

# DOCUMENTOS DE TRABAJO

# BILTOKI

D.T. 2010.10

Conditional beta pricing models: A nonparametric approach.

Eva Ferreira, Javier Gil-Bazo y Susan Orbe

eman ta zabal zazu



Universidad Euskal Herriko  
del País Vasco Unibertsitatea

Facultad de Ciencias Económicas.  
Avda. Lehendakari Aguirre, 83  
48015 BILBAO.

**Documento de Trabajo BILTOKI DT2010.10**

Editado por el Departamento de Economía Aplicada III (Econometría y Estadística)  
de la Universidad del País Vasco.

Depósito Legal No.: BI-3160/2010

ISSN: 1134-8984

URN: RePEc:ehu:biltok:201010

# Conditional beta pricing models: A nonparametric approach

Eva Ferreira<sup>a</sup>      Javier Gil-Bazo<sup>b</sup>      Susan Orbe<sup>a,\*</sup>

<sup>a</sup>University of the Basque Country

<sup>b</sup>University Pompeu Fabra

---

## Abstract

We propose a two-stage procedure to estimate conditional beta pricing models that allow for flexibility in the dynamics of assets' covariances with risk factors and market prices of risk (MPR). First, conditional covariances 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 covariances using the entire sample. We prove the consistency and asymptotic normality of the estimators. Results from a Monte Carlo simulation for the three-factor model of Fama and French (1993) suggest 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 MPR associated with the three Fama-French factors exhibit substantial variation through time. Finally, the flexible version of the three-factor model beats alternative parametric specifications in terms of forecasting future returns.

*JEL classification:* G12; C14; C32

*Keywords:* Kernel estimation; Conditional asset pricing models; Fama-French three-factor model; Locally stationary processes

---

\* Corresponding author: email: javier.gil.bazo@uc3m.es

The authors would like to thank Eduardo Schwartz, Alfredo Ibáñez, Alvaro Cartea, and Martin Seim, as well as seminar participants at the 2007 Cemapre Conference on Advances in Semiparametric Methods and Applications, 2007 International Conference on Applied Stochastic Models and Data Analysis, 2008 European Economic Association & Econometric Society Meetings, 2010 FMA European Conference, 2010 EFMA Conference, and Basque Centre of Applied Mathematics for their comments and suggestions on previous versions of this paper. All errors are the authors' responsibility. This work has been supported by the Spanish Government grants ECO2008-00777/ECON and SEJ2007-67448, and Grupo MACLAB (IT-241-07)-Basque Government.

## 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 covariance of the asset's return with one or more factors capturing market-wide sources of risk. Re-scaled by the variance of each risk factor, covariances are referred to as *betas*, and are interpreted as the asset's exposure to risks that cannot be eliminated through diversification. The slopes of the linear relation, which must be equal for all assets, are interpreted as the rewards per unit of covariance 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), Fama and French (1989), Ferson (1989), 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 time-varying factor sensitivities and MPR are allowed to vary through time.<sup>1</sup> A drawback of conditional models is that estimation requires additional assumptions about the dynamics of risk exposures and/or MPR. For instance, Bollerslev, Engle and Wooldridge (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 the 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 allow for flexibility in the dynamics of covariances 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

---

<sup>1</sup>It is worth noting that similar conditional asset pricing relations for option and bond returns are also obtained in arbitrage-free models, such as Black and Scholes (1973), Cox, Ingersoll and Ross (1985), and their extensions, with discrete returns replaced by instantaneous returns.

models.<sup>2</sup> In the first stage of the Fama-MacBeth method, asset betas are computed for every asset and period using a time series regression of several periods of previous 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 covariances nonparametrically for each asset and period using previous information. However, unlike the Fama-MacBeth procedure, conditional covariances are assumed to be smooth (but possibly nonlinear) functions of the state variables. In the second stage, time-varying MPR are estimated at each point in time from the cross-section of returns and estimated covariances (the regressors), but instead of running a single cross-sectional regression, 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 conditional factor sensitivities and MPR. Our method exhibits a number of important advantages 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 (covariances) ours is more efficient when betas (covariances) are believed to be functions of a set of variables capturing the state of the system.<sup>3</sup> 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 for 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 covariances. Finally, we assume locally stationary variables as defined in Dalhaus (1997), which permit time-varying mean and, therefore, enable us

---

<sup>2</sup>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.

<sup>3</sup>For instance, the 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.

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 considered by Wang (2002, 2003) risk factors are portfolio returns, so conditional market prices of risk equal conditional expected excess returns on factor portfolios and 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) 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, Ferreira and Rodriguez-Póo (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, Ferreira

and Rodriguez-Póo (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 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 conditional covariances) 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 covariances 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 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. 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 nonparametrically estimated MPR exhibit substantial time variation, which supports the use of flexible estimation methods. Further, the nonparametric method proposed in this paper is clearly superior to different parametric alternatives in terms of its ability to forecast the cross-section of future returns. A purely empirical model, however, appears to dominate even our flexible version of the Fama-French model.

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 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

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

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  $\ell$ 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.,  $E(F_{\ell t}) = 0$ . The error term  $\nu_{it}$  is serially independent with zero mean and nonsingular covariance matrix, conditional on factor realizations. 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 equation (1).  $\beta_{i\ell}$  represents the sensitivity of asset  $i$ 's return to the  $\ell$ th risk factor and, under the assumption that the risk factors are orthogonal to each other and to the error term, it equals the covariance between the factor and the asset return re-scaled by the variance of the risk factor. The coefficient  $\gamma_\ell$ , 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  $\ell$ .

The first stage of the two-pass estimation procedure of Fama and MacBeth (1973), consists of estimating betas in equation (1) for each asset and time from a time-series regression. In the second stage,  $\gamma$ '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.

The asset pricing relation (2) can be equivalently rewritten as

$$E(R_{it}) = \gamma_1 \frac{Cov(R_{it}, F_{1t})}{Var(F_{1t})} + \dots + \gamma_p \frac{Cov(R_{it}, F_{pt})}{Var(F_{pt})}. \quad (3)$$

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

$$R_{it} = E(R_{it}|I_{t-1}) + \frac{Cov(R_{it}, F_{1t}|I_{t-1})}{Var(F_{1t}|I_{t-1})}F_{1t} + \dots + \frac{Cov(R_{it}, F_{pt}|I_{t-1})}{Var(F_{pt}|I_{t-1})}F_{pt} + \nu_{it}, \quad (4)$$

$$E(R_{it}|I_{t-1}) = \gamma_{1t} \frac{Cov(R_{it}, F_{1t}|I_{t-1})}{Var(F_{1t}|I_{t-1})} + \dots + \gamma_{pt} \frac{Cov(R_{it}, F_{pt}|I_{t-1})}{Var(F_{pt}|I_{t-1})}, \quad (5)$$

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  $X_{t-1} = (X_{1t-1} \dots X_{mt-1})'$ .

Following Harvey (1989), we are interested in estimating the reward per unit of covariance risk or market price of risk associated with the  $\ell$ th factor. Denoting by  $\sigma_{\ell t}^2$  the conditional variance of the  $\ell$ th factor, by  $c_{i\ell t}$  the conditional covariance between the  $i$ th asset return and the  $\ell$ th risk factor, and by  $\lambda_{\ell t} \equiv \gamma_{\ell t}/\sigma_{\ell t}^2$  the market price of risk, equation (5) can be rewritten as

$$E(R_{it}|X_{t-1}) = \lambda_{1t}c_{i1t} + \dots + \lambda_{pt}c_{ipt} \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T. \quad (6)$$

If we define  $\varepsilon_{it} \equiv R_{it} - E(R_{it}|X_{t-1})$ , then we may write

$$R_{it} = \lambda_{1t}c_{i1t} + \dots + \lambda_{pt}c_{ipt} + \varepsilon_{it} \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T. \quad (7)$$

It follows from (4) that the errors  $\varepsilon_{it}$  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  $\nu_{it}$  are serially independent for all  $\ell, i$  and  $t$ , so  $E(\varepsilon_{it}\varepsilon_{js}) = 0$ , for all  $i, j$  and  $t \neq s$ .

To estimate  $\lambda$ 's in (7), we form the system of equations

$$\begin{aligned} R_{1t} &= \lambda_{1t} c_{11t} + \lambda_{2t}c_{12t} + \dots + \lambda_{pt}c_{1pt} + \varepsilon_{1t} \\ R_{2t} &= \lambda_{1t} c_{21t} + \lambda_{2t}c_{22t} + \dots + \lambda_{pt}c_{2pt} + \varepsilon_{2t} \\ &\vdots \\ R_{Nt} &= \lambda_{1t} c_{N1t} + \lambda_{2t}c_{N2t} + \dots + \lambda_{pt}c_{Npt} + \varepsilon_{Nt}, \end{aligned} \quad (8)$$

where  $\{\lambda_{\ell t}\}_{\ell=1}^p$  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  $\lambda_{\ell t}$  in (8). For a better description of the procedure, consider some extra notation:  $C_s = (C_{1s} \ C_{2s} \ \dots \ C_{Ns})'$  is a  $N \times p$ -order matrix where each term  $C_{is}$  denotes the  $N$ -order column vector  $(c_{i1s} \ c_{i2s} \ \dots \ c_{ips})'$ , for  $i = 1, \dots, N$ ;  $R_s$  is the  $N$ -order column vector  $(R_{1s} \ R_{2s} \ \dots \ R_{Ns})'$ ; and  $\lambda_t = (\lambda_{1t} \ \dots \ \lambda_{pt})'$  is the  $p$ -order vector of prices of risk. Finally, the state-variable  $p$ -order column vector is denoted by  $X_t = (X_{1t} \ \dots \ X_{mt})'$ . According to this notation, model (8) can be compactly written as

$$R_t = C_t \lambda_t + \varepsilon_t \quad t = 1, \dots, T. \quad (9)$$

Within this framework, we propose to estimate the time-varying vector of market prices of risk at each time  $t$ ,  $\lambda_t$ , taking into account the structure of the error 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:

$$\min_{\lambda_t} \sum_{s=1}^T K_{h,ts} (R_s - C_s \lambda_t)' \Omega_s^{-1} (R_s - C_s \lambda_t), \quad (10)$$

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

$$\widehat{\lambda}_t = \left( \sum_{s=1}^T K_{h,ts} C_s' \Omega_s^{-1} C_s \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s' \Omega_s^{-1} R_s. \quad (11)$$

With the usual standardization,  $R_s^* = V_s^{-1} R_s$  and  $C_s^* = V_s^{-1} C_s$ , where  $V_s$  is the matrix such that  $V_s V_s' = \Omega_s$ , the optimization problem (10) can be written as

$$\min_{\lambda_t} \sum_{s=1}^T K_{h,ts} (R_s^* - C_s^* \lambda_t)' (R_s^* - C_s^* \lambda_t), \quad (12)$$

which allows us to express the estimator of the market prices of risk (11) in a more compact form

$$\widehat{\lambda}_t = \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} R_s^*. \quad (13)$$

**Remark 1** For large enough  $h$ , the estimator (13) leads to the same estimates as those obtained in classical SURE model estimation with constant coefficients, subject to the equality constraints:

$$\widehat{\lambda} = \left( \sum_{s=1}^T C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T C_s^{*'} R_s^* = \left( \sum_{s=1}^T C_s' \Omega_s^{-1} C_s \right)^{-1} \sum_{s=1}^T C_s' \Omega_s^{-1} R_s. \quad (14)$$

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

$$\widehat{\lambda}_t = (C_t^{*'} C_t^*)^{-1} C_t^{*'} R_t^* = (C_t' \Omega_t^{-1} C_t)^{-1} C_t' \Omega_t^{-1} R_t, \quad (15)$$

which is equivalent to estimating  $\lambda_t$  from a cross-sectional regression at time  $t$ .

**Remark 2** There is a close relation between the estimator in (13) and the estimator proposed in Shanken (1985), and asymptotically studied in Shanken (1992). Considering constant coefficients ( $h \rightarrow \infty$ ) as in (14) and assuming that the covariances between returns and the risk factors are time invariant, i.e.,  $C_s = C \ \forall s$ , the resulting estimator is

$$\widehat{\lambda} = \left( \sum_{s=1}^T C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T C_s^{*'} R_s^* = (C' \Omega^{-1} C)^{-1} C' \Omega^{-1} \bar{R}, \quad (16)$$

and substituting  $C$  and  $\Omega$  by their estimators, respectively, then  $\widehat{\lambda}$  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 (13), we study the mean square error of  $\hat{\lambda}_t$ , which we define as the sum of mean square errors of  $\hat{\lambda}_{\ell t}$  for all  $\ell$ , under the assumptions below. We denote this sum by  $MSE$ :

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

**Assumption (A1)** The market prices of risk are smooth functions of the time index; that is,  $\lambda_{\ell t} = \lambda_{\ell}(t/T)$  where each  $\lambda_{\ell}$  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 covariance can only vary with time through the state vector at time  $t-1$ ,  $X_{t-1}$ . That is,  $c_{i\ell t} = c_{i\ell}(X_{t-1})$ , where it is assumed that  $c_{i\ell}$  is at least twice differentiable for all partial derivatives.

**Assumption (A4)** Both  $C_{it}$  and  $X_{it}$  are statistically independent of  $\varepsilon_{is}$ , for all  $s \geq t$ . Moreover, we assume the process (9) with finite distributions such that the sequence  $\{X_{it}, C_{it}, \varepsilon_{it}\}$  is strong  $\alpha$ -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  $\theta$ .

**Assumption (A5)** At each time  $t$ , the unconditional expectation  $E(C_t^{*'} C_t^*) = G_t$  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  $\varepsilon_t$  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  $\sigma^{ijt}$  be a generic term in  $\Omega_t^{-1}$ . The  $p$ -order matrix

$$\left( \sum_{s=1}^T \sum_{i,j=1}^N K_{h,ts} \sigma^{ijt} C_{i,s} C'_{j,s} \right), \quad (17)$$

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. (A2) holds for technical reasons in kernel estimation. (A3) imposes smoothness on the explanatory variables. (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), and Dalhaus (2000)). We also assume smoothness in errors' covariances. (A6) excludes equations with exploiting variances or with lineary dependent error terms and (A7) ensures that the estimator is identified. (A8) is standard in nonparametric estimation.

**Theorem 1** *Under the set of assumptions (A1) to (A8), the MSE for the estimator defined in (13), has bias and variance,*

$$S^2(\hat{\lambda}_t) = \frac{h^4 d_K^2}{4} \left\| \frac{\partial^2 \lambda_t}{\partial t \partial t} + 2G_t^{-1} \frac{\partial G_t}{\partial t} \frac{\partial \lambda_t}{\partial t} \right\|_2^2 + o(h^4) \quad (18)$$

and

$$V(\hat{\lambda}_t) = \frac{c_K}{Th} \text{tr}(G_t^{-1}) + o((Th)^{-1}) \quad (19)$$

where  $G_t = E(C_t^{*'} C_t^*)$ , 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) to (A8), the asymptotic order and the leading terms are the same considering either stationary or locally stationary variables.

**Corollary 1** *Consider model (9) and a consistent estimator  $\hat{\Omega}_s = \hat{V}_s \hat{V}_s'$  of  $\Omega_s = V_s V_s'$ . Then, under the same assumptions of Theorem 1, and if either*

(i)  $\hat{\Omega}_s - \Omega_s = o(MSE(\hat{\lambda}_t))$ , or

(ii) *the entries in  $C_s$  are bounded,*

*the Feasible Generalized Least Squares (FGLS) estimator*

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

*has the same asymptotic properties as the estimator (13).*

All previous asymptotic results have been 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 that each element  $c_{i\ell t}$  of  $C_t$  measures the covariance between the return on the  $i$ th asset and the  $\ell$ th risk factor, we propose to estimate  $c_{i\ell t}$  as a conditional rolling smoothed sample covariance

$$\widehat{c}_{i\ell t} = \widehat{c}_{i\ell}(X_{t-1}) = \left( \sum_{s=t-r}^{t-1} K_B(X_{s-1} - X_{t-1}) \right)^{-1} \sum_{s=t-r}^{t-1} K_B(X_{s-1} - X_{t-1}) P_{i\ell s}, \quad (20)$$

where we recall that  $X_s = (X_{1s} \dots X_{ms})'$  denotes the vector of state variables. We define  $P_{i\ell s} = R_{is} F_{\ell s} - \mu_{R_{is}}(X_{s-1}) \mu_{F_{\ell s}}(X_{s-1})$ , where  $\mu_{R_{is}}(X_{s-1})$  and  $\mu_{F_{\ell s}}(X_{s-1})$  denote the estimated means of  $R_{is}$  and  $F_{\ell s}$  conditional on  $X_{s-1}$ , respectively.  $K_B$  is a  $m$ -variate kernel  $K_B(u) = |B|^{-1/2} K(B^{-1/2}u)$ , with smoothing matrix  $B$ . That is, (20) 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  $c_{i\ell t}$  that uses past information only.

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

$$\widehat{\lambda}_t^{SGLS} = \left( \sum_{s=1}^T K_{h,ts} \widehat{C}_s^{*'} \widehat{C}_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} \widehat{C}_s^{*'} R_s^*, \quad (21)$$

is similar to (13) with  $C$  replaced by  $\widehat{C}$  and  $\widehat{C}_s^* = V_s^{-1} \widehat{C}_s$ . In order to reach the desirable asymptotic results some additional assumptions are required:

**Assumption (C1)** The  $m$ -variate kernel  $K$  is compactly supported such that  $\int K(u) du = 1$  and  $\int uu' K(u) du = \mu_K I_p$ , where  $\mu_K$  is a nonnegative scalar and  $\int$  and  $du$ , are the shorthands for  $\int \dots \int_{\mathbb{R}^p}$  and  $du_1 \dots du_p$ , respectively.

**Assumption (C2)** Consider a sequence of positively definite diagonal bandwidth matrices  $B = \text{diag}(b_{i1}^2, b_{i2}^2, \dots, b_{im}^2)$  for  $i = 1, \dots, N$ , such that  $|B|^{1/2}$  and  $r/T$  go to zero,  $T|B|^{1/2}$  and  $r|B|^{1/2}$  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 (C3)** The distribution of  $X_t$  has an order-one-Lipschitz time-varying density,  $f_t(x) = f(\tau, x)$ , where  $\tau = t/T$ .

The following proposition states the properties of estimator (20).

**Proposition 1** *Consider the set of assumptions (A3) to (A6), and (C1) to (C3) then, the estimator defined by (20) is a consistent estimator of  $c_{i\ell}(X_{t-1})$ , with asymptotic bias and variance:*

$$\begin{aligned} \text{Bias}(\hat{c}_{i\ell}(X_{t-1})|X_{t-1} = x_{t-1}) &= O(\text{trace}(B)) \\ \text{Var}(\hat{c}_{i\ell}(X_{t-1})|X_{t-1} = x_{t-1}) &= O\left(\frac{1}{r|B|^{1/2}}\right). \end{aligned}$$

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 covariances described above are employed.

**Theorem 2** *Under the set of assumptions (A1) to (A8) and (C1) to (C3), using the covariance estimator  $(\hat{C}_t)$  whose elements are defined in (20), the SGLS estimator for the market prices of risk defined in (21) is consistent, with the same asymptotic results for the two components of the MSE as in Theorem 1.*

**Corollary 2** *Consider model (9) with the consistent estimator of  $C_t$  defined in (20) and a consistent estimator  $\hat{\Omega} = \hat{V}_s \hat{V}_s'$  for  $\Omega_s = V_s V_s'$ . Then, under the assumptions in Theorem 2, and if either*

(i)  $\hat{\Omega}_s - \Omega_s = o(\text{MSE}(\hat{\lambda}_t^{SGLS}))$ , or

(ii) *the entries in  $C$  are bounded,*

*the Smoothed Feasible Generalized Least Squares (SFGLS) estimator*

$$\hat{\lambda}_t^{SFGLS} = \left( \sum_{s=1}^T K_{h,ts} \hat{C}'_s \hat{\Omega}_s^{-1} \hat{C}_s \right)^{-1} \sum_{s=1}^T K_{h,ts} \hat{C}'_s \hat{\Omega}_s^{-1} R_s \quad (22)$$

*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 (22).

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

$$\hat{\sigma}_{ijt} = \left( \sum_{s=1}^T K_G(X_{s-1} - X_{t-1}) \right)^{-1} \sum_{s=1}^T K_G(X_{s-1} - X_{t-1}) (R_{it} - \hat{\lambda}_t \hat{C}_{it})' (R_{jt} - \hat{\lambda}_t \hat{C}_{jt}) \quad (23)$$

with  $K_G(u) = |G|^{-1/2}K(G^{-1/2}u)$ , being  $G$  the  $m$ -order smoothing matrix and  $K$  a  $m$ -variate second order kernel. Under assumptions (A1)-(A8), (C3) and the kernel  $K_G$  satisfying (C1) and (C2) (although no assumption for  $r$  is needed here), (23) provides a consistent estimator for a generic term of  $\Omega_t$ , for each  $t$ .

The next (pointwise) asymptotic distribution for the estimator of  $\lambda_t$ , 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 (A1)-(A8) and (C1)-(C3), 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  $\lambda_t$  at  $k$  different locations  $t_1, \dots, t_k$  converges in distribution to the multivariate normal as,*

$$\left( (Th)^{1/2} (\hat{\lambda}_{t_j}^{SGLS} - \lambda_{t_j}) \right)_{j=1}^k \xrightarrow{p} N(0, c_K G_{t_j}^{-1}). \quad (24)$$

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  $\lambda$ '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 covariances between risk factors and asset returns,  $B$ ; 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  $B$  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

$$(NT)^{-1} \sum_{t=1}^T (R_t - \hat{C}_t(B) \hat{\lambda}_t(h))' (R_t - \hat{C}_t(B) \hat{\lambda}_t(h)) \mathcal{G}(h, B), \quad (25)$$



where  $\mathcal{G}(h, B)$  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  $C_t$  defined in (20) does not include the observation at time  $t$ , there is no need to penalize the selection of  $B$ . Thus, we will use a function that only penalizes low values of  $h$ , i.e.,  $\mathcal{G}(h, B) = \mathcal{G}(h)$ . In particular, in the Generalized Cross Validation method, the penalty is

$$\mathcal{G}(h) \approx (1 - (NT)^{-1} \text{trace} P(h))^{-2}, \quad (26)$$

where  $P(h)$  is the projection matrix

$$\frac{K(0)}{Th} \sum_{i=1}^N \sum_{t=1}^T \left[ C'_{is} (C'_i \mathcal{K}_{h,t} C_i)^{-1} C'_i \right] Z_t, \quad (27)$$

where  $C_{is} = (c_{i1t} \ c_{i2t} \ \dots \ c_{ipt})'$  is defined above, the  $T \times p$ -order matrix  $C_i = (C_{i1} \ C_{i2} \ \dots \ C_{iT})'$  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 the smoothing parameters  $B$  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

$$\sum_{t=1}^T (R_t - \hat{C}_t \hat{\lambda}_t)' \hat{\Omega}_t^{-1}(\mathcal{H}) (R_t - \hat{C}_t \hat{\lambda}_t) \quad (28)$$

where the estimator of any generic term of  $\hat{\Omega}$ ,  $\hat{\sigma}_{ijt}$ , is given by (23).

The  $\hat{\lambda}$ 's estimated with the selected  $\mathcal{H}$  are then used again to obtain the model residuals, and reestimate conditional covariances, the residual covariance matrix and the new  $\hat{\lambda}$ '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  $\hat{\lambda}$ 's. However, changing  $B$  and  $h$  in this iterative procedure does not 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 or covariances are believed to be functions

of a set of variables capturing the state of the system. Another advantage 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:

$$R_{it} = \beta_{im,t}R_{mt} + \beta_{ismb,t}R_{smbt} + \beta_{ihml,t}R_{hmlt} + \epsilon_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T, \quad (29)$$

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  $N = 1$ , so we can omit the asset subscript, and assume constant betas with respect to SMB and HML, i.e.,  $\beta_{smb,t} = \beta_{smb}$  and  $\beta_{hml,t} = \beta_{hml}$  for all  $t$ .<sup>4</sup> We model  $\beta_{m,t}$  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 dividends on the S&P500 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. Data used to compute the dividend yield and the default spread were obtained from CRSP and the Federal Reserve Economic Data database.

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 four different specifications:

$$\beta_{m,t} = 1 - 20X_{1,t-1} \quad (30)$$

$$\beta_{m,t} = 1 - 20X_{1,t-1} + 900X_{1,t-1}^2 \quad (31)$$

$$\beta_{m,t} = 1 - 20X_{1,t-1} + 900X_{1,t-1}^2 + 30X_{2,t-1} - 200x_{2,t-1}^2 - 2500X_{1,t-1}X_{2,t-1} \quad (32)$$

$$\beta_{m,t} = 1 - 8X_{1,t-1}e^{-100X_{1,t-1}} - 20X_{2,t-1}e^{-100X_{2,t-1}} \quad (33)$$

---

<sup>4</sup>We set  $\beta_{smb} = 0.8$  and  $\beta_{hml} = 0.4$  in all simulations.

For each series of  $\beta_{m,t}$  we then simulate 1,000 paths, indexed by  $j$ , of asset return realizations,  $R_t^j = \beta_{im,t}R_{mt} + \beta_{ismb}R_{smbt} + \beta_{ihml}R_{hmlt} + \epsilon_t^j$ , for  $t = 1, \dots, T$ . Henceforth, we refer to each one of the four models corresponding to (30)-(33) as Model 1-Model 4. Realizations of the three Fama-French factors were downloaded from Kenneth French's website and are orthogonalized.<sup>5</sup> Random errors,  $\epsilon_t^j$ , are drawn independently from  $N(0, \sigma = 0.04)$ .

We estimate the vector of conditional betas for each simulated path using a kernel estimator that is analogous to the nonparametric estimator of the conditional covariances proposed above (20):

$$\hat{\beta}_t^{NP,j} = \left( \sum_{s=t-60}^{t-1} K_B(X_{s-1} - X_{t-1})F_s F_s' \right)^{-1} \sum_{s=t-60}^{t-1} K_B(X_{s-1} - X_{t-1})F_s (R_s^j)', \quad (34)$$

where  $F_s$  denotes the vector of the orthogonalized three Fama-French factors and simulated asset returns,  $R_s^j$ , are demeaned.

To compare the accuracy of the nonparametric estimator with that of the rolling betas, i.e., betas estimated in overlapping rolling samples of prior data, we also estimate for each one of the simulated paths:

$$\hat{\beta}_t^{ROLL,j} = \left( \sum_{s=t-60}^{t-1} F_s F_s' \right)^{-1} \sum_{s=t-60}^{t-1} F_s (R_s^j)'. \quad (35)$$

Finally, we compute the mean square error of each series of estimated betas as well as the average mean square error for all simulated paths within each model. That is, we compute  $MSE_j = (1/T) \sum_{t=1}^T (\hat{\beta}_{mt}^j - \beta_{mt})^2$  and  $\overline{MSE} = (1/1000) \sum_{j=1}^{1000} MSE_j$ .

Simulation results are presented in Table 1. In column one, we report for each model the percentage of simulated paths for which the MSE is lower for the nonparametric estimator of market beta than for the rolling estimator. Columns two and three report mean square errors averaged across all simulations for each model and each estimator. Results in Table 1 indicate that the nonparametric estimator is more accurate for all models: The percentage of times that the nonparametric estimator is superior to the rolling estimator ranging between 81.5% of the times (Model 4) and 97.4% (Model 1). Average MSE are also lower for the nonparametric estimator for all four models.

Figure 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 quar-

---

<sup>5</sup>We orthogonalize the size factor by regressing  $R_{smb}$  on  $R_m$ . We then define the orthogonalized size factor as the residuals from that regression. Next, we regress  $R_{hml}$  on  $R_m$  and the orthogonalized size factor, and define the orthogonalized book-to-market factor analogously.

tile are all lower for the nonparametric estimator than for the rolling estimator. The most striking difference is achieved for Model 1: 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 Figure 2 we plot the median, the first and the third quartiles of estimated betas for both estimators under the four specifications together with the true betas. The figure shows that the nonparametric estimator performs remarkably well under the four specifications, especially in the pre-1990 period. The median estimated beta tracks closely the true beta and the interquartile range is very narrow. The rolling estimator, in contrast, appears to respond slower to changes in the true beta. Also the interquartile range is substantially wider in all cases than for the nonparametric estimator.

## 6. Empirical application

In this section we apply the non-parametric method presented above to estimate conditional covariances and MPR in a flexible conditional version of the Fama and French (1993) three-factor model. By providing consistent estimates of time-varying market prices of risk as well as confidence intervals, our method makes it possible to identify time variation in the price of common risk factors. This is a desirable advantage over parametric methods that assume constant MPR since a statistically insignificant constant MPR associated with a risk factor may hide the fact that the risk factor is actually priced for certain subperiods or that the sign of the price has changed throughout the full sample period. We then compare the performance of the nonparametric approach to that of several alternatives that have been proposed in the literature in terms of their ability to predict the cross-section of future returns. This comparison enables us to establish whether the nonparametric approach leads to better estimates not only of factor sensitivities but also of MPR and, ultimately, asset expected returns.

### 6.1. Model and Data

We consider a particular case of the asset pricing relation (6) where the marketwide factors are the orthogonalized three Fama-French factors described in the previous section:

$$E(R_{it}|X_{t-1}) = \lambda_{m,t}C_{im,t} + \lambda_{smb,t}C_{ismb,t} + \lambda_{hml,t}C_{ihml,t} \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T. \quad (36)$$

Like other studies, we use data on the 25 equity portfolios formed by sorting individual stocks on market capitalization and book-to-market (Fama and French (1993)). Monthly data on the 25 portfolio returns and the one-month risk-free rate were downloaded from Kenneth French’s website.

We follow closely Ferson and Harvey (1999) and select five conditioning variables ( $X_{t-1}$ ) 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*”). *DP*, *Tb1m*, and *hb3* were constructed from data obtained from CRSP. Data on *term* and *def* were obtained from the Federal Reserve’s FRED database.

Our final sample contains 510 monthly observations of factor realizations, portfolio returns and lagged state variables in the July 1963-December 2005 period.

## 6.2. Estimation results

We start by estimating nonparametrically the conditional covariances of returns with the three risk factors and for each one of the 25 portfolios. We then estimate the time-varying market prices of risk associated with each factor.

To estimate conditional covariances at time  $t$ , we use 60 months of prior data on portfolio returns, factor realizations and conditioning variables. This results in a loss of 60 observations in the estimation of MPR. 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. More specifically, since *DP* has the most predictive power over future returns in univariate predictive regressions (unreported), we consider four pairs of conditioning variables, each one combining *DP* with one of the remaining four state variables.

In Table 2 we report summary statistics of nonparametrically estimated MPR associated with the three Fama-French factors. For the sake of brevity, we report results only when the two conditioning variables are *DP* and *term*. Results for other choices of conditioning variables are similar and available from the authors upon request. Each panel in Table 2 shows the minimum, the mean, the standard deviation, and maximum values of  $\hat{\lambda}_m$ ,  $\hat{\lambda}_{smb}$ , and  $\hat{\lambda}_{hml}$  in a different subperiod. We also report the fraction of months in each subperiod in which estimated MPR are positive, and both positive and statistically significant at the 5% level. Results reveal that the prices associated with exposure to the market and the book-to-market risk factors are positive most of the time and with similar frequency: 84% and 84.67% of all months, respectively. However, the price of book-

to-market risk is statistically significant more often (about 29% of the time in the full sample) than the price of market risk (only 6% of all observations). The prices of risk associated with both risk factors are statistically significant more often in the 1996-2005 subperiod (15.83% and 35.83%, respectively) than in other subperiods. In contrast with these results, the price of the risk factor associated with size is negative on average in all subperiods with the exception of the 1976-1985 subperiod, and is positive only for 38.22% of all observations. Moreover, the price of size risk is never both positive and statistically significant. This price exhibits the highest standard deviation in the full sample. Interestingly, these results are broadly consistent with the finding by Ferson and Harvey (1999) that only the price of book-to-market risk is statistically significant in the 1963-1994 sample period.

Figure 3 displays estimated MPR associated with the three Fama-French factors as well as 95% confidence intervals. Again, we report results obtained when  $DP$  and  $term$  are the only conditioning variables. The figure reveals that market prices of risk have varied substantially through time. Although the price of market risk has been consistently positive, it is not statistically significant most of the time, an important exception being the late nineties, when  $\hat{\lambda}_m$  reached its maximum value. Another interesting insight revealed by Figure 3 is that the price of size risk appears to have risen in the post-1999 period. Finally, although the book-to-market risk factor has been consistently priced by investors, the price associated with this risk factor appears to have varied through time: It achieved a peak in the early nineties and stayed at high levels again in the 2001-2005 period.

Taken together, our estimation results indicate that lack of significance of prices of the Fama-French market and size risk factors found by Ferson and Harvey (1999) survives a flexible specification of the conditional FF model: Only the price of book-to-market risk is often positive and statistically significant. However, our results also suggest that MPR exhibit substantial variation through time. More specifically, the risk factor identified with the return on the market portfolio has been positive and significant in certain subperiods, even though the average price of this factor does not appear to be distinguishably different from zero.

### 6.3. Forecasting results

Ferson and Harvey (1999) test whether constant or linear FF factor betas can explain the cross-section of conditional expected returns against the alternative that conditional expected returns are captured by forecasts based on predictive OLS regressions of returns on lagged state variables. They find that including the OLS-forecast in cross-sectional regressions reduces the significance of Fama-MacBeth coefficients on both SMB and HML betas. Moreover, the coefficient on the OLS-forecast variable remains highly significant. They interpret these results as evidence against both

the unconditional and conditional versions of the FF model.

To evaluate the ability of the nonparametric approach to capture the cross-section of conditional expected returns, we perform a test that is similar to that employed by Ferson and Harvey (1999). First, for each month  $t$  we use information up to  $t - 1$  to estimate nonparametrically conditional covariances between asset returns and the risk factors. We also estimate the MPR corresponding to the last observation, i.e.,  $\hat{\lambda}_{t-1}$ , and use them together with conditional covariances to compute conditional expected returns, which we denote by  $\hat{E}_{t-1}^{NP}(R_{it})$ . Like Ferson and Harvey (1999), for each month  $t$  and each portfolio  $i$  in our sample, we also run the predictive OLS regression:

$$R_{is} = \delta'_{it-1} X_{s-1} + u_{is} \quad s = 1, 2, \dots, t - 1, \quad (37)$$

and compute the fitted conditional expected return from the empirical model,  $\hat{E}_{t-1}^{OLS}(R_{it}) \equiv \hat{\delta}'_{it-1} X_{t-1}$ . We then estimate a series of cross-sectional regressions at each point in time:

$$R_{it} = \phi_{0t} + \phi_{NPt} \hat{E}_{t-1}^{NP}(R_{it}) + \phi_{OLSt} \hat{E}_{t-1}^{OLS}(R_{it}) + \eta_{it} \quad i = 1, 2, \dots, N. \quad (38)$$

Finally, the Fama-MacBeth coefficients are computed from the time series of regression coefficients.<sup>6</sup> The idea of the test is that under the null hypothesis that the conditional FF holds, we would expect the estimated coefficient on  $\hat{E}_{t-1}^{NP}(R_{it})$  to be close to unity provided that our nonparametric approach yields accurate estimates of both conditional covariances and market prices of risk. Like Ferson and Harvey (1999), we include  $\hat{E}_{t-1}^{OLS}(R_{it})$  in the regression to confront the model with a powerful empirically motivated alternative.

To put the results for the nonparametric model in perspective, we also estimate conditional expected returns from the parametric model. More specifically, we estimate the four parametric versions of the FF model considered by Ferson and Harvey (1999): (1) an unconditional FF model with betas estimated using expanding samples;<sup>7</sup> (2) an unconditional FF model with betas estimated using 60-month rolling samples; (3) a conditional FF model with betas that are linear in the lagged state variables estimated using expanding samples;<sup>8</sup> and (4) a conditional FF model with linear betas estimated using 60-month rolling samples. In all cases, rewards per unit of beta risk are estimated as the Fama-MacBeth coefficients from cross-sectional regressions of returns on betas

<sup>6</sup>Regression (38) is run 330 months, which means that 180 observations are lost: 60 months of prior data are used to estimate conditional covariances; a minimum of 60 months of prior data are then used to estimate market prices of risk; and the first 60 return forecasts are used to choose the bandwidths that minimize forecasting errors.

<sup>7</sup>The expanding sample used to estimate factor betas at time  $t$  includes all observations from the first month to month  $t - 1$ .

<sup>8</sup>Linear betas are estimated from the time series regression:  $R_{it} = a_{0i} + (b_{m0i} + b'_{m1i} X_{t-1}) R_{m,t} + (b_{smb0i} + b'_{smb1i} X_{t-1}) R_{smb,t} + (b_{hml0i} + b'_{hml1i} X_{t-1}) R_{hml,t} + \nu_{it}$ .

over the previous 60 months.<sup>9</sup> We use the notation  $\hat{E}_{t-1}^P(R_{it})$  to denote return forecasts implied by the parametric implementation of the FF model.

Tables 3 and 4 report results for the nonparametric and the parametric estimation methods, respectively. To complete the analysis, we also report results from univariate regressions. The regression coefficient on the forecast based on the nonparametric FF model is positive, with values ranging from 0.67 to 0.91, and statistically significant at least at the 5% significance level in all cases. Moreover, none of the intercepts is statistically significant. We interpret these results as evidence in favor of the flexible version of the conditional three-factor model. When the OLS-based forecast is included in the regression, the coefficient on  $\hat{E}_{t-1}^{NP}(R_{it})$  is still positive in all cases but its value decreases by half and becomes statistically insignificant. The coefficient on the OLS-based forecast, however, is statistically significant, although both its value and its statistical significance are lower than in the case in which  $\hat{E}_{t-1}^{OLS}(R_{it})$  is the only regressor.

When returns are regressed on forecasts obtained from different parametric implementations of the FF model (Table 4), estimated slope coefficients are positive in all cases but lower than the coefficients obtained from the univariate regressions of Table 3, and they are never statistically significant. Interestingly, the lowest coefficients are obtained when betas are estimated using rolling samples rather than expanding samples. Further, the conditional model in which betas are allowed to vary linearly with the conditioning variables estimated with rolling samples provides the poorest fit: The slope coefficient takes the lowest value and the intercept becomes significant at the 5% level. If model-based forecasts are confronted with the OLS-based forecast, the coefficient on  $\hat{E}_{t-1}^P(R_{it})$  becomes negative in all cases although not significant at any conventional significance level, while the coefficient on  $\hat{E}_{t-1}^{OLS}(R_{it})$  remains high and statistically significant at the 1% level in all cases.

We may conclude from these results that our nonparametric approach to estimating the conditional FF model improves substantially upon alternative parametric methods in terms of forecasting the cross-section of future returns. However, nonparametric FF-based forecasts are still dominated by forecasts obtained from simple OLS regressions of returns on lagged state variables.

## 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 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

---

<sup>9</sup>In unreported results we also compute prices of beta risk as the average of the previous 30 regression coefficients and obtain qualitatively similar results.



models.

Like previously proposed nonparametric estimation methods, the method presented in this paper is not subject to Ghysels' critique (Ghysels (1998), 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 empirical analysis, we first carry out a simulation study and then apply the method to data on equity returns. Both analyses are based on the 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 are consistent with prior evidence on the significance of market prices associated with the Fama-French risk factors. However, our results also suggest that inference based on constant market prices of risk may hide that fact that risk factors are significantly priced in specific subperiods. Finally, our nonparametric version of the Fama-French model does a much better job at forecasting the cross-section of future stock returns than previously proposed parametric implementations, although the performance of the Fama-French model declines when confronted with purely empirical forecasts.

## Appendix

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

**Lemma 1** *Under Assumptions (A2) to (A5), and (A8), it holds that*

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

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

### Proof of Lemma 1

To simplify notation we denote any generic scalar term of  $TK_{h,ts}C_s^{*'}C_s^*$  as  $Z_s = TK_{h,ts}C_s^{*'}C_s^*$ .  $Z_s$  is a  $\alpha$ -mixing sequence of size  $6/5$  with the proper bounded moments, and  $E(Z_s) = \frac{1}{h}K\left(\frac{t-s}{Th}\right)g_s$ , that tends to  $g_t$ . 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 Square Error

$$\begin{aligned} MSE(\widehat{\lambda}_t) &= trE[(\widehat{\lambda}_t - \lambda_t)(\widehat{\lambda}_t - \lambda_t)'] \\ &= \|Bias(\widehat{\lambda}_t)\|_2^2 + trVar(\widehat{\lambda}_t) = S^2(\widehat{\lambda}_t) + V(\widehat{\lambda}_t). \end{aligned}$$

Then, note that

$$\begin{aligned} \widehat{\lambda}_t - \lambda_t &= \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} R_s^* - \lambda_t \\ &= \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* (\lambda_s - \lambda_t) \\ &\quad + \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} \varepsilon_s \end{aligned}$$

has a random denominator. We overcome this problem redefining the bias and variance terms using the weight  $W_t^* = G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^*$ . Hence, the redefined bias is  $Bias^*(\widehat{\lambda}_t) = Bias(W_t^* \widehat{\lambda}_t)$ . For

technical reasons, we use different bandwidths for  $W_t^*$  and for  $\widehat{\lambda}_t$ , say  $h^*$  and  $h$  respectively, such that the following condition holds:

$$\frac{E \| W_t^* - I \|^2}{E \| \widehat{\lambda}_t - \lambda_t \|^2} = o(1), \quad (39)$$

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

Considering the term defined by  $Bias^*$

$$\begin{aligned} Bias^*(\widehat{\lambda}_t) &= G_t^{-1} \sum_{s=1}^T K_{h,ts} E(C_s^{*'} C_s^*) (\lambda_s - \lambda_t) + G_t^{-1} \sum_{s=1}^T K_{h,ts} E(C_s^{*'} \varepsilon_s^*) \\ &= G_t^{-1} \sum_{s=1}^T K_{h,ts} E(C_s^{*'} C_s^*) (\lambda_s - \lambda_t), \end{aligned}$$

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

$$\begin{aligned} Bias^*(\widehat{\lambda}_t) &= G_t^{-1} \sum_{s=1}^T K_{h,ts} G_s (\lambda_s - \lambda_t) \\ &= G_t^{-1} \int K(u) \left[ G_t - hu \frac{\partial G_t}{\partial t} + o(h^2) \right] \left[ -\frac{\partial \lambda_t}{\partial t} hu + \frac{1}{2} \frac{\partial^2 \lambda_t}{\partial t \partial t} (hu)^2 + o(h^2) \right] \\ &= \frac{1}{2} d_k h^2 \left( \frac{\partial^2 \lambda_t}{\partial t \partial t} + 2G_t^{-1} \frac{\partial G_t}{\partial t} \frac{\partial \lambda_t}{\partial t} \right) + o(h^2). \end{aligned}$$

Thus,

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

The variance term is given by

$$\begin{aligned} Var(\widehat{\lambda}_t) &= Var(\widehat{\lambda}_t - \lambda_t) = Var \left[ \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} R_s^* \right] \\ &= Var \left[ \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* (\lambda_s - \lambda_t) \right. \\ &\quad \left. + \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* \right)^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} \varepsilon_s \right], \end{aligned}$$

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

$$Var^*(\widehat{\lambda}_t) = Var \left[ G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* (\lambda_s - \lambda_t) \right] + Var \left[ G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} \varepsilon_s \right]. \quad (40)$$

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

$$\begin{aligned} V(\widehat{\lambda}_t) &= tr Var^*(\widehat{\lambda}_t) = tr Var \left[ G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* (\lambda_s - \lambda_t) \right] \\ &+ tr Var \left[ G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} \varepsilon_s \right] = V_1 + V_2. \end{aligned} \quad (41)$$

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

$$\begin{aligned} V_1 &= tr Var \left( G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* (\lambda_s - \lambda_t) \right) = tr G_t^{-1} Var \left( \sum_{s=1}^T K_{h,ts} C_s^{*'} C_s^* (\lambda_s - \lambda_t) \right) G_t^{-1} \\ &= tr G_t^{-1} E \left[ \sum_{s=1}^T K_{h,ts} (C_s^{*'} C_s^* - G_s) (\lambda_s - \lambda_t) \left( \sum_{s=1}^T K_{h,ts} (C_s^{*'} C_s^* - G_s) (\lambda_s - \lambda_t) \right)' \right] G_t^{-1} \\ &= tr G_t^{-1} E \left( \sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} (C_s^{*'} C_s^* - G_s) (\lambda_s - \lambda_t) (\lambda_{s'} - \lambda_t)' (C_{s'}^{*'} C_{s'}^* - G_{s'}) \right) G_t^{-1} \\ &= tr \sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} (\lambda_s - \lambda_t) (\lambda_{s'} - \lambda_t)' E [(C_{s'}^{*'} C_{s'}^* - G_{s'}) G_t^{-1} G_t^{-1} (C_s^{*'} C_s^* - G_s)] \\ &= tr \sum_{s=1}^T \sum_{s'=1}^T K_{h,ts} K_{h,ts'} (\lambda_s - \lambda_t) (\lambda_{s'} - \lambda_t)' Q_{s,s'}, \end{aligned} \quad (42)$$

where  $Q_{s,s'} = E [(C_{s'}^{*'} C_{s'}^* - G_{s'}) G_t^{-1} G_t^{-1} (C_s^{*'} C_s^* - G_s)]$  is a bounded  $p$ -order square matrix. Expression (42) can be divided in two parts, those corresponding to same terms and the cross terms.

When  $s = s'$

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

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

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

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

$$tr \sum_{\substack{s,s'=1 \\ s \neq s'}}^T K_{h,ts} K_{h,ts'} (\lambda_s - \lambda_t) (\lambda_{s'} - \lambda_t)' Q_{s,s'}$$

has same order as

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

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

For the second term in (41) and taking into account that  $E(\varepsilon_s^* | C_s^*) = 0$ :

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

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

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

Then, as  $E(\varepsilon_s^* \varepsilon_s^* | C_s^*) = I$  and using the result (39) of Lemma 1 we have that

$$V_2 = \text{trVar} \left[ G_t^{-1} \sum_{s=1}^T K_{h,ts} C_s^{*'} \varepsilon_s^* \right] = \frac{c_k}{Th} \text{tr} G_t^{-1} G_t G_t^{-1} + o((Th)^{-1}) = \frac{c_k}{Th} \text{tr} G_t^{-1} + o((Th)^{-1}).$$

Finally, since the order of  $V_1$  is negligible with respect to  $V_2$  and (39) holds, we have

$$V(\hat{\lambda}) = \frac{c_k}{Th} \text{tr} G_t^{-1} + o((Th)^{-1}),$$

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  $\Omega_s$  implies the equivalence between the asymptotic properties of  $\hat{\lambda}_t^{FGLS}$  and  $\hat{\lambda}_t$ .  $\square$

### Proof of Proposition 1

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

$$\text{Bias}^*(\hat{c}_{il}(X_{t-1}) | X_{t-1} = x_{t-1}) = \text{Bias} \left[ \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} K_B(X_{s-1} - x_{t-1}) \hat{c}_{il}(x_{t-1}) \right]$$

with  $\tau = t/T$ . Then

$$\text{Bias}^*(\hat{c}_{il}(X_{t-1}) | X_{t-1} = x_{t-1}) = \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} (E[K_B(X_{s-1} - x_{t-1}) P_{ils}] - f(\tau, x'_{t-1}) c_{il}(x_{t-1}))$$

and, since  $P_{ils} = c_{il}(X_{s-1}) + u_s$  with  $E(u_s | X_{s-1}) = 0$ :

$$\begin{aligned} \text{Bias}^*(\hat{c}_{il}(X_{t-1}) | X_{t-1} = x_{t-1}) &= \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} (E[K_B(X_{s-1} - x_{t-1}) c_{il}(X_{s-1})] - f(\tau, x'_{t-1}) c_{il}(x_{t-1})) \\ &= \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \left[ \int K_B(\omega - x_{t-1}) c_{il}(\omega) f(s/T, \omega') d\omega - f(t/T, x'_{t-1}) c_{il}(x_{t-1}) \right]. \end{aligned}$$

Now, define

$$f_s(x) \equiv f(s/T, x),$$

$$\mathcal{D}_c(x_{t-1}) \equiv \left( \frac{\partial c_{i\ell}(x_{t-1})}{\partial x_{1t-1}} \cdots \frac{\partial c_{i\ell}(x_{t-1})}{\partial x_{pt-1}} \right)',$$

$$\mathcal{D}_{f_s}(x'_{t-1}) \equiv \left( \frac{\partial f_s(x'_{t-1})}{\partial x_{1t-1}} \cdots \frac{\partial f_s(x'_{t-1})}{\partial x_{pt-1}} \right),$$

and the  $(p \times p)$ -order matrices  $\mathcal{H}_c(x_{t-1})$  and  $\mathcal{H}_{f_s}(x'_{t-1})$  having as generic terms,  $(j, j')$ ,  $\frac{\partial^2 c_{i\ell}(x_{t-1})}{\partial x_{jt} \partial x_{j't}}$  and  $\frac{\partial^2 f_s(x'_{t-1})}{\partial x_{jt} \partial x_{j't}}$  respectively.

Using a standard multivariate kernel of order two and the Lipschitz condition for the density  $f$ ; we have that

$$\begin{aligned} & \text{Bias}^*(\widehat{c}_{i\ell}(X_{t-1}) | X_{t-1} = x_{t-1}) = \\ &= \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \left[ \int K(z) c_{i\ell}(x_{t-1} + B^{1/2}z) f_s(x'_{t-1} + B^{1/2}z') dz - f_t(x'_{t-1}) c_{i\ell}(x_{t-1}) \right] \\ &= \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \left[ \int K(z) \left( c_{i\ell}(x_{t-1}) + (B^{1/2}z)' \mathcal{D}_c(x_{t-1}) + \frac{1}{2} (B^{1/2}z)' \mathcal{H}_c(x_{t-1}) (B^{1/2}z) + o(\text{trace}(B)) \right) \times \right. \\ & \quad \left. \left( f_s(x'_{t-1}) + (B^{1/2}z')' \mathcal{D}_{f_s}(x'_{t-1}) + \frac{1}{2} (B^{1/2}z')' \mathcal{H}_{f_s}(x'_{t-1}) (B^{1/2}z') + o(\text{trace}(B)) \right) dz - f_t(x'_{t-1}) c_{i\ell}(x_{t-1}) \right] \\ &= \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \frac{1}{2} \text{tr} \left( B H_c(x_{t-1}) \int K(z) z z' dz \right) f_s(x'_{t-1}) \\ &+ \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} c_{i\ell}(x_{t-1}) (f_s(x'_{t-1}) - f_t(x'_{t-1})) + O(\text{trace}(B)) \\ &= \frac{1}{rf(\tau, x'_{t-1})} \frac{\mu_K}{2} \text{tr}(B H_c(x_{t-1})) \sum_{s=t-r}^{t-1} f_s(x'_{t-1}) \\ &+ \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} c_{i\ell}(x_{t-1}) (f_s(x'_{t-1}) - f_t(x'_{t-1})) + O(\text{trace}(B)) \\ &= \frac{\mu_K}{2} \text{tr}(B H_c(x_{t-1})) + \frac{1}{rf(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} c_{i\ell}(x_{t-1}) O\left(\frac{r}{T}\right) + O(\text{trace}(B)) \\ &= O(\text{trace}(B)) + O\left(\frac{r}{T}\right). \end{aligned}$$

Next, we obtain the redefined variance for a generic term  $\widehat{c}_{i\ell}(X_{t-1})$ :

$$\begin{aligned}
\text{Var}^*(\widehat{c}_{il}(X_{t-1})|X_{t-1} = x_{t-1}) &= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \text{Var} \left[ \sum_{s=t-r}^{t-1} K_B(X_{s-1} - x_{t-1}) p_{ils} \right] \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} \text{Var} (K_B(X_{s-1} - x_{t-1}) p_{ils} | x_{t-1}) \right. \\
&\quad \left. + \sum_{\substack{s, s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} (K_B(X_{s-1} - x_{t-1}) p_{ils}, K_B(X_{s'-1} - x_{t-1}) p_{ils'} | x_{t-1}) \right] \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} \text{Var} (K_B(X_{s-1} - x_{t-1}) (c_{il}(x_{s-1}) + u_{ils}) | x_{t-1}) \right. \\
&\quad \left. + \sum_{\substack{s, s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} (K_B(X_{s-1} - x_{t-1}) (c_{il}(x_{s-1}) + u_{ils}), K_B(X_{s'-1} - x_{t-1}) (c_{il}(x_{s'-1}) + u_{ils'}) | x_{t-1}) \right] \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} \text{Var} (K_B(X_{s-1} - x_{t-1}) u_{ils} | x_{t-1}) + \sum_{s=t-r}^{t-1} \text{Var} (K_B(X_{s-1} - x_{t-1}) c_{il}(x_{s-1}) | x_{t-1}) \right. \\
&\quad \left. + \sum_{\substack{s, s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} (K_B(X_{s-1} - x_{t-1}) c_{il}(X_{s-1}), K_B(X_{s'-1} - x_{t-1}) c_{il}(X_{s'-1}) | x_{t-1}) \right. \\
&\quad \left. + \sum_{\substack{s, s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} (K_B(X_{s-1} - x_{t-1}) u_{ils}, K_B(X_{s'-1} - x_{t-1}) u_{ils'} | x_{t-1}) \right] \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} \text{Var} (K_B(X_{s-1} - x_{t-1}) u_{ils} | x_{t-1}) + \sum_{s=t-r}^{t-1} \text{Var} (K_B(X_{s-1} - x_{t-1}) c_{il}(x_{s-1}) | x_{t-1}) \right. \\
&\quad \left. + \sum_{\substack{s, s'=t-r \\ s \neq s'}}^{t-1} \text{Cov} (K_B(X_{s-1} - x_{t-1}) c_{il}(X_{s-1}), K_B(X_{s'-1} - x_{t-1}) c_{il}(X_{s'-1}) | x_{t-1}) \right] = T_1 + T_2 + T_3,
\end{aligned}$$

since for  $s \neq s'$  the conditional expectation  $E(u_{ils} u_{ils'})$  cancels and, therefore, only the diagonal



terms remain. For  $T_1$

$$\begin{aligned}
T_1 &= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} E(K_B^2(X_{s-1} - x_{t-1}) E(u_{i\ell s}^2 | X_{s-1}) | x_{t-1}) \right] \\
&= \frac{\sigma_{u_{i\ell}}^2}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} E(K_B^2(X_{s-1} - x_{t-1}) | x_{t-1}) \right] \\
&= \frac{\sigma_{u_{i\ell}}^2}{r^2 f^2(\tau, x'_{t-1})} \left[ \sum_{s=t-r}^{t-1} \int K_B^2(z - x_{t-1}) f(s/T, z') dz \right] \\
&= \frac{\sigma_{u_{i\ell}}^2}{r^2 f^2(\tau, x'_{t-1}) |B|^{1/2}} \sum_{s=t-r}^{t-1} \int K^2(u) \left( f(\tau, x'_{t-1}) + O(\text{trace} B^{1/2}) + O\left(\frac{r}{T}\right) \right) du \\
&= \frac{\sigma_{u_{i\ell}}^2}{r f(\tau, x'_{t-1}) |B|^{1/2}} \int K^2(u) du + h.o.t. = O\left(\frac{1}{r |B|^{1/2}}\right) + h.o.t.
\end{aligned}$$

For  $T_2$

$$\begin{aligned}
T_2 &= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \text{Var}(K_B(X_{s-1} - x_{t-1}) c_{i\ell}(x_{s-1}) | x_{t-1}) \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \left[ \int K_B^2(w - x_{t-1}) c_{i\ell}^2(w) f_s(w') dw - \left( \int K_B(w - x_{t-1}) c_{i\ell}(w) f_s(w') dw \right)^2 \right] \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1})} \sum_{s=t-r}^{t-1} \left[ |B|^{-1/2} \int K^2(z) c_{i\ell}^2(x_{t-1} + B^{1/2} z) f_s(x'_{t-1} + B^{1/2} z') dz \right. \\
&\quad \left. - \left( \int K(z) c_{i\ell}(x_{t-1} + B^{1/2} z) f_s(x'_{t-1} + B^{1/2} z') dz \right)^2 \right] \\
&= \frac{1}{r^2 f^2(\tau, x'_{t-1}) |B|^{1/2}} \sum_{s=t-r}^{t-1} \left[ \left( c_{i\ell}^2(x_{t-1}) f_s(x'_{t-1}) \int K^2(u) du + O(\text{trace} B^{1/2}) \right) \right. \\
&\quad \left. - |B|^{1/2} \left( c_{i\ell}(x_{t-1}) f_s(x'_{t-1}) + O(\text{trace} B^{1/2}) \right)^2 \right] \\
&= \frac{c_{i\ell}^2(x_{t-1})}{r^2 f^2(\tau, x'_{t-1}) |B|^{1/2}} \int K^2(u) du \sum_{s=t-r}^{t-1} \left( f(\tau, x'_{t-1}) + O(\text{trace} B^{1/2}) + O\left(\frac{r}{T}\right) \right) + h.o.t \\
&= \frac{c_{i\ell}^2(x_{t-1})}{r f(\tau, x'_{t-1}) |B|^{1/2}} \int K^2(u) du + h.o.t. = O\left(\frac{1}{r |B|^{1/2}}\right) + h.o.t.
\end{aligned}$$

And finally for the third term,  $T_3$ ,

$$T_3 = \frac{1}{r^2 f^2(\tau, x'_{t-1})} \sum_{\substack{s, s'=t-r \\ s \neq s'}}^{t-1} \text{Cov}[K_B(X_{s-1} - x_{t-1}) c_{i\ell}(X_{s-1}), K_B(X_{s'-1} - x_{t-1}) c_{i\ell}(X_{s'-1}) | x_{t-1}].$$

Using (A4)

$$\sum_{k=1}^r Cov [K_B(X_{s-1} - x_{t-1})c_{i\ell}(X_{s-1}), K_B(X_{s+k} - x_{t-1})c_{i\ell}(X_{s+k})|x_{t-1}]$$

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 final expression for each  $(i, \ell)$  variance term is

$$Var^*(\widehat{c}_{i\ell}(X_{t-1})|X_{t-1} = x_{t-1}) = \frac{c_{i\ell}^2(x_{t-1}) + \sigma_{u_{i\ell}}^2}{rf(\tau, x'_{t-1})|B|^{1/2}} \int K^2(u)du + h.o.t.$$

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 covariances instead of the real ones. First, note that (A4) holds for the estimated covariances ( $\widehat{C}$ ) and that (A5) holds up to order  $o(1)$ ; that is,  $E(\widehat{C}_t^{*'}\widehat{C}_t^*) = E(C_t^{*'}C_t^*) + o(1) = G_t + o(1)$ .

Now, the steps of the proof of Theorem 1 follow straightforward using  $\widehat{C}$  instead of  $C$ . Only the second term for the variance (41) need an extra step.

The second term for the variance can be written as,

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

since  $\varepsilon_s$  is independent of the past information. Using the fact that  $E(\widehat{C}_t^{*'}\widehat{C}_t^*) = E(C_t^{*'}C_t^*) =$

$G_t + o(1)$ , it finally holds

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

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) to (A5) and (C1) to (C3); it holds that*

$$\frac{1}{r} \sum_{s=t-r}^{t-1} K_G(X_{s-1} - x_{t-1}) \xrightarrow{\text{a.s.}} f(\tau, x_{t-1}), \quad (43)$$

where  $\tau = t/T$ .

### Proof of Lemma 2

Following similar steps than in Lemma 1, define  $Z_s = K_G(X_{s-1} - x_{t-1})$ . The sequence  $Z_s$  has mean  $f(s/T, x_{s-1})$  and therefore  $E \left( (1/r) \sum_{s=t-r}^{t-1} Z_s \right) = f(\tau, x_{t-1}) + o(1)$ . 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

$$Z_t = \sum_{s=1}^T K_{h,ts} \widehat{C}_s^{*t} \varepsilon_s^*. \quad (44)$$

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. and J. Chen, 2007, CAPM over the Long Run: 1926–200. *Journal of Empirical Finance* 14, 1–40.
- Black, F. and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy* 81, 637–654.
- Bollerslev, T., R. Engle, and J. Wooldridge, 1988, A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy* 96, 116—131.
- Brandt, M. and D. Chapman, 2006, Linear Approximations and Tests of Conditional Pricing Models, National Bureau of Economic Research Cambridge, Mass., USA.
- Cai, Z., 2007, Trending Time-Varying Coefficient Time Series Models with Serial Correlated Errors. *Journal of Econometrics* 136, 163–188.
- Cheridito, P., D. Filipović, and R. Kimmel, 2007, Market Price of Risk Specifications for Affine Models: Theory and Evidence. *Journal of Financial Economics* 83, 123–170.
- Cox, J., J. Ingersoll, and S. Ross, 1985, An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica* 53, 363–384.
- Dalhaus, R., 1997, Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics* 25, 1–37.
- Dalhaus, R., 2000, A Likelihood Approximation for Locally Stationary Processes. *The Annals of Statistics* 28, 1762—1794.
- Fama, E. and K. French, 1989, Business Conditions and the Expected Returns on Bonds and Stocks. *Journal of Financial Economics* 25, 23–50.
- Fama, E. and K. French, 1993, Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fama, E. and J. MacBeth, 1973, Risk, Return, and Equilibrium: Empirical Test. *Journal of Political Economy* 81, 607–636.
- Fan, J. and I. Gijbels, 1996, Local Polynomial Modelling and its Applications, Chapman and Hall, London.

- Ferson, W., 1989, Changes in Expected Security Returns, Risk, and the Level of Interest Rates. *Journal of Finance* 44, 1191–1217.
- Ferson, W. and C. Harvey, 1991, The Variation of Economic Risk Premiums. *Journal of Political Economy* 99, 385–415.
- Ferson, W. and C. Harvey, 1999, Conditioning Variables and the Cross Section of Stock Returns. *Journal of Finance* 54, 1325–1360.
- Ferson, W. and R. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance* 51, 425–461.
- Grauer, R.R., and J.A. Janmaat, 2009, On the Power of Cross-Sectional and Multivariate Tests of the CAPM. *Journal of Banking & 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. and Z. Wang, 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 SandP 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. and R. Stambaugh, 1986, Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics* 17, 357–390.
- Lewellen, J. and S. Nagel, 2006, The Conditional CAPM does not Explain Asset-Pricing Anomalies. *Journal of Financial Economics* 82, 289–314.
- Lettau, M. and S. Ludvigson, 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.
- Orbe, S., E. Ferreira, and J. Rodriguez-Póo, 2005, Nonparametric Estimation of Time Varying Parameters under Shape Restrictions. *Journal of Econometrics* 126, 53–77.
- Orbe, S., E. Ferreira, and J. Rodriguez-Póo, 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 P. Hackl, ed., *Statistical Analysis and Forecasting Economic Structural Change*. Springer-Verlag, 253–264.
- 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. and G. Zhou, 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.
- Wand, M. and M. Jones, 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. and I. Domowitz, 1984, Nonlinear Regression with Dependent Observations. *Econometrica* 52, 143–161.

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.

**Table 1**

Monte Carlo simulation results.

	$\% (MSE^{NP} < MSE^{ROLL})$	$\overline{MSE}^{NP}$	$\overline{MSE}^{ROLL}$
Model 1	97.4	0.0135	0.0298
Model 2	84.9	0.0182	0.0278
Model 3	82.5	0.0275	0.0356
Model 4	81.5	0.0221	0.0300

*Note:* This table reports results from a Monte Carlo simulation study of the nonparametric (NP) and the rolling (ROLL) estimators of market betas in a conditional Fama-French three factor model. Market betas are generated under four different parametric models: 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). For each model, column 2 reports the percentage of simulations for which Mean Square Errors (MSE) are lower for the nonparametric estimator of beta than for the rolling beta, column 3 reports MSE for the nonparametric estimator averaged across all simulations, and column 4 reports MSE for the rolling estimator averaged across all simulations.



**Table 2**

Estimation results: nonparametric market prices of risk.

	$\hat{\lambda}_m$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$
July 1968 - December 1975			
Min	-0.012	-0.090	-0.061
Mean	0.027	-0.047	0.019
S.D.	0.034	0.018	0.055
Max	0.078	-0.017	0.095
% Positive	62.22	0.00	55.56
% Positive and Significant (5% level)	1.11	0.00	28.89
January 1976 - December 1985			
Min	-0.011	-0.042	-0.005
Mean	0.011	0.041	0.059
S.D.	0.018	0.037	0.044
Max	0.054	0.093	0.126
% Positive	68.33	81.67	88.33
% Positive and Significant (5% level)	0.00	0.00	31.67
January 1986 - December 1995			
Min	0.019	-0.062	-0.004
Mean	0.040	-0.035	0.068
S.D.	0.020	0.019	0.040
Max	0.089	0.004	0.122
% Positive	100.00	5.83	96.67
% Positive and Significant (5% level)	5.83	0.00	19.17
January 1996 - December 2005			
Min	0.026	-0.117	-0.014
Mean	0.043	-0.023	0.057
S.D.	0.024	0.052	0.032
Max	0.102	0.026	0.097
% Positive	100.00	55.83	90.83
% Positive and Significant (5% level)	15.83	0.00	35.83
Full Sample			
Min	-0.012	-0.117	-0.061
Mean	0.031	-0.014	0.053
S.D.	0.027	0.049	0.046
Max	0.102	0.093	0.126
% Positive	84.00	38.22	84.67
% Positive and Significant (5% level)	6.00	0.00	28.89

*Note:* This table reports summary statistics of nonparametrically estimated market prices of risk in a flexible conditional Fama-French model. Conditional covariances between asset returns (the 25 size and book-to-market portfolios) and risk factors (market, size, and book-to-market) are estimated nonparametrically as a function of the lagged dividend yield and the slope of the term structure of interest rates.

**Table 3**

Forecasting results: nonparametric approach.

Constant	$\hat{E}_{t-1}^{NP}(R_{it})$	$\hat{E}_{t-1}^{OLS}(R_{it})$
Panel A. OLS-based forecast		
0.612** (0.284)		0.457*** (0.133)
Panel B. Conditioning variables for nonparametric FF: <i>DP, term</i>		
0.377 (0.362)	0.909*** (0.333)	
0.294 (0.381)	0.461 (0.309)	0.302** (0.129)
Panel C. Conditioning variables for nonparametric FF: <i>DP, def</i>		
0.424 (0.362)	0.668** (0.313)	
0.338 (0.369)	0.331 (0.271)	0.255** (0.125)
Panel D. Conditioning variables for nonparametric FF: <i>DP, Tb1m</i>		
0.447 (1.243)	0.741** (2.189)	
0.420 (0.368)	0.350 (0.299)	0.261** (0.127)
Panel E. Conditioning variables for nonparametric FF: <i>DP, hb3</i>		
0.443 (0.350)	0.719** (0.310)	
0.296 (0.363)	0.360 (0.272)	0.274** (0.126)

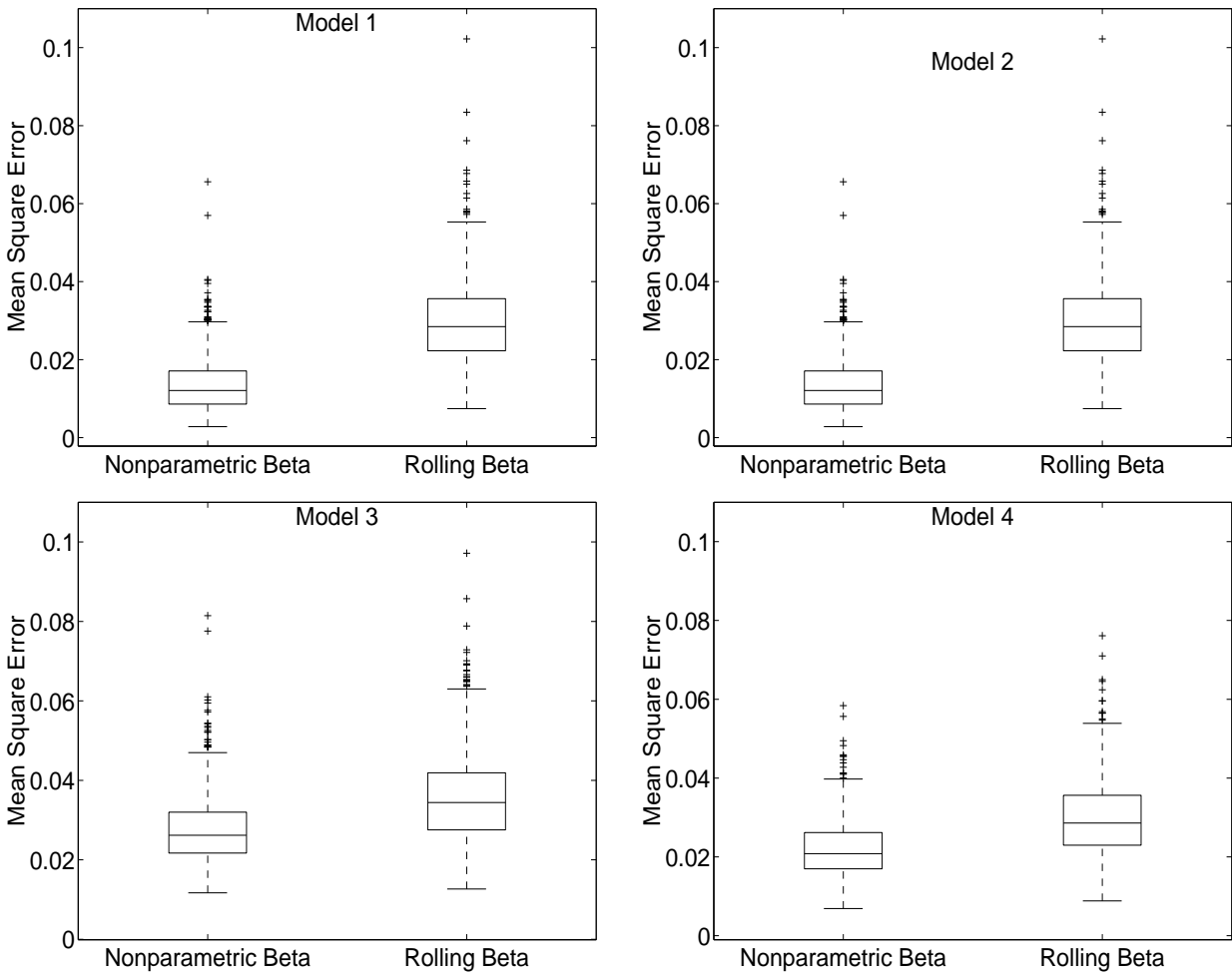
*Note:* This table reports Fama-MacBeth coefficients and standard errors computed from cross-sectional regressions of monthly returns on one-month-ahead return forecasts computed using past information.  $\hat{E}_{t-1}^{OLS}(R_{it})$  denotes the return forecast obtained from an OLS time-series regression of asset returns on lagged values of the five conditioning variables: *DP*; *term*; *def*; *Tb1m*; and *hb3*.  $\hat{E}_{t-1}^{NP}(R_{it})$  denotes the return forecast implied by the conditional Fama-French factor estimated nonparametrically using *DP* and *term* as conditioning variables. The sample period is from July 1978 to December 2005. One, two, and three asterisks denote statistical significance at the 10%, 5%, and 1% significance level, respectively.

**Table 4**

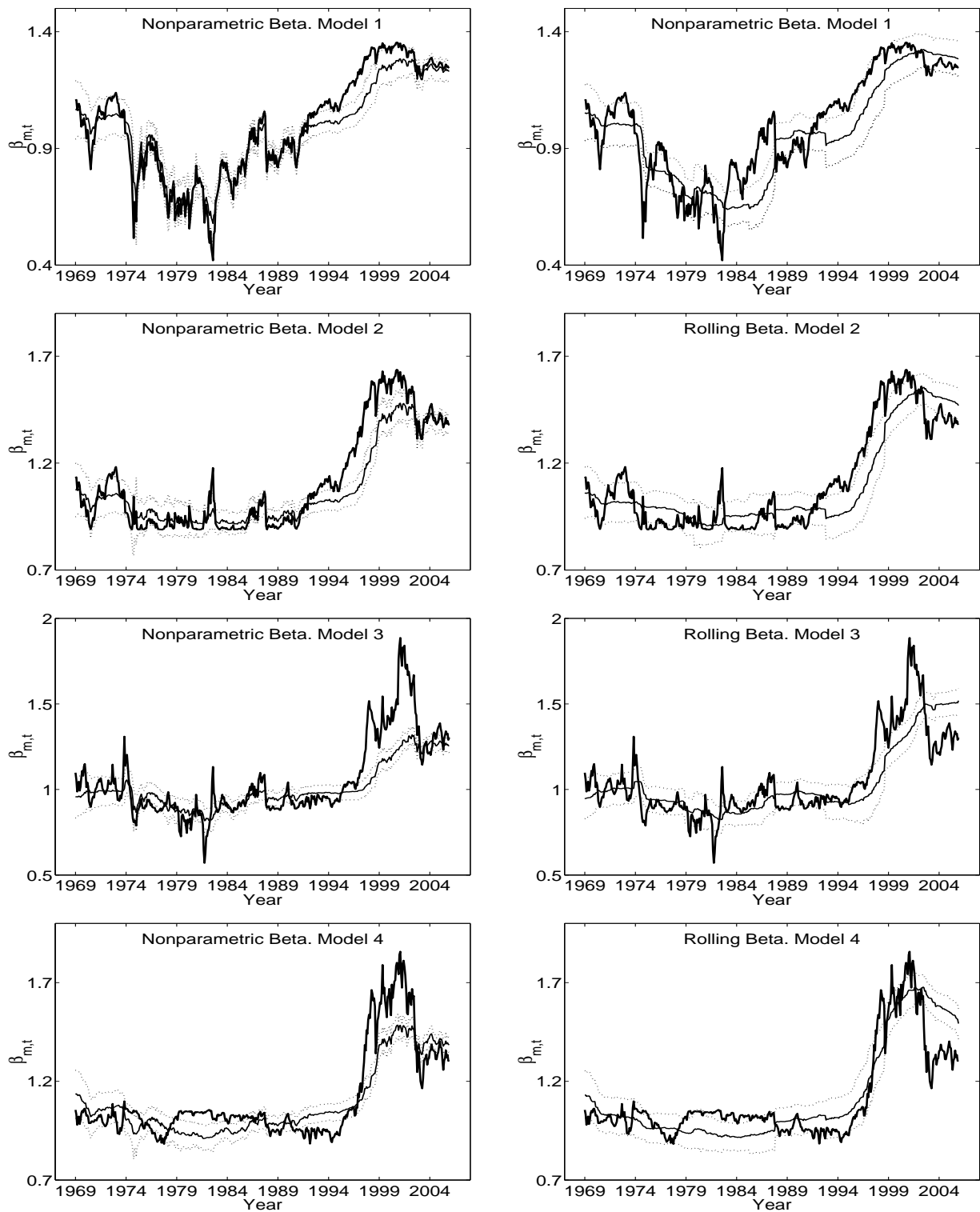
Forecasting results: parametric approach.

Constant	$\hat{E}_{t-1}^P(R_{it})$	$\hat{E}_{t-1}^{OLS}(R_{it})$
Panel A. OLS-based forecast		
0.612** (0.284)		0.457*** (0.133)
Panel B. Constant betas, expanding samples		
0.536 (0.469)	0.417 (0.398)	
0.742 (0.449)*	-0.056 (0.362)	0.346*** (0.127)
Panel C. Constant betas, rolling samples		
0.676* (0.403)	0.289 (0.329)	
0.833** (0.392)	-0.061 (0.296)	0.362*** (0.127)
Panel C. Linear betas, expanding samples		
0.623 (0.417)	0.335 (0.323)	
0.804 (0.401)	-0.091 (0.293)	0.367*** (0.132)
Panel E. Linear betas, rolling samples		
0.780** (0.328)	0.150 (0.208)	
0.797** (0.319)	-0.152 (0.183)	0.431*** (0.131)

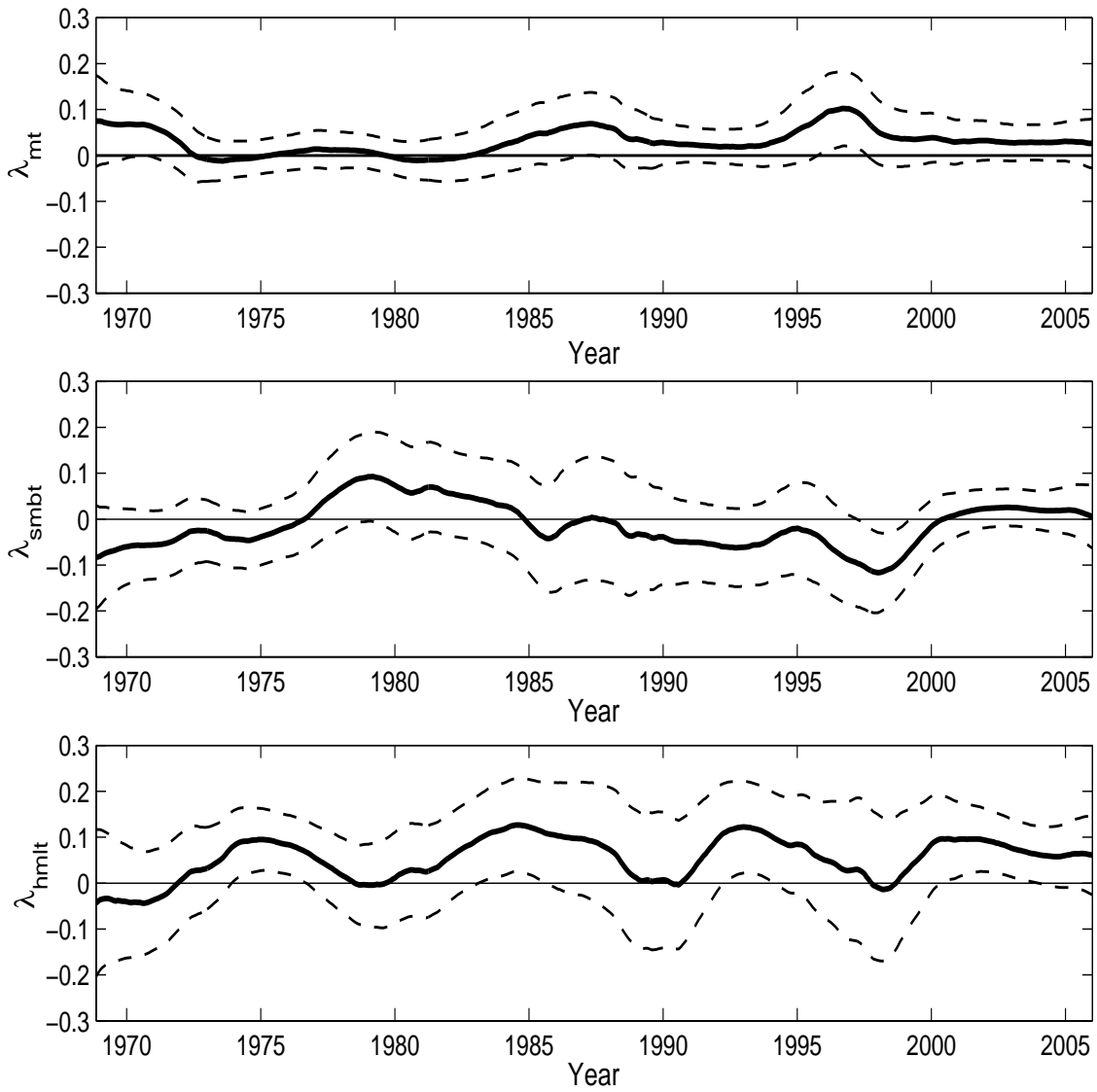
*Note:* This table reports Fama-MacBeth coefficients and standard errors computed from cross-sectional regressions of monthly returns on return forecasts computed using past information.  $\hat{E}_{t-1}^{OLS}(R_{it})$  denotes the one-month-ahead return forecast obtained from an OLS time-series regression of asset returns on lagged values of five conditioning variables: *DP*; *term*; *def*; *Tb1m*; and *hb3*.  $\hat{E}_{t-1}^P(R_{it})$  denotes the return forecast implied by a parametric version of the Fama-French model. Returns are expressed in percentages. The sample period is from July 1978 to December 2005. One, two, and three asterisks denote statistical significance at the 10%, 5%, and 1% significance level, respectively.



**Fig. 1.** Monte Carlo Study: Boxplots of Mean Square Errors of nonparametric and rolling beta estimators. Market betas are generated under four different parametric models: 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).



**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 under four different models: 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).



**Fig. 3.** Nonparametric estimates of market prices of risk associated with the market risk factor,  $\lambda_{m,t}$  (top panel), the size risk factor,  $\lambda_{smb,t}$  (middle panel), and the book-to-market risk factor,  $\lambda_{hml,t}$  (bottom panel) in the period from July 1968 to December 2005. Dashed lines represent 95% confidence bands.