



Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Conditional beta pricing models: A nonparametric approach

Eva Ferreira^a, Javier Gil-Bazo^{b,*}, Susan Orbe^a^a *Departamento de Economía Aplicada III (Econometría y Estadística), University of the Basque Country, Bilbao, Spain*^b *Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain*

ARTICLE INFO

Article history:

Received 28 June 2009

Accepted 20 May 2011

Available online 27 May 2011

JEL classification:

G12

C14

C32

Keywords:

Kernel estimation

Conditional beta pricing models

Fama–French three-factor model

Locally stationary processes

ABSTRACT

We propose a two-stage procedure to estimate conditional beta pricing models that allows for flexibility in the dynamics of asset betas and market prices of risk (MPR). First, conditional betas are estimated nonparametrically for each asset and period using the time-series of previous data. Then, time-varying MPR are estimated from the cross-section of returns and betas. We prove the consistency and asymptotic normality of the estimators. We also perform Monte Carlo simulations for the conditional version of the three-factor model of Fama and French (1993) and show that nonparametrically estimated betas outperform rolling betas under different specifications of beta dynamics. Using return data on the 25 size and book-to-market sorted portfolios, we find that the nonparametric procedure produces a better fit of the three-factor model to the data, less biased estimates of MPR and lower pricing errors than the Fama–MacBeth procedure with betas estimated under several alternative parametric specifications.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Beta pricing models, such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) or the Arbitrage Pricing Theory (APT) of Ross (1976), are used extensively in portfolio management, risk management, and capital budgeting applications. In these models, an asset's risk premium (its expected return in excess of the risk-free interest rate) is linearly related to the asset's betas with respect to factors capturing market-wide sources of risk. The slopes of the linear relation, which must be equal for all assets, are interpreted as the rewards per unit of beta risk or market prices of risk (MPR) associated with each factor.

The implementation of beta pricing models has traditionally relied on the assumption of constant betas and constant MPR. This assumption contradicts the mounting empirical evidence that risk premia vary through time (e.g., Keim and Stambaugh, 1986; Ferson and Harvey, 1991). As an alternative, some researchers have proposed conditional beta pricing models in which the linear relation holds period by period and both factor sensitivities and MPR are allowed to vary through time. A drawback of conditional models is that estimation requires additional assumptions about the dynamics of risk exposures and/or MPR. For instance, Bollerslev

et al. (1988) model conditional covariances as an ARCH process. Harvey (1989) assumes that conditional expected returns are a fixed linear function of a vector of lagged state variables capturing conditioning information. Similarly, Jagannathan and Wang (1996) assume that the conditional market risk premium is linear in one state variable. Ferson and Schadt (1996), Ferson and Harvey (1999), and Lettau and Ludvigson (2001), among others, assume that betas are a fixed linear function of some state variables. More recently, Ang and Chen (2007) assume that conditional betas follow a first-order autoregressive process. To the extent that such assumptions fail to capture the true dynamics of risk premia, the pricing errors of conditional models may be larger than those of unconditional models (Ghysels, 1998; Brandt and Chapman, 2006). In this paper, we propose a new nonparametric procedure to estimate conditional beta pricing models that allows for flexibility in the dynamics of betas and MPR and, therefore, reduces misspecification error. The method we develop in this paper can be seen as an extension of the popular Fama–MacBeth two-pass method (Fama and MacBeth, 1973), originally developed in the context of unconditional models.¹ In the first stage of the Fama–MacBeth method, asset betas are computed for every asset and period using a time series regression with several periods of previous

* Corresponding author. Tel.: +34 93 542 2718; fax: +34 93 542 1746.

E-mail addresses: eva.ferreira@ehu.es (E. Ferreira), javier.gil-bazo@upf.es (J. Gil-Bazo), susan.orbe@ehu.es (S. Orbe).¹ Shanken and Zhou (2007) and Grauer and Janmaat (2009) study the small-sample properties of the two-pass approach and alternative estimation and testing procedures.

data, typically spanning between three and five years. In the second stage, a cross sectional regression of returns on betas is run at every period, which produces a time series of estimated slope coefficients. The constant slope estimator is finally obtained as the sample mean of the corresponding series of estimated slope coefficients. Similarly, we propose to estimate conditional betas nonparametrically for each asset and period using previous information. However, unlike the Fama–MacBeth procedure, conditional betas are assumed to be smooth (but possibly nonlinear) functions of the state variables. In the second stage, time-varying MPR are estimated at each period from the assets' returns and estimated betas (the regressors), but instead of using the cross section of returns and betas of that period only, the method uses the entire sample. More specifically, in the second pass we use a Seemingly Unrelated Regression Equations (SURE) model, introduced by Zellner (1962), with each equation in the system corresponding to one asset. Time-varying slope coefficients (MPR) are treated as free parameters that vary smoothly through time and are estimated nonparametrically subject to the constraint of equality of slopes across assets, allowing for heteroscedastic and cross-sectionally correlated errors. The method, therefore, enables us to estimate time-varying MPR in conditional models under no specific parametric structure.

Although the Fama–MacBeth procedure was derived to estimate and test unconditional asset pricing models, it also yields a time series of factor sensitivities and MPR. Our method exhibits a number of important differences with respect to Fama–MacBeth. First, in our method the weight of observations used in the estimation process is driven by the data, that is, it is determined optimally for each data set rather than established ex-ante by the researcher. Second, although both methodologies allow for time variation in betas, ours is more efficient when betas are believed to be functions of a set of variables capturing the state of the system.² Third, we derive the asymptotic distribution of the time-varying MPR, rather than that of the constant MPR, which enables us to conduct inference on MPR at each point in time and not only on the constant MPR. Fourth, under the assumption that MPR vary smoothly through time, there is a substantial efficiency gain in our estimators of MPR relative to the time series of slope coefficients since, in order to estimate MPR at each point in time, we use the entire sample rather than a single cross section of asset returns and betas. Finally, we assume locally stationary variables as defined in Dalhaus (1997), which permit time-varying mean and, therefore, enable us to drop the usual strong hypothesis of stationarity.

Our work is closely related to that of Stanton (1997), Jones (2006), Wang (2002, 2003), and Lewellen and Nagel (2006). These authors also estimate flexible conditional beta pricing models in different contexts. Stanton (1997) first estimates conditional covariances and conditional expected returns nonparametrically, and then obtains MPR by solving directly the system of equations imposed by the conditional asset pricing model for two assets at each point in time. One problem with this approach is that it can generate highly unstable estimates of the MPR. Furthermore, the method does not enable formal inference to be conducted on MPR. Jones (2006) uses Legendre polynomials to approximate conditional expected returns and betas, which are estimated in a Bayesian framework. He then solves for the parameters of the polynomial for the price of risk that minimize mean squared pricing errors for the whole panel of returns. An advantage of our method is that inference can be conducted on the basis of the closed-form asymptotic distribution of the estimators instead of the numerically obtained posterior distribution of the model parameters. Wang (2002, 2003) proposes a test statistic for the null hypothesis

that conditional expected pricing errors from a conditional asset pricing model are zero. The test is based on the idea that a regression of pricing errors on a vector of instruments should yield zero coefficients. In the models he considers, risk factors are portfolio returns, so pricing errors can be estimated directly as the intercepts from time series regressions of excess returns on the risk factors. In contrast, the method we propose does not require that risk factors be portfolio returns, so it can be applied to models where factors are identified with any aggregate variable. Moreover, while the focus of Wang (2002, 2003) is on model testing, our focus is on the estimation of MPR, which may be used together with estimates of factors sensitivities, to estimate expected returns for the purpose of asset allocation or cost-of-capital computation. Finally, Lewellen and Nagel (2006) have recently used rolling-window regressions to test the conditional CAPM. In particular, they use short windows (ranging from one quarter to one year) and high-frequency returns to estimate both time-varying betas and pricing errors associated with individual portfolios. Then, they test the null hypothesis that pricing errors are zero. Like Wang (2002, 2003), Lewellen and Nagel (2006) consider only models in which risk factors are portfolio returns and the focus of their work is not on the estimation of MPR.

The method proposed in this paper builds on previous econometric research in the context of nonparametric time-varying regression models, that extends the original work by Robinson (1989). Orbe et al. (2005) analyze a single equation regression model under the assumptions of time-varying coefficients with seasonal pattern and locally stationary variables, although neither a two-step procedure nor a multi-equation model is considered. In Orbe et al. (2006) a local constrained least squares estimation method is studied for a single equation regression under the usual assumption of ergodicity. Cai (2007) proposes to estimate a model with time-varying coefficients using local polynomial regression under stationarity of the state variables. Kapetanios (2007) also uses the properties of locally stationary variables to estimate time-varying variances for the error term in the regression model. As mentioned above, in this paper a SURE model is nonparametrically estimated with time-varying coefficients subject to constraints across coefficients corresponding to different equations for each time period. Further, the highest difficulty is related to the fact that, in practice, the explanatory variables (the betas) are not observed and must be estimated in advance. Hence, we deal with generated regressors that have been widely studied by Zellner (1970) or Pagan (1984), among others, for the classical parametric regression model. In order to avoid the inconsistency problems for the coefficient's estimator derived from the potential correlation between the estimated regressor and the error term, conditional betas are estimated at each date using only past information.

To evaluate the performance of the method in practice, we first carry out a Monte Carlo simulation and then apply the method to data on US stock returns. We base both analyses on the Fama and French (1993) three-factor model. More specifically, for the purpose of the simulation study we consider different specifications for the dynamics of beta, all of which assume that beta is a function of observable state variables. Results indicate that the nonparametric estimator clearly outperforms the traditional rolling estimator under all specifications considered. When we apply the method to the 25 Fama–French portfolios sorted on size and book-to-market for the 1963–2005 period, we find that the nonparametric procedure provides a better model fit than the Fama–MacBeth procedure with betas estimated under different parametric specifications: Intercepts of the cross-sectional regressions are closer to zero, MPR are less biased estimates of average factor realizations, and pricing errors are lower than those obtained using the Fama and MacBeth method. Unlike the Fama–MacBeth approach with time-varying parametric betas, the

² For instance, a large amount of empirical evidence on stock return predictability suggests that equity risk premia vary with observable market-wide variables such as the dividend yield or the slope of the term structure of interest rates.

nonparametric procedure leads to lower pricing errors than the unconditional model. We take these results as evidence that misspecification error in parametric estimation rather than time-varying betas and MPR, hurts the performance of the conditional Fama–French model. Finally, while the MPR associated with market or size are not significant under the Fama and MacBeth procedure, the nonparametric MPR associated with these risk factors are significant about one third of the sample period.

The rest of the paper is organized as follows. Section 2 presents the general conditional beta pricing model; Section 3 describes the estimation method and presents the main asymptotic results; Section 4 deals with the practical implementation of the method; Section 5 describes the Monte Carlo simulation and discusses the results; Section 6 contains the empirical application of the method to equity return data; and, finally, Section 7 concludes. The Appendix contains the proofs.

2. The model

In unconditional beta pricing models, asset returns are assumed to be driven by a set of common risk factors

$$R_{it} = \mu_i + \beta_{i1}F_{1t} + \dots + \beta_{ip}F_{pt} + u_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T, \quad (1)$$

where R_{it} denotes the return on asset i from time $t - 1$ to t in excess of the risk-free interest rate and $F_{\ell t}$ denotes the realization of the ℓ th risk factor at time t , for $\ell = 1, \dots, p$. Risk factors are assumed to be orthogonal to each other. Without loss of generality, we assume that factor realizations have zero mean, i.e., $E(F_{\ell t}) = 0$. $\beta_{i\ell}$ represents the sensitivity of asset i 's return to the ℓ th risk factor. The error term u_{it} is serially independent with zero mean and nonsingular covariance matrix, conditional on factor realizations, with variance $\sigma_{u_i}^2 \leq \sigma_u^2$, for some $0 < \sigma_u^2 < \infty$. The sample size of the time series is T , and N is the sample size of the cross section. The standard asset pricing relation is

$$E(R_{it}) = \mu_i = \gamma_1\beta_{i1} + \dots + \gamma_p\beta_{ip}, \quad (2)$$

where $E(R_{it})$ is the expected return on the i th asset and $\beta_{i1}, \dots, \beta_{ip}$ are the coefficients from Eq. (1). The coefficient γ_ℓ , which is equal across assets, is interpreted as the reward (in terms of increase in expected return) per unit of beta risk associated with factor ℓ .

The first stage of the two-pass estimation procedure of Fama and MacBeth (1973), consists of estimating betas in Eq. (1) for each asset and time from a time-series regression. In the second stage, γ 's are estimated as the slope coefficients of a cross-sectional regression of returns on estimated betas. See Shanken (1992) for an analysis of different aspects of the two-pass procedure and a derivation of the asymptotic distribution of the second-pass estimators, and Shanken and Zhou (2007) for a study of the small-sample properties of the methods and a comparison with alternative approaches.

In conditional beta pricing models, such as those studied by Harvey (1989), Jagannathan and Wang (1996) or Lettau and Ludvigson (2001), the asset pricing relation is assumed to hold period by period, unconditional expected returns and betas are replaced by conditional moments, and the rewards per unit of risk are allowed to change over time. Therefore, the conditional beta pricing model is given by

$$R_{it} = E(R_{it}|I_{t-1}) + \beta_{i1}(I_{t-1})F_{1t} + \dots + \beta_{ip}(I_{t-1})F_{pt} + u_{it}, \quad (3)$$

$$E(R_{it}|I_{t-1}) = \gamma_1\beta_{i1}(I_{t-1}) + \dots + \gamma_p\beta_{ip}(I_{t-1}), \quad (4)$$

where I_{t-1} represents investors' information set at time $t - 1$. In empirical applications, the conditioning information set is replaced by an m -dimensional vector of observable state variables $X_{t-1} = (X_{1t-1} \dots X_{mt-1})'$.

If we define $\varepsilon_{it} \equiv R_{it} - E(R_{it}|X_{t-1})$ and denote $\beta_{i\ell}(I_{t-1})$ by $\beta_{i\ell t}$, for $\ell = 1, \dots, p$, then we may write

$$R_{it} = \gamma_1\beta_{i1t} + \dots + \gamma_p\beta_{ipt} + \varepsilon_{it} \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T. \quad (5)$$

It follows from (3) that the errors ε_{it} are heteroscedastic and cross-sectionally related conditional on X_{t-1} , i.e., $E(\varepsilon_{it}\varepsilon_{jt}|X_{t-1}) \neq 0$ for $i \neq j$. We further assume that $F_{\ell t}$ and u_{it} are serially independent for all ℓ, i and t , so $E(\varepsilon_{it}\varepsilon_{js}) = 0$, for all i, j and $t \neq s$.

To estimate γ 's in (5), we form the system of equations

$$\begin{aligned} R_{1t} &= \gamma_1\beta_{11t} + \gamma_2\beta_{12t} + \dots + \gamma_p\beta_{1pt} + \varepsilon_{1t} \\ R_{2t} &= \gamma_1\beta_{21t} + \gamma_2\beta_{22t} + \dots + \gamma_p\beta_{2pt} + \varepsilon_{2t} \\ &\vdots \\ R_{Nt} &= \gamma_1\beta_{N1t} + \gamma_2\beta_{N2t} + \dots + \gamma_p\beta_{Npt} + \varepsilon_{Nt}, \end{aligned} \quad (6)$$

where $\{\gamma_{\ell t}\}_{\ell=1}^p$ are the market prices of risk to be estimated. The error term of the system, $\varepsilon_t = [\varepsilon_{1t}\varepsilon_{2t}\dots\varepsilon_{Nt}]'$ has zero mean and covariance matrix given by

$$E(\varepsilon_t\varepsilon_t'|X_{t-1}) = \Omega_t = \begin{bmatrix} \sigma_{11t} & \sigma_{12t} & \dots & \sigma_{1Nt} \\ \sigma_{21t} & \sigma_{22t} & \dots & \sigma_{2Nt} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1t} & \sigma_{N2t} & \dots & \sigma_{NNt} \end{bmatrix},$$

where $\sigma_{ijt} = E(\varepsilon_{it}\varepsilon_{jt}|X_{t-1})$ denotes the covariance, conditional on the value of the state variables, between the error terms corresponding to different equations i and j at time t . Note that this context allows for heteroscedasticity ($\sigma_{iit} = E(\varepsilon_{it}^2|X_{t-1})$) in each equation and for contemporaneous correlations ($\sigma_{ijt} = E(\varepsilon_{it}\varepsilon_{jt}|X_{t-1})$). As mentioned above, all other correlations are zero.

3. Estimation procedure and main results

This section describes the proposed estimator for the coefficients $\gamma_{\ell t}$ in (6). For a better description of the procedure, consider some extra notation: $B_s = (B_{1s} \ B_{2s} \ \dots \ B_{Ns})'$ is a $N \times p$ -order matrix where each term B_{is} denotes the p -order vector $(\beta_{i1s} \ \beta_{i2s} \ \dots \ \beta_{ips})$, for $i = 1, \dots, N$; R_s is the N -order column vector $(R_{1s}R_{2s}\dots R_{Ns})'$; and $\gamma_t = (\gamma_{1t} \ \dots \ \gamma_{pt})'$ is the p -order vector of prices of risk. Finally, the p -order column vector of state variables is denoted by $X_t = (X_{1t} \ \dots \ X_{mt})'$. According to this notation, model (6) can be compactly written as

$$R_t = B_t\gamma_t + \varepsilon_t \quad t = 1, \dots, T. \quad (7)$$

Within this framework, we propose to estimate the time-varying vector of market prices of risk at each time t , γ_t , taking into account the structure of the errors' covariance matrix, the equality constraints on the coefficients across assets, and the assumed smoothness of the coefficients. In order to achieve this goal, we minimize the weighted sum of squared residuals using all available observations:

$$\min_{\gamma_t} \sum_{s=1}^T K_{h,ts} (R_s - B_s\gamma_t)' \Omega_s^{-1} (R_s - B_s\gamma_t), \quad (8)$$

where $K_{h,ts} = (Th)^{-1}K((t-s)/(Th))$, $K(\cdot)$ denotes the kernel weight used to introduce smoothness in the path of coefficients and $h > 0$ is the bandwidth that regulates the degree of smoothness. Solving the normal equations, the resulting estimator has the following closed form

$$\hat{\gamma}_t = \left(\sum_{s=1}^T K_{h,ts} B_s' \Omega_s^{-1} B_s \right)^{-1} \sum_{s=1}^T K_{h,ts} B_s' \Omega_s^{-1} R_s. \quad (9)$$

With the usual standardization, $R_s^* = V_s^{-1}R_s$ and $B_s^* = V_s^{-1}B_s$, where V_s is the matrix such that $V_s V_s' = \Omega_s$, the optimization problem (8) can be rewritten as

$$\min_{\gamma_t} \sum_{s=1}^T K_{h,ts} (R_s^* - B_s^* \gamma_t)' (R_s^* - B_s^* \gamma_t), \quad (10)$$

which allows us to express the estimator of the market prices of risk (9) in a more compact form

$$\hat{\gamma}_t = \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} R_s^*. \quad (11)$$

Remark 1. For large enough h , the estimator (11) leads to the same estimates as those obtained in classical SURE model estimation with constant coefficients, subject to the equality constraints:

$$\hat{\gamma} = \left(\sum_{s=1}^T B_s^{*'} B_s^* \right)^{-1} \sum_{s=1}^T B_s^{*'} R_s^* = \left(\sum_{s=1}^T B_s' \Omega_s^{-1} B_s \right)^{-1} \sum_{s=1}^T B_s' \Omega_s^{-1} R_s. \quad (12)$$

On the other hand, when h is small enough no smoothness is imposed and the estimator of each γ_t takes into account only the N observations corresponding to the same time period ($s = t$). That is,

$$\hat{\gamma}_t = (B_t^{*'} B_t^*)^{-1} B_t^{*'} R_t^* = (B_t' \Omega_t^{-1} B_t)^{-1} B_t' \Omega_t^{-1} R_t, \quad (13)$$

which is equivalent to estimating γ_t from a cross-sectional regression at time t .

Remark 2. There is a close relation between the estimator in (11) and the estimator proposed in Shanken (1985), and asymptotically studied in Shanken (1992). Considering constant coefficients ($h \rightarrow \infty$) as in (12) and assuming that the betas between returns and the risk factors are time invariant, i.e., $B_s = B \forall s$, the resulting estimator is

$$\hat{\gamma} = \left(\sum_{s=1}^T B^{*'} B^* \right)^{-1} \sum_{s=1}^T B^{*'} R_s^* = (B' \Omega^{-1} B)^{-1} B' \Omega^{-1} \bar{R}, \quad (14)$$

where \bar{R} denotes the sample mean of R_s . Substituting B and Ω by their estimators, respectively, $\hat{\gamma}$ coincides with the Generalized Least Squares (GLS) estimator proposed by Shanken (1985).

In order to analyze the properties of consistency and asymptotic normality of the general estimator (11), we study the mean squared error of $\hat{\gamma}_t$, which we define as the sum of mean squared errors of $\hat{\gamma}_{t\ell}$ for all ℓ , under the assumptions below. We denote this sum by *MSE*:

$$MSE(\hat{\gamma}_t) = \sum_{\ell=1}^p \left(Bias^2(\hat{\gamma}_{t\ell}) + Var(\hat{\gamma}_{t\ell}) \right) \equiv S^2(\hat{\gamma}_t) + V(\hat{\gamma}_t).$$

Assumption A1. The market prices of risk are smooth functions of the time index; that is, $\gamma_{t\ell} = \gamma_{\ell}(t/T)$ where each γ_{ℓ} is a smooth function in $C^2[0, 1]$.

Assumption A2. The weight function $K(u)$ is a symmetric second order kernel with compact support $[-1, 1]$, Lipschitz continuous, and its Fourier transform is absolutely integrable, such that $\int u^2 K^2(u) du$ and $\int K^4(u) du$ are bounded.

Assumption A3. The conditional betas can only vary with time through the state vector at time $t - 1$, X_{t-1} . That is, $\beta_{it} = \beta_{it}(X_{t-1})$, where it is assumed that β_{it} is at least twice differentiable for all partial derivatives.

Assumption A4. Both B_{it} and X_{it} are statistically independent of ε_{is} , for all $s \geq t$. Moreover, we assume the process (7) with finite distributions such that the sequence $\{X_{it}, B_{it}, \varepsilon_{it}\}$ is strong α -mixing with coefficients $\alpha(k)$ of order $6/5$; that is $\alpha(k) = O(k^{-\delta})$, with $\delta > 6/5$. All moments up to order $12 + \theta$ exist and they are uniformly bounded, for some positive θ .

Assumption A5. At each time t , the unconditional expectation $E(B_{it}^* B_{it}^*) = G_t$ is symmetric and strictly positive definite, and it can be decomposed as a smooth function of t/T , at least twice differentiable and uniformly bounded, plus a term of order $O(T^{-1})$.

Assumption A6. The error term ε_t has zero mean conditional on X_{t-1} and conditional covariance matrix $\Omega_t = E(\varepsilon_t \varepsilon_t' | X_{t-1})$, symmetric and positive definite.

Assumption A7. Let σ^{ijt} be a generic term in Ω_t^{-1} . The p -order matrix

$$\left(\sum_{s=1}^T \sum_{ij=1}^N K_{h,ts} \sigma^{ijt} B_{is} B_{js}' \right), \quad (15)$$

is positive definite and uniformly bounded from above and below.

Assumption A8. The smoothing parameter h goes to zero and Th goes to infinity, as the sample size T goes to infinity.

Assumption A1 imposes smoothness on the market prices of risk. **Assumption A2** holds for technical reasons in kernel estimation. **Assumption A3** imposes smoothness on the explanatory variables. **Assumptions A4** and **A5** ensure that the generating distribution process for the data is locally stationary, which allows for time-varying means, variances and also serial correlations. These types of processes are very useful and realistic since they can help model nonstationary variables with a nonexplosive behavior (see Dalhaus, 1997). We also assume smoothness in errors' covariances. **Assumption A6** excludes equations with exploiting variances or with linearly dependent error terms and **Assumption A7** ensures that the estimator is identified. **Assumption A8** is standard in nonparametric estimation.

Theorem 1. Under the set of Assumptions A1–A8, the MSE for the estimator defined in (11), has bias and variance,

$$S^2(\hat{\gamma}_t) = \frac{h^4 d_K^2}{4} \left\| \frac{\partial^2 \gamma_t}{\partial t \partial t} + 2G_t^{-1} \frac{\partial G_t}{\partial t} \frac{\partial \gamma_t}{\partial t} \right\|_2^2 + o(h^4) \quad (16)$$

and

$$V(\hat{\gamma}_t) = \frac{c_K}{Th} \text{tr}(G_t^{-1}) + o((Th)^{-1}) \quad (17)$$

where $G_t = E(B_{it}^* B_{it}^*)$, and the constants related to the kernel, d_K and c_K , are defined as $d_K = \int u^2 K(u) du$ and $c_K = \int K^2(u) du$, respectively.

Remark 3. It is important to observe that under Assumptions A1–A8, the asymptotic order and the leading terms are the same considering either stationary or locally stationary variables.

Corollary 1. Consider model (7) and a consistent estimator $\widehat{\Omega}_s = \widehat{V}_s \widehat{V}'_s$ of $\Omega_s = V_s V'_s$. Then, under the same assumptions of Theorem 1, and if either

- (i) $\widehat{\Omega}_s - \Omega_s = o(MSE(\hat{\gamma}_t))$, or
- (ii) the entries in B_s are bounded,

the Feasible Generalized Least Squares (FGLS) estimator

$$\hat{\gamma}_t^{FGLS} = \left(\sum_{s=1}^T K_{h,ts} B'_s \widehat{\Omega}_s^{-1} B_s \right)^{-1} \sum_{s=1}^T K_{h,ts} B'_s \widehat{\Omega}_s^{-1} R_s$$

has the same asymptotic properties as the estimator (9).

All previous asymptotic results are obtained under the assumption that the explanatory variables are observable and, therefore, they can be used directly in the estimation. However, this is not the case in the context of beta pricing models, in which explanatory variables are not directly observable and must be replaced by proxies. Moreover, the procedure to obtain them should ensure that the properties of the true unobserved variables are preserved.

Taking into account the regression Eq. (1), we propose the estimator

$$\widehat{\beta}_{it}(X_{t-1}) = \left(\sum_{s=t-r}^{t-1} K_H(X_{t-1} - X_s) F_s F'_s \right)^{-1} \sum_{s=t-r}^{t-1} K_H(X_{t-1} - X_s) F_s R_{is}, \quad (18)$$

where we recall that $X_s = (X_{1s}, \dots, X_{ms})'$ denotes the vector of state variables. K_H is a m -variate kernel $K_H(u) = |H|^{-1/2} K(H^{-1/2}u)$, with smoothing matrix H and $F_s = (F_{1s}, \dots, F_{ps})'$. That is, (18) can be seen as a one-sided conditional nonparametric estimator in a time series model. To avoid inconsistency in the estimation of MPR, it is crucial that we employ a truncated estimator of β_{it} that only uses past information.

Thus, the resulting Smoothed Generalized Least Squares (SGLS) estimator for the market prices of risk,

$$\hat{\gamma}_t^{SGLS} = \left(\sum_{s=1}^T K_{h,ts} \widehat{B}'_s \widehat{B}_s \right)^{-1} \sum_{s=1}^T K_{h,ts} \widehat{B}'_s R_s^*, \quad (19)$$

is similar to (11) with B_s replaced by $\widehat{B}_s^* = V_s^{-1} \widehat{B}_s$. In order to reach the desirable asymptotic results some additional assumptions are required:

Assumption B1. The m -variate kernel K_H is compactly supported such that $\int K(u)du = 1$ and $\int uu'K(u)du = \mu_K J_p$, where μ_K is a nonnegative scalar and \int and du , are shorthands for $\int \dots \int_{\mathbb{R}^p}$ and $du_1 \dots du_p$, respectively.

Assumption B2. Consider a sequence of positively definite diagonal bandwidth matrices $H = \text{diag}(h_{1i}^2, h_{2i}^2, \dots, h_{mi}^2)$ for $i = 1, \dots, N$, such that $|H|^{1/2}$ and r/T go to zero, $T|H|^{1/2}$ and $r|H|^{1/2}$ go to infinity as the subsample size (r) and the sample size (T) go to infinity. Note that the bandwidth matrices are considered to be equal for all i , to simplify notation and without loss of generality.

Assumption B3. The distribution of X_t has a one-order-Lipschitz time-varying density, $f_t(x) = f(\tau, x)$, where $\tau = t/T$.

Assumption B4. Assume that $E(F_{it}^2 | X_{t-1} = x)$ and $E(F_{it}^4 | X_{t-1} = x)$ exist and are smooth enough, and $E(F_{it} F_{\ell t} | X_{t-1} = x) = 0$ for $\ell \neq i$.

The following proposition states the properties of estimator (18).

Proposition 1. Consider the set of Assumptions A3–A6, and Assumptions B1–B4 then, the estimator defined by (18) is a consistent estimator of $\beta_i(X_{t-1})$, with asymptotic bias and variance:

$$\begin{aligned} \text{Bias}(\widehat{\beta}_i(X_{t-1}) | X_{t-1} = x) &= O(\text{trace}(H)) \\ \text{Var}(\widehat{\beta}_i(X_{t-1}) | X_{t-1} = x) &= O\left(\frac{1}{r|H|^{1/2}}\right). \end{aligned}$$

We are now in a position to derive the asymptotic results for the estimator of the market prices of risk when the estimators of the conditional betas described above are employed.

Theorem 2. Under the set of Assumptions A1–A8 and B1–B4, using the estimator (\widehat{B}_t) whose elements are defined in (18), the SGLS estimator for the market prices of risk defined in (19) is consistent, with the same asymptotic results for the two components of the MSE as in Theorem 1.

Corollary 2. Consider model (7) with the consistent estimator of B_t defined in (18) and a consistent estimator $\widehat{\Omega}_s = \widehat{V}_s \widehat{V}'_s$ for $\Omega_s = V_s V'_s$. Then, under the assumptions in Theorem 2, and if either

- (i) $\widehat{\Omega}_s - \Omega_s = o(MSE(\hat{\gamma}_t^{SGLS}))$, or
- (ii) the entries in B_s are bounded,

the Smoothed Feasible Generalized Least Squares (SFGLS) estimator

$$\hat{\gamma}_t^{SFGLS} = \left(\sum_{s=1}^T K_{h,ts} \widehat{B}'_s \widehat{\Omega}_s^{-1} \widehat{B}_s \right)^{-1} \sum_{s=1}^T K_{h,ts} \widehat{B}'_s \widehat{\Omega}_s^{-1} R_s \quad (20)$$

has the same asymptotic properties as in the previous theorems.

The following proposition provides a consistent estimator for the error covariance matrix that must be estimated in advance in order to compute the estimated market prices of risk defined in (20).

Proposition 2. Consider the estimator for a generic element of the covariance matrix,

$$\begin{aligned} \hat{\sigma}_{ijt} &= \left(\sum_{s=t-r}^{t-1} K_{\mathcal{H}}(X_{t-1} - X_s) \right)^{-1} \text{um}_{s=t-r}^{t-1} K_{\mathcal{H}}(X_{t-1} - X_s) \\ &\quad \times (R_{is} - \hat{\gamma}_s \hat{\beta}_{is})(R_{js} - \hat{\gamma}_s \hat{\beta}_{js}) \end{aligned} \quad (21)$$

with $K_{\mathcal{H}}(u) = |\mathcal{H}|^{-1/2} K(\mathcal{H}^{-1/2}u)$, being \mathcal{H} the m -order smoothing matrix and $K_{\mathcal{H}}$ a m -variate second order kernel. Under same assumptions than in Proposition 1 and the kernel $K_{\mathcal{H}}$ satisfying Assumptions B1 and B2, (21) provides a consistent estimator for a generic term of Ω_t , for each t .

The following (pointwise) asymptotic distribution for the estimator of γ_t , allows us to test for invariance of the prices of the risk factors through time or to test whether the price of risk factors is different from zero.

Theorem 3. Assume Assumptions A1–A8 and B1–B4, consider $h = o(T^{-1/5})$, such that the bias tends to zero faster than the variance, and that either (i) or (ii) in Corollary 2 holds. Then, the SGLS estimator of γ_t at k different locations t_1, \dots, t_k converges in distribution to the multivariate normal as,

$$\left((Th)^{1/2} \left(\hat{\gamma}_{t_j}^{SGLS} - \gamma_{t_j} \right) \right)_{j=1}^k \xrightarrow{D} N\left(0, C_K G_t^{-1}\right). \quad (22)$$

Finally, using the consistent estimator for G_{t_j} defined in Lemma 1 (in the Appendix), we can obtain confidence intervals for the k selected γ 's.

4. Implementation

The proposed estimator for the SURE model with unknown explanatory variables requires the selection of several smoothing parameters: the matrix of bandwidths used to estimate conditional betas between risk factors and asset returns, H ; the smoothing parameter used to estimate time-varying market prices of risk, h ; and the matrix of bandwidths used to estimate the residuals' conditional covariance matrix, \mathcal{H} .

In general situations, the bandwidths are selected using data-driven methods like cross-validation, penalized sum of squared residuals or plug-in methods. For a detailed discussion of each see Härdle (1990), Wand and Jones (1995) or Fan and Gijbels (1996), among others. For multivariate cases, the penalty methods, such as Rice or Generalized Cross-Validation, are appropriate, easy to interpret and faster to compute than the others.

To solve the selection problem in this specific context, we proceed in two steps. In the first step, we address the selection of H and h jointly. In the second step, we select the smoothing parameter matrix \mathcal{H} . For the first step, we propose to minimize a penalized sum of squared residuals

$$(NT)^{-1} \sum_{t=1}^T (R_t - \hat{B}_t(H)\hat{\gamma}_t(h))' (R_t - \hat{B}_t(H)\hat{\gamma}_t(h)) \mathcal{G}(h, H), \quad (23)$$

where $\mathcal{G}(h, H)$ denotes the penalizing function. It is well known that a sum of squared residuals equal to zero is easily obtained for bandwidths very close to zero. Note, however, that since the estimator of B_t defined in (18) does not include the observation at time t , there is no need to penalize the selection of H . Thus, we will use a function that only penalizes low values of h , i.e., $\mathcal{G}(h, H) = \mathcal{G}(h)$. In particular, in the Generalized Cross Validation method, the penalty is

$$\mathcal{G}(h) \approx (1 - (NT)^{-1} \text{trace}P(h))^{-2}, \quad (24)$$

where $P(h)$ is the projection matrix

$$\frac{K(0)}{Th} \sum_{i=1}^N \sum_{t=1}^T [B'_{is} (B'_{i\mathcal{K}_{h,t}} B_i)^{-1} B'_i] Z_t, \quad (25)$$

where $B_{is} = (\beta_{i1t} \beta_{i2t} \dots \beta_{ipt})'$ is defined above, the $T \times p$ -order matrix $B_i = (B_{i1} B_{i2} \dots B_{iT})'$ is the data matrix corresponding to the i th equation, $\mathcal{K}_{h,t} = \text{diag}\{K_{h,ts}\}_{s=1}^T$ is a T -order diagonal matrix with kernel weights, and Z_t is a T order column vector with t th element equal to one and rest of elements equal to zero.

Once H and h have been selected, the second step is to select the smoothing parameter matrix \mathcal{H} for the errors' covariance matrix. In particular, for fixed B and h , we estimate MPR and obtain the

model's errors. Then, we select \mathcal{H} that minimizes the weighted sum of squared residuals

$$\sum_{t=1}^T (R_t - \hat{B}_t\hat{\gamma}_t)' \hat{\Omega}_t^{-1}(\mathcal{H}) (R_t - \hat{B}_t\hat{\gamma}_t) \quad (26)$$

where the estimator of any generic term of $\hat{\Omega}$, $\hat{\sigma}_{ijt}$, is given by (21).

The $\hat{\gamma}$'s estimated with the selected \mathcal{H} are then used again to obtain the model residuals, and reestimate conditional betas, the residual covariance matrix and the new $\hat{\gamma}$'s. For this reason, it is possible that the smoothing parameters selected in the first step are not optimal. This problem suggests the need to iterate in order to attempt to converge to final $\hat{\gamma}$'s. However, changing H and h in this iterative procedure does not necessary provide a convergent method. Therefore, we cannot ensure optimality of all smoothing parameters.

5. Monte Carlo study

As explained above, the nonparametric estimator of factor sensitivities is more efficient than the traditional Fama–MacBeth rolling estimator when betas are believed to be functions of a set of variables capturing the state of the system. Another advantage of the method is the fact that the weight of past observations used in the estimation process is optimally determined for each data set rather than established ex-ante by the researcher. It is interesting to study whether these features lead to more accurate estimates of betas in small samples for two reasons. First, estimated betas are important by themselves. For instance, they are necessary inputs in risk management problems. Second, more accurate betas can improve the estimation of market prices of risk in the second-pass estimation. To evaluate the ability of the nonparametric approach to capture the dynamics of time-varying betas relative to that of the more traditional rolling estimator, we conduct a Monte Carlo simulation study. In particular, we focus on the conditional version of the three-factor model of Fama and French (FF) (1993) in which betas are allowed to vary through time:

$$R_{it} = \beta_{im,t} R_{mt} + \beta_{ismb,t} R_{smbt} + \beta_{ihml,t} R_{hmlt} + u_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T, \quad (27)$$

where R_m , R_{smb} and R_{hml} denote the returns on the market portfolio (in excess of the risk-free rate), on the size factor-mimicking portfolio (SMB) and on the book-to-market factor-mimicking portfolio (HML), respectively. To simplify the analysis, we set $N = 1$ and omit the asset subscript.

The purpose of the analysis is to investigate whether the method enables us to capture different dynamics of betas. Time

Table 1
Monte Carlo simulation results.

	% (MSE ^{NP} < MSE ^{ROLL})			MSE ^{NP}			MSE ^{ROLL}		
	$\hat{\beta}_m$ (1)	$\hat{\beta}_{smb}$ (2)	$\hat{\beta}_{hml}$ (3)	$\hat{\beta}_m$ (4)	$\hat{\beta}_{smb}$ (5)	$\hat{\beta}_{hml}$ (6)	$\hat{\beta}_m$ (7)	$\hat{\beta}_{smb}$ (8)	$\hat{\beta}_{hml}$ (9)
Model 1	97.4	95.3	91.6	0.0135	0.0213	0.0297	0.0298	0.0444	0.0532
Model 2	84.9	94.2	90.7	0.0182	0.0212	0.0305	0.0278	0.0434	0.0530
Model 3	82.5	97.8	96.0	0.0275	0.0250	0.0250	0.0356	0.0435	0.0534
Model 4	81.5	98.0	94.8	0.0221	0.0166	0.0270	0.0300	0.0429	0.0535
Model 5	75.3	89.7	76.2	0.0209	0.0315	0.0456	0.0280	0.0548	0.0624
Model 6	58.8	92.3	73.2	0.0323	0.0395	0.0563	0.0352	0.0636	0.0709

Note: This table reports results from a Monte Carlo simulation study of the nonparametric (NP) and the rolling (ROLL) estimators of betas in a conditional Fama–French three factor model. Coefficients are generated according to different functions. In Model 1 the market beta is a linear univariate function of the dividend yield. In Model 2 a quadratic univariate function of the dividend yield is assumed for the market beta. A quadratic bivariate function of the dividend yield and the default spread is considered for the market beta in Model 3. In Model 4 the market beta is an exponential bivariate function of the dividend yield and the default spread and remaining coefficients are constant. In Models 1–4 betas with respect to the size factor and book-to-market factor are constant. In Model 5 all betas are generated as quadratic univariate functions of the dividend yield. Finally, in Model 6 all betas are quadratic bivariate functions of the dividend yield and the default spread. For each model, Columns (1)–(3) report the percentage of simulations for which Mean Squared Errors (MSE) ($MSE(\hat{\beta}_i)$ $i = m, smb, hml$) are lower for the nonparametric estimator than for the rolling estimator. Columns (4)–(6) report MSE averaged across all simulations for the nonparametric estimator. Columns (7)–(9) reports MSE averaged across all simulations for the rolling estimator.

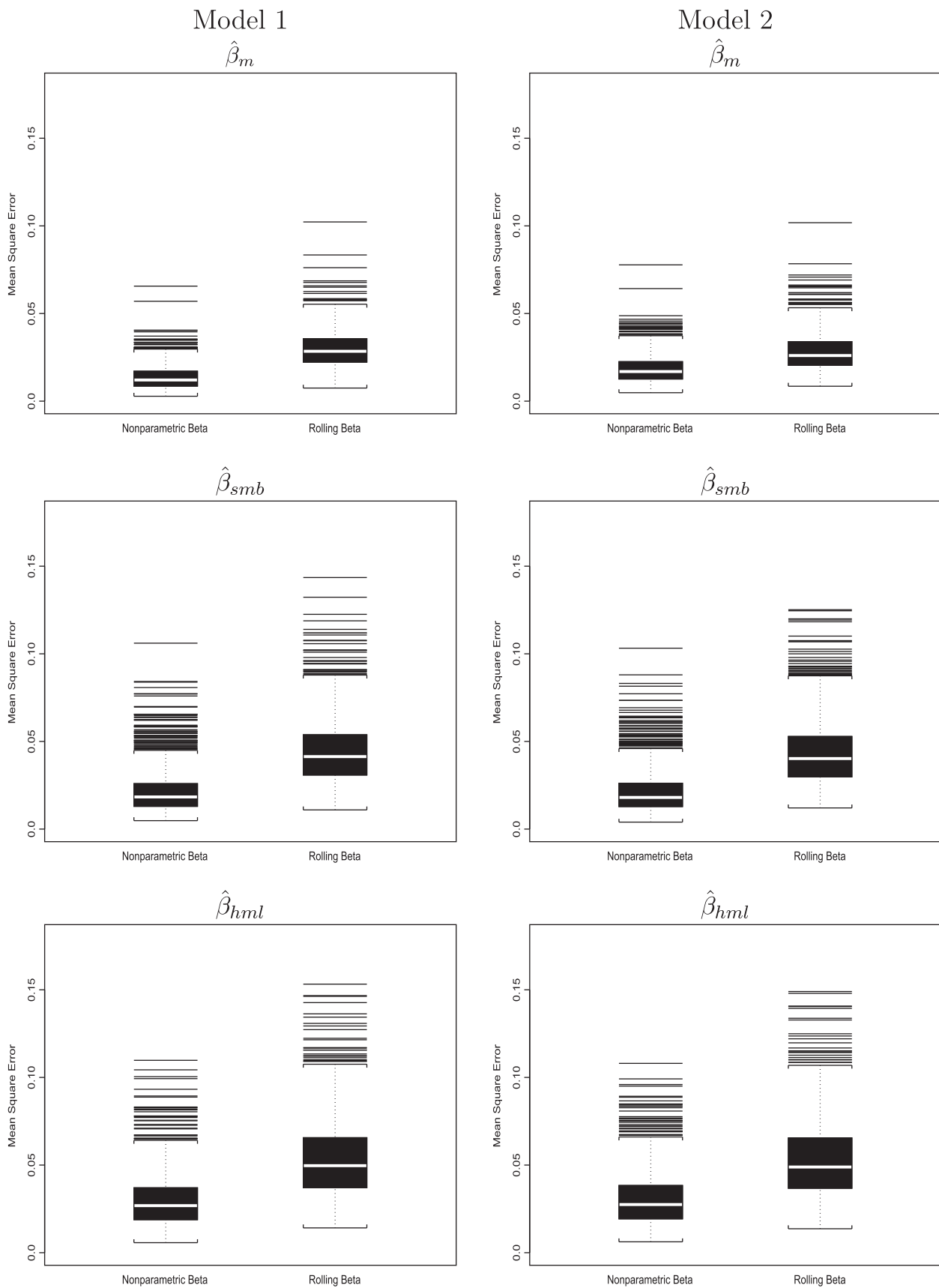


Fig. 1. Monte Carlo study: Boxplots of Mean Squared Errors of nonparametric and rolling beta estimators for the three coefficients considered in the model. The market beta is generated as a linear univariate function of the dividend yield in Model 1 and a quadratic univariate function of the dividend yield in Model 2. The remaining coefficients are constant in time and equal in both models. The market beta is generated as a quadratic bivariate function of the dividend yield and the default spread in Model 3 while in Model 4 it is an exponential bivariate function of the dividend yield. The remaining coefficients are constant in time and equal in both models. In Model 5 all betas are generated as quadratic univariate functions of the dividend yield. In Model 6 betas are quadratic bivariate functions of the dividend yield and the default spread.

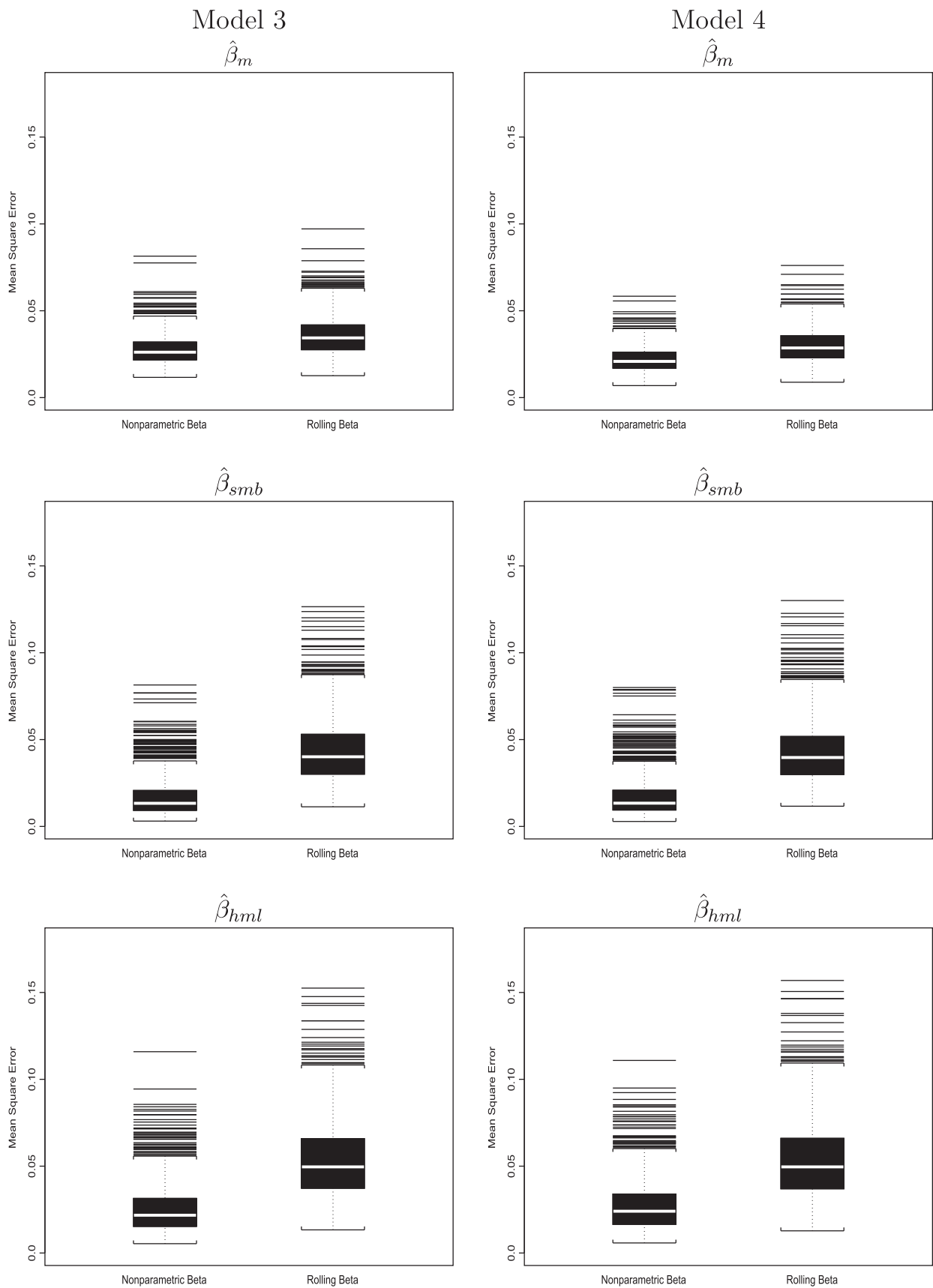


Fig. 1 (continued)

variation is modeled as a function of lagged values of two state variables, denoted by X_1 and X_2 . More specifically, we employ

two variables that are commonly used in the literature: the dividend yield on the market portfolio, computed as the sum of

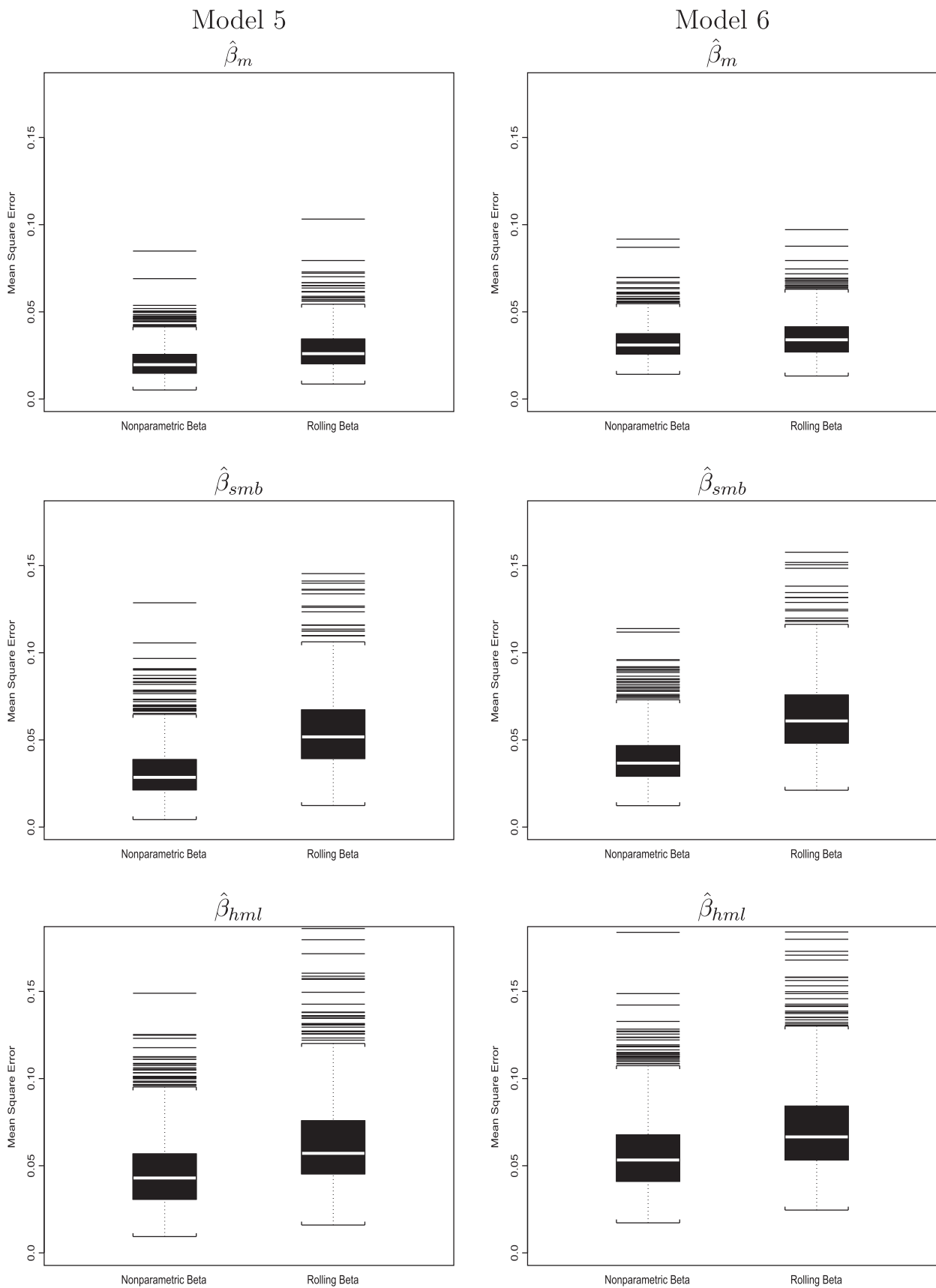


Fig. 1 (continued)

dividends on the S&P 500 index in the last 12 months divided by the index level at the end of the year, and the default spread as

proxied by the difference between the average rates of Moody's Baa- and Aaa-rated corporate debt. The data used to compute the

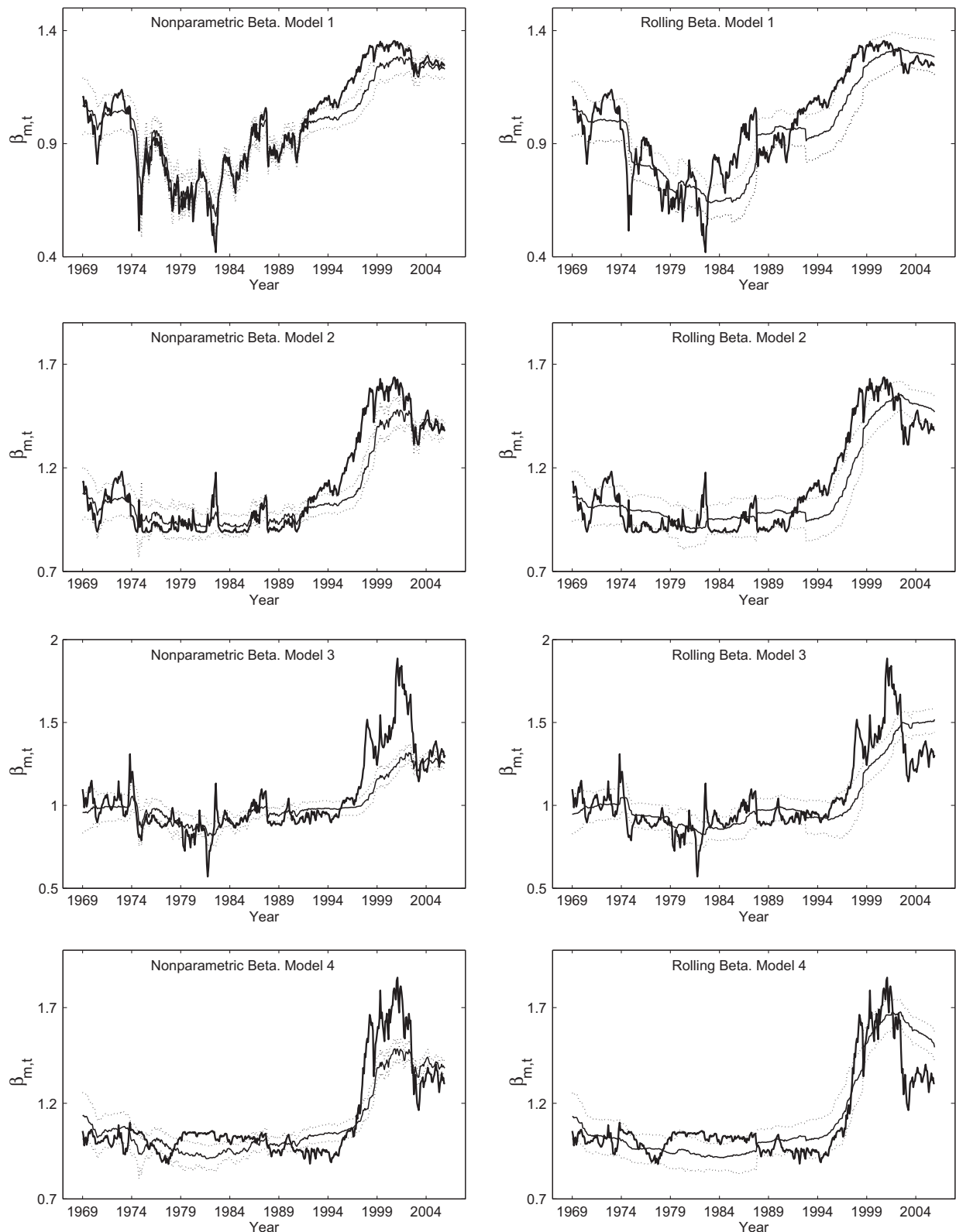


Fig. 2. Monte Carlo study: Nonparametric versus rolling beta estimators. Each graph displays the generated market beta, $\beta_{m,t}$ (thick solid line), the median estimated beta (thin solid line), and the first and third quartiles of estimated betas (dotted lines). Market betas are generated using four different specifications: a linear univariate function of the dividend yield (Model 1); a quadratic univariate function of the dividend yield (Model 2); a quadratic bivariate function of the dividend yield and the default spread (Model 3); and an exponential bivariate function of the dividend yield and the default spread (Model 4). The other two betas are held constant.

dividend yield and the default spread are obtained from CRSP and the Federal Reserve Economic Data database, respectively.

To simulate returns, we start by generating market betas, $\beta_{m,t}$, for $t = 1 \dots T$, using data on the lagged state variables in the January

1964–December 2005 period and according to six different specifications. In the first four models, the asset's beta with respect to the market factor is time varying while betas with respect to SMB and HML factors are held constant. More specifically,

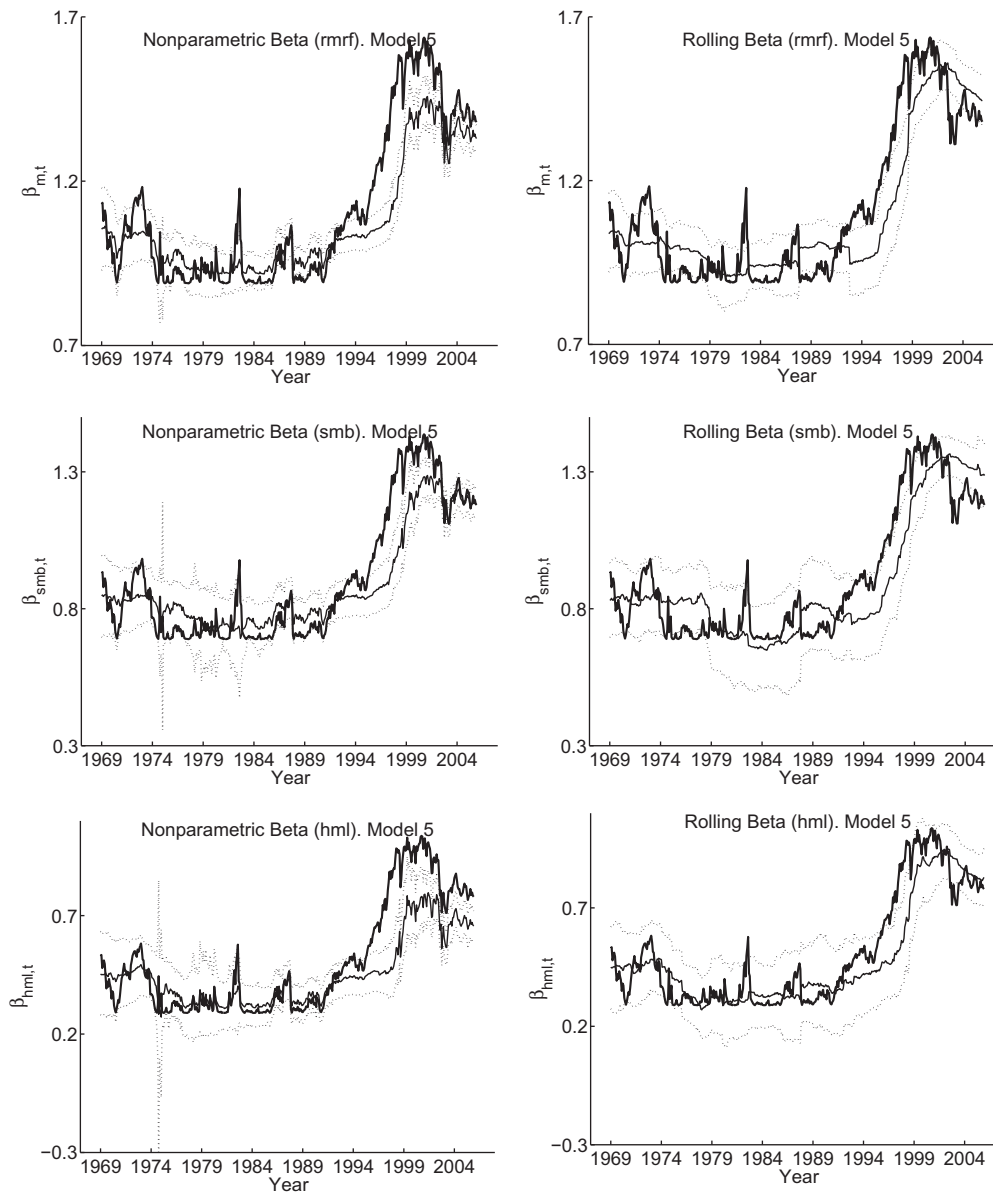


Fig. 3. Monte Carlo study: Nonparametric versus rolling beta estimators. The graphics display the generated coefficient, ($\beta_{m,t}$, $\beta_{smb,t}$ or $\beta_{hml,t}$) (thick solid line), the median estimated beta (thin solid line), and the first and third quartiles of estimated betas (dotted lines). All betas correspond to Model 5 and are simulated as quadratic univariate functions of the dividend yield.

market beta is simulated according to the following functions, respectively:

$$\beta_{m,t} = 1 - 20X_{1,t-1} \tag{28}$$

$$\beta_{m,t} = 1 - 20X_{1,t-1} + 900X_{1,t-1}^2 \tag{29}$$

$$\beta_{m,t} = 1 - 20X_{1,t-1} + 900X_{1,t-1}^2 + 30X_{2,t-1} - 200X_{2,t-1}^2 - 2500X_{1,t-1}X_{2,t-1} \tag{30}$$

$$\beta_{m,t} = 1 - 8X_{1,t-1}e^{-100X_{1,t-1}} - 20X_{2,t-1}e^{-100X_{2,t-1}}, \tag{31}$$

Henceforth, we refer to each one of the four models corresponding to (28)–(31) as models 1, 2, 3, and 4, respectively.³ In all cases, we set $\beta_{smb} = 0.8$ and $\beta_{hml} = 0.4$.

³ The alternative specifications for the dynamics of betas attempt to capture a rich set of forms of dependence on the state variables. We choose quadratic functions, which are a second-order Taylor series approximation to nonlinear functions. To allow for higher-order effects, we also consider an exponential function. Coefficients are chosen to yield realistic values of betas.

In Model 5 and Model 6 all betas vary conditional on the state variables. In particular, in Model 5 the coefficients are quadratic functions of the dividend yield:

$$\beta_{m,t} = 1 - 20X_{1,t-1} + 900X_{1,t-1}^2 \tag{32}$$

$$\beta_{smb,t} = 0.8 - 20X_{1,t-1} + 900X_{1,t-1}^2 \tag{33}$$

$$\beta_{hml,t} = 0.4 - 20X_{1,t-1} + 900X_{1,t-1}^2. \tag{34}$$

Finally, in Model 6 all the coefficients are quadratic bivariate functions of the dividend yield and the default spread:

$$\beta_{m,t} = 1 - 20X_{1,t-1} + 900X_{1,t-1}^2 + 30X_{2,t-1} - 200X_{2,t-1}^2 - 2500X_{1,t-1}X_{2,t-1} \tag{35}$$

$$\beta_{smb,t} = 0.8 - 20X_{1,t-1} + 900X_{1,t-1}^2 + 30X_{2,t-1} - 200X_{2,t-1}^2 - 2500X_{1,t-1}X_{2,t-1} \tag{36}$$

$$\beta_{hml,t} = 0.4 - 20X_{1,t-1} + 900X_{1,t-1}^2 + 30X_{2,t-1} - 200X_{2,t-1}^2 - 2500X_{1,t-1}X_{2,t-1}. \tag{37}$$

Using the series of time-varying betas, we then simulate 1,000 paths, indexed by j , of asset return realizations, $R_t^j = \beta_{m,t}R_{mt} + \beta_{smb,t}R_{smb,t} + \beta_{hml,t}R_{hml,t} + \epsilon_t^j$, for $t = 1, \dots, T$. Realizations of the three Fama–French factors have been downloaded from Kenneth French's website. Random errors, ϵ_t^j , are drawn independently from $N(0, \sigma = 0.04)$.

We estimate the vector of conditional betas for each simulated path using (18) and we compare the accuracy of the nonparametric estimator with that of the rolling betas, i.e., betas estimated in overlapping rolling samples of sixty months of prior data.

Finally, we compute the mean squared error of each series of estimated betas as well as the average mean squared error for all simulated paths within each model. That is, we compute for each coefficient (i) and simulation (j), $MSE(\hat{\beta}_i^j) = (1/T) \sum_{t=1}^T (\hat{\beta}_{it}^j - \beta_{it})^2$ $i = m, smb, hml$ and sum over the simulations for the average $\overline{MSE}(\hat{\beta}_i) = (1/1000) \sum_{j=1}^{1000} MSE(\hat{\beta}_i^j)$.

Simulation results are presented in Table 1. Columns (1)–(3) report for each model and each beta the percentage of simulated paths for which the MSE is lower for the nonparametric estimator than for the rolling estimator. Columns (4)–(9) report mean squared errors averaged across all simulations for each model and coefficient for both estimators. Results in Table 1 indicate that the nonparametric estimator of factor loadings is more accurate than the rolling estimator for all models. When true β_{smb} and β_{hml} are constant over the sample (Models 1–4), the percentage of simulations for which the MSE is lower for the nonparametric estimator of β_m than the MSE for the rolling estimator ranges from 81.5% (Model 4) to 97.4% (Model 1). Inspection of columns (4)–(9) indicates that the average MSE is also lower for the nonparametric estimator for the three betas. Importantly, the nonparametric estimator of the two betas that are held constant in the simulations also performs better than the rolling estimator. When all betas are simulated as functions of the state variables (Models 5 and 6), the

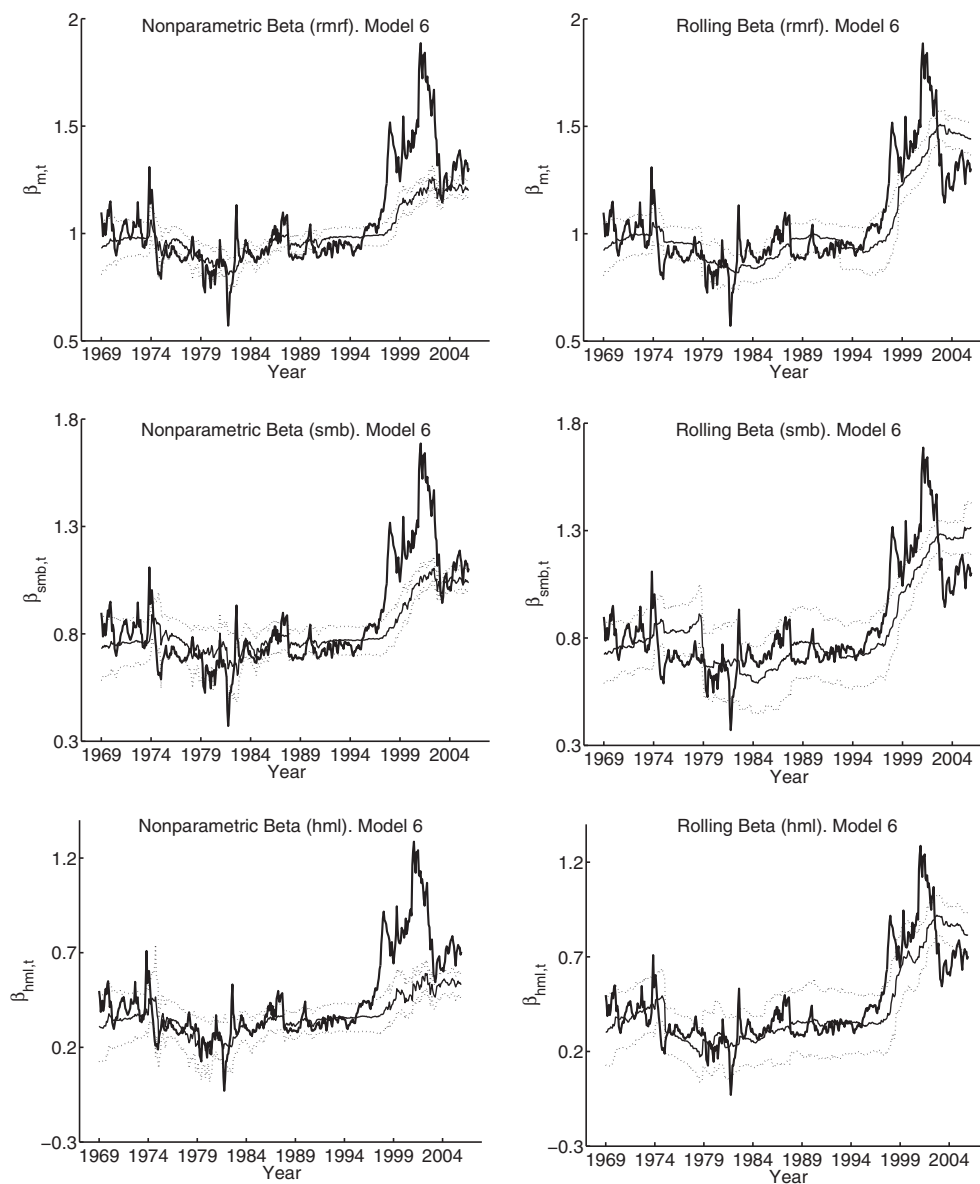


Fig. 4. Monte Carlo study: Nonparametric versus rolling beta estimators. The graphics display the generated coefficient, ($\beta_{m,t}$, $\beta_{smb,t}$ or $\beta_{hml,t}$) (thick solid line), the median estimated beta (thin solid line), and the first and third quartiles of estimated betas (dotted lines). All betas correspond to Model 6 and are simulated as quadratic bivariate functions of the dividend yield and the default spread.

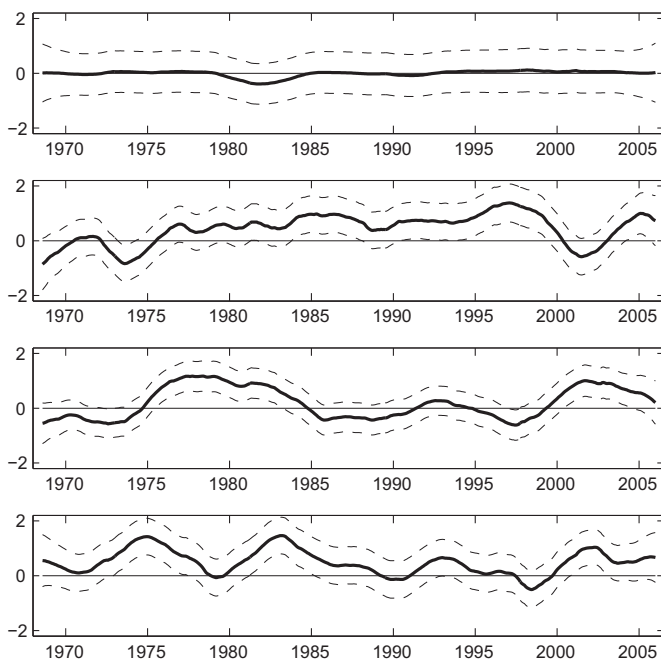


Fig. 5. Intercept and market prices of risk obtained from nonparametric cross-sectional regressions of monthly excess returns on the 25 Fama–French size and book-to-market portfolios and the three factor-mimicking portfolios (market, size, and book-to-market) on betas between asset returns and the risk factors. Betas are estimated nonparametrically as a function of the lagged values of the dividend yield, “DP,” and the one-month Treasury bill yield, “Tb1m.” The dashed lines represent 95% confidence bands. Monthly returns are expressed in % units.

nonparametric estimator still outperforms the rolling estimator, although the gap in accuracy between the two decreases relative to Models 1–4.

Fig. 1 displays boxplots of the empirical distribution of MSE for each model and both estimators. More specifically, boxplots show graphically the median MSE, as well as the first quartile (q_1) and third quartile (q_3), the limits $q_3 + w(q_3 - q_1)$ and $q_1 - w(q_3 - q_1)$ with $w = 1.5$, and values outside those limits. The figure shows that the first quartile, the median MSE and the third quartile are all lower for the nonparametric estimator than for the rolling estimator in all cases. One of the most striking differences is achieved in Model 1 for the market beta: The third quartile of the empirical MSE distribution is lower for the nonparametric estimator than the first quartile for the rolling estimator. We may therefore conclude that under the specifications considered, the nonparametric estimator clearly outperforms the rolling estimator in terms of providing more accurate estimates of betas.

To gain further insight on the performance of the nonparametric estimator relative to the rolling estimator, in Figs. 2–4 we plot the median, the first and the third quartiles of estimated betas for both estimators under the six specifications together with the true betas. The figures show that the nonparametric estimator performs remarkably well under the six specifications, especially in the pre-1990 period. The median estimated beta tracks closely the true beta and the interquartile range is narrow. The rolling estimator, in contrast, appears to respond slower to changes in true beta. Also the interquartile range is substantially wider than that of the nonparametric estimator in all cases (see Fig. 5).

6. Empirical application

In this section we apply the non-parametric method presented above to estimate conditional betas and MPR in a flexible conditional version of the Fama and French (1993) three-factor model.

We compare estimation results to those obtained when applying the Fama–MacBeth procedure under alternative parametric specifications of beta dynamics.

6.1. Model and data

We consider a particular case of the asset pricing relation (4) where the marketwide factors are the three Fama–French factors described in the previous section:

$$E(R_{it}|X_{t-1}) = \gamma_{m,t}\beta_{im,t} + \gamma_{smb,t}\beta_{ismb,t} + \gamma_{hml,t}\beta_{ihml,t} \quad (38)$$

$$i = 1, 2, \dots, N \quad t = 1, 2, \dots, T.$$

To estimate conditional betas at time t , we require at least 60 months of prior data on portfolio returns, factor realizations, and conditioning variables. This results in the loss of the 60 initial observations in the estimation of MPR. To capture mispricing, we estimate the above equation with an intercept.

We follow closely Ferson and Harvey (1999) and select five conditioning variables that have been used in the literature on stock return predictability: (1) the annual dividend yield of the S&P 500 index (“DP”); (2) the slope of the term structure (“term”) as proxied by the difference between the yield on the ten-year Treasury bond and the yield on a one-year Treasury bill; (3) the default spread (“def”); (4) the one-month Treasury bill yield (“Tb1m”); and (5) the difference between the monthly returns of a three-month and a one-month Treasury bill (“hb3”).⁴ “DP,” “Tb1m,” and “hb3” are constructed from data obtained from CRSP. Data on “term” and “def” are obtained from the Federal Reserve’s FRED database.

The test assets in our cross-sectional regressions are the 25 equity portfolios formed by sorting individual stocks on market capitalization and book-to-market (Fama and French, 1993). Since risk factors in the Fama–French model are returns on traded portfolios, we follow the advice in Lewellen et al. (2010) and include the risk factors as test assets to be priced by the model. Monthly data on the 25 portfolio returns and the one-month risk-free rate were downloaded from Kenneth French’s website. Our final sample contains 510 monthly observations of factor realizations, portfolio returns and lagged state variables in the July 1963–December 2005 period.

To mitigate the effects of the well-known curse of dimensionality that affects non-parametric estimation, we use only two conditioning variables at a time when estimating betas nonparametrically, although we report results for all ten possible combinations. One possible way of establishing ex-ante which pair of instruments to select is to run time-series regressions of returns on the risk factors allowing betas to depend linearly on two instruments, and then select the pair of instruments that results in the highest R^2 . Although the linear specification is only a first-order approximation to the true dynamics of betas, this simple approach will lead to the right choice of instruments if second-order effects on R^2 coefficients are similar across pairs of instruments. In particular, for each portfolio, we run the time series regression:

$$R_{it} = a_{0i} + (\beta_{m0i} + \beta'_{m1i}X_{t-1})R_{m,t} + (\beta_{smb0i} + \beta'_{smb1i}X_{t-1})R_{smb,t} + (\beta_{hml0i} + \beta'_{hml1i}X_{t-1})R_{hml,t} + \epsilon_{it} \quad (39)$$

where X_{t-1} denotes the vector of lagged values of the instruments. Table 2 reports the adjusted R^2 coefficients for each pair of instruments and for each portfolio. It also shows adjusted R^2 coefficients for the unconditional model in which betas are constant parameters

⁴ All of these variables have been shown to be good predictors of future stock returns (e.g., Keim and Stambaugh, 1986; Ferson and Harvey, 1991). Given their predictive power, these variables have been extensively used in the literature as state variables in conditional asset pricing models (e.g., Ferson and Schadt, 1996; Wu, 2002).

Table 2
Time series regressions with linear betas.

Portfolio	Instruments											
	None	DP, term	DP, def	DP, Tb1m	DP, hb3	term, def	term, Tb1m	term, hb3	Tb1m, def	Tb1m, hb3	def, hb3	ALL
S1/B1	92.14	92.18	92.27	92.34	92.11	92.22	92.26	92.19	92.23	92.19	92.20	92.41
S1/B2	94.34	94.50	94.44	94.49	94.40	94.45	94.49	94.41	94.46	94.35	94.42	94.60
S1/B3	95.01	95.18	95.15	95.15	95.18	95.08	95.13	95.04	95.05	95.05	95.05	95.19
S1/B4	94.18	94.37	94.38	94.39	94.47	94.21	94.27	94.27	94.22	94.30	94.29	94.45
S1/B5	94.31	94.74	94.61	94.83	94.59	94.55	94.51	94.36	94.61	94.45	94.48	94.80
S2/B1	95.27	95.46	95.50	95.42	95.47	95.42	95.44	95.39	95.39	95.35	95.39	95.59
S2/B2	93.92	94.89	94.52	94.73	94.49	94.53	94.60	94.47	94.06	93.86	94.02	94.93
S2/B3	93.34	94.47	94.33	94.60	94.31	93.83	93.97	93.75	93.57	93.36	93.52	94.81
S2/B4	93.53	94.24	94.13	94.27	94.13	93.89	94.02	93.82	93.76	93.64	93.62	94.42
S2/B5	94.33	94.69	94.59	94.63	94.58	94.62	94.67	94.58	94.59	94.51	94.41	94.78
S3/B1	94.91	94.92	94.95	94.90	94.87	94.99	94.94	94.92	94.93	94.89	94.93	94.95
S3/B2	90.39	91.70	91.47	91.83	91.44	91.05	91.10	91.13	90.59	90.37	90.55	92.09
S3/B3	89.20	91.20	90.94	91.33	90.90	90.18	90.31	89.99	89.71	89.27	89.59	91.69
S3/B4	89.57	91.50	91.15	91.48	91.00	90.75	90.81	90.52	90.18	89.65	90.01	91.81
S3/B5	89.44	90.53	90.39	90.61	90.45	89.83	89.74	89.84	89.66	89.64	89.71	90.75
S4/B1	93.63	93.70	93.81	93.71	93.73	93.69	93.62	93.61	93.73	93.66	93.67	93.83
S4/B2	88.37	90.56	90.46	90.63	90.48	89.50	90.02	88.96	89.24	88.92	89.09	91.08
S4/B3	87.78	90.08	89.34	89.78	89.29	89.29	89.45	89.12	88.32	87.80	88.25	90.31
S4/B4	88.29	89.13	88.67	88.79	88.45	89.09	89.22	88.95	88.42	88.32	88.32	89.55
S4/B5	85.94	86.67	86.60	86.64	86.43	86.25	86.27	86.12	86.01	85.83	85.97	86.89
S5/B1	93.47	93.85	93.88	94.25	93.74	93.73	93.65	93.60	93.69	93.57	93.60	94.44
S5/B2	89.78	90.97	90.99	91.07	91.17	90.29	90.74	90.06	90.44	90.36	90.49	91.30
S5/B3	84.64	85.48	85.54	85.85	85.54	85.47	85.53	84.97	85.52	85.34	85.27	86.01
S5/B4	87.72	88.37	88.07	88.19	88.03	88.23	88.22	88.21	87.75	87.65	87.71	88.47
S5/B5	79.12	79.54	79.55	79.49	79.70	79.64	79.48	79.80	79.62	79.97	79.72	80.10
Mean	90.90	91.72	91.59	91.74	91.56	91.39	91.46	91.28	91.19	91.05	91.13	91.97

Note: This table reports adjusted R^2 statistics (in %) from time-series regressions of excess returns of each of the 25 Fama–French portfolios on the three Fama–French factors and the products of the factors with a set of lagged instrumental variables. The instrumental variables are “hb3,” the difference between three-month and one-month T-bill returns; “DP,” the annual dividend yield of the S&P 500 index; “def,” the default spread; “term,” the spread between the 10-year and three-month Treasury yields; “Tb1m,” the one-month Treasury bill yield. The sample is July 1963 to December 2005 and the number of observations is 510. Returns on 25 value-weighted portfolios formed on size and the ratio of book value to market value are measured in excess of the return on a 30-day Treasury bill. S1 denotes the lowest 20% of market capitalization, S5 denotes the largest 20%, B1 denotes the lowest 20% of the book/market ratios and B5 is the highest 20%.

and for a conditional model that uses all five instruments. The results suggest that there exists a gain in term of adjusted R^2 from allowing betas to vary linearly with the instruments. If only two instruments are considered, the highest adjusted R^2 coefficients correspond to pairs of instruments that include the dividend yield. In particular, the combination of “DP” with “Tb1m” gives the highest adjusted R^2 averaged across portfolios, 91.74%, which is very close to the average adjusted R^2 obtained with the five instruments, 91.97%.

We compare estimation results for the nonparametric approach to those of the Fama–MacBeth approach for four parametric versions of the FF model considered by Ferson and Harvey (1999), each one of which corresponds to a different way of estimating betas: (1) as constant parameters estimated using an expanding sample⁵; (2) as constant parameters estimated using rolling samples; (3) as linear functions of the lagged state variables using an expanding sample; and (4) as linear functions of the lagged state variables using rolling samples. Model (1) produces betas that are almost constant. Since Fama–MacBeth estimates of MPR are also constant, model (1) can be interpreted as the unconditional model. We consider rolling samples of two different lengths: 36 and 60 months. Once betas have been estimated, we run OLS cross-sectional regressions of returns on betas for every month in the sample starting in July 1968:

$$R_{it} = \gamma_{0t} + \gamma_{mt} \hat{\beta}_{im,t} + \gamma_{smbt} \hat{\beta}_{ismb,t} + \gamma_{hmlt} \hat{\beta}_{ihml,t} + \varepsilon_{it}, \quad (40)$$

which produces a time series of intercepts and slope coefficients. Fama–MacBeth coefficients are then estimated as the sample averages of the monthly estimates. We also report Fama–MacBeth stan-

⁵ The expanding sample used to estimate betas at time t includes all observations from the first month to month $t - 1$.

dard errors corrected for the errors-in-variable problem as proposed by Shanken (1992).

We also compute the GLS-weighted sum of squared pricing errors, which is the basis of several tests of asset pricing models proposed in the literature, such as Shanken's (1985) cross-sectional T^2 test statistic⁶:

$$T\alpha' \hat{\Omega}^{-1} \alpha, \quad (41)$$

where α is the N -dimensional vector of pricing errors defined as the difference between the vector of sample mean returns and the vector of unconditional mean returns implied by the model (\hat{R}). For the parametric specifications, each element in \hat{R} is computed as the product of mean betas and the constant Fama–MacBeth MPR coefficients (excluding the intercept). For the nonparametric procedure, \hat{R} is the sample mean of the product of nonparametric betas and estimated time-varying MPR. In all cases, we use the sample covariance matrix of asset returns as an estimator of $\hat{\Omega}$.

To conclude, we compare the performance of the parametric and nonparametric procedures in terms of: (i) estimated intercepts in the cross-sectional regression, which under the null hypothesis that the Fama–French model holds should be zero; (ii) the bias of each estimated market price of risk with respect to the mean realization of the corresponding factor; (iii) the statistical significance of the MPR⁷; and (iv) the GLS-weighted sum of squared unconditional pricing errors.

⁶ See Balvers and Huang (2009) for a theoretical study of Shanken's test statistic and its relation to alternative model evaluation criteria.

⁷ As pointed out by Lozano and Rubio (2010), an estimator of market prices of risk has desirable properties if it satisfies the following three conditions: (1) it has low standard error, (2) it is statistically different from zero, and (3) it has low bias relative to the observable factor.

Table 3
Fama–MacBeth cross-sectional regressions with parametric betas.

Betas	γ_0			γ_m			Pricing errors
	Coeff.	s.e.	t-stat	Coeff.	s.e.	t-stat	
Constant betas within expanding sample	0.04	0.02	2.03	0.41	0.22	1.85	
Constant betas 5-year rolling sample	0.13	0.03	4.62	0.31	0.21	1.43	
Constant betas 3-year rolling sample	0.14	0.03	4.09	0.31	0.21	1.45	
Linear betas within expanding sample	0.11	0.05	2.12	0.35	0.20	1.76	
Linear betas 5-year rolling sample	0.30	0.09	3.53	0.21	0.17	1.20	
Linear betas 3-year rolling sample	0.41	0.12	3.45	0.13	0.14	0.88	
Mean	0.19	0.06	3.31	0.28	0.19	1.43	

Betas	γ_{smb}			γ_{hml}			Pricing errors
	Coeff.	s.e.	t-stat	Coeff.	s.e.	t-stat	
Constant betas within expanding sample	0.11	0.16	0.69	0.50	0.15	3.21	84.93
Constant betas 5-year rolling sample	0.13	0.16	0.81	0.45	0.15	3.00	96.19
Constant betas 3-year rolling sample	0.12	0.16	0.79	0.42	0.15	2.91	100.94
Linear betas within expanding sample	0.10	0.15	0.68	0.43	0.14	3.10	94.12
Linear betas 5-year rolling sample	0.03	0.13	0.24	0.35	0.12	2.89	104.49
Linear betas 3-year rolling sample	0.00	0.11	0.03	0.27	0.11	2.57	113.55
Mean	0.08	0.14	0.54	0.40	0.14	2.95	99.04

Note: This table reports Fama–MacBeth coefficients, their associated Shanken-corrected standard errors (s.e.), and *t*-statistics (*t*-stat) estimated from cross-sectional regressions of monthly excess returns on the 25 Fama–French size and book-to-market portfolios and the three factor-mimicking portfolios (market, size, and book-to-market) on betas between asset returns and risk factors. Betas are estimated as constant parameters or as fixed linear functions of the lagged values of the five instruments in either an expanding or a rolling sample. “Pricing errors” is the GLS-weighted sum of squared pricing errors, where pricing errors are calculated as the difference between each of the test asset’s sample mean excess return and the mean excess return implied by the model. Monthly returns are expressed in % units.

6.2. Estimation results for the Fama–MacBeth approach

Table 3 contains estimation results for the parametric specifications described above. The large values of the Fama–MacBeth estimates of γ_0 , which range from 4 bp to 41 bp per month, can be interpreted as evidence against the parametric versions of the Fama–French model. In particular, all intercepts are statistically significant at the 5% level. Interestingly, allowing for more variation in betas not only fails to improve the model’s fit, but also leads to larger mispricing: All versions of time-varying betas have higher values of $\hat{\gamma}_0$ than the constant betas, expanding sample version. The second lower intercept corresponds to the model that uses an expanding sample but allows betas to vary linearly with the conditioning variables. Rolling samples lead to higher intercepts, which are statistically significant at the 1% level. The worst model fit corresponds to the use of shorter windows, especially when rolling windows are combined with betas that are linear in the instruments.

Fama–MacBeth estimates of γ_m range from 0.13% to 0.41% per month and are below the actual mean value of the market excess return in the sample period (0.43%). The price associated with the market risk factor is statistically significant only for the unconditional model and only marginally so. Consistently with the results for the intercepts, betas estimated with rolling samples as linear functions of the instruments yield lower values of $\hat{\gamma}_m$ (larger bias). Estimates of γ_{smb} range from 0% to 0.13%, below the mean value of SMB, 0.15%. None of them is statistically significant. Estimated values of the HML factor, which range from 0.27% to 0.50%, are in all cases but one, lower than the actual mean factor realization of 0.47%. Allowing for more variation in estimated betas (linear specification, rolling samples) leads to lower values of $\hat{\gamma}_{hml}$. However, in contrast to parametric estimates of the other MPR, all estimates of γ_{hml} are statistically significant at conventional significance levels.

Inspection of the last column in Table 3 containing the GLS-weighted sum of squared pricing errors confirms the conclusion that allowing for time variation in betas leads to larger pricing errors than the unconditional model in the context of the Fama–MacBeth procedure with parametrically estimated betas. The largest mispricing corresponds to the case in which betas are estimated as linear functions of the state variables using short rolling windows.

To summarize the results of Table 3, conditional parametric versions of the Fama–French model result in large and significant intercepts, biased estimates of MPR, insignificant MPR associated with the market and the size factors, and larger pricing errors than the unconditional model.

6.3. Estimation results for the nonparametric approach

In Table 4 we report estimation results for the nonparametric method. The table shows the values of $\hat{\gamma}_{mt}$, $\hat{\gamma}_{smbt}$, and $\hat{\gamma}_{hmlt}$ averaged across the whole sample, together with their average standard errors, and their average *t*-statistics. We also report the fraction of months in which estimated MPR are positive, and both positive and statistically significant at the 5% level. As in the parametric case, we estimate an intercept (which is allowed to vary with time) and include the risk factors as test assets. The results of Table 3 are clearly more favorable to the conditional Fama–French model than those from its parametric versions. For all combinations of the predictive variables, absolute values of the intercepts are lower on average than the intercepts from all the parametric implementations of the model reported in Table 2 excluding the unconditional model. Average estimated values of γ_m are very close to the mean value of the excess return on the market portfolio (0.42% on average versus 0.43%). The market factor is positively priced between 76% and 80% of the time and the price is statistically significant between 30% and 40% of the time. Therefore, we may reject the null hypothesis of a constant and zero price for the market risk factor. In other words, ignoring the price of market risk (as the results from the parametric methods suggest) would have been inadequate about one third of the time. The estimated prices of size risk are between 0.19% and 0.22% per month, and therefore slightly larger than the mean value of the return on the SMB portfolio (0.15%). The price of size risk is statistically significant between 31.1% and 33.5% of the time, despite the lack of significance of the price of this risk factor obtained with the Fama–MacBeth procedure. Finally, the estimated price of the HML risk factor exhibits no bias on average with respect to its realization, and is positive about 85% of the time and statistically significant between 27% and 30% of all months.

It is worth remarking that the average standard errors of the nonparametric estimates of MPR reported in Table 4 are larger

Table 4
Nonparametric cross-sectional regressions with nonparametric betas.

Instruments	γ_0					γ_m					Pricing errors
	Mean coeff.	Mean s.e.	Mean t-stat	% Pos.	% Sign.	Mean coeff.	Mean s.e.	Mean t-stat	% Pos.	% Pos and sign.	
DP, term	0.02	0.39	0.04	71.78	0.00	0.43	0.34	1.27	79.11	38.44	
DP, def	0.05	0.39	0.13	86.67	0.00	0.36	0.34	1.05	76.67	33.33	
DP, Tb1m	-0.01	0.40	-0.03	64.67	0.00	0.44	0.35	1.27	78.67	39.56	
DP, hb3	0.00	0.39	-0.02	63.33	0.00	0.41	0.34	1.22	78.89	39.33	
term, def	0.04	0.40	0.10	78.22	0.00	0.40	0.35	1.17	78.44	32.00	
term, Tb1m	-0.01	0.40	-0.03	55.11	0.00	0.46	0.35	1.32	79.56	38.89	
term, hb3	-0.02	0.39	-0.05	53.33	0.00	0.45	0.35	1.31	80.00	36.44	
Tb1m, def	0.00	0.40	-0.01	67.56	0.00	0.43	0.35	1.21	77.56	37.33	
Tb1m, hb3	-0.03	0.40	-0.08	54.44	0.00	0.45	0.35	1.31	79.11	40.89	
def, hb3	0.02	0.39	0.05	75.11	0.00	0.40	0.35	1.16	79.11	30.22	
Mean	0.01	0.40	0.01	67.02	0.00	0.42	0.35	1.23	78.71	36.64	

Instruments	γ_{smb}					γ_{hml}					Pricing errors
	Mean coeff.	Mean s.e.	Mean t-stat	% Pos.	% Sign.	Mean coeff.	Mean s.e.	Mean t-stat	% Pos.	% Pos and sign.	
DP, term	0.19	0.28	0.66	53.56	31.78	0.45	0.34	1.33	86.22	28.00	82.01
DP, def	0.17	0.28	0.59	53.33	31.11	0.43	0.34	1.26	81.11	26.89	81.02
DP, Tb1m	0.19	0.28	0.65	53.56	31.78	0.49	0.35	1.44	87.11	29.11	79.59
DP, hb3	0.21	0.28	0.73	53.56	32.89	0.46	0.34	1.36	84.67	28.89	81.00
term, def	0.19	0.28	0.65	53.33	31.78	0.44	0.35	1.29	82.44	27.11	86.95
term, Tb1m	0.20	0.28	0.70	54.44	31.78	0.50	0.35	1.43	86.67	30.00	84.08
term, hb3	0.22	0.28	0.81	54.22	33.56	0.48	0.34	1.41	86.44	30.44	86.84
Tb1m, def	0.18	0.28	0.62	53.56	31.33	0.49	0.35	1.43	85.11	30.22	82.11
Tb1m, hb3	0.21	0.28	0.76	54.00	32.89	0.49	0.35	1.43	87.33	28.67	82.71
def, hb3	0.19	0.28	0.68	53.11	32.44	0.43	0.34	1.30	81.56	29.78	84.62
Mean	0.19	0.28	0.69	53.67	32.13	0.47	0.34	1.37	84.87	28.91	83.09

Note: This table reports results from nonparametrically estimated cross-sectional regressions of monthly excess returns on the 25 Fama–French size and book-to-market portfolios and the three factor-mimicking portfolios (market, size, and book-to-market) on betas between asset returns and risk factors. Betas are estimated nonparametrically as a function of the lagged values of the instruments. s.e. denotes the asymptotic standard error of the estimated coefficient, t-stat denotes the t-statistic, % pos. denotes the fraction of months in which the corresponding estimated coefficient is positive, % sign. denotes the fraction of months in which the corresponding estimated coefficient is statistically significant at the 5% level, % pos. and sign. denotes the fraction of months in which the corresponding estimated coefficient is both positive and statistically significant at the 5% level. “Pricing errors” is the GLS-weighted sum of squared pricing errors, where pricing errors are calculated as the difference between each of the test asset’s sample mean excess return and the mean excess return implied by the model. Monthly returns are expressed in % units.

than the standard errors of the Fama–MacBeth estimates reported in Table 3. However, it should be noted that mean standard errors are not the standard errors of the mean but mean values of pointwise estimates’ standard errors. Like any local estimator, our nonparametric estimator of time-varying MPR at any given point in time is necessarily less precise than a global estimator, such as the Fama–MacBeth estimator of constant MPR. What is important is the fact that despite the higher standard errors of the nonparametric MPR estimators, we are able to reject the null of zero constant MPR for all three risk factors, whereas the parametric approach with constant MPR suggests that only the risk associated with HML is positively and significantly priced by investors.

The last column of Table 4 shows that for all pairs of predictive variables but two, pricing errors are smaller than those obtained for the unconditional model as reported in Table 3. Moreover, all pricing errors obtained with the nonparametric approach are lower than pricing errors obtained with the parametric specification with time-varying betas. We take these results as evidence that misspecification error in parametric estimation rather than time-varying betas and MPR, hurts the performance of the conditional Fama–French model. It is also interesting to note that there exists certain correspondence between pricing errors in Table 4 and adjusted R² coefficients reported in Table 2. The instrument that leads to the best fit in time-series regressions with linear betas, “DP,” also leads to the smallest pricing errors. Furthermore, the lowest pricing errors in Table 4 and the highest adjusted R² in Table 2 are achieved with the same pair of predictors: “DP” and “Tb1m”.

In addition to producing lower pricing errors, the nonparametric procedure proposed in this paper enables us to investigate time-variation in gammas. In Fig. 3, we display the values of nonparametrically estimated MPR, as well as 95% confidence intervals,

for the pair of predictive variables that leads to the lowest pricing errors: “DP” and “Tb1m”. An inspection of the top panel, displaying the evolution of $\hat{\gamma}_0$ suggests that this measure of mispricing has been historically low (in absolute value) on average with the exception of the 1981–1982 period, in which the intercept reached -0.39%. The figure also reveals that all market prices of risk have varied substantially through time. Consistently with the results in Table 4, compensation for bearing market risk has been positive most of the time in the 1968–2005 period. Although it is often difficult to reject the null of a zero γ_m in a given point in time, the price of risk associated with the market risk factor was statistically significant in the mid-eighties, in the 1995–2000 period, and towards the end of the sample. Another interesting insight revealed by Fig. 3 is the fact that the price of risk associated with the size factor has changed signs several times during the sample period. This is consistent with the evidence reported by van Dijk (2011) that the small-size effect has experienced several reversals through time. Such reversals can explain why tests based on constant MPR, such as Fama–MacBeth, cannot reject the null of a zero price for the size risk factor. However, assuming that size risk bears no price would have been inadequate in the late seventies or around 2002, when our estimate of the market price of size risk was highly significant and reached levels of 1.2% per month. Finally, the price associated with book-to-market risk appears to have experienced substantial variation through time with maximum values reaching 1.5% per month and minimum values of -0.5%.

6.4. Fama–MacBeth with nonparametric betas

The results presented above indicate that the nonparametric approach leads to lower pricing errors than the Fama–MacBeth

procedure. The superiority of our approach could be the result of nonparametrically estimated betas being less biased and more precise than parametrically estimated betas. Moreover, as Fama and French (1997) point out, better estimates of betas improve the estimates of MPR in the second pass. However, the better performance of the nonparametric approach could also be attributable to the procedure employed to estimate MPR itself. First, the nonparametric procedure uses more information than the monthly cross-sectional regressions of the Fama–MacBeth procedure, which can explain the smaller bias of MPR estimates of the nonparametric procedure. Moreover, even if the Fama–MacBeth constant slope coefficients were unbiased estimates of average MPR, we would still expect the nonparametric procedure to result in smaller pricing errors if the true model is conditional. The reason is that mean unconditional expected returns are not equal to the product of mean betas and mean MPR if betas and MPR are both time-varying and correlated, as noted by Jagannathan and Wang (1996).

To disentangle the effect of nonparametric betas from that of nonparametric MPR on the ability of the conditional model to fit the data, we apply the Fama–MacBeth procedure to nonparametrically estimated betas. Results, reported in Table 5, suggest that nonparametric betas alone represent a large improvement upon parametrically estimated betas in terms of providing less biased estimates of MPR and lower pricing errors. First, all intercepts are lower than intercepts obtained when parametric betas are used with the exception of the unconditional model. Second, all estimates of γ_m are higher (and closer to the average factor realizations) than all Fama–MacBeth estimates, with the same exception as above. Third, all estimates of γ_{hml} are closer to the average factor realization than estimates obtained with parametric betas. Finally,

the GLS-weighted sum of squared pricing errors is lower for all pairs of instruments than for any parametric betas excluding the unconditional model. Importantly, betas estimated as nonparametric functions of “DP” and “Tb1m” result in lower pricing errors than the unconditional model. On the other hand, the model fit of the Fama–MacBeth approach with nonparametric betas is somewhat worse than the nonparametric procedure with nonparametric betas. Taken together, these findings indicate that the better performance of the nonparametric approach proposed in this paper can be attributed to a large extent, but not completely, to the superiority of nonparametric betas over parametric betas.

7. Summary and conclusions

In this paper we show how to estimate consistently time-varying market prices of risk in a general conditional beta pricing model without imposing any parametric structure on the dynamics of factor sensitivities or market prices of risk. The method can be seen as a nonparametric analogue of the two-pass approach developed by Fama and MacBeth (1973) to estimate and test unconditional beta pricing models.

Like previously proposed nonparametric estimation methods, the method presented in this paper is not subject to Ghysels's (1998) critique that misspecification of time-varying conditional moments and market prices of risk may induce larger pricing errors than those obtained by unconditional beta pricing models. Unlike previous proposals, however, ours does not assume that risk factors can be identified with portfolio returns, so it can be applied to a more general family of models. Moreover, our method provides estimates of both factor sensitivities and market prices of risk, which can be used to estimate expected returns for the purposes of forecasting future returns or estimating the cost of capital.

To evaluate the performance of the method in the empirical analysis, we first carry out a simulation study and then apply the method to data on US equity returns. Both analysis are based on the conditional Fama–French three-factor model. Simulation results suggest that the nonparametric methodology provides more accurate estimates of conditional betas than the traditional rolling-sample approach when beta depends on observable state variables. Estimation results using data on the 25 size and book-to-market sorted portfolios suggest that inference based on constant market prices of risk may hide the fact that risk factors are significantly priced in specific subperiods. Further, the nonparametric version of the Fama–French model does a better job at delivering unbiased estimates of MPR and low pricing errors than the Fama–MacBeth method. Moreover, the nonparametric procedure yields in most cases lower pricing errors than the unconditional model, unlike the Fama–MacBeth method with time-varying parametric betas. Given its ability to overcome many of the difficulties associated with the use of parametric methods, we believe that the nonparametric procedure proposed in this paper has great potential in empirical applications that require the estimation of factor sensitivities and MPR in conditional beta pricing models.

Acknowledgments

The authors thank Ike Mathur (the editor), and an anonymous referee, as well as Cesare Robotti, Francisco Peñaranda, Eduardo Schwartz, Alfredo Ibáñez, Alvaro Cartea, Martin Seim, and seminar participants at the 2007 Cemapre Conference on Advances in Semi-parametric Methods and Applications, 2007 International Conference on Applied Stochastic Models and Data Analysis, 2008 European Economic Association and Econometric Society Meetings, 2010 FMA European Conference, 2010 EFMA Conference, 2010 FMA Conference, and Basque Centre of Applied Mathematics

Table 5
Fama–MacBeth cross-sectional regressions with nonparametric betas.

Instruments	γ_0			γ_m		
	Coeff.	s.e.	t-stat	Coeff.	s.e.	t-stat
DP, term	0.07	0.02	2.89	0.38	0.22	1.73
DP, def	0.07	0.03	2.93	0.36	0.22	1.66
DP, Tb1m	0.07	0.02	3.16	0.37	0.22	1.69
DP, hb3	0.08	0.03	2.90	0.36	0.22	1.66
Term, def	0.09	0.03	3.30	0.38	0.22	1.76
Term, Tb1m	0.07	0.03	2.78	0.39	0.22	1.80
Term, hb3	0.06	0.03	2.50	0.39	0.21	1.82
Tb1m, def	0.09	0.03	3.56	0.38	0.22	1.72
Tb1m, hb3	0.07	0.03	2.73	0.37	0.21	1.75
def, hb3	0.08	0.03	3.09	0.36	0.21	1.70
Mean	0.08	0.03	2.98	0.38	0.22	1.73

Instruments	γ_{smb}			γ_{hml}			Pricing errors
	Coeff.	s.e.	t-stat	Coeff.	s.e.	t-stat	
DP, term	0.11	0.16	0.67	0.47	0.15	3.09	85.32
DP, def	0.12	0.16	0.78	0.47	0.15	3.12	85.56
DP, Tb1m	0.11	0.16	0.72	0.47	0.15	3.09	84.61
DP, hb3	0.12	0.16	0.78	0.47	0.15	3.12	85.35
Term, def	0.10	0.16	0.65	0.45	0.15	2.97	88.77
Term, Tb1m	0.10	0.16	0.63	0.46	0.15	3.04	86.00
Term, hb3	0.11	0.16	0.72	0.47	0.15	3.06	87.54
Tb1m, def	0.10	0.16	0.65	0.45	0.15	3.01	86.98
Tb1m, hb3	0.12	0.16	0.73	0.46	0.15	3.04	86.14
def, hb3	0.13	0.16	0.80	0.46	0.15	3.03	88.53
Mean	0.11	0.16	0.71	0.46	0.15	3.06	86.48

Note: This table reports Fama–MacBeth coefficients and their associated Shanken-corrected standard errors (s.e.) and t-statistics (t-stat) estimated from cross-sectional regressions of monthly excess returns on the 25 Fama–French size and book-to-market portfolios and the three factor-mimicking portfolios (market, size, and book-to-market) on betas between asset returns and risk factors. Betas are estimated nonparametrically as a function of the lagged values of the instruments. “Pricing errors” is the GLS-weighted sum of squared pricing errors, calculated as the difference between each of the test asset’s sample mean excess return and the mean excess return implied by the model. Monthly returns are expressed in % units.

for their comments and suggestions. All errors are the authors' responsibility. This work has been supported by the Spanish Government Grants ECO2008-00777/ECON and ECO2010-17158, by the Basque Government Grant Grupo MACLAB (IT-241-07) and by the Madrid Government Grant CCG10-UC3M/HUM-5611. A previous version of this paper was circulated under the title "Nonparametric estimation of conditional beta pricing models."

Appendix A

In order to prove Theorem 1 the following lemma are needed.

Lemma 1. Under Assumptions A2–A5 and A8, it holds that

$$\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* \xrightarrow{a.s.} G_t,$$

$$(Th) \sum_{s=1}^T K_{h,ts}^2 B_s^{*'} B_s^* \xrightarrow{a.s.} c_K G_t.$$

Proof of Lemma 1. To simplify notation we denote any generic scalar term of $(Th)K_{h,ts}B_s^{*'}B_s^*$ as $Z_{ijs} = (Th)K_{h,ts}\beta_{is}^*\beta_{js}^*$. Z_{ijs} is a α -mixing sequence of size $6/5$ with the proper bounded moments, and $E(Z_{ijs}) = K\left(\frac{t-s}{Th}\right)g_{ijs}$, that tends to $g_{ij}(t/T)$. Therefore, the first result follows from the Strong Law of Large Numbers in White (1984), Corollary 3.48, for dependent variables under mixing conditions. The second result can be proven following similar steps. \square

Proof of Theorem 1. First we write the Mean Squared Error

$$MSE(\hat{\gamma}_t) = trE[(\hat{\gamma}_t - \gamma_t)(\hat{\gamma}_t - \gamma_t)'] = \|Bias(\hat{\gamma}_t)\|_2^2 + trVar(\hat{\gamma}_t)$$

$$= S^2(\hat{\gamma}_t) + V(\hat{\gamma}_t).$$

Then, note that

$$\hat{\gamma}_t - \gamma_t = \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^*\right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} R_s^* - \gamma_t$$

$$= \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^*\right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* (\gamma_s - \gamma_t)$$

$$+ \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^*\right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s$$

has a random denominator. We overcome this problem redefining the bias and variance terms using the weight $W_t^* = G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^*$. Hence, the redefined bias is $Bias^*(\hat{\gamma}_t) = Bias(W_t^* \hat{\gamma}_t)$. For technical reasons, we use different bandwidths for W_t^* and for $\hat{\gamma}_t$, say h^* and h respectively, such that the following condition holds:

$$\frac{E\|W_t^* - I\|^2}{E\|\hat{\gamma}_t - \gamma_t\|^2} = o(1), \tag{42}$$

as T goes to infinity. This condition establishes that W_t^* goes to the identity at a faster rate than the mean squared error goes to zero, and this implies that the rate of convergence for the mean squared error must be suboptimal, which in this case means slower than $T^{-4/5}$.

Considering the term defined by $Bias^*$

$$Bias^*(\hat{\gamma}_t) = G_t^{-1} \sum_{s=1}^T K_{h,ts} E(B_s^{*'} B_s^*) (\gamma_s - \gamma_t) + G_t^{-1} \sum_{s=1}^T K_{h,ts} E(B_s^{*'} \varepsilon_s^*)$$

$$= G_t^{-1} \sum_{s=1}^T K_{h,ts} E(B_s^{*'} B_s^*) (\gamma_s - \gamma_t),$$

since $E(B_s^{*'} \varepsilon_s^*) = E(B_s^{*'} E(\varepsilon_s^* | B_s^*)) = 0$. Using the Taylor expansion with $t - s = Thu$,

$$Bias^*(\hat{\gamma}_t) = G_t^{-1} \sum_{s=1}^T K_{h,ts} G_s (\gamma_s - \gamma_t)$$

$$= G_t^{-1} \int K(u) \left[G_t - hu \frac{\partial G_t}{\partial t} + o(h^2) \right]$$

$$\times \left[-\frac{\partial \gamma_t}{\partial t} hu + \frac{1}{2} \frac{\partial^2 \gamma_t}{\partial t^2} (hu)^2 + o(h^2) \right]$$

$$= \frac{1}{2} d_k h^2 \left(\frac{\partial^2 \gamma_t}{\partial t^2} + 2G_t^{-1} \frac{\partial G_t}{\partial t} \frac{\partial \gamma_t}{\partial t} \right) + o(h^2).$$

Thus,

$$S^2(\hat{\gamma}_t) = \frac{1}{4} d_k^2 h^4 \left\| \frac{\partial^2 \gamma_t}{\partial t^2} + 2G_t^{-1} \frac{\partial G_t}{\partial t} \frac{\partial \gamma_t}{\partial t} \right\|_2^2 + o(h^4).$$

The variance term is given by

$$Var(\hat{\gamma}_t) = Var(\hat{\gamma}_t - \gamma_t) = Var \left[\left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} R_s^* \right]$$

$$= Var \left[\left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* (\gamma_s - \gamma_t) \right]$$

$$+ \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s,$$

and using the redefined variance term, $Var^*(\hat{\gamma}_t) = Var(W_t^* \hat{\gamma}_t)$, it follows

$$Var^*(\hat{\gamma}_t) = Var \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* (\gamma_s - \gamma_t) \right]$$

$$+ Var \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s \right]. \tag{43}$$

Since the cross terms cancel because $E(\varepsilon_s | B_s^*) = 0$, the sum of variances can be split into two terms:

$$V(\hat{\gamma}_t) = trVar^*(\hat{\gamma}_t) = trVar \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* (\gamma_s - \gamma_t) \right]$$

$$+ trVar \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s \right] = V_1 + V_2. \tag{44}$$

For the first term and taking into account that $G_s = E(B_s^{*'} B_s^*)$ it follows

$$V_1 = trVar \left(G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* (\gamma_s - \gamma_t) \right)$$

$$= trG_t^{-1} Var \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} B_s^* (\gamma_s - \gamma_t) \right) G_t^{-1}$$

$$= trG_t^{-1} E \left[\sum_{s=1}^T K_{h,ts} (B_s^{*'} B_s^* - G_s) (\gamma_s - \gamma_t) \right]$$

$$\times \left(\sum_{s=1}^T K_{h,ts} (B_s^{*'} B_s^* - G_s) (\gamma_s - \gamma_t) \right)' G_t^{-1}$$

$$= trG_t^{-1} E \left(\sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} (B_s^{*'} B_s^* - G_s) (\gamma_s - \gamma_t) (\gamma_{s'} - \gamma_t)' \right)$$

$$\times (B_{s'}^{*'} B_{s'}^* - G_{s'}) G_t^{-1} = tr \sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} (\gamma_s - \gamma_t) (\gamma_{s'} - \gamma_t)' E$$

$$\times [(B_{s'}^{*'} B_{s'}^* - G_{s'}) G_t^{-1} G_t^{-1} (B_s^{*'} B_s^* - G_s)] = tr \sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} (\gamma_s - \gamma_t)$$

$$\times (\gamma_{s'} - \gamma_t)' Q_{s,s'}, \tag{45}$$

where $Q_{s,s'} = E[(B_s^{*'} B_{s'}^{*} - G_s) G_t^{-1} G_t^{-1} (B_s^{*'} B_{s'}^{*} - G_s)]$ is a bounded p -order square matrix. Expression (45) can be divided in two parts, those corresponding to same terms and the cross terms. When $s = s'$

$$tr \sum_{s=1}^T K_{h,ts}^2 (\gamma_s - \gamma_t) (\gamma_s - \gamma_t)' Q_{s,s},$$

where Q_{ss} is bounded and has same order as

$$\begin{aligned} \sum_{s=1}^T K_{h,ts}^2 (\gamma_s - \gamma_t) (\gamma_s - \gamma_t)' &= (Th)^{-1} \\ &\times \int K^2(u) \left(-hu \frac{\partial \gamma_t}{\partial t} + o(h) \right) \left(-hu \frac{\partial \gamma_t}{\partial t} + o(h) \right)' du \\ &= (Th^{-1}) h^2 \frac{\partial \gamma_t}{\partial t} \left(\frac{\partial \gamma_t}{\partial t} \right)' \left(\int u^2 K^2(u) du \right) + o(h^2) = O\left(\frac{h}{T}\right). \end{aligned}$$

For the cross terms, $s \neq s'$

$$tr \sum_{\substack{s,s'=1 \\ s \neq s'}}^T K_{h,ts} K_{h,ts'} (\gamma_s - \gamma_t) (\gamma_{s'} - \gamma_t)' Q_{s,s'}$$

has same order as

$$\sum_{\substack{s,s'=1 \\ s \neq s'}}^T K_{h,ts} K_{h,ts'} (\gamma_s - \gamma_t) (\gamma_{s'} - \gamma_t)' = O\left(\frac{h}{T} + \frac{1}{T^2}\right).$$

Thus $V_1 = O\left(\frac{h}{T} + \frac{1}{T^2}\right)$.

For the second term in (44) and taking into account that $E(\varepsilon_s^* | B_s^*) = 0$:

$$\begin{aligned} V_2 &= trVar \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s^* \right] = trG_t^{-1} Var \left[\sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s^* \right] G_t^{-1} \\ &= trG_t^{-1} E \left[\sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s^* \left(\sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s^* \right)' \right] G_t^{-1} \\ &= trG_t^{-1} E \left[\sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} B_s^{*'} \varepsilon_s^* \varepsilon_{s'}^{*'} B_{s'}^* \right] G_t^{-1}. \end{aligned}$$

Now, since $E(\varepsilon_{is} \varepsilon_{js}') = 0$ for all $s \neq s'$, $E(\varepsilon_{is}^* \varepsilon_{js'}^{*'}) = 0$ and

$$\begin{aligned} V_2 &= trVar \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s^* \right] = trG_t^{-1} E \left[\sum_{s=1}^T K_{h,ts}^2 B_s^{*'} \varepsilon_s^* \varepsilon_s^{*'} B_s^* \right] G_t^{-1} \\ &= trG_t^{-1} E \left[\sum_{s=1}^T K_{h,ts}^2 B_s^{*'} E(\varepsilon_s^* \varepsilon_s^{*'} | B_s^*) B_s^* \right] G_t^{-1}. \end{aligned}$$

Then, as $E(\varepsilon_s^* \varepsilon_s^{*'} | B_s^*) = I$ and using the result (42) of Lemma 1 we have that

$$\begin{aligned} V_2 &= trVar \left[G_t^{-1} \sum_{s=1}^T K_{h,ts} B_s^{*'} \varepsilon_s^* \right] = \frac{C_k}{Th} trG_t^{-1} G_t G_t^{-1} + o((Th)^{-1}) \\ &= \frac{C_k}{Th} trG_t^{-1} + o((Th)^{-1}). \end{aligned}$$

Finally, since the order of V_1 is negligible with respect to V_2 and (42) holds, we have

$$V(\hat{\gamma}) = \frac{C_k}{Th} trG_t^{-1} + o((Th)^{-1}),$$

from where it follows that the order of the leading term in the variance coincides with the order of the variance term in standard results. \square

Proof of Corollary 1. Either condition (i) or (ii) provides, together with the rest of assumptions, the sufficient conditions of regularity to check that the convergence of $\hat{\Omega}_s$ to Ω_s implies the equivalence between the asymptotic properties of $\hat{\gamma}_t^{FGLS}$ and $\hat{\gamma}_t$. \square

Proof of Proposition 1. For the sake of simplicity and without loss of generality, assume $p = 1$ that is, β_{it} is one-dimensional.

In order to deal with the random denominator, we define the modified bias

$$Bias^*(\hat{\beta}_i(X_{t-1}) | X_{t-1} = x) = Bias \left[\frac{1}{(r-1)J_t(x)} \sum_{s=t-r}^{t-1} K_H(x - X_s) F_s R_{is} \right]$$

with $J_t(x) = E(F_t^2 | X_{t-1} = x) f_t(x)$. Then since $R_{is} = F_s \beta_i(X_s) + u_{is}$ with $E(u_{is} | X_{s-1}) = 0$:

$$\begin{aligned} Bias^*(\hat{\beta}_i(X_{t-1}) | X_{t-1} = x) &= \frac{1}{(r-1)J_t(x)} \\ &\times \sum_{s=t-r}^{t-1} \left(E[K_H(x - X_s) F_s^2 \beta_i(X_s)] - J_t(x) \beta_i(x) \right) \\ &= \frac{1}{(r-1)J_t(x)} \sum_{s=t-r}^{t-1} \left[\int K_H(x - \omega) F^2(\omega) \beta_i(\omega) f_s(\omega) d\omega - J_t(x) \beta_i(x) \right]. \end{aligned}$$

Using a standard multivariate kernel of order two,

$$\begin{aligned} Bias^*(\hat{\beta}_i(X_{t-1}) | X_{t-1} = x) &= \frac{1}{(r-1)J_t(x)} \\ &\times \sum_{s=t-r}^{t-1} \left[\int K(z) \beta_i(x - H^{1/2} z) f_s(x - H^{1/2} z) F_s^2(x - H^{1/2} z) dz - J_t(x) \beta_i(x) \right] \\ &= \frac{1}{(r-1)J_t(x)} \sum_{s=t-r}^{t-1} \left[\int K(z) (\beta_i(x) J_t(x) + O(\text{trace}(H))) dz - J_t(x) \beta_i(x) \right] \\ &= O(\text{trace}(H)). \end{aligned}$$

Next, we obtain the redefined variance for a generic term $\hat{\beta}_i(X_{t-1})$:

$$\begin{aligned} Var^*(\hat{\beta}_i(X_{t-1}) | X_{t-1} = x) &= \frac{1}{(r-1)^2 J_t^2(x)} Var \left[\sum_{s=t-r}^{t-1} K_H(x - X_s) F_s R_{is} | X_{t-1} = x \right] \\ &= \frac{1}{(r-1)^2 J_t^2(x)} \left[\sum_{s=t-r}^{t-1} Var(K_H(x - X_s) F_s R_{is} | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov(K_H(x - X_s) F_s R_{is}, K_H(x - X_{s'}) F_{s'} R_{is'}) | X_{t-1} = x \right] \\ &= \frac{1}{(r-1)^2 J_t^2(x)} \left[\sum_{s=t-r}^{t-1} Var(K_H(x - X_s) (F_s^2 \beta_i(X_s) + F_s u_{is}) | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov(K_H(x - X_s) (F_s^2 \beta_i(X_s) + F_s u_{is}), K_H(x - X_{s'}) (F_{s'}^2 \beta_i(X_{s'}) \right. \\ &\quad \left. + F_{s'} u_{is'}) | X_{t-1} = x) \right] \\ &= \frac{1}{(r-1)^2 J_t^2(x)} \left[\sum_{s=t-r}^{t-1} Var(K_H(x - X_s) F_s u_{is} | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{s=t-r}^{t-1} Var(K_H(x - X_s) F_s^2 \beta_i(X_s) | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov(K_H(x - X_s) F_s^2 \beta_i(X_s), K_H(x - X_{s'}) F_{s'}^2 \beta_i(X_{s'}) | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} Cov(K_H(x - X_s) F_s u_{is}, K_H(x - X_{s'}) F_{s'} u_{is'}) | X_{t-1} = x) \right] \\ &= \frac{1}{(r-1)^2 J_t^2(x)} \left[\sum_{s=t-r}^{t-1} Var(K_H(x - X_s) F_s u_{is} | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{s=t-r}^{t-1} Var(K_H(x - X_s) F_s^2 \beta_i(X_s) | X_{t-1} = x) \right. \\ &\quad \left. + \sum_{s=t-r}^{t-1} Var(K_H(x - X_s) F_s \beta_i(X_s) | X_{t-1} = x) \right] \end{aligned}$$

$$\begin{aligned}
 & + \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} \left(K_H(x - X_s) F_s^2 \beta_i(X_s), K_H(x - X_{s'}) F_{s'}^2 \beta_i(X_{s'}) | X_{t-1} = x \right) \\
 & = T_1 + T_2 + T_3,
 \end{aligned}$$

since for $s \neq s'$ the conditional expectation $E(u_{is} u_{is'})$ cancels and, therefore, only the diagonal terms remain. For T_1

$$\begin{aligned}
 T_1 & = \frac{1}{(r-1)^2 J_t^2(x)} \left[\sum_{s=t-r}^{t-1} E(K_H^2(x - X_s) F_s^2 E(u_{is}^2) | X_{t-1} = x) \right] \\
 & \leq \frac{\sigma_u^2}{(r-1)^2 J_t^2(x)} \left[\sum_{s=t-r}^{t-1} \int K_H^2(x - z) J_s(z) dz \right] \\
 & = \frac{\sigma_u^2}{(r-1)^2 J_t^2(x) |H|^{1/2}} \sum_{s=t-r}^{t-1} \\
 & \quad \times \int K^2(u) (J_t(x) + O(\text{trace} H^{1/2}) + O(\frac{1}{T})) du \\
 & = \frac{\sigma_u^2}{(r-1)^2 J_t^2(x) |H|^{1/2}} \int K^2(u) du + h.o.t. = O\left(\frac{1}{r|H|^{1/2}}\right)
 \end{aligned}$$

For T_2 , denoting $L(X_s) = E(F_s^2 | X_s)$ and $Q(X_s) = E(F_s^4 | X_s)$, we can write $F_s^2 = L(X_s) + u_{ls}$ and $F_s^4 = Q(X_s) + u_{qs}$ and, therefore

$$\begin{aligned}
 T_2 & = \frac{1}{(r-1)^2 J_t^2(x)} \sum_{s=t-r}^{t-1} \text{Var} \left(K_H(x - X_s) F_s^2 \beta_i(X_s) | X_{t-1} = x \right) \\
 & = \frac{1}{(r-1)^2 J_t^2(x)} \sum_{s=t-r}^{t-1} \left[\int K_H^2(x - w) \beta_i^2(w) (Q(w) + \sigma_Q^2) + f_s(w) dw \right. \\
 & \quad \left. - \left(\int K_H(x - w) \beta_i(w) L(w) f_s(w) dw \right)^2 \right] = \frac{1}{(r-1)^2 J_t^2(x)} \\
 & \quad \times \sum_{s=t-r}^{t-1} \left[|H|^{-1/2} \int K^2(z) \beta_i^2(x - H^{1/2} z) (Q(x - H^{1/2} z) + \sigma_Q^2) \right. \\
 & \quad \times f_s(x - H^{1/2} z) dz - \left(\int K(z) \beta_i(x - H^{1/2} z) L(x - H^{1/2} z) \right. \\
 & \quad \left. \times f_s(x - H^{1/2} z) dz \right)^2 \Big] = \frac{1}{(r-1)^2 J_t^2(x) |H|^{1/2}} \\
 & \quad \times \sum_{s=t-r}^{t-1} \left[(\beta^2 \cdot Q \cdot f_s(x) + \sigma_Q^2 \beta^2 \cdot f_s(x)) \int K^2(u) du + o(1) \right] \\
 & = O\left(\frac{1}{r|H|^{1/2}}\right)
 \end{aligned}$$

And finally for the third term, T_3 ,

$$\begin{aligned}
 T_3 & = \frac{1}{(r-1)^2 J_t^2(x)} \\
 & \quad \times \sum_{\substack{s,s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} \left[K_H(x - X_s) F_s^2 \beta_i(X_s), K_H(x - X_{s'}) F_{s'}^2 \beta_i(X_{s'}) | X_{t-1} = x \right].
 \end{aligned}$$

Using Assumption A4

$$\sum_{k=1}^{r-1} \text{Cov} \left[K_H(x - X_s) F_s^2 \beta_i(X_s), K_H(x - X_{s+k}) F_{s+k}^2 \beta_i(X_{s+k}) | X_{t-1} = x \right]$$

is uniformly bounded and, hence, the order of T_3 is $O(r^{-1})$, negligible with respect to T_1 and T_2 .

Therefore, the leading term for the variance term is $O((r|H|^{1/2})^{-1})$ and the proof is complete. \square

Proof of Theorem 2. It is sufficient to check that the proof of Theorem 1 follows for the estimated betas instead of the real ones. First, note that Assumption A4 holds for the estimated betas (\hat{B})

and that Assumption A5 holds up to order $o(1)$; that is, $E(\hat{B}_t^* \hat{B}_t^*) = E(B_t^* B_t^*) + o(1) = G_t + o(1)$.

Now, the steps of the proof of Theorem 1 follow straightforward using \hat{B} instead of B . Only the second term for the variance (44) need an extra step.

The second term for the variance can be written as,

$$\begin{aligned}
 \text{Var} \left(G_t^{-1} \sum_{s=1}^T K_{h,ts} \hat{B}_s^* \varepsilon_s^* \right) & = G_t^{-1} E \left[E \left(\sum_s K_{h,ts}^2 \hat{B}_s^* \varepsilon_s^* \varepsilon_s^{*\prime} \hat{B}_s^* \right. \right. \\
 & \quad \left. \left. + \sum_{s \neq s'} K_{h,ts} K_{h,ts'} \hat{B}_s^* \varepsilon_s^* \varepsilon_{s'}^{*\prime} \hat{B}_{s'}^* \right) | G_t^{-1} = G_t^{-1} E \left[\sum_s K_{h,ts}^2 \hat{B}_s^* \varepsilon_s^* \varepsilon_s^{*\prime} \hat{B}_s^* \right. \right. \\
 & \quad \left. \left. + \sum_{s < s'} K_{h,ts} K_{h,ts'} \hat{B}_s^* \varepsilon_s^* E(\varepsilon_{s'}^{*\prime} | \hat{B}_s^*, \hat{B}_{s'}^*, \varepsilon_s) \hat{B}_{s'}^* \right. \right. \\
 & \quad \left. \left. + \sum_{s > s'} K_{h,ts} K_{h,ts'} \hat{B}_s^* E(\varepsilon_s^{*\prime} | \hat{B}_s^*, \hat{B}_{s'}^*, \varepsilon_s) \varepsilon_{s'}^{*\prime} \hat{B}_{s'}^* \right) \right] G_t^{-1} \\
 & = G_t^{-1} E \left[\sum_s K_{h,ts}^2 \hat{B}_s^* \varepsilon_s^* \varepsilon_s^{*\prime} \hat{B}_s^* \right] G_t^{-1},
 \end{aligned}$$

since ε_s is independent of the past information. Using the fact that $E(\hat{B}_t^* \hat{B}_t^*) = E(B_t^* B_t^*) = G_t + o(1)$, it finally holds

$$\text{Var} \left(G_t^{-1} \sum_{s=1}^T K_{h,ts} \hat{B}_s^* \varepsilon_s^* \right) = \frac{C_k}{Th} G_t^{-1} + o((Th)^{-1})$$

and this step completes the proof. \square

Proof of Corollary 2. Apply the same arguments than in Corollary 1. \square

Lemma 2. Under Assumptions A3–A5 and B1–B3; it holds that

$$\frac{1}{(r-1)} \sum_{s=t-r}^{t-1} K_{\tau t}(X_s - x) \xrightarrow{a.s.} f(\tau, x), \tag{46}$$

where $\tau = t/T$.

Proof of Lemma 2. Following similar steps than in Lemma 1, define $Z_s = K_{\tau t}(X_s - x)$. The sequence Z_s has mean $f(\tau, x)$ and therefore $E\left(\frac{1}{(r-1)} \sum_{s=t-r}^{t-1} Z_s\right) = f(\tau, x) + o(1)$. A direct application of White (1984), Corollary 3.48, leads to the result. \square

Proof of Proposition 2. It holds following the proof of Proposition 1. \square

Proof of Theorem 3. Consider the sequence of variables Z_t defined as

$$Z_t = \sum_{s=1}^T K_{h,ts} \hat{B}_s^* \varepsilon_s^*. \tag{47}$$

Using White and Domowitz (1984), it is sufficient to verify that, since their Assumption A holds, the result in their Theorem 2.4 applies. Since the bias is negligible with respect to the variance term, the result follows straightforward by applying Crammer. \square

References

Ang, A., Chen, J., 2007. CAPM over the long run: 1926–2001. *Journal of Empirical Finance* 14, 1–40.
 Balvers, R.J., Huang, D., 2009. Evaluation of linear asset pricing models by implied portfolio performance. *Journal of Banking and Finance* 33, 1586–1596.
 Bollerslev, T., Engle, R., Wooldridge, J., 1988. A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96, 116–131.

- Brandt, M., Chapman, D., 2006. Linear Approximations and tests of conditional pricing models. NBER Working Paper 12513.
- Cai, Z., 2007. Trending time-varying coefficient time series models with serial correlated errors. *Journal of Econometrics* 136, 163–188.
- Dalhaus, R., 1997. Fitting time series models to nonstationary processes. *The Annals of Statistics* 25, 1–37.
- Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fama, E., French, K., 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153–193.
- Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical test. *Journal of Political Economy* 81, 607–636.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Ferson, W., Harvey, C., 1991. The variation of economic risk premiums. *Journal of Political Economy* 99, 385–415.
- Ferson, W., Harvey, C., 1999. Conditioning variables and the cross section of stock returns. *Journal of Finance* 54, 1325–1360.
- Ferson, W., Schadt, R., 1996. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* 51, 425–461.
- Grauer, R.R., Janmaat, J.A., 2009. On the power of cross-sectional and multivariate tests of the CAPM. *Journal of Banking and Finance* 33, 775–787.
- Ghysels, E., 1998. On stable factor structures in the pricing of risk: do time varying betas help or hurt? *Journal of Finance* 53, 549–574.
- Härdle, W., 1990. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Harvey, C., 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24, 289–317.
- Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51, 3–54.
- Jones, C., 2006. A nonlinear factor analysis of S and P 500 index option returns. *Journal of Finance* 41, 2325–2363.
- Kapetanios, G., 2007. Estimating deterministically time-varying variances in regression models. *Economics Letters* 97, 97–104.
- Keim, D., Stambaugh, R., 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357–390.
- Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96, 175–194.
- Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82, 289–314.
- Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109, 1238–1287.
- Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47, 13–27.
- Lozano, M., Rubio, G., 2010. Evaluating alternative methods for testing asset pricing models with historical data. *Journal of Empirical Finance* 18, 136–146.
- Orbe, S., Ferreira, E., Rodriguez-Póo, J., 2005. Nonparametric estimation of time varying parameters under shape restrictions. *Journal of Econometrics* 126, 53–77.
- Orbe, S., Ferreira, E., Rodriguez-Póo, J., 2006. On the estimation and testing of time varying constraints in econometric models. *Statistica Sinica* 16, 1313–1333.
- Pagan, A., 1984. Econometric Issues in the analysis of regressions with generated regressors. *International Economic Review* 25, 221–247.
- Robinson, P., 1989. Nonparametric estimation of time varying parameters. In: Hackl, P. (Ed.), *Statistical Analysis and Forecasting Economic Structural Change*. Springer-Verlag, Berlin.
- Ross, S., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.
- Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. *Journal of Financial Economics* 14, 327–348.
- Shanken, J., 1992. On the estimation of beta-pricing models. *Review of Financial Studies* 5, 1–33.
- Shanken, J., Zhou, G., 2007. Estimating and testing beta pricing models: alternative methods and their performance in simulations. *Journal of Financial Economics* 84, 40–86.
- Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 33, 885–901.
- Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.
- van Dijk, M.A., 2011. Is size dead? A review of the size effect in equity returns. *Journal of Banking and Finance* 35, 3263–3274.
- Wand, M., Jones, M., 1995. *Kernel Smoothing*. Chapman and Hall, London.
- Wang, K., 2002. Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio. *Journal of Empirical Finance* 9, 133–169.
- Wang, K., 2003. Asset Pricing with conditioning information: a new test. *Journal of Finance* 58, 161–196.
- White, H., 1984. *Asymptotic Theory for Econometricians*. Academic Press, London.
- White, H., Domowitz, I., 1984. Nonlinear regression with dependent observations. *Econometrica* 52, 143–161.
- Wu, X., 2002. A conditional multifactor analysis of return momentum. *Journal of Banking and Finance* 26, 1675–1696.
- Zellner, A., 1962. An efficient method for estimating seemingly unrelated regressions and tests for aggregate bias. *Journal of the American Statistical Association* 57, 348–368.
- Zellner, A., 1970. Estimation of regression relationships containing unobserved variables. *International Economic Review* 11, 441–454.