = 1, \ldots, J$, have continuous derivatives up to the second order.

Assumption 2. The process of the conditioning variables $\left\{X_t = (X_{1t}, \ldots, X_{Jt})^\intercal\right\}$ is strictly stationary and α-mixing with the mixing coefficient decaying at a geometric rate, $\alpha_X(k) \sim \gamma_0^k$, $0 < \gamma_0 < 1$. Each component variable X_{jt} has a continuous marginal density function $f_j(\cdot)$ on a compact support denoted by $\mathcal{X}_j$. For all $t > 1$, the joint density function of (X_1, X_t) exists and is uniformly bounded.

Assumption 3. The process of the asset returns $\left\{R_t = (R_{1t}, \ldots, R_{nt})^\intercal\right\}$ is strictly stationary and α-mixing with the mixing coefficient decaying at a geometric rate, $\alpha_R(k) \sim \gamma_0^k$, $0 < \gamma_0 < 1$. Furthermore, there exists a $\delta > 0$ such that

$$\max_{1\leq j\leq J} \mathsf{E}\left[\left\|R_t^*(R_t^*)^\intercal \ddot{u}\left(w_j^\intercal(X_{j,t-1})R_t\right)\right\|^{2+\delta} + \left\|R_t^* \dot{u}\left(w_j^\intercal(X_{j,t-1})R_t\right)\right\|^{2+\delta}\right] < \infty,$$

where R_t^* is defined in Section 3. Let

$$\mathsf{E}\left[R_t^*(R_t^*)^\intercal \ddot{u}\left(w_j^\intercal(x_j)R_t\right) | X_{j,t-1} = x_j\right]$$

be non-singular uniformly for $x_j \in \mathcal{X}_j$, $j = 1, \ldots, J$.

Assumption 4. The kernel function $K(\cdot)$ is positive, continuous and symmetric with a compact support and $\int K(z)dz = 1$.

Assumption 5. The bandwidth h satisfies $h \to 0$,

$$Th^4 = o(1) \quad \text{and} \quad \frac{T^{1-1/(2+\delta)}h}{\log T} \to \infty.$$

Assumption 5′. The bandwidth h satisfies $h \to 0$,

$$Th^4 = o(1) \quad \text{and} \quad \frac{T^{1-1/(2+\delta)}h}{J_T^{1/(2+\delta)}\log T} \to \infty.$$

The above assumptions are mild and justifiable. Some of the assumptions are similar to those in Brandt (1999). In this paper, we impose the stationarity and mixing dependence condition on the processes of the returns of the risky assets and the conditioning variables, see, for example, Assumptions 2 and 3. The methodology and theory developed in the present paper are also applicable to the more general dependence structure, say the near epoch dependent process (Li, Lu and Linton, 2012). To facilitate our proofs, we assume that the mixing coefficients decay at a geometric rate, which can be relaxed to a polynomial rate at the cost of more lengthy proofs. The bandwidth conditions in Assumptions 5 and 5′ indicate that there is a trade-off between the moment conditions and the bandwidth restriction. And the condition $Th^4 = o(1)$ shows that certain under-smoothing is needed in the asymptotic analysis, which is not uncommon in semiparametric estimation.

B Proofs of the theoretical results

We next give the proofs of the theoretical results stated in Sections 3 and 4. In this appendix, we let C be a positive constant whose value may change from line to line.

Proof of Theorem 3.1. Throughout this proof, $J = J_0$ is a fixed positive integer. By the definition of $\widehat{w}_j^*(x_j) = [\widehat{w}_{1j}(x_j), \ldots, \widehat{w}_{n-1,j}(x_j)]^\intercal$ or $\widehat{w}_j(x_j) = [\widehat{w}_{1j}(x_j), \ldots, \widehat{w}_{nj}(x_j)]^\intercal$, we have

$$\frac{1}{Th}\sum_{t=1}^{T}(R_{it} - R_{nt})\dot{u}\left(\widehat{w}_j^\intercal(x_j)R_t\right)K\left(\frac{X_{j,t-1} - x_j}{h}\right) = 0 \tag{B.1}$$

for $i = 1, \ldots, n-1$ and $j = 1, \ldots, J_0$. By Assumption 1 and using the Taylor's expansion for $\dot{u}(\cdot)$,

$$\dot{u}\left(\widehat{w}_j^\intercal(x_j)R_t\right) = \dot{u}\left(w_j^\intercal(x_j)R_t\right) + \ddot{u}\left(w_\diamond^\intercal(x_j)R_t\right)\left\{(R_t^*)^\intercal\left[\widehat{w}_j^*(x_j) - w_j^*(x_j)\right]\right\},$$

where $w_\diamond(x_j)$ lies between $\widehat{w}_j(x_j)$ and $w_j(x_j)$, and $w_j^*(x_j) = [w_{1j}(x_j), \ldots, w_{n-1,j}(x_j)]^\intercal$. Then we may prove that

$$\widehat{w}_j^*(x_j) - w_j^*(x_j) = \mathcal{A}_{nj}^{-1}(x_j)\mathcal{B}_{nj}(x_j) \tag{B.2}$$

with

$$\begin{aligned}
\mathcal{A}_{nj}(x_j) &= \frac{1}{Th}\sum_{t=1}^{T}R_t^*(R_t^*)^\intercal\ddot{u}\left(w_\diamond^\intercal(x_j)R_t\right)K\left(\frac{X_{j,t-1} - x_j}{h}\right), \\
\mathcal{B}_{nj}(x_j) &= \frac{1}{Th}\sum_{t=1}^{T}R_t^*\dot{u}\left(w_j^\intercal(x_j)R_t\right)K\left(\frac{X_{j,t-1} - x_j}{h}\right).
\end{aligned}$$

By Assumptions 2–5 in Appendix A and following the standard argument in nonparametric kernel-based smoothing in time series (c.f., Robinson, 1983), we can show that

$$\mathcal{A}_{nj}(x_j) = \mathbf{\Lambda}_j(x_j) + o_P(1) \tag{B.3}$$

when $\widehat{w}_j(x_j)$ is sufficiently close to $w_j(x_j)$, where $\mathbf{\Lambda}_j(x_j)$ is defined in Section 3. The convergence in (B.3) holds uniformly for $x_j \in \mathcal{X}_j$ and $j = 1, \ldots, J_0$ (c.f., the proof of Proposition 4.1 below). On the other hand, we recall that $Z_{jt}(x_j) = R_t^*\dot{u}\left(w_j^\intercal(X_{j,t-1})R_t\right)K\left(\frac{X_{j,t-1}-x_j}{h}\right)$. By Assumptions 1(i)(ii) and the Taylor's expansion for $\dot{u}\left(w_j^\intercal(\cdot)R_t\right)$, we may show that

$$\mathcal{B}_{nj}(x_j) = \frac{1}{Th}\sum_{t=1}^{T}Z_{jt}(x_j) + O_P(h^2). \tag{B.4}$$

Noting that $nh^4 = o(1)$ in Assumption 5 and by (B.2)–(B.4),

$$\sqrt{Th}\left[\widehat{w}_j^*(x_j) - w_j^*(x_j)\right] = \mathbf{\Lambda}_j^{-1}(x_j)\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} Z_{jt}(x_j) + o_P(1). \tag{B.5}$$

Then, using the central limit theorem for the stationary α-mixing sequence (e.g., Section 2.6.4 in Fan and Yao, 2003), we can complete the proof of (3.2) in Theorem 3.1(i).

As in Section 3, let

$$W_{jt}(x_j) = \mathbf{\Lambda}_j^{-1}(x_j)Z_{jt}(x_j), \quad W_t(x|a) = \sum_{j=1}^{J_0} a_j W_{jt}(x_j).$$

By (B.5) and the definitions of $\widehat{w}_a^*(x)$ and $w_a^*(x)$, we have

$$\sqrt{Th}\left[\widehat{w}_a^*(x) - w_a^*(x)\right] = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} W_t(x|a) + o_P(1). \tag{B.6}$$

Using (B.6), we can readily prove (3.3) in Theorem 3.1(ii). ■

Proof of Theorem 3.2. The proof of this theorem is similar to the proof of Theorem 3.1 above. The only difference that the stronger bandwidth condition in Assumption 5′ is needed when we prove (B.3) uniformly for $x_j \in \mathcal{X}_j$ and $j = 1, \ldots, J_T$. ■

Proof of Proposition 4.1. We only consider the proof of (4.6) for the case when $J = J_T$ is diverging, as the case of $J = J_0$ is similar and simpler. Noting that $\widehat{w}_{nj}(x_j) = 1 - \sum_{i=1}^{n-1}\widehat{w}_{ij}(x_j)$ and using (B.2) and (B.3) in the proof of Theorem 3.1, we only need to show that

$$\max_{1\leq j\leq J_T}\sup_{x_j\in\mathcal{X}_j}\left\|\frac{1}{Th}\sum_{t=1}^{T} R_t^*\dot{u}\left(w_j^{\intercal}(x_j)R_t\right)K\left(\frac{X_{j,t-1}-x_j}{h}\right)\right\| = O_P\left(h^2 + \sqrt{\log T/(Th)}\right), \tag{B.7}$$

as $\mathbf{\Lambda}_j(x_j)$ is nonsingular uniformly for $x_j \in \mathcal{X}_j$, $1 \leq j \leq J_T$ (see Assumption 3). Note that the convergence result in (B.4) can be strengthened from the point-wise convergence to the uniform convergence over $x_j \in \mathcal{X}_j$, $1 \leq j \leq J_T$. Hence, in order to prove (B.7), we only need to show that

$$\max_{1\leq j\leq J_T}\sup_{x_j\in\mathcal{X}_j}\left\|\frac{1}{Th}\sum_{t=1}^{T} Z_{jt}(x_j)\right\| = O_P\left(\sqrt{\log T/(Th)}\right), \tag{B.8}$$

where $Z_{jt}(x_j)$ is defined in the proof of Theorem 3.1.

For simplicity, denote $\xi_T = \left(\frac{\log T}{Th}\right)^{1/2}$. The main idea of proving (B.8) is to consider covering the compact support $\mathcal{X}_j$ by a finite number of disjoint subsets $\mathcal{X}_j(k)$ which are centered at x_{jk} with radius $r_T = \xi_T h^2$, $k = 1, \ldots, \mathcal{N}_j$. It is easy to show that $\max_{1\leq j\leq J_T} \mathcal{N}_j = O(r_T^{-1}) = O(\xi_T^{-1}h^{-2})$ and

$$\begin{aligned}\max_{1\leq j\leq J_T} \sup_{x_j\in\mathcal{X}_j} \left\| \frac{1}{Th}\sum_{t=1}^{T} Z_{jt}(x_j)\right\| &\leq \max_{1\leq j\leq J_T}\max_{1\leq k\leq \mathcal{N}_j} \left\| \frac{1}{Th}\sum_{t=1}^{T} Z_{jt}\left(x_{jk}\right)\right\| + \\ &\quad \max_{1\leq j\leq J_T}\max_{1\leq k\leq \mathcal{N}_j}\sup_{x_j\in\mathcal{X}_j(k)} \left\| \frac{1}{Th}\sum_{t=1}^{T} Z_{jt}(x_j) - \frac{1}{Th}\sum_{t=1}^{T} Z_{jt}\left(x_{jk}\right)\right\| \\ &\equiv \Pi_{T1} + \Pi_{T2}. \end{aligned} \tag{B.9}$$

By the continuity condition on $K(\cdot)$ in Assumption 4 and using the definition of r_T, we readily have

$$\Pi_{T2} = O_P\left(\frac{r_T}{h^2}\right) = O_P(\xi_T). \tag{B.10}$$

For Π_{T1}, we apply the truncation technique and the Bernstein-type inequality for the α-mixing dependent random variables which can be found in Bosq (1998) and Fan and Yao (2003) to obtain the convergence rate. Let $M_T = M_1(TJ_T)^{1/(2+\delta)}$,

$$\overline{Z}_{jt}\left(x_{jk}\right) = Z_{jt}\left(x_{jk}\right)\cdot \mathrm{I}\left\{\left\|R_t^*\dot{u}\left(R_{jt}^w\right)\right\| \leq M_T\right\}$$

and

$$\widetilde{Z}_{jt}\left(x_{jk}\right) = Z_{jt}\left(x_{jk}\right)\cdot \mathrm{I}\left\{\left\|R_t^*\dot{u}\left(R_{jt}^w\right)\right\| > M_T\right\},$$

where $\mathrm{I}\{\cdot\}$ is an indicator function and $R_{jt}^w = w_j^\intercal(X_{j,t-1})R_t$ as in Section 4. Then we have

$$\begin{aligned}\Pi_{T1} &\leq \max_{1\leq j\leq J_T}\max_{1\leq k\leq \mathcal{N}_j} \left\| \frac{1}{Th}\sum_{t=1}^{T} \left\{\overline{Z}_{jt}(x_{jk}) - \mathsf{E}\left[\overline{Z}_{ij}(x_{jk})\right]\right\}\right\| + \\ &\quad \max_{1\leq j\leq J_T}\max_{1\leq k\leq \mathcal{N}_j} \left\| \frac{1}{Th}\sum_{i=1}^{n} \left\{\widetilde{Z}_{jt}(x_{jk}) - \mathsf{E}\left[\widetilde{Z}_{jt}(x_{jk})\right]\right\}\right\| \\ &\equiv \Pi_{T3} + \Pi_{T4}. \end{aligned} \tag{B.11}$$

Note that for $M_2 > 0$ and any $\epsilon > 0$, by Assumption 3 and the Markov inequality,

$$\begin{aligned}\mathsf{P}\Big(\Pi_{T4} > M_2\xi_T\Big) &\leq \mathsf{P}\left(\max_{1\leq j\leq J_T}\max_{1\leq k\leq \mathcal{N}_j}\max_{1\leq t\leq T}\left\|\widetilde{Z}_{jt}(x_{jk})\right\| > 0\right) \\ &\leq \sum_{j=1}^{J_T}\sum_{t=1}^{T}\mathsf{P}\left(\left\|R_t^*\dot{u}\left(R_{jt}^w\right)\right\| > M_T\right) \\ &\leq M_1^{-(2+\delta)}\cdot\max_{1\leq j\leq J_T}\mathsf{E}\left[\left\|R_t^*\dot{u}\left(R_{jt}^w\right)\right\|^{2+\delta}\right] < \epsilon, \end{aligned}$$

if we choose

$$M_1 > \left\{ \max_{1\le j\le J_T} \mathsf{E}\left[\left\| R_t^* \dot{u}(R_{jt}^w) \right\|^{2+\delta} \right] \right\}^{1/(2+\delta)} \epsilon^{-1/(2+\delta)}.$$

Then, by letting ϵ be arbitrarily small, we can show that

$$\Pi_{T4} = O_P(\xi_T). \tag{B.12}$$

On the other hand, note that

$$\left\| \overline{Z}_{jt}(x_{jk}) - \mathsf{E}\left[\overline{Z}_{jt}(x_{jk}) \right] \right\| \le C_0 M_T, \tag{B.13}$$

and

$$\mathsf{Var}\left[\overline{Z}_{jt}(x_{jk}) \right] \le C_0 h \tag{B.14}$$

where C_0 is a positive constant. By (B.13), (B.14) and Theorem 1.3(2) in Bosq (1998) with $p = [(M_2 M_T \xi_T/4)^{-1}]$ which tends to infinity by (4.7), we have

$$\begin{aligned}
\mathsf{P}(\Pi_{T3} > M_2\xi_T) &= \mathsf{P}\left(\max_{1\le j\le J_T} \max_{1\le k\le \mathcal{N}_j} \left\| \frac{1}{Th} \sum_{t=1}^{T} \left\{ \overline{Z}_{jt}(x_{jk}) - \mathsf{E}\left[\overline{Z}_{ij}(x_{jk}) \right] \right\} \right\| > M_2 \xi_T \right) \\
&= \sum_{j=1}^{J_T} \mathcal{N}_j \left(4\exp\left\{ \frac{-q M_2^2 \xi_T^2}{4C_0 M_2 M_T \xi_T/h + 16 C_0/(ph)} \right\} + 22\left(1 + 4C_0 M_T/(M_2 h \xi_T)\right) q \gamma_0^p \right) \\
&\le C \sum_{j=1}^{J_T} \mathcal{N}_j \left[\exp\left\{ -M_2 \log T \right\} + T M_T^2 \gamma_0^p \right] = o(1),
\end{aligned}$$

where M_2 is chosen sufficiently large and $q = T/(2p)$. Hence we have

$$\Pi_{T3} = O_P(\xi_T). \tag{B.15}$$

In view of (B.10)–(B.12) and (B.15), we have shown (B.8), completing the proof of Proposition 4.1. ■

PROOF OF THEOREM 4.2. Recall that

$$\begin{aligned}
&\widehat{a}^* = (\widehat{a}_1, \ldots, \widehat{a}_{J-1})^\intercal, \quad a_0^* = (a_{10}, \ldots, a_{J-1,0})^\intercal, \\
&R_t(w) = (R_{1t}^w, \ldots, R_{Jt}^w)^\intercal, \quad R_t^*(w) = \left(R_{1t}^w, \ldots, R_{J-1,t}^w\right)^\intercal, \\
&\widehat{R}_t(w) = \left(\widehat{R}_{1t}^w, \ldots, \widehat{R}_{Jt}^w\right)^\intercal, \quad \widehat{R}_t^*(w) = \left(\widehat{R}_{1t}^w, \ldots, \widehat{R}_{J-1,t}^w\right)^\intercal.
\end{aligned}$$

As in the proof of Theorem 3.1, throughout this proof, we let $J = J_0$ be a fixed positive integer. As $\widehat{a}_{J_0} = 1 - \sum_{j=1}^{J_0-1} \widehat{a}_j$ and $a_{J_0 0} = 1 - \sum_{j=1}^{J_0-1} a_{j0}$, by Proposition 4.1(i), Assumption 1(i) and the Taylor's expansion for $\dot{u}(\cdot)$, we may show that

$$
\begin{aligned}
& \dot{u}\big[\sum_{j=1}^{J_0} \widehat{a}_j \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] - \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] \\
= \;& \ddot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] \sum_{j=1}^{J_0} (\widehat{a}_j - a_{j0})\, \widehat{w}_j^\intercal (X_{j,t-1}) R_t + o_P(\|\widehat{a} - a_0\|) \\
= \;& \ddot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] (\widehat{a} - a_0)^\intercal \widehat{R}_t(w) + o_P(\|\widehat{a} - a_0\|) \\
= \;& \ddot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] (\widehat{a}^* - a_0^*)^\intercal \left[\widehat{R}_t^*(w) - \widehat{R}_{J_0 t}^w \mathbf{1}_{J_0-1}\right] + o_P(\|\widehat{a}^* - a_0^*\|)
\end{aligned}
$$

and

$$
\begin{aligned}
& \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] - \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] \\
= \;& \ddot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] \sum_{j=1}^{J_0} a_{j0} \left[\widehat{w}_j(X_{j,t-1}) - w_j(X_{j,t-1})\right]^\intercal R_t + O_P\left(h^4 + \frac{\log T}{Th}\right) \\
= \;& \ddot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] a_0^\intercal \left[\widehat{R}_t(w) - R_t(w)\right] + O_P\left(h^4 + \frac{\log T}{Th}\right).
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\dot{u}\big[\sum_{j=1}^{J_0} \widehat{a}_j \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] &= \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] + \dot{u}\big[\sum_{j=1}^{J_0} \widehat{a}_j \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] - \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] \\
&= \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] + \dot{u}\big[\sum_{j=1}^{J_0} \widehat{a}_j \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] - \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] \\
&\quad + \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] - \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big] \\
&= \eta_t + \eta_t^* \left[\widehat{R}_t^*(w) - \widehat{R}_{J_0 t}^w \mathbf{1}_{J_0-1}\right]^\intercal (\widehat{a}^* - a_0^*) + \eta_t^* \left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0 \\
&\quad + O_P\left(h^4 + \frac{\log T}{Th}\right) + o_P(\|\widehat{a}^* - a_0^*\|), \qquad \text{(B.16)}
\end{aligned}
$$

where $\eta_t = \dot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big]$ and $\eta_t^* = \ddot{u}\big[\sum_{j=1}^{J_0} a_{j0} w_j^\intercal (X_{j,t-1}) R_t\big]$.

By (4.4) and (B.16), we have

$$
\begin{aligned}
\mathbf{0} &= \frac{1}{T}\sum_{t=1}^{T}\left[\widehat{R}_t^*(w) - \widehat{R}_{J_0 t}^w \mathbf{1}_{J_0-1}\right] \dot{u}\big[\sum_{j=1}^{J} \widehat{a}_j \widehat{w}_j^\intercal (X_{j,t-1}) R_t\big] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\widehat{R}_t^*(w) - \widehat{R}_{J_0 t}^w \mathbf{1}_{J_0-1}\right] \left\{\eta_t + \eta_t^* \left[\widehat{R}_t^*(w) - \widehat{R}_{J_0 t}^w \mathbf{1}_{J_0-1}\right]^\intercal (\widehat{a}^* - a_0^*)\right. \\
&\quad \left. + \eta_t^* \left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0\right\} + O_P\left(h^4 + \frac{\log T}{Th}\right) + o_P(\|\widehat{a}^* - a_0^*\|). \qquad \text{(B.17)}
\end{aligned}
$$

By (B.17), we readily have

$$
\sqrt{T}\,(\widehat{a}^* - a_0^*) \overset{P}{\sim} \left[\frac{1}{T}\sum_{t=1}^{T}\eta_t^* \widehat{V}_t^* (\widehat{V}_t^*)^\intercal\right]^{-1} \left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^* \eta_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^* \eta_t^* \left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0\right\}, \qquad \text{(B.18)}
$$

where $\widehat{V}_t^* = \widehat{R}_t^*(w) - \widehat{R}_{J_0 t}^w \mathbf{1}_{J_0-1}$ and $\alpha_n \overset{P}{\sim} \beta_n$ denotes that $\lim_{n\to\infty} \alpha_n/\beta_n = 1$.

By Proposition 4.1 and the law of larger numbers, we readily have

$$
\frac{1}{T}\sum_{t=1}^{T}\eta_t^* \widehat{V}_t^* (\widehat{V}_t^*)^\intercal = \frac{1}{T}\sum_{t=1}^{T}\eta_t^* V_t^* (V_t^*)^\intercal + o_P(1) = \mathbf{\Delta}_1 + o_P(1), \qquad \text{(B.19)}
$$

where $\mathbf{\Delta}_1$ is defined in Section 4. Note that

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^* \eta_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^* \eta_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\widehat{V}_t^* - V_t^*\right)\eta_t.
$$

By Assumptions 2 and 3 and following the argument in the proof of Lemma B.3 in Li, Linton and Lu (2014), we may show that the second term on the right hand side of (B.19) is asymptotically negligible. Hence, we have

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^* \eta_t \overset{P}{\sim} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^* \eta_t. \qquad \text{(B.20)}
$$

We next consider $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^* \eta_t^* \left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0$. It is easy to see that

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^* \eta_t^* \left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0 \overset{P}{\sim} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^* \eta_t^* \left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0 \qquad \text{(B.21)}
$$

by using Proposition 4.1. Let $\mathbf{W}$ be an $n \times (n-1)$ matrix which is defined by

$$\mathbf{W} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1 \\ -1 & \cdots & -1 \end{pmatrix}.$$

It is easy to show that for any $j = 1, \ldots, J_0$ and $x_i \in \mathcal{X}_j$,

$$\widehat{w}_j(x_j) - w_j(x_j) = \mathbf{W}\left[\widehat{w}_j^*(x_j) - w_j^*(x_j)\right]. \tag{B.22}$$

Hence, by (B.22) and using the argument in the proofs of Theorem 3.1 and Proposition 4.1, we may show that

$$\begin{aligned}
\widehat{R}_{jt}^w - R_{jt}^w &= \left[\widehat{w}_j(X_{j,t-1}) - w_j(X_{j,t-1})\right]^\intercal R_t \\
&= \left[\widehat{w}_j^*(X_{j,t-1}) - w_j^*(X_{j,t-1})\right]^\intercal \mathbf{W}^\intercal R_t \\
&\overset{P}{\sim} R_t^\intercal \mathbf{W}\boldsymbol{\Lambda}_j^{-1}(X_{j,t-1}) \cdot \left[\frac{1}{Th}\sum_{s=1}^{T} Z_{js}(X_{j,t-1})\right] \\
&= R_t^\intercal \mathbf{W}\boldsymbol{\Lambda}_j^{-1}(X_{j,t-1}) \cdot \left[\frac{1}{Th}\sum_{s=1}^{T} R_s^* \dot{u}\left(w_j^\intercal(X_{j,s-1})R_s\right) K\left(\frac{X_{j,s-1} - X_{j,t-1}}{h}\right)\right] \\
&= R_t^\intercal \mathbf{W}\boldsymbol{\Lambda}_j^{-1}(X_{j,t-1}) \cdot \left[\frac{1}{Th}\sum_{s=1}^{T} R_s^* \varepsilon_{js} K\left(\frac{X_{j,s-1} - X_{j,t-1}}{h}\right)\right],
\end{aligned} \tag{B.23}$$

where $\varepsilon_{js} = \dot{u}\left(w_j^\intercal(X_{j,s-1})R_s\right) = \dot{u}\left(R_{js}^w\right)$ and $Z_{js}(\cdot)$ is defined in the proof of Theorem 3.1. By (B.23), we readily have

$$\left[\widehat{R}_t(w) - R_t(w)\right]^\intercal a_0 = R_t^\intercal \mathbf{W}\boldsymbol{\Lambda}_j^{-1}(X_{j,t-1}) \cdot \left[\frac{1}{Th}\sum_{s=1}^{T}\sum_{j=1}^{J_0} R_s^* \varepsilon_{js} a_{j0} K\left(\frac{X_{j,s-1} - X_{j,t-1}}{h}\right)\right],$$

which indicates that

$$\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^*\eta_t^*\left[\widehat{R}_t(w)-R_t(w)\right]^{\intercal} a_0 \\
\overset{P}{\sim} \;& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^*\eta_t^* R_t^{\intercal}\mathbf{W}\mathbf{\Lambda}_j^{-1}(X_{j,t-1})\cdot\left[\frac{1}{Th}\sum_{s=1}^{T}\sum_{j=1}^{J_0} R_s^*\varepsilon_{js}a_{j0}K\left(\frac{X_{j,s-1}-X_{j,t-1}}{h}\right)\right] \\
= \;& \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sum_{j=1}^{J_0}\varepsilon_{js}a_{j0}\left[\frac{1}{Th}\sum_{t=1}^{T}\eta_t^* V_t^* R_t^{\intercal}\mathbf{W}\mathbf{\Lambda}_j^{-1}(X_{j,t-1})K\left(\frac{X_{j,s-1}-X_{j,t-1}}{h}\right)\right] R_s^* \\
\overset{P}{\sim} \;& \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\sum_{j=1}^{J_0}\varepsilon_{js}a_{j0}Q_{js}, \qquad\qquad (B.24)
\end{aligned}$$

where

$$Q_{js}=\left\{\mathsf{E}\left[\eta_t^* V_t R_t^{\intercal}\mathbf{W}\mathbf{\Lambda}_j^{-1}(X_{j,t-1})|X_{j,t-1}=X_{j,s-1}\right]\right\}R_s^*.$$

Recall that $\varepsilon_t=\left[\varepsilon_{1t}a_{10},\dots,\varepsilon_{J_0t}a_{J_00}\right]^{\intercal}$ and $Q_t=(Q_{1t},\dots,Q_{J_0t})^{\intercal}$. Then we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^*\eta_t^*\left[\widehat{R}_t(w)-R_t(w)\right]^{\intercal} a_0 \overset{P}{\sim} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Q_t^{\intercal}\varepsilon_t \qquad (B.25)$$

By (B.20), (B.21) and (B.25), we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^*\eta_t+\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_t^*\eta_t^*\left[\widehat{R}_t(w)-R_t(w)\right]^{\intercal} a_0 \overset{P}{\sim} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(V_t^*\eta_t+Q_t^{\intercal}\varepsilon_t\right). \qquad (B.26)$$

By the central limit theorem for the α-mixing sequence, we can prove that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(V_t^*\eta_t+Q_t^{\intercal}\varepsilon_t\right)\xrightarrow{d}\mathsf{N}\left(\mathbf{0},\mathbf{\Delta}_2\right) \qquad (B.27)$$

Then, we can complete the proof of Theorem 4.2 by (B.18)–(B.21), (B.26) and (B.27). ■

PROOF OF THEOREM 4.3. The main idea in this proof is similar to the proof of Theorem 4.2 above with some modifications. Hence, we next only sketch the proof.

Following the proof of (B.18) and using the condition (4.9), we may show that

$$\begin{aligned}
\sqrt{T}\mathbf{\Psi}_T\mathbf{\Delta}_T^{-\frac{1}{2}}\left(\widehat{a}^*-a_0^*\right) \;=\; & \mathbf{\Psi}_T\mathbf{\Delta}_T^{-\frac{1}{2}}\left[\frac{1}{T}\sum_{t=1}^{T}\eta_t^*\widehat{V}_t^*(\widehat{V}_t^*)^{\intercal}\right]^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^*\eta_t\right. \\
& \left.+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^*\eta_t^*\left(\widehat{R}_t(w)-R_t(w)\right)^{\intercal} a_0\right]. \qquad (B.28)
\end{aligned}$$

Note that (B.19) still holds by using Proposition 4.1(ii) and (4.9). Hence, by (B.28), we have

$$\sqrt{T}\mathbf{\Psi}_T\mathbf{\Delta}_T^{-\frac{1}{2}}\left(\widehat{a}^* - a_0^*\right) \overset{P}{\sim} \mathbf{\Psi}_T\mathbf{\Delta}_T^{-\frac{1}{2}}\mathbf{\Delta}_1^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^*\eta_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{V}_t^*\eta_t^*\left(\widehat{R}_t(w) - R_t(w)\right)^\intercal a_0\right]. \quad \text{(B.29)}$$

Furthermore, using the argument in proving (B.26), we obtain

$$\sqrt{T}\mathbf{\Psi}_T\mathbf{\Delta}_T^{-\frac{1}{2}}\left(\widehat{a} - a_0\right) \overset{P}{\sim} \mathbf{\Psi}_T\mathbf{\Delta}_T^{-\frac{1}{2}}\mathbf{\Delta}_1^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(V_t^*\eta_t + Q_t^\intercal\varepsilon_t\right)\right]. \quad \text{(B.30)}$$

By the central limit theorem and using the condition that $\mathbf{\Psi}_T\mathbf{\Psi}_T^\intercal \to \mathbf{\Psi}$, we can complete the proof of Theorem 4.3. ■

References

[1] Back, K. E., 2010. *Asset Pricing and Portfolio Choice Theory.* Oxford University Press.

[2] Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (2nd Edition), Springer.

[3] Brandt, M. W., 1999. Estimating portfolio and consumption choice: a conditional Euler equations approach. *Journal of Finance* **54**, 1609–1645.

[4] Brandt, M. W., 2010. Portfolio choice problems. *Handbook of Financial Econometrics, Volume 1: Tools and Techniques* (Editors: Aït-Sahalia Y. and Hansen L. P.), 269-336.

[5] Fama, E. F., 1970. Multiperiod consumption-investment decisions. *American Economic Review* **60**, 163–174.

[6] Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics* **32**, 928–961.

[7] Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

[8] Györfi, L., G. Ottucsák, and A. Urbán (2011). Empirircal Log-Optimal Portfolio Selections: A Survey. World Scientific Review, 79-116

[9] Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

[10] Hansen, L. P., Singleton, K. J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* **50**, 1269–1286.

[11] Kristensen, D., 2009. Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory* **25**, 1433–1445.

[12] Li, D., Linton, O., Lu, Z., 2014. A flexible semiparametric forecasting model for time series. Under revision for *Journal of Econometrics*.

[13] Li, D., Lu, Z., Linton, O., 2012. Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory* **28**, 935–958.

[14] Markowitz, H. M., 1952. Portfolio selection. *Journal of Finance* **7**, 77–91.

[15] Merton, R. C., 1969. Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics* **51**, 247–257.

[16] Robinson, P. M., 1983. Nonparametric estimators for time series. *Journal of Time Series Analysis* **4**, 185–207.