Econometric Theory, **31**, 2015, 560–580.
doi:10.1017/S0266466614000589

PRICING KERNEL ESTIMATION: A LOCAL ESTIMATING EQUATION APPROACH

ZONGWU CAI
University of Kansas and Xiamen University

YU REN
Xiamen University

LINMAN SUN
Wells Fargo Bank

This paper investigates a general semiparametric stochastic discount factor formulation that avoids functional form misspecification. A new semiparametric estimation procedure is proposed which combines orthogonality conditions and local linear fitting to give a semiparametric generalized estimating equation approach. Asymptotic properties of the estimators are established and we explore the empirical usefulness of the proposed approach to value-weighted stock returns.

1. INTRODUCTION

Asset pricing models are a cornerstone of finance. They reveal how portfolio returns are determined and which factors affect returns. They also imply stochastic discount factors (SDFs) that can be used to determine current market prices of portfolios by discounting future payoffs, state by state. Since many asset pricing models assume a simple and stable linear relationship between assets' systematic risks and their expected returns, the SDFs in these models are a linear combination of the systematic factors with time-invariant coefficients.

However, this simple and stable relationship assumption in asset pricing models has been challenged by several recent studies based on empirical evidence of time variation in betas and expected returns (as well as return volatilities). Related works include Bansal, Hsieh, and Viswanathan (1993), Bansal and Viswanathan (1993), Cochrane (1996), Jaganathan and Wang (1996, 2002), Reyes (1999), Ferson and Harvey (1991, 1993, 1998, 1999), Cho and Engle (2000),

The authors would like to thank the editor, Professor Peter C.B. Phillips, the co-editor, Professor Oliver Linton, and the anonymous referees for their comments. Cai's research was supported, in part, by the National Nature Science Foundation of China Grants #71131008 (Key Project), #70871003, and #70971113. Ren's research was supported by a Natural Science Foundation of Fujian Province Grant (#2011J01384), the Fundamental Research Funds for the Central Universities (#2013221022), and the Natural Science Foundation of China Grants (#71301135, #71203189, #71131008). Address correspondence to Zongwu Cai, Department of Economics, University of Kansas, Lawrence, KS 66045, USA and The Wang Yanan Institute for Studies in Economics and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China; e-mail: zongwucai@gmail.com.

 © Cambridge University Press 2014

Wang (2002, 2003), Akdeniz, Altay-Salih, and Caner (2003), Ang and Liu (2004), Fraser, Hamelink, Hoesli, and MacGregor (2004), Gagliardini, Ossola, and Scaillet (2011), and the references therein. In particular, Ferson (1989), Harvey (1989), Ferson and Harvey (1991, 1993, 1998, 1999), Ferson and Korajczyk (1995), and Jaganathan and Wang (1996) concluded that beta and market risk premia indeed vary over time. Therefore, asset pricing models should incorporate time variation in their parameters.

Although there is a vast amount of empirical evidence on time variation in betas and risk premia, there is little theoretical guidance on how SDFs vary with time or what variables represent conditioning information. Bansal et al. (1993) and Bansal and Viswanathan (1993) were the first to advocate a flexible SDF model in empirical asset pricing and focused on nonlinear arbitrage pricing theory (APT) models by assuming that the SDF is a nonlinear function of some underlying state variables. Dumas and Solnik (1995) and Dittmar (2002) treated the coefficients of the factors in an SDF as linear functions of some instrument variables. Cochrane (1996) justified this specification by the so-called scaling factors.

Parametric models for time-varying betas and a nonlinear pricing kernel can be most efficient if the underlying models are correctly specified. However, a misspecification may cause serious bias and model constraints may distort the model in the local area, as discussed by Ghysels (1998). Nonparametric or semiparametric modeling is appealing in these situations. One of their main advantages is that little or no restrictive prior information on pricing kernels is needed. In a mean-covariance efficiency framework, Wang (2002, 2003) explored a nonparametric form of the SDF model by using the Nadaraya–Watson kernel regressions. More recently, Gourieroux and Monfort (2007) considered a class of nonlinear parametric and semiparametric SDF models for derivative pricing by assuming that the SDFs are exponential-affine functions of an underlying state variable.

In this paper, we estimate SDFs based on the conditional capital asset pricing model (CAPM). We propose a new semiparametric estimation procedure, termed the semiparametric generalized estimating equations (SPGEEs), which combines two techniques: local linear fitting and generalized estimating equations. This procedure avoids the potential model misspecification associated with the strong assumption of linearity. In addition, in contrast to Wang (2002, 2003), our approach does not require a mean-covariance efficiency framework. Using a simulation study and an empirical analysis, we compare our model with the GMM approach discussed in Dittmar (2002), which treats the smoothing variables as instruments of the conditional moments. We find that the proposed estimators are superior to the GMM estimators in terms of the mean square error (MSE) of pricing errors. This highlights the practical usefulness of our model and its estimation procedure.

Our paper is related to the work of Chen and Ludvigson (2009), Escanciano and Hoderlein (2010), Lewbel, Linton, and Srisuma (2011), Fang, Ren, and Yuan (2011), and Chen, Favilukis, and Ludvigson (2012), among others. However, ours differs in the following aspects. First, they estimate SDFs using the Euler equations, whereas the SDFs in our paper are derived from the conditional CAPM.

Also, our model and its estimating procedure can be easily extended to other APT models. Second, our proposed estimation method is in the spirit of local generalized estimating equations (GEEs) originally proposed by Carrol, Ruppert, and Welsh (1998), Cai (2003), and Cai and Li (2008). In contrast, their estimation methods are based on local GMM or the sieve minimum distance procedure. Local GEEs do not use the sandwich formula and use the kernel weight only once, whereas local GMM uses the sandwich formula and hence uses the kernel weight twice.

The rest of this paper is organized as follows. Section 2 describes our estimation model and presents the asymptotic results. Section 3 presents numerical results based on simulation and an empirical application. Section 4 concludes the paper. All the mathematical proofs are contained in the Appendix.

2. ECONOMETRIC MODEL AND ESTIMATION PROCEDURE

2.1. Model

We generalize the models studied by Bansal et al. (1993), Bansal and Viswanathan (1993), Ghysels (1998), Jagannathan and Wang (1996, 2002), Wang (2002, 2003), and some other models in the finance literature, under a very flexible framework. Specifically, we assume that the nonlinear pricing kernel has the form $m_{t+1} = 1 - m(Z_t) r_{p,t+1}$,[1] where $m(\cdot)$ is unspecified. Our approach focuses on estimating the semiparametric model

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

where $m(\cdot)$ is an unknown function of Z_t and Z_t is an $L \times 1$ vector of conditioning variables from Ω_t, and $r_{p,t+1}$ is the factor. Indeed, model (1) can be regarded as a conditional moment (orthogonal) condition. It is unnecessary to require $r_{p,t+1}$ to satisfy mean-variance efficiency, unlike in Wang (2002, 2003) and others. Hence, one of our main interests is to identify and estimate the nonlinear function $m(\cdot)$. Clearly, an alternative expression for (1) when $m(\cdot)$ is a scalar function is

$$m(Z_t) = E(r_{i,t+1} | \Omega_t) / E(r_{p,t+1} r_{i,t+1} | \Omega_t), \tag{2}$$

and under mean-variance efficiency, $m(Z_t)$ reduces to $b(Z_t) = E(r_{p,t+1} | Z_t) / E(r_{p,t+1}^2 | Z_t)$ as in Wang (2002, 2003).

Remark 1 (Extension to multifactor models). It is easy to extend model (1) to cover multiple factor models. In such a case, $r_{p,t+1}$ is a vector. Then, model (1) becomes

$$E[\{1 - m(Z_t)^\top r_{p,t+1}\} r_{i,t+1} \,|\, \Omega_t] = 0. \tag{3}$$

For simplicity, our focus here is on the one-dimensional case.

In the famous three-factor model of Fama and French (1993), $r_{p,t+1}$ can be expressed as

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

Then, the model becomes a special case of the model

$$E[\{1 - m^*(Z_t)^\top r_{mf,t+1}\} \, r_{i,t+1} \,|\, \Omega_t] = 0, \tag{4}$$

where $m^*(Z_t) = m(Z_t)\theta(t)$. Here, θ is allowed to vary over time and can be identified and estimated in a fully semiparametric way. Indeed, a simple version of model (4) was considered by Wang (2003).

Note that the estimation of model (1) and its econometric theory as applied in the next two sections to a single market portfolio also hold for models (3) and (4); however, the details are omitted due to their similarity. Also, note that, unfortunately, the results do not extend to the nonparametric pricing kernel in (2).

2.2. Semiparametric GEE

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

$$E[\{1 - m(Z_t) r_{p,t+1}\} \, r_{i,t+1} \,|\, I_t] = 0, \tag{5}$$

which can be regarded as an approximation of model (1). It follows from the orthogonality condition in (5) that, for any vector function $Q(I_t) \equiv Q_t$ with dimension d_q (specified later),

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

with sample version

$$\frac{1}{T} \sum_{t=1}^{T} Q_t \{1 - m(Z_t) r_{p,t+1}\} \, r_{i,t+1} = 0. \tag{6}$$

Therefore, this is an estimation approach similar to the GMM of Hansen (1982) for parametric models and the estimation equations in Cai (2003) for nonparametric models. We propose a semiparametric estimation procedure to combine the orthogonality conditions given in (6) with the local linear fitting scheme of Fan and Gijbels (1996) to estimate the unknown function $m(\cdot)$. This estimation approach is termed the semiparametric generalized estimating equations (SPGEEs).

It is well known in the literature (see, for example, Fan and Gijbels, 1996) that local linear fitting has several nice properties, when compared to the classical Nadaraya–Watson (local constant) method, such as high statistical efficiency in an asymptotic minimax sense, design-adaptation, and automatic edge correction. To estimate $m(\cdot)$ using local linear fitting from observations $\left\{(r_{i,t+1}, r_{p,t+1}, Z_t)\right\}_{t=1}^{T}$, we assume throughout that $m(\cdot)$ is twice continuously differentiable. Then, for a given point z_0 and for $\{Z_t\}$ in a neighborhood of z_0, by a Taylor expansion, $m(Z_t)$ is approximated by the linear function $a+b^\top(Z_t-z_0)$ with $a=m(z_0)$ and $b=m'(z_0)$ (the derivative of $m(z_0)$), so that model (6) is approximated by the orthogonality condition

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

Therefore, for $\{Z_t\}$ in a neighborhood of z_0, the orthogonality conditions in (6) can be approximated by the locally weighted orthogonality conditions

$$\sum_{t=1}^{T} Q_t\,[1-(a+b^\top(Z_t-z_0))r_{p.t+1}]\, r_{i,t+1}\, K_h(Z_t-z_0)=0, \tag{7}$$

where $K_h(\cdot)=h^{-L}K(\cdot/h)$, $K(\cdot)$ is a kernel function in $\mathbb{R}^L$, and $h=h_T>0$ is a bandwidth which controls the amount of smoothing used in the estimation. Equation (7) can be viewed as a generalization of the nonparametric estimation equations in Cai (2003) and the locally weighted version of (9.2.29) in Hamilton (1994, p. 243) or (14.2.20) in Hamilton (1994, p. 419) for parametric instrumental variable models. To ensure that the equation in (7) has a unique solution, the dimension of $Q(\cdot)$ must be greater than $L+1$, the number of parameters in (7). Therefore, solving the above equation leads to the so-called SPGEE estimate of $m(z_0)$, denoted by $\widehat{m}(z_0)$, and the SPGEE estimate of $m'(z_0)$, denoted by $\widehat{m}'(z_0)$. That is,

$$\begin{pmatrix}\widehat{m}(z_0)\\ \widehat{m}'(z_0)\end{pmatrix}=\begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix}=(S_T^\top S_T)^{-1}\,S_T^\top L_T, \tag{8}$$

where $S_T=\frac{1}{T}\sum_{t=1}^{T}Q_t\,Q_t^{*\top}\,r_{p,t+1}K_h(Z_t-z_0)\,r_{i,t+1}$, $L_T=\frac{1}{T}\sum_{t=1}^{T}Q_t\,K_h$ $(Z_t-z_0)\,r_{i,t+1}$, and $Q_t^*=\begin{pmatrix}r_{p,t+1}r_{i,t+1}\\ r_{p,t+1}r_{i,t+1}(Z_t-z_0)\end{pmatrix}$. Clearly, when $d_q=L+1$ and S_T is nonsingular, $\begin{pmatrix}\widehat{m}(z_0)\\ \widehat{m}'(z_0)\end{pmatrix}$ becomes $S_T^{-1}\,L_T$. Note that (8) provides a formula for computational implementation, which can be carried out by any standard statistical package. By moving z_0 over the whole domain of $m(\cdot)$, the entire estimated curve of $m(\cdot)$ is obtained.

We now turn to the choice of Q_t in (7). Motivated by the estimation equations in Cai (2003) and following a similar idea in Cai and Li (2008), we choose Q_t as

$$Q_t=Q_t^*; \tag{9}$$

see Remark 5 later for further discussion on this choice. Thus, $\begin{pmatrix} \widehat{m}(z_0) \\ \widehat{m}'(z_0) \end{pmatrix}$ becomes $S_T^{-1} L_T$ for $Q_t = Q_t^*$. Finally, we note that the method proposed in Cai (2003) can be regarded as a special case of the SPGEE estimation procedure.

2.3. Asymptotic Theory

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

Assumption A.

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

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

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

(iii) Assume that $\sigma_0^2(z)$ satisfies the Lipschitz condition. Also, assume that there exists some $\delta > 0$ such that $E\{|e_t|^{2+\delta} | Z_t = z\}$ is continuous at z_0.

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

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

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

A5. Assume that $T h^{L[1+2/(1+\delta)]} \to \infty$.

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

Before stating our main asymptotic result, we need to introduce some notation, which will be used throughout the paper. Define $\mu_2(K) = \int u u^\top K(u) du$ and $\nu_0(K) = \int K^2(u) du$. Define $H = \text{diag}\{1, h^2 I_L\}$, where I_L is an $L \times L$ identity matrix. Finally, define $S(z) = M(z)\text{diag}\{1, \mu_2(K)\}$ and $S^*(z) = \text{diag}\{\nu_0(k), h^2\mu_2(K^2)\}\sigma_0^2(z)$. The asymptotic normality of the SPGEE estimator is established in Theorem 1 with its proof relegated to the Appendix.

THEOREM 1. *Under Assumptions A1–A5, for any grid point z_0, we have*

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

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

$$\sqrt{Th^L}\left[\widehat{m}(z_0) - m(z_0) - \frac{h^2}{2}tr(\mu_2(K)m''(z_0))\right] \to N(0,\, \sigma_m^2(z_0)), \qquad \textbf{(10)}$$

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

Remark 3 (Consistent estimate of asymptotic variance). The first consequence of Theorem 1 is to provide an easy way to obtain a consistent estimate of the asymptotic variance $\sigma_m^2(z_0)$. After estimating the semiparametric pricing kernel, we can obtain the estimated pricing error as $\widehat{e_t} = [1 - \widehat{m}(Z_t) r_{p,t+1}] r_{i,t+1}$. Then, any nonparametric kernel smoothing method, say the local linear technique, can be applied to obtain a consistent estimate for $\sigma_0^2(z_0)$, $f(z_0)$, and $M(z_0)$. One can apply some existing optimal bandwidth selectors, such as plugging in, cross-validation, generalized cross-validation, the nonparametric Akaike information criterion, etc. Therefore, a consistent estimate for $\sigma_m^2(z_0)$ is $\widehat{\sigma}_m^2(z_0) = \nu_0(K)\widehat{\sigma}_0^2(z_0)\widehat{f}^{-1}(z_0)\widehat{M}^{-2}(z_0)$. Thus, for a given grid point z_0, a 95% pointwise confidence interval for $m(z_0)$ with bias ignored can be constructed as

$$\widehat{m}(z_0) \pm 1.96 \times \frac{\widehat{\sigma}_m(z_0)}{\sqrt{Th^L}},$$

which will be used in computing the confidence interval for the empirical application presented in Section 3.

Remark 4 (A rule of thumb for bandwidth selection). It is well known that bandwidth plays an essential role in the trade-off between bias and variance. Oftentimes, one would like to have a rough idea of how large the amount of smoothing should be, so a rule of thumb is very appealing. While somewhat crude, such a rule is simple and requires little programming effort compared to other methods. To this end, from Theorem 1, we see that the weighted integrated asymptotic mean squared error (AMSE) is given by

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

where $C_1 = E\left[\sigma_m^2(Z_t)\right]$ and $C_2 = E\left[\text{tr}(\mu_2(K)m''(Z_t))\right]$. By minimizing the AMSE with respect to h, we obtain the optimal theoretical bandwidth

$$h_{opt} = \left(\frac{LC_1}{C_2}\right)^{1/(L+4)} T^{-1/(L+4)} \equiv C_3 T^{-1/(L+4)}. \quad \textbf{(11)}$$

With this choice of h_{opt}, we see that the optimal AMSE has the order of $O(T^{-4/(L+4})$. Clearly, the formulation in (11) provides an easy way to find a data-driven bandwidth selection method; say, a plugging in method. So, we need to estimate C_3 consistently, which can be done as follows. First, take a pilot bandwidth h_0 which is much smaller than $T^{-1/(L+4)}$, say $h_\sigma = 0.1 \times T^{-1/(L+4)}$ or smaller. Using this pilot bandwidth, we can estimate $\sigma_m^2(z_0)$, so that we obtain $\widehat{C}_1$ using the average.

We use a simple method to estimate $m''(z_0)$ consistently and easily. We fit a multivariate polynomial of a certain order L_m (say $L_m = \log(T)$ or larger) globally to $m(z)$, leading to a parametric fit. Other global parametric approaches, including series and spline methods, can also be used. Then, the GMM of Hansen (1982) can be used to estimate the parameters. The choice of a global fit results in a derivative function $\widehat{m}''(z)$ which is a multivariate polynomial of order $L_m - 2$. Thus, we obtain $\widehat{C}_2$ by average and get $\widehat{C}_3$ and $\widehat{h}_{opt} = \widehat{C}_3 T^{-1/(L+4)}$. This rule of thumb bandwidth selector will be used in our computation for the empirical application discussed in Section 3.

Remark 5 (Choice of instruments). After we establish the asymptotic property of the estimator, we now turn to the choice of $Q_t = Q(Z_t)$. For now, we assume that $Q(Z_t) = \begin{pmatrix} Q_0(Z_t) \\ Q_0(Z_t)(Z_t - z_0) \end{pmatrix}$, where $Q_0(Z_t)$ is an unknown scale function. By following the same proofs used in the proof of Theorem 1, we can show that the asymptotic normality in (10) also holds true for our choice of Q_t with the asymptotic variance

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

where $S_1(z) = Q_0(z)S(z)$ and $S_1^*(z) = Q_0^2(z)S^*(z)$. It is clear that the asymptotic variance $\triangle_{m,0}(z) = \triangle_m(z)$, which is not related to the choice of $Q_0(\cdot)$. Hence, we assume that $Q(\cdot)$ has the form given in (9).

2.4. Multiple Assets

In economics and finance applications, risky portfolios are usually multiple, so it may not be appropriate to focus on only one particular portfolio. In this section, we discuss how to extend our estimation procedure to a multiple-portfolio scenario.

As shown earlier, the orthogonal condition is

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

or

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

Here $r_{i,t+1}$ is the i-th excess return of the risky assets. If we have N excess returns, then this orthogonal condition is valid for $i = 1, 2, \ldots, N$.

We can use the following strategy to consider these N orthogonal conditions. First note that the linear combination of excess returns can also be regarded as an excess return. That is to say, if

$$\theta_1 + \theta_2 + \cdots + \theta_N = 1,$$

then we can define $\tilde{r}_{t+1}$ as

$$\tilde{r}_{t+1} = \theta_1 r_{1,t+1} + \theta_2 r_{2,t+1} + \cdots + \theta_N r_{N,t+1}.$$

It is clear that $\tilde{r}_{t+1}$ is still an excess return. If we impose the orthogonal condition on $\tilde{r}_{t+1}$, then we have

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

which is equivalent to

$$E\left[Q_t\, \{1-m(Z_t)r_{p,t+1}\}\, \{\theta_1 r_{1,t+1} + \theta_2 r_{2,t+1} + \cdots + \theta_N r_{N,t+1}\} \,|\, I_t\right] = 0,$$

or

$$\begin{aligned} &\theta_1 E\left[Q_t\, \{1-m(Z_t)r_{p,t+1}\}\, r_{1,t+1} \,|\, I_t\right] + \cdots \\ &\quad + \theta_N E\left[Q_t\, \{1-m(Z_t)r_{p,t+1}\}\, r_{N,t+1} \,|\, I_t\right] = 0. \end{aligned}$$

This equation can be explained as the linear combination of N orthogonal conditions with specific weight $(\theta_1, \theta_2, \ldots, \theta_N)$.

We cannot choose θ_i to optimize the estimator. However, in practice, the parametric GMM used for estimation of the SDFs does not adopt the optimal weighting matrix either. Most papers just use the identity matrix to estimate SDFs. So we can safely set $\theta_i = 1/N$ for simplicity.

3. NUMERICAL ANALYSIS

3.1. Monte Carlo Simulation

To illustrate the validity of our method, we conduct some Monte Carlo simulations. For simplicity of implementation, we choose only one variable, z_t, which has the autoregressive data generating process

$$z_t = 0.02 z_{t-1} + 0.01\epsilon_{t,1},$$

where $\epsilon_{t,1}$ has a standard normal distribution. The conditional mean of $r_{p,t+1}$ takes the form $r_{p,t+1} = 0.01g(z_t) + 0.05\epsilon_{t,2}$, where $g(z_t) = 0.1 + 0.1 z_t^2$ and $\epsilon_{t,2}$ has a standard normal distribution. We choose $m(z_t)$ as

$$m(z_t) = 0.01g(z_t)/[(0.05)^2 + 0.01^2 g(z_t)^2].$$

Thus, $r_{i,t+1}$ is determined by (1) as $r_{i,t+1} = e_{i,t}/[1 - m(z_t)r_{p,t+1}]$, where

$$e_{i,t} = 0.05 e_{i,t-1} + 0.01 \upsilon_{i,t},$$

where $\upsilon_{i,t}$ is also a standard normal variable. Random variable $\upsilon_{i,t}$ is i.i.d. (independent and identically distributed) across both i and t for all $1 \leq i \leq N$ and $1 \leq t \leq T$. Here, $N = 25$ and $T = 250$, 500, and 1,000.

Based on the simulated data, SDFs are estimated using the proposed method. For each setting and each sample size, we perform 1,000 iterations and collect the MSE of the SDFs. In order to highlight the performance of point estimation, we only report the results when $z_t = 0$ in Table 1.

In addition, we compare our estimators with the ones provided by a nonlinear parametric model. The nonlinear parametric model can be expressed as

$$m_{t+1} = F_{t+1}b,$$

where $F_t = [1, r_{p,t+1}, z_t, r_{p,t+1}z_t]$. We can use GMM to estimate b as

$$\min_b g_T(b) \equiv \omega'_T(b) W \omega_T(b).$$

If we define $D_T = T^{-1}\sum_{t=1}^{T} R'_t F_t$, then $\omega_T = D_T b - 1_N$, where 1_N is a N-vector of ones. Thus, b can be estimated as

$$\widehat{b} = (D'_T W D_T)^{-1} D'_T W 1_N.$$

TABLE 1. MSEs of SDF estimation

	T = 250	T = 500	T = 1,000
SPGEE	0.0226	0.0170	0.0067
GMM	0.0628	0.0540	0.0486

Here, the weighting matrix W is chosen to be an identity matrix since, in our data, T is small. The inverse of the sample covariance matrix may not be a good estimator for the optimal weighting matrix. Altonji and Segal (1996) and Cochrane (2001) recommend the identity matrix for this scenario. We can see from Table 1 that our SDF estimators have much smaller MSEs than the GMM estimators because GMM estimators suffer from model misspecification. Also, one can observe clearly that the MSEs of our estimators decrease as the sample size increases, which is consistent with our theory.

3.2. Empirical Application

Our test assets are the monthly returns of the Fama–French 25 portfolios from July 1963 to November 2010. The return on the benchmark portfolio is the

TABLE 2. Mean square errors. We use our method, SPGEE, and the nonlinear parametric method, GMM, to estimate SDFs of the Fama–French 25 portfolios. This table reports the MSEs of the pricing errors. The portfolios are sorted by size (S) and book-to-market ratio (B). S1 (and B1) denotes the lowest order while S5 (and B5) denotes the highest order

	SPGEE	GMM
S1/B1	0.0060	0.0162
S1/B2	0.0045	0.0135
S1/B3	0.0033	0.0085
S1/B4	0.0029	0.0073
S1/B5	0.0033	0.0074
S2/B1	0.0050	0.0135
S2/B2	0.0032	0.0074
S2/B3	0.0026	0.0063
S2/B4	0.0025	0.0059
S2/B5	0.0031	0.0070
S3/B1	0.0043	0.0114
S3/B2	0.0027	0.0062
S3/B3	0.0022	0.0050
S3/B4	0.0022	0.0049
S3/B5	0.0027	0.0062
S4/B1	0.0035	0.0092
S4/B2	0.0024	0.0052
S4/B3	0.0023	0.0051
S4/B4	0.0021	0.0051
S4/B5	0.0027	0.0060
S5/B1	0.0022	0.0056
S5/B2	0.0019	0.0043
S5/B3	0.0017	0.0042
S5/B4	0.0018	0.0046
S5/B5	0.0023	0.0054

value-weighted return on all NYSE, AMEX, and NASDAQ stocks. The risk-free rate is the one-month Treasury bill rate. The chosen condition variable is the default spread of the return on BAA and AAA corporate bonds. As required by the model, this variable is lagged one period.

After obtaining the SDF estimators by two different methods, we report the MSEs of the pricing errors delivered by the two models. The results are summarized in Table 2. We can see that for each portfolio, the MSE of our method is much smaller than that for the GMM method, which implies that although we can use the cross-product between the information variable and the factor to capture the conditional information, the potential for model misspecification jeopardizes the performance of the GMM method.

Finally, we plot the estimated SDFs with the excess returns and the default spreads. Figure 1 shows that when the excess return is large, the SDF is low. This is reasonable since the SDF can be explained as the ratio of the marginal utility of consumption of the next period over the marginal utility of current consumption. When there is high excess return in the next period, the consumption of the next period is going to increase accordingly. So the marginal utility of the next period decreases and the SDF is low. We also observe that the default spread in the current period is usually associated with a high SDF of the next period. This is because a high SDF of the next period implies that consumers anticipate low consumption in the next period. Low consumption usually occurs in a recession. In a recession, consumers need higher premia to compensate for larger risk in the markets. Therefore, to attract consumers to take risky assets in the current period, the default spread is going to be high.

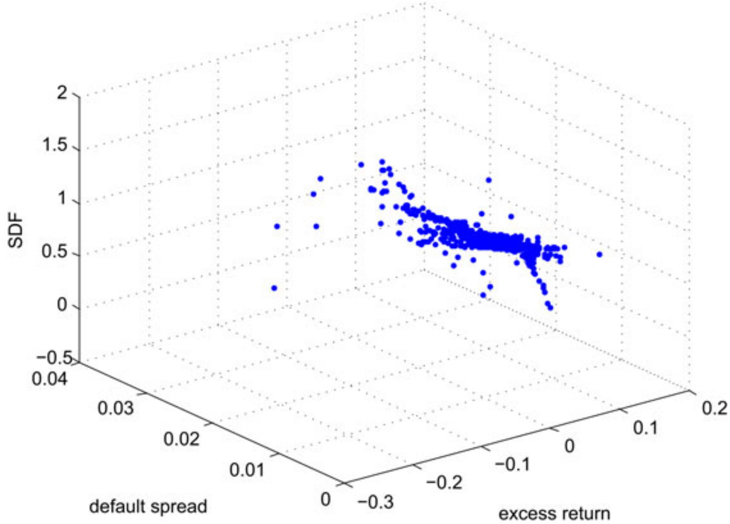

FIGURE 1. Estimated SDFs.

4. CONCLUSION

This paper proposes a semiparametric method to estimate SDFs. When the conventional SDF holds conditionally on the current information set, the coefficients of the factors in the SDF are time-varying. We assume that the coefficients are unknown functions of some observable variables and estimate those functions semiparametrically. Our estimator avoids model misspecification and the assumption that the benchmark portfolio is mean-variance efficient. In both simulations and our empirical application, we find that our estimator delivers smaller pricing errors than alternative estimators in the literature.

NOTE

1. A detailed explanation about the form of $m_{t+1} = 1 - m(Z_t) r_{p,t+1}$ can be found in Wang (2003).

REFERENCES

Akdeniz, L., A. Altay-Salih, & M. Caner (2003) Time-varying betas help in asset pricing: The threshold CAPM. *Studies in Nonlinear Dynamics & Econometrics* 6(4), Article 1.

Altonji, J. & L. Segal (1996) Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics* 14, 353–366.

Ang, A. & J. Liu (2004) How to discount cashflows with time-varying expected return. *Journal of Finance* 59, 2745–2783.

Bansal, R., D.A. Hsieh, & S. Viswanathan (1993) A new approach to international arbitrage pricing. *Journal of Finance* 48, 1719–1747.

Bansal, R. & S. Viswanathan (1993) No arbitrage and arbitrage pricing: A new approach. *Journal of Finance* 47, 1231–1262.

Cai, Z. (2002) Regression quantiles for time series data. *Econometric Theory* 18, 169–192.

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

Cai, Z. & Q. Li (2008) Nonparametric estimation of varying coefficient dynamic panel models. *Econometric Theory* 24, 1321–1342.

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

Carrol, R., D. Ruppert, & A. Welsh (1998) Local estimating equations. *Journal of the American Statistical Association* 93, 214–227.

Chen, X., J. Favilukis, & S. Ludvigson (2012) An Estimation of Economic Models with Recursive Preferences. Working paper, New York University.

Chen, X. & S. Ludvigson (2009) Land of addicts? An empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics* 24, 1057–1093.

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

Cho, Y.-H. & R.F. Engle (2000) Time-varying Betas and Asymmetric Effects of News: Empirical Analysis of Blue Chip Stocks. Working paper, New York University.

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. (2001) *Asset Pricing*. Princeton University Press.

Dittmar, R. (2002) Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance* 57, 369–403.

Dumas, B. & B. Solnik (1995) The world price of foreign exchange rate risk. *Journal of Finance* 50, 445–479.

Escanciano, J.-C. & S. Hoderlein (2010) Nonparametric Identification of Euler Equations. Working paper, Boston College.

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

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fang, Y., Y. Ren, & Y. Yuan (2011) Nonparametric estimation and testing of stochastic discount factor. *Finance Research Letters* 8, 196–205.

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

Ferson, W.E. & C.R. Harvey (1991) The variation of economic risk premiums. *Journal of Political Economy* 99, 385–415.

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

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

Ferson, W.E. & C.R. Harvey (1999) Conditional variables and the cross section of stock return. *Journal of Finance* 54, 1325–1360.

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

Fraser, P., F. Hamelink, M. Hoesli, & B. MacGregor (2004) Time-varying betas and cross-sectional return-risk relation: Evidence from the UK. *European Journal of Finance* 10, 255–276.

Gagliardini, P., E. Ossola, & O. Scaillet (2011) Time-varying risk premia in large cross-sectional equity datasets. *Swiss Finance Institute Research Paper No. 11–40*.

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.

Gourieroux, C. & A. Monfort (2007) Econometric specification of stochastic discount factor models. *Journal of Econometrics* 136, 509–530.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

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

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

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

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

Lewbel, A., O. Linton, & S. Srisuma (2011) Nonparametric Euler Equation Identification and Estimation. Working paper, Boston College.

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

Shao, Q. & H. Yu (1996) Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* 24, 2098–2127.

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

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

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

APPENDIX: Theoretical Proofs

To prove Theorem 1, we need the following four lemmas, which are stated below without proof. For detailed proofs of these lemmas, we refer the reader to Hall and Heyde (1980) for Lemma 1, Volkonskii and Rozanov (1959) for Lemma 2, and Shao and Yu (1996) for Lemma 3. We use the same notation as introduced in Section 2. Throughout this Appendix, we denote C as a generic positive constant, which may take different values at different times.

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

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

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

LEMMA A.2. *Let $V_1,\ldots,V_{L_1}$ be α-mixing stationary random variables that are $\mathbb{F}_{i_1}^{j_1},\ldots,\mathbb{F}_{i_{L_1}}^{j_{L_1}}$-measurable, respectively, with $1 \leq i_1 < j_1 < \cdots < j_{L_1}, i_{l+1} - j_l \geq \tau$, and $|V_l| \leq 1$ for $l = 1,\ldots L_1$. Then,*

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

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

LEMMA A.3. *Let V_t be an α-mixing process with $E(V_t) = 0$ and $\| V_t \|_r < \infty$ for $2 < p < r \leq \infty$. Define $S_n = \sum_{t=1}^{n} V_t$ and assume that $\alpha(\tau) = O(\tau^{-\theta})$ for some $\theta > pr/(2(r-p))$. Then,*

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

where K is a finite positive constant.

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

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

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

We define some notations as follows. Let $\tilde{S}_T = H^{-1} S_T$, where $S_T = Q_t Q_t^\top K_h (Z_t - z_0) r_{i,t+1} r_{p,t+1}$. Also, set $B_T = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2} Q_t r_{i,t+1} r_{p,t+1} (Z_t - z_0)^\top m''(z_0)(Z_t - z_0) K_h (Z_t - z_0)$ and $R_t = m(Z_t) - m(z_0) - m'(z_0)^\top (Z_t - z_0) - \frac{1}{2}(Z_t - z_0)^\top m''(z_0)(Z_t - z_0)$. Finally, set $R_T^* = \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t - z_0) Q_t r_{i,t+1} r_{p,t+1} \big[m(Z_t) - m(z_0) - m'(z_0)^\top (Z_t - z_0) - \frac{1}{2}(Z_t - z_0)^\top m''(z_0)(Z_t - z_0)\big]$. Now, we have the following asymptotic results for the above quantities.

PROPOSITION A.1. *Under Assumptions A1–A5, we have*

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

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

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

Proof. By the stationarity assumption and Assumptions A1–A5,

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

and

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

By Assumptions A1 and A2,

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

which implies that

$$I_1 = O(1).$$

Next we prove that $I_2 \to 0$. To this end, reformulate I_2 as $I_2 = I_3 + I_4$, where $I_3 = 2h^L/T\sum_{t=1}^{d_T}(\cdots)$ and $I_4 = 2h^L/T\sum_{t>d_T}(\cdots)$. Let $d_T \to \infty$ be a sequence of integers such that $d_T h^L \to 0$. First, we show that $I_3 \to 0$. Conditional on Z_1, Z_t, and using Assumption A2, we obtain

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

Thus, it follows that $I_3 \leq d_T h^L \to 0$. We now consider the contribution of I_4. For an α-mixing process, we use Davydov's inequality (see Lemma 1) to obtain the following:

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

By Assumption A2,

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

Thus,

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

and

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

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

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

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

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

Using similar arguments, we can show that

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

and

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

Therefore, by (A.1) and (A.2),

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

This proves (i).

Next, we show (ii). Note that by the stationarity assumption and Assumption A2,

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

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

Finally,

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

where, by Assumption A2,

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

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

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

Similarly, we can show that $\text{Var}[h^{-2} R_T^*] = o(1)$. ■

PROPOSITION A.2. *Under Assumptions A1–A5,*

$$
T h^L\, \mathit{Var}(G_T) \;\to\; f(z_0) S^*,
$$

where $G_T = \frac{1}{T} \sum_{t=1}^{T} Q_t\, e_t\, K_h(Z_t - z)$.

Proof. By the orthogonal condition in (1), $E(G_T) = 0$ and

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

By Assumption A2, similar to the proof of Proposition 1,

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

Similar to I_2, we split I_6 into two parts $I_6 = I_7 + I_8$, where $I_7 = 2h^L/T\sum_{t=1}^{d_T}(\cdots) \leq d_T h^L \rightarrow 0$ and $I_8 = 2h^L/T\sum_{t>d_T}(\cdots)$. By taking v_1 and v_2 as 0 or 1, by Davydov's inequality, we obtain

$$|\text{Cov}(e_1 K_h(Z_1 - z_0)(Z_1 - z_0)^{v_1}, e_t K_h(Z_t - z_0)(Z_t - z_0)^{v_2})|$$
$$\leq C[\alpha(t)]^{\frac{\delta}{2+\delta}} \| e_1 K_h(Z_1 - z_0)(Z_1 - z_0)^{v_1} \|_{2+\delta} \| e_t K_h(Z_t - z_0)(Z_t - z_0)^{v_2} \|_{2+\delta}$$

and

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

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

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

and

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

This proves Proposition 2. ■

Proof of Theorem 1. Recall that

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

It follows from Propositions 1 and 2 that

$$H\left\{\begin{pmatrix}\widehat{m}(z_0)\\ \widehat{m}'(z_0)\end{pmatrix} - \begin{pmatrix}m(z_0)\\ m'(z_0)\end{pmatrix}\right\} - B(z_0) + o_p(h^2) = f^{-1}(z_0)S^{-1}G_T\{1 + o_p(1)\}. \quad \textbf{(A.3)}$$

To prove Theorem 1, it suffices to establish the asymptotic normality of $\sqrt{Th^L}G_T$, which we do using the Wold–Cramér device, so that we may consider a linear combination with a unit vector $d^\top G_T$. It is easy to show by simple algebra that

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

where $w_t = \sqrt{h^L}d^\top\left\{Q_t r_{i,t+1}K_h(Z_t - z_0)(1 - m(Z_t)r_{p,t+1})\right\}$. It is clear that the problem reduces to proving the asymptotic normality of $\sum_{t=1}^{T} w_t/\sqrt{T}$. By Proposition 2, we show that

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

Therefore,

$$\mathrm{Var}\left(\sqrt{Th^L}\,d^\top G_T\right) = \theta^2(z_0)(1+o(1)). \tag{A.4}$$

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

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

Then,

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

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

$$\frac{1}{T}E[Q_{T,2}]^2 \to 0, \quad \frac{1}{T}E[Q_{T,3}]^2 \to 0, \tag{A.5}$$

$$\left| E\left[\exp(itQ_{T,1})\right] - \prod_{j=0}^{q_T-1} E\left[\exp(it\eta_j)\right]\right| \to 0, \tag{A.6}$$

$$\frac{1}{T}\sum_{j=0}^{q_T-1} E(\eta_j^2) \to \theta^2(z_0), \tag{A.7}$$

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

$$\frac{1}{T}\sum_{j=0}^{q_T-1} E\left[\eta_j^2 I\{|\eta_j| \geq \epsilon^*\theta(z_0)\sqrt{T}\}\right] \to 0. \tag{A.8}$$

Clearly, these four statements imply that the sums over small and residual blocks, $Q_{T,2}/\sqrt{T}$ and $Q_{T,3}/\sqrt{T}$, are asymptotically negligible in probability, and that $\{\eta_j\}$ in $Q_{T,1}$ is asymptotically independent. Also, (A.7) and (A.8) are standard Lindeberg–Fellow conditions for the asymptotic normality of $Q_{T,1}/\sqrt{T}$. To show the asymptotical normality of $d^\top G_T$, it suffices to establish the four statements in (A.5)–(A.8). First, we choose the block sizes

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

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

$$s_T/r_T \to 0, \quad r_T/T \to 0, \quad \text{and} \quad q_T\,\alpha(s_T) \to 0. \tag{A.9}$$

Now we establish (A.5) and (A.7). Clearly,

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

By stationarity and (A.4),

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

and

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

Hence, by (A.9),

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

It follows from the stationarity condition, (A.9), and Proposition 2 that

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

From Lemma 2,

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

This proves (A.6). It remains to show that

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

It follows from Lemma 3 that

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

It can be easily shown that

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

By plugging this into the right-hand side of the previous equation, we obtain

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

by Assumption A5. This proves Theorem 1. ∎