

Article

Hierarchical Time-Varying Estimation of Asset Pricing Models

Richard T. Baillie ^{1,2,3,*}, Fabio Calonaci ^{4,†} and George Kapetanios ^{1,†}

¹ King's Business School, King's College London, 30 Aldwych, London WC2B 4BG, UK; g.kapetanios@kcl.ac.uk

² Department of Economics, Michigan State University, East Lansing, MI 48825, USA

³ Rimini Center for Economic Analysis, Via Angherà 22, 47921 Rimini, Emilia-Romagna, Italy

⁴ School of Economics and Finance, Queen Mary University of London, London E1 4NS, UK; f.calonaci@qmul.ac.uk

* Correspondence: baillie@msu.edu

† These authors contributed equally to this work.

Abstract: This paper presents a new hierarchical methodology for estimating multi factor dynamic asset pricing models. The approach is loosely based on the sequential Fama–MacBeth approach and developed in a kernel regression framework. However, the methodology uses a very flexible bandwidth selection method which is able to emphasize recent data and information to derive the most appropriate estimates of risk premia and factor loadings at each point in time. The choice of bandwidths and weighting schemes are achieved by a cross-validation procedure; this leads to consistent estimators of the risk premia and factor loadings. Additionally, an out-of-sample forecasting exercise indicates that the hierarchical method leads to a statistically significant improvement in forecast loss function measures, independently of the type of factor considered.

Keywords: asset pricing model; Fama–MacBeth model; estimation of beta; kernel-weighted regressions; cross-validation; time-varying parameter regressions



Citation: Baillie, Richard T., Fabio Calonaci, and George Kapetanios. 2022. Hierarchical Time-Varying Estimation of Asset Pricing Models. *Journal of Risk and Financial Management* 15: 14. <https://doi.org/10.3390/jrfm15010014>

Academic Editor: Thanasis Stengos

Received: 11 September 2021

Accepted: 18 November 2021

Published: 4 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concept of a time-varying risk premium is a standard idea in the literature of asset pricing finance. For example, see [Campbell and Shiller \(1988\)](#), [Ferson and Harvey \(1991\)](#), [Lewellen and Nagel \(2006\)](#) and many others. The fundamental method in empirical finance is due to [Fama and MacBeth \(1973\)](#), who estimated equity risk premia by a cross-sectional regression method, where the pricing of different types of risk factors is to be assumed constant.

This paper extends the [Fama and MacBeth \(1973\)](#) approach by developing a sequential hierarchical structure, which allows the risk factor estimates to change over time in a flexible yet tractable manner, inside a kernel-weighted regression framework. This method maintains the [Fama and MacBeth \(1973\)](#) stages of estimating risk factors (or betas) from a time-series regression and then the factor loadings (or gammas) from cross-sectional regressions. An additional aspect of the methodology presented here is an additional stage for the selection of optimal bandwidths via a cross-validation procedure. One contribution is to employ a flexible approach for bandwidth selection, which essentially determines the speed of updates of the betas (risk factors). A similar methodology is also applied to determine the factor loadings, identifying an optimal time-varying bandwidth level optimised for each asset at each time point. This avoids imposing any a priori structure and allows us to incorporate economic and financial change, that is relevant for the pricing of assets, in a natural, data-orientated way. The approach is really hierarchical, since there is an initial consideration of structural change by permitting time variation in the estimation of the parameters and then, in a second stage, the bandwidth is also allowed to be internally changed. The method can also be seen as an extension of the least squares rolling window regression approach, which has extensively been used in empirical finance (e.g.,

Jagannathan and Wang (1996) and Lewellen and Nagel (2006)). The empirical results in this study overwhelmingly indicate the importance of removing the restriction of constant betas, in line with Fama and French (2020); the full superiority of the hierarchical methodology becomes apparent in terms of prediction of out-of-sample returns for a wide range of assets. The results show that the time variation of risk associated with stocks and portfolios can be captured with the flexible methods described in this paper. It is shown that the hierarchical approach in this paper is able to produce an increase in the forecasting performance between 4% and 7% greater than the conventional methods.

The remainder of this paper is organized as follows: Section 2 provides a discussion of the contribution of this paper and also describes the standard Fama and MacBeth (1973) approach. Section 3 presents the hierarchical procedure. After describing the data in Section 4, the empirical application of the methodology is described in the next section, which also includes a series of robustness checks.

2. Background Literature

The benchmark Capital Asset Pricing Model, or (CAPM), by Sharpe (1964), Lintner (1965) and Markowitz (1968), implies that the expected excess return on any asset is influenced by its sensitivity to the market, which is measured by the beta coefficient, times the market risk premia. Traditionally, this beta is considered invariant over time and represents the covariance between the return of the asset and the return on the market portfolio. The basic model has been criticised by Black et al. (1972), Fama and French (1992) and Fama and MacBeth (1973), among others, on the grounds that only one factor, the market beta, is inadequate to describe the systemic risk. Hence, many researchers have attempted to improve the basic CAPM by the introduction of other factors. Most notably, there is the three-factor model by Fama and French (1992), which introduced the size, or SMB factor (positive returns are related to small size), and the high minus low, HML, factor (high book-to-market ratios are associated with higher returns). On the other hand, Carhart (1997) introduced a fourth momentum factor, MOM, which describes the tendency of a stock price to continue recent trends. Several other factors have been proposed and investigated in the literature; see Harvey et al. (2016) and Harvey and Liu (2019) for more details.

Further developments with extending the basic CAPM have centred on implementing more flexible estimation strategies where the beta coefficient(s) are not necessarily assumed to be constant across time or space. For example, see Harvey (1989), Ferson and Harvey (1991), Bollerslev et al. (1988) and Fama and French (1997), Fama and French (2006), who have suggested that a constant beta estimated using OLS does not capture the dynamics of the beta and is unable to satisfactorily explain the cross-section of average returns on equities.

Adrian and Franzoni (2005) argue that models without time-evolving betas fail to capture investor characteristics and may lead to inaccurate estimates of the true underlying risk. There are numerous factors that contribute to the variation in beta, including regulation, economic and monetary policies, and exchange rates. Many researchers, such as Zolotoy (2011), show that variations in betas are more evident around important news announcements. Jagannathan and Wang (1996), Lettau and Ludvigson (2001b) and Beach (2011) show that the conditional CAPM with time-varying beta generally outperforms an unconditional CAPM with a constant beta.

One technique that is often used is to take into account changes in the systematic risk of an asset through a rolling window OLS regression (e.g., Fama and MacBeth (1973) and Lewellen and Nagel (2006)). While the former paper uses monthly returns over a five-year window, the latter employs returns at different horizons to capture the different rate of variation of risk over a variety of interval lengths (monthly, quarterly and semi-annually). The main difficulty of the rolling window regression approach is the attempt to capture local variations by having short intervals of data, which is incompatible with the desire of having tight standard errors, hence tight confidence intervals on the estimated beta parameters. Other researchers have directly exploited the covariation between the market

and other assets; e.g., [Engle \(2002\)](#) and [Bali and Engle \(2010\)](#) estimated time-varying betas using multivariate dynamic conditional correlation methods to exploit correlations between cross-sectional average returns of various factor portfolios. The usage of a realized beta allows us to adjust information instantaneously.

As previously described, the variation in the beta coefficients can be modelled through the evolution of the conditional distribution returns as a function of lagged state variables (see [Jagannathan and Wang \(1996\)](#), [Ferson and Harvey \(1999\)](#) and [Adrian and Franzoni \(2005\)](#), among others). In all cases, the authors explicitly specify the covariance between the market and portfolio returns as affine functions of pre-determined state variables. [Jagannathan and Wang \(1996\)](#), instead, develop a conditional version of the CAPM, augmented by a human–capital factor, and show that it explains a substantial fraction of the cross-sectional variation in the returns on 100 portfolios sorted by size and book-to-market ratio. Further, [Adrian and Franzoni \(2005\)](#) admit unobservable long-run changes in risk factor loadings, given by a learning process of rational investors. Recently, [Fama and French \(2020\)](#) have shown that models that use only cross-sectional factors provide better descriptions of average returns than time-series models that use time-series factors. This has been proven to be valid when considering prespecified and optimised time-varying loadings. The main drawback of these parametric approaches is that they require the correct specification for the functional form of the betas, or, in other words, they need to identify the right state variables. As pointed out by [Ghysels \(1998\)](#) and [Harvey \(2001\)](#), models with misspecified betas often feature larger pricing errors than models with constant betas.

Recent non-parametric approaches have been proposed to allow the CAPM parameters to evolve smoothly over time. [Ang and Kristensen \(2012\)](#) used this methodology to investigate the distributions for conditional and for long-run alphas and betas, averaged over time. They used different bandwidths for conditional and long estimates in order for any finite-sample biases and variances to vanish. In addition, kernel-smoothing estimators have the appealing feature that they nest, as a special case, rolling window estimates of betas (see, for example, [Ferson and Harvey 1991](#); [Petkova and Zhang 2005](#), among many others).

The methodology uses [Giraitis et al. \(2014, 2015, 2018\)](#). They provide a rigorous justification for using kernel methods to estimate structural change when the parameters that undergo change are not governed by a deterministic function of time, allowing a wide class of stochastic processes that are characterised by persistence to be performed.

Fama and MacBeth Formulation

The seminal paper by [Fama and MacBeth \(1973\)](#) advocates a two-step procedure to estimate risk premia in the multi-factor asset pricing setting. The model assumes the coefficients are constant and estimates them using ordinary least square regression. The first step regresses the excess risk-free return of each asset, or portfolio, on various factors over time to determine the exposure of each factor; hence, it estimates the beta parameters. The second step consists of a cross-sectional regression of the excess return of the assets against the factor exposures, or betas, at each point in time, in order to obtain a time series of risk premium coefficients, or gammas, for each factor. The method by [Fama and MacBeth \(1973\)](#) averages these coefficients to obtain the expected premium for a unit of each risk factor and testing if these are appropriately priced by the market. More details are available in [Appendix A](#).

This two-pass cross-sectional method is subject to an error-in-variable (EIV) problem, due to using estimated betas in the second step. The procedure can produce consistent risk premium estimators. However, as the time-series sample size tends to infinity and the cross-sectional size is fixed, the traditional *FMcB* standard errors are not consistent, requiring an asymptotic bias correction. Recently, [Adrian et al. \(2015\)](#) used the weighted kernel estimator by [Ang and Kristensen \(2012\)](#) to propose a methodology more robust to misspecification errors. Their empirical application features good pricing properties across stocks and bonds and shows notable time variation of expected returns associated

with highly significant dynamic price of risk parameters. Moreover, they showed that the Gaussian kernel estimator yields smaller pricing errors than simple rolling window regressions for both specifications with constant and time-varying prices of risk.

3. Hierarchical Methodology

The main contribution of this paper is to develop a flexible methodology, inside the kernel regression framework, to easily allow users to have time variation in both the betas and gammas of the baseline [Fama and MacBeth \(1973\)](#) approach.

The main tool to achieve this is to have a flexible bandwidth parameter which essentially controls the weight given to local information for updating the beta and gamma coefficients. This paper optimises the choice of bandwidth and is based on out-of-sample cross-validation methods, which allow the bandwidth to change over time. The novelty of the approach is to identify an optimal time-varying bandwidth for each asset that helps to reduce the forecast errors of the risk premia via a more accurate estimation of the factor loadings. In the first step, we use a cross-validation approach to identify the optimal asset-specific bandwidth. In the second step, asset returns are regressed in the time series on risk factors, using the bandwidths obtained before, generating the time-varying risk betas for each asset. In the final step, the price of the risk parameters are computed by regressing the excess return on the betas from the time-series regression, cross-sectionally.

3.1. Cross-Validation—Bandwidth Choice

As previously mentioned, an important aspect of this paper is the use of cross-validation to search for the most appropriate bandwidth in the kernel function that sets the degree of smoothness of the estimates. This parameter turns out to be critical in providing the appropriate degree of persistence in determining the *memory* of the window used for the estimation of the time-varying coefficient of the model. Following previous literature, this paper considers the classical three factors of the model proposed by [Fama and French \(1995\)](#): market factor, *MRKT*, size factor, *SML*, and book-to-market factor, *HML*¹.

The first part of the hierarchical approach is to calculate the time-varying parameters (*TVP*) associated with the coefficients of the factors (β s). The method used here is based on a kernel-weighted regression; hence,

$$\left(R_{i,t} - R_{f,t}\right)_h = \beta_{1,t,i,h} F_{MRKT,t} + \beta_{2,t,i,h} F_{SMB,t} + \beta_{3,t,i,h} F_{HML,t} + u_{i,t,h}, \quad (1)$$

where $i \in [1 : N]$ is the number of assets, $t \in [1 : T]$ is the period of time, $k \in [1 : 3]$ is the number factors and h is the bandwidth parameter, to be discussed later, such that $h \in [0.05, 0.95]$ with an interval of 0.05. Further, it is generally assumed throughout the paper that $u_{n,t+1}$ is *i.i.d.*($0, \sigma^2$). The β parameters are estimated by an extension of the methodology by [Giraitis et al. \(2014\)](#), summarized in the Appendix B of this paper. Hence, the beta for the k^{th} factor is estimated by

$$\hat{\beta}_{k,t,i,h} = \frac{\sum_{t=1}^T K\left(\frac{t-j}{H}\right) \left(R_{i,t} - R_{f,t}\right) F_{k,t}}{\sum_{t=1}^T K\left(\frac{t-j}{H}\right) F_{k,t}^2}, \quad (2)$$

where $K\left(\frac{t-j}{H}\right)$ is assumed to be a Gaussian kernel function. The bandwidth, H , represents the degree of smoothness of the estimates. In other terms, if the bandwidth is small, the estimates are under-smoothed, with high variability; otherwise, if the value of H is big, the resulting estimators are over-smoothed and farther from the real function. [Ang and Kristensen \(2012\)](#) suggested to optimise the choice of the bandwidth for conditional and long estimates in order to reduce any finite-sample biases and variances. [Giraitis et al. \(2014, 2018\)](#), instead, proved, under very mild condition, that, if the bandwidth is $H = T^h$, with the bandwidth parameter $h = 0.5$, the estimator shows desirable properties, such as consistency and asymptotic normality; additionally, it provides valid standard errors.

The method in this paper is agnostic on the choice of the parameter h and then on the bandwidth. An additional insight is to allow the parameter to vary across time and assets. A cross-validation procedure is used to identify the time-varying bandwidth optimised for each asset. Therefore, the optimal parameter $h_{i,t}^{opt}$ is found for each asset and time period, selected from an out-of-sample, one-step ahead forecasting comparison over a grid search of h , which incorporates 19 different values of h , for the grid of $h \in [0.05;0.95]$, with an interval of 0.05 for each grid.

For the remaining of the paper, the optimisation of the bandwidth is meant to relate to the choice of the parameter h inside the bandwidth formula $H = T^h$.

At the end of this stage, the process generates, for each asset i , a time series of beta estimates for different values of the bandwidth parameter h . These estimated betas allow us to identify the price of risk factor loadings for different values of h , γ_h using the following equation:

$$\left(R_{i,t} - R_{f,t}\right)_h = \gamma_{0,h} + \hat{\beta}'_{1,t,i,h} \gamma_{1,t,h} + \hat{\beta}'_{2,t,i,h} \gamma_{2,t,h} + \hat{\beta}'_{3,t,i,h} \gamma_{3,t,h} + \varepsilon_{n,t,h}, \tag{3}$$

where $\varepsilon_{n,t,h}$ is assumed to be *i.i.d.*($0, \sigma_\varepsilon^2$). This process generates $k + 1$ series of γ s (including the constant) for every value of the bandwidth parameter h . Then, the cross-validation procedure compares the forecasting performance of the competing models via the computation of the forecast errors $e_{i,t+1,h}$. The initial T_0 observations are the *training period*, while the remaining ones, $(T - T_0)$, define the out-of-sample period. The training period is fixed at 60 observations, or 5 years of data; we also performed robustness tests with different values of T_0 , 120 and 180. Then, the one-step ahead forecast for each asset is obtained from the following regression:

$$\widehat{\left(R_{i,t+1} - R_{f,t+1}\right)}_h = \hat{\gamma}_{0,h} + \hat{\beta}'_{1,t,i,h} \hat{\gamma}_{1,t,h} + \hat{\beta}'_{2,t,i,h} \hat{\gamma}_{2,t,h} + \hat{\beta}'_{3,t,i,h} \hat{\gamma}_{3,t,h}. \tag{4}$$

The forecast errors $e_{i,t+1,h}$ are computed for each period and for each of the eighteen different values of h . The time-varying *RMSE* is calculated at each point in time and for each asset, while the value of h is chosen via a minimization procedure. Several different criteria and approaches were investigated to compute this measure, including the rolling window and non-parametric kernel-smoothed technique. The former approach refers to the classical rolling window method with a different window w , such that $w \in [12; 24]$. Hence, the unadjusted rolling *RMSE* is given by

$$RMSE_t^{roll} = \sqrt{\frac{1}{w} \sum_{j=1}^w e_{i,t+j,h}^2} \tag{5}$$

while the kernel-weighted *RMSE* is instead computed as

$$RMSE_t^{kern} = \sqrt{\sum_{j=1}^T W\left(\frac{t-j}{H^{(i)}}\right) e_{i,t+j,h}^2} \tag{6}$$

where $H^{(i)} = T^{h'}$ and $h' \in [0.05;0.95]$. Clearly, when $W(H) = 1$, the formula reduces to the regular $RMSE_t$ formula in Equation (5). Both approaches generate a matrix of 18 columns and $(T - T_0 - w)$ or $(T - T_0)$ rows, according to the method used, for each asset. Then, this matrix of $RMSE_t$ is used to determine the optimal values of h for each asset $h_{i,t}^{opt}$, such as the value that produces the lowest *RMSE*. The approach generates a time series of optimal values of h , that are used in the second step of this procedure to obtain a more accurate estimation of β coefficients.

3.2. Estimation of Factor Risk Loadings

Once the matrix with the optimal values of h is obtained, the time-varying factor risk loading is calculated. The point of this procedure is to allow the parameters to be fully liberalised, optimizing the choice of a time-varying bandwidth for each asset. The forecasting performance of the new method was compared with different approaches for the computation of β :

- (i) The classical [Fama and MacBeth \(1973\)](#) approach, where the betas are computed using an OLS rolling window approach with a five-year window.
- (ii) A kernel-weighted approach with $h = 0.5$. As showed by [Giraitis et al. \(2014\)](#), this bandwidth allows us to obtain smooth estimates with desirable properties such as consistency and asymptotic normality; in addition, it provides asymptotically valid standard errors. This model is used as a benchmark.
- (iii) The alternative kernel approach, where h is fixed for each asset and time and is determined from a poll average of the optimal bandwidth parameters $h_{i,t}^{opt}$, as follows:

$$\bar{h}^{Polling} = (NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N h_{t,i}. \tag{7}$$

- (iv) A further kernel regression approach, with h computed by averaging the optimal bandwidth parameters across assets. While the parameter varies over time, it is not asset specific:

$$\bar{h}_t^{Average} = (N)^{-1} \sum_{i=1}^N h_{t,i}. \tag{8}$$

- (v) A kernel approach that uses the optimal $h_{i,t}^{opt}$, which is different for each asset to give; this method is named *Specific*.

These five approaches are all implemented in the three-factor [Fama and French \(1992\)](#) model:

$$R_{i,t} - R_{f,t} = \beta_{1,t,i,m} F_{MRKT,t} + \beta_{2,t,i,m} F_{SMB,t} + \beta_{3,t,i,m} F_{HML,t} + u_{i,t,m}, \tag{9}$$

where $m \in [1 : 5]$ represents one of the aforementioned approaches used for the computation of the factor loadings, β s. The coefficients are computed according to Equation (2).

3.3. Estimation of Risk Premia

The time-varying estimates of beta $\hat{\beta}_t$ are then used in the third step to facilitate the computation of risk premia associated with the factors under investigations γ s. Then, the hierarchical methodology replaces the assets' excess returns by their corresponding time-varying kernel-weighted average $\left(R_{n,t} - R_{f,t} \right)$ for coherence in terms of degree of smoothness. Indeed, they are computed using the bandwidth h that was selected in the previous step for the computation of the β s. Then, the kernel-weighted averages for the excess returns are

$$\left(\widehat{R_{i,t} - R_{f,t+1}} \right) = \sum_{k=1}^T K\left(\frac{t-k}{H^*}\right) \left(R_{i,k} - R_{f,k} \right), \tag{10}$$

where $K\left(\frac{t-k}{H^*}\right)$ is the same continuously bounded kernel function and $H^* = T^{h^m}$, where h^m identifying the bandwidth used at the previous step for the computation of the coefficients $m \in [1 : 5]$. These smoothed excess returns are then used for the OLS regressions, to identify the risk premia,

$$\left(\widehat{R_{i,t} - R_{f,t}} \right) = \gamma_{0,m} + \hat{\beta}_{1,t,i,m} \gamma_{1,t,m} + \hat{\beta}_{2,t,i,m} \gamma_{2,t,m} + \hat{\beta}_{3,t,i,m} \gamma_{3,t,m} + \varepsilon_{i,t,m}. \tag{11}$$

This results in $m + 1$ series of $\hat{\gamma}$ (including the constant), for each of the five different approaches previously considered for the estimation of the β s.

The last stage of the hierarchical approach is to select the best methodology in terms of *RMSE* minimization for an out-of-sample forecasting exercise. This is achieved by forecasting the average excess return across all assets using the average of the estimated betas, which realizes the time series of forecasts of the average,

$$\overline{R_{t+1} - R_{f,t+1}} = \hat{\gamma}_{0,m} + \hat{\gamma}_{1,t,m} \hat{\beta}_{1,t,m} + \hat{\gamma}_{2,t,m} \hat{\beta}_{2,t,m} + \hat{\gamma}_{3,t,m} \hat{\beta}_{3,t,m}, \tag{12}$$

where

$$\overline{R_{t+1} - R_{f,t+1}} = \frac{1}{N} \sum_{i=1}^N (R_{t+1} - R_{f,t+1}), \tag{13}$$

and

$$\bar{\beta}_{j,t,m} = \frac{1}{N} \sum_{i=1}^N \beta_{j,t,i,m}.$$

Then, the *RMSE* are computed for each method and compared to identify the estimation method, using the [Diebold and Mariano \(1995\)](#) test.

4. Data

The new hierarchical methodology was applied to three different financial return datasets. The first dataset contained $N = 25$ portfolios sorted by size and book-to-market ratio, while the second one contained $N = 55$ portfolios (25 portfolios from the first dataset and 30 portfolios sorted by industry, both available from Ken French’s on-line data library). A further 200 Standard and Poor’s constituents from the Center for Research in Securities Prices (CRSP) were included, so that $N = 200^2$. The excess returns over the 30-day Treasury bill yield were computed with the total series covering the period from August 1973 to January 2020, for a total of $T = 514$ observations (again, these are available from Kenneth French’s on-line web site).

The following set of factors was used in the subsequent analysis: excess return on the market, *MRKT*; value-weight return of all *CRSP* firms incorporated in the *US* and listed on either the *NYSE*, *AMEX*, or the *NASDAQ*. The small minus big, *SMB*, factor and the high minus low, *HML*, factor are derived in the same way as in [Fama and French \(1992\)](#) and are available from Ken French’s on-line data library³.

5. Empirical Results of the Hierarchical Analysis

Following the details of the above methodological framework, [Table 1](#) provides the descriptive statistics for the optimal bandwidth parameters h^{opt} for all the different datasets. The results are categorized in terms of the methodology used to compute the time-varying *RMSE_t* measure; $h_{w=12}^{opt}$ and $h_{w=24}^{opt}$ refer to the conventional rolling window approach with windows of 12 and 24 observations, [Equation \(5\)](#), while h_{kern}^{opt} refers to the kernel approach ([Equation \(6\)](#)). From the analysis of the panels containing the portfolio results, it can be seen that the cross-validation procedure was remarkably consistent in choosing an h near 0.50 and a standard deviation of the estimates relatively small, lying in the range from 0.044 to 0.076 for all the methodologies. Regular t -tests were unable to reject the hypothesis that $h = 0.50$ for any of the portfolio classifications. This finding is particularly interesting, since $h = 0.50$ is the theoretical value identified by [Giraitis et al. \(2014\)](#) as being the optimal value for h in terms of achieving an appropriate rate of convergence to an asymptotic distribution of the *TVP*. However, the averages for h^{opt} for the individual stock data were higher than the ones for the two portfolios, being around $h = 0.65$. This can be interpreted as the need to increase the degree of smoothness when using data with high levels of heterogeneity. In addition, the analysis of across the different methods for computing the *RMSE_t* shows that the non-parametric *kernel* approach provided the highest values for the standard deviations for each portfolio.

Table 1. Descriptive statistics for the optimal bandwidth parameter.

	Obs.	Mean	St. Dev.	Min	Max	Skew	Kurt
25 Portfolios							
$h_{w=12}$	443	0.51975	0.04937	0.37800	0.64600	−0.03670	−0.04397
$h_{w=24}$	431	0.55826	0.04407	0.42600	0.68800	−0.01419	−0.32630
h_{Kern}	454	0.52804	0.07684	0.33400	0.69600	0.11501	−0.55003
55 Portfolios							
$h_{w=12}$	443	0.55331	0.04655	0.37273	0.68455	0.05771	0.02067
$h_{w=24}$	431	0.58815	0.04586	0.47636	0.72727	−0.05941	−0.35675
h_{Kern}	454	0.55532	0.06884	0.37182	0.71909	0.23544	−0.48401
200 Stocks							
$h_{w=12}$	443	0.65671	0.06742	0.40100	0.76000	−1.22993	1.83550
$h_{w=24}$	431	0.70924	0.05967	0.44625	0.81100	−1.31970	1.52070
h_{Kern}	454	0.65222	0.07571	0.50425	0.81475	0.14502	−0.16497

Note: The table reports the descriptive statistics of the optimal bandwidth parameters considering different datasets (25 portfolios, 55 portfolios or 200 individual stocks) and obtained using different methods for computing the $RMSE_t$, as discussed in Section 3.1. The former method refers to the classical rolling window approach with different w such that $w \in [12; 24]$ (Equation (5)), while the latter involves a kernel average method (Equation (6)).

Figure 1 plots the selected optimal bandwidth parameters, averaged across assets as in Equation (8), for each of the different methodologies for computing the time-varying $RMSE_t$ and also for different datasets. All the methods provided an erratic mean-reverting path, centred around 0.5, where the kernel approach confirms to be the most volatile in all the data combination. In general, the non-parametric approach appears to be the most volatile and is the only one that increased in the global financial crisis, *GFC*.

Tables 2 and 3 provide details of the estimated beta coefficients for representative assets for each dataset. The portfolio datasets were estimated for the median portfolios, named *ME3.BM3*, while, for the constituents of the *SandP500*, they were analysed by the *Ford* index. Standard errors are given for each of the three factor loadings: *MRKT*, *SMB* and *HML*. The estimated market beta $\hat{\beta}_{MRKT}$ was close to the unity for all the portfolio datasets, while it was around 0.7 for the *Ford* stock, in line with previous literature. The standard errors provided by the *Specific* approach were the smallest and were very important for subsequent efficient estimation of risk premia⁴. Figure 2 presents the factor risk loading estimates. Each one of the nine separate panels shows five *TVP* beta estimates derived from the methodologies presented in Section 3. In particular, the rolling window is displayed with a green line, the kernel estimate with constant bandwidth parameter of $h = 0.5$ in black, the constant bandwidth parameter from poll average, *Polling*, in purple; further, the time-varying h , set equal for all the asset, *Average*, is represented by a blue line and the time-varying h optimised for each asset, *Specific*, by a red line. The last three methods all use the Gaussian kernel.

In all the scenarios, the time-varying estimates were centred around the constant ones, highlighting the correctness of the methodology. Further, a similar path can be seen for all the kernel estimates with the ones produced by the classical rolling window approach. In accordance with [Adrian et al. \(2015\)](#), the estimates produced with the classical approach exhibited, overall, a higher variation than the one produced with kernel approaches. Although these estimates follow a path in line with the others, they are characterised by numerous sudden changes along the sample period. These changes appear to be asset-specific, hence, different asset by asset.

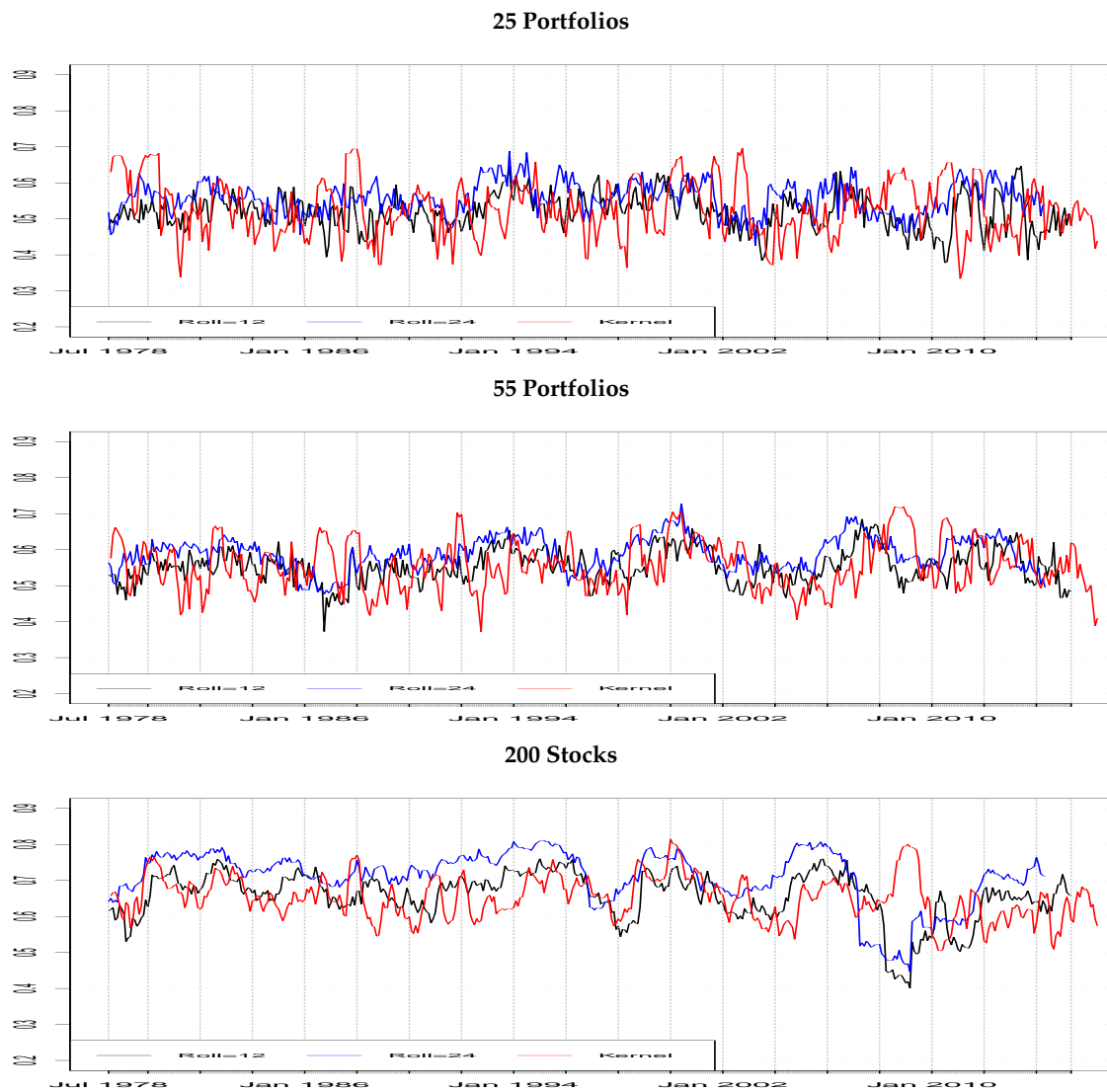


Figure 1. Time-varying optimal bandwidth parameters. The figure reports the plots of the optimal bandwidth parameters considering different datasets and methods for computing the $RMSE_t$, as discussed in Section 3.1. The bandwidths reported in red are computed according Equation (5) using the classical rolling window approach with $w = 12$, while, for those in blue, $w = 24$. On the other hand, the bandwidths in black are computed using the kernel average method discussed in Equation (6).

Table 2. Factor risk loading estimates for Ford stocks.

	Constant β s			Rolling			$h = 0.5$		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.7331 (0.1164)	0.6690 (0.1728)	0.8870 (0.1744)	0.6008 (0.2993)	0.7378 (0.4719)	0.5933 (0.4822)	0.6404 (0.1148)	0.7413 (0.2719)	0.1441 (0.2926)
$h_{w=24}$	0.7379 (0.1164)	0.6988 (0.1728)	0.8895 (0.1744)	0.6249 (0.2778)	0.7443 (0.4361)	0.6284 (0.4428)	0.6435 (0.1199)	0.7615 (0.2864)	0.1341 (0.3057)
h_{Kern}	0.7449 (0.1164)	0.6571 (0.1728)	0.8763 (0.1744)	0.5719 (0.3260)	0.7405 (0.5146)	0.5524 (0.5292)	0.6460 (0.1129)	0.7135 (0.2622)	0.6401 (0.2848)

Table 2. Cont.

	<i>h Pooling</i>			<i>Average h</i>			<i>h Specific</i>		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.7267 (0.0438)	0.7116 (0.1041)	0.7683 (0.1067)	0.7375 (0.0466)	0.7060 (0.1217)	0.7757 (0.1176)	0.6333 (0.0482)	0.7759 (0.0896)	0.8200 (0.0916)
$h_{w=24}$	0.7410 (0.0343)	0.7150 (0.0742)	0.7986 (0.0782)	0.7804 (0.0376)	0.6880 (0.0964)	0.8053 (0.0921)	0.6197 (0.0401)	0.7704 (0.0679)	0.8294 (0.0828)
h_{Kern}	0.7214 (0.0438)	0.6726 (0.1020)	0.7392 (0.1067)	0.7105 (0.0471)	0.6767 (0.1163)	0.7621 (0.1212)	0.6441 (0.0721)	0.6066 (0.1281)	0.7476 (0.1505)

Note: Average estimates of factor risk loadings for Ford stock using the 200 stock dataset for the computation of the optimal bandwidth. There are 6 different methodologies: simple ordinary least square regression (*Constant*), *Rolling* window approach (with a 5-year window) and kernel-weighted regressions using 4 different optimal bandwidths; $h = 0.5$; *Polling* a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time. In parenthesis, there are the averages of the standard errors.

Table 3. Factor risk loading estimates for ME3.BM3—25 portfolios.

Panel A	<i>Constant βs</i>			<i>Rolling</i>			$h = 0.5$		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.9943 (0.0175)	0.4265 (0.0259)	0.4017 (0.0262)	1.0034 (0.0413)	0.5275 (0.0619)	0.3101 (0.0651)	1.0142 (0.0119)	0.5363 (0.0282)	0.3035 (0.0303)
$h_{w=24}$	0.9935 (0.0175)	0.4211 (0.0259)	0.4059 (0.0262)	1.0022 (0.0384)	0.5230 (0.0571)	0.3165 (0.0596)	1.0138 (0.0122)	0.5342 (0.0291)	0.2962 (0.0311)
h_{Kern}	0.9969 (0.0175)	0.4282 (0.0259)	0.4029 (0.0262)	1.0052 (0.0453)	0.5296 (0.0681)	0.3036 (0.0721)	1.0135 (0.0120)	0.5304 (0.0279)	0.3024 (0.0303)
	<i>h pooling</i>			<i>Average h</i>			<i>h specific</i>		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	1.0132 (0.0104)	0.5322 (0.0249)	0.3072 (0.0265)	1.0114 (0.0109)	0.5281 (0.0267)	0.3085 (0.0289)	1.0123 (0.0048)	0.4670 (0.0103)	0.3830 (0.0158)
$h_{w=24}$	1.0099 (0.0083)	0.5204 (0.0203)	0.3101 (0.0208)	1.0080 (0.0086)	0.5200 (0.0217)	0.3101 (0.0221)	1.0041 (0.0053)	0.4684 (0.0080)	0.3792 (0.0167)
h_{Kern}	1.0116 (0.0099)	0.5241 (0.0233)	0.3074 (0.0250)	1.0089 (0.0115)	0.5199 (0.0268)	0.3067 (0.0304)	1.0126 (0.0067)	0.4461 (0.0165)	0.4163 (0.0176)
Panel B	<i>Constant βs</i>			<i>Rolling</i>			$h = 0.5$		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.9943 (0.0175)	0.4265 (0.0259)	0.4017 (0.0262)	1.0034 (0.0413)	0.5275 (0.0619)	0.3101 (0.0651)	1.0142 (0.0119)	0.5363 (0.0282)	0.3035 (0.0303)
$h_{w=24}$	0.9935 (0.0175)	0.4211 (0.0259)	0.4059 (0.0262)	1.0022 (0.0384)	0.5230 (0.0571)	0.3165 (0.0596)	1.0138 (0.0122)	0.5342 (0.0291)	0.2962 (0.0311)
h_{Kern}	0.9969 (0.0175)	0.4282 (0.0259)	0.4029 (0.0262)	1.0052 (0.0453)	0.5296 (0.0681)	0.3036 (0.0721)	1.0135 (0.0120)	0.5304 (0.0279)	0.3024 (0.0303)
	<i>h pooling</i>			<i>Average h</i>			<i>h specific</i>		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	1.0105 (0.0084)	0.5234 (0.0203)	0.3145 (0.0212)	1.0093 (0.0088)	0.5192 (0.0214)	0.3221 (0.0231)	1.0078 (0.0078)	0.4694 (0.0132)	0.3856 (0.0204)
$h_{w=24}$	1.0068 (0.0069)	0.5104 (0.0169)	0.3203 (0.0171)	1.0051 (0.0072)	0.5069 (0.0179)	0.3277 (0.0190)	1.0071 (0.0071)	0.4741 (0.0125)	0.3925 (0.0219)
h_{Kern}	1.0091 (0.0083)	0.5164 (0.0197)	0.3133 (0.0209)	1.0071 (0.0095)	0.5105 (0.0217)	0.3203 (0.0242)	1.0058 (0.0107)	0.4703 (0.0200)	0.3864 (0.0237)

Note: Average estimates of factor risk loadings for the portfolio ME3.BM3 using 25 (Panel) and 55 Portfolios for the computation of the optimal bandwidth, respectively, in Panel A and Panel B. There are 6 different methodologies: simple ordinary least square regression (*Constant*), *Rolling* window approach (with a 5-year window) and kernel-weighted regressions using 4 different optimal bandwidths; $h = 0.5$; *Polling* a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time. In parenthesis, there are the averages of the standard errors.



Figure 2. A dynamic comparison of the factor loading estimates. **Note:** The figure provides the estimates of factor risk loadings computed using the normal approach *Constant* β (grey line), *Rolling* window, with a 5-year estimation period (green line) and kernel-weighted regressions using 4 different optimal bandwidths; $h = 0.5$ (black line); *Pooling* a single value of h coming from the poll average across assets and time, as shown in Equation (7), (purple line); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)), (blue line); *Specific*, multiple time-varying bandwidths, one for each asset and time (red line). The choice of the optimal bandwidth parameter h_t^{opt} was made using the kernel approach, as discussed in Equation (6).

As expected, the beta estimates for portfolio datasets exhibited a lower degree of variation than those that employed stock indexes. This is presumably due to noise using stock data and the loss of information induced by grouping stocks to build a portfolio (Lo and MacKinlay (1990)). In general, the betas on the *MKT* and *HML* factors were the ones that most often switch sign, while the *SMB* appears to be the most stable factor. Table 4 provides the estimates and the respective standard errors of the risk premium parameters γ_i , with $i \in [0, 3]$, also including the constant term. The Newey West standard errors are also displayed in the last column. Further, it presents results for h^{opt} , computed using

$RMSE_t$ with the kernel-averaging approach. The results for the other two parametric approaches are available online.

The average prices of risk appear to be very similar across the different methods and within each dataset. The *Specific* method shows the smallest standard errors despite the sample considered. The sample size appears to matter and affects the significance of the price of all the factors. In particular, *SMB* was priced only considering individual stocks. This result is consistent with other studies showing that *SMB* is not priced in the cross-section of portfolios sorted by size and book to market (see [Adrian et al. \(2015\)](#) and [Lettau and Ludvigson \(2001a\)](#)).

Table 4. Descriptive statistics of risk premium estimates.

	Obs.	Mean	St. Dev.	Min	Max	Skew	Kurt	SE	NW SE
25 Portfolio									
Rolling									
$\hat{\gamma}_0$	454	0.0086	0.0083	-0.0123	0.0371	0.4184	0.5990	0.0060	0.0054
$\hat{\gamma}_{\beta_{MRKT}}$	454	-0.0022	0.0088	-0.0226	0.0233	0.0766	-0.5119	0.0057	0.0050
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0021	0.0054	-0.0085	0.0145	0.4392	-0.7676	0.0010	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0032	0.0042	-0.0058	0.0168	0.2544	-0.2611	0.0011	0.0010
h = 0.5									
$\hat{\gamma}_0$	454	0.0092	0.0082	-0.0067	0.0344	0.9116	1.3169	0.0059	0.0048
$\hat{\gamma}_{\beta_{MRKT}}$	454	-0.0024	0.0086	-0.0216	0.0129	-0.2629	-0.8884	0.0057	0.0046
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0010	0.0042	-0.0045	0.0099	0.5480	-0.7747	0.0010	0.0009
$\hat{\gamma}_{\beta_{HML}}$	454	0.0028	0.0036	-0.0027	0.0099	0.2900	-1.0656	0.0011	0.0008
Polling									
$\hat{\gamma}_0$	454	0.0096	0.0074	-0.0044	0.0316	0.9550	1.0650	0.0059	0.0047
$\hat{\gamma}_{\beta_{MRKT}}$	454	-0.0028	0.0079	-0.0202	0.0103	-0.3081	-0.9680	0.0056	0.0045
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0010	0.0037	-0.0043	0.0087	0.5381	-0.8492	0.0010	0.0009
$\hat{\gamma}_{\beta_{HML}}$	454	0.0028	0.0032	-0.0021	0.0086	0.1337	-1.1173	0.0011	0.0008
Average									
$\hat{\gamma}_0$	454	0.0098	0.0069	-0.0048	0.0350	0.8941	1.6039	0.0058	0.0046
$\hat{\gamma}_{\beta_{MRKT}}$	454	-0.0025	0.0073	-0.0216	0.0150	-0.1567	-0.6690	0.0056	0.0044
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0007	0.0037	-0.0113	0.0118	0.4682	0.2256	0.0009	0.0008
$\hat{\gamma}_{\beta_{HML}}$	454	0.0031	0.0032	-0.0069	0.0132	0.1595	-0.2499	0.0010	0.0008
Specific									
$\hat{\gamma}_0$	454	0.0064	0.0281	-0.1892	0.1286	-0.2239	5.9720	0.0161	0.0151
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0005	0.0275	-0.1089	0.1753	0.1478	4.6923	0.0155	0.0147
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0010	0.0061	-0.0158	0.0267	0.4351	1.6323	0.0036	0.0029
$\hat{\gamma}_{\beta_{HML}}$	454	0.0030	0.0086	-0.0284	0.0407	0.0598	2.1849	0.0039	0.0036
55 Portfolio									
Rolling									
$\hat{\gamma}_0$	454	0.0029	0.0060	-0.0083	0.0206	0.8131	0.4169	0.0032	0.0034
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0037	0.0074	-0.0137	0.0214	0.2648	-0.6650	0.0031	0.0034
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0019	0.0058	-0.0102	0.0146	0.3154	-0.8416	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0012	0.0052	-0.0127	0.0163	0.0905	0.1267	0.0012	0.0014
h = 0.5									
$\hat{\gamma}_0$	454	0.0031	0.0060	-0.0080	0.0185	0.5166	-0.1746	0.0032	0.0031
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0039	0.0068	-0.0088	0.0177	0.3021	-0.7173	0.0031	0.0031
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0007	0.0047	-0.0055	0.0106	0.4547	-0.8907	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0004	0.0052	-0.0097	0.0091	0.0824	-0.8843	0.0012	0.0015
Polling									
$\hat{\gamma}_0$	454	0.0038	0.0052	-0.0049	0.0156	0.4892	-0.5169	0.0031	0.0029
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0032	0.0057	-0.0075	0.0142	0.2503	-0.8742	0.0030	0.0030
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0006	0.0038	-0.0049	0.0083	0.4662	-0.9538	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0007	0.0043	-0.0090	0.0076	-0.2592	-0.3218	0.0012	0.0014
Average									
$\hat{\gamma}_0$	454	0.0043	0.0051	-0.0074	0.0179	0.2525	-0.4356	0.0030	0.0030
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0030	0.0057	-0.0107	0.0171	0.3331	-0.4746	0.0029	0.0030
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0004	0.0039	-0.0076	0.0097	0.2985	-0.7415	0.0010	0.0010
$\hat{\gamma}_{\beta_{HML}}$	454	0.0010	0.0044	-0.0098	0.0106	-0.2347	-0.2759	0.0012	0.0013
Specific									
$\hat{\gamma}_0$	454	0.0053	0.0109	-0.0385	0.0406	-0.0563	1.4520	0.0062	0.0063
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0020	0.0125	-0.0380	0.0517	-0.0727	1.9992	0.0061	0.0065
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0002	0.0066	-0.0220	0.0306	0.6872	1.7246	0.0026	0.0028
$\hat{\gamma}_{\beta_{HML}}$	454	0.0007	0.0076	-0.0228	0.0252	-0.0234	-0.0214	0.0028	0.0032

Table 4. Cont.

	Obs.	Mean	St. Dev.	Min	Max	Skew	Kurt	SE	NW SE
200 Stocks									
Rolling									
$\hat{\gamma}_0$	454	0.0013	0.0049	−0.0121	0.0110	−0.8412	0.5448	0.0011	0.0013
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0069	0.0095	−0.0198	0.0260	−0.6055	−0.3970	0.0024	0.0032
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0018	0.0062	−0.0125	0.0223	0.9650	0.6839	0.0013	0.0016
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0003	0.0062	−0.0164	0.0181	0.2282	−0.3997	0.0015	0.0017
h = 0.5									
$\hat{\gamma}_0$	454	0.0023	0.0049	−0.0097	0.0114	−0.6056	0.5540	0.0011	0.0011
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0071	0.0101	−0.0117	0.0226	−0.4628	−1.2147	0.0024	0.0031
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0011	0.0045	−0.0056	0.0125	0.7757	0.0968	0.0013	0.0014
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0018	0.0062	−0.0140	0.0106	0.2524	−0.8254	0.0015	0.0017
Polling									
$\hat{\gamma}_0$	454	0.0026	0.0034	−0.0049	0.0085	−0.4529	−0.4068	0.0008	0.0008
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0068	0.0100	−0.0091	0.0179	−0.2393	−1.6461	0.0022	0.0025
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0024	0.0021	−0.0013	0.0083	0.4873	−0.0664	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0030	0.0023	−0.0062	0.0007	0.2737	−1.3960	0.0013	0.0015
Average									
$\hat{\gamma}_0$	454	0.0027	0.0034	−0.0070	0.0114	−0.2455	0.1043	0.0008	0.0008
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0065	0.0098	−0.0111	0.0182	−0.2348	−1.5546	0.0022	0.0025
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0023	0.0023	−0.0042	0.0107	0.3858	1.0965	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0028	0.0028	−0.0086	0.0082	1.0775	1.4751	0.0013	0.0015
Specific									
$\hat{\gamma}_0$	454	0.0021	0.0049	−0.0145	0.0187	0.2124	0.1013	0.0001	0.0015
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0066	0.0157	−0.0493	0.0464	−0.5967	1.1184	0.0003	0.0042
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0009	0.0080	−0.0240	0.0234	−0.2290	−0.0516	0.0001	0.0022
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0004	0.0093	−0.0319	0.0304	−0.0736	0.2182	0.0002	0.0025

Note: Descriptive statistics of the estimated risk premia, computed for the classical *FMCB* approach and 4 different bandwidth specifications: $h = 0.5$; *Polling* a single value of h coming from the poll average across assets and time (as shown in Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time. The Newey West standard errors are also displayed in the last column. The choice of the optimal bandwidth parameter h_i^{opt} was made using the kernel approach for the computation of the time-varying *RMSE*, as discussed in Equation (6).

Despite most of the factors were not statistically different from zero on average—hence not priced—they exhibited a statistically significant time variation and fluctuated a lot between positive and negative values. Further, the significance of the constant term (γ_0), mostly throughout the entire sample, is in line with the literature, suggesting that the factors considered by Fama and French (1992) only partially describe the excess returns⁵. This time variation of the price of risk is well documented by the set of Figures 3–6. Figure 3 plots, by columns, the γ s for the three different samples, with the top panel relating to the 25 portfolios, the central panel to the 55 portfolios and the bottom panel individual stocks. As before, the value of $h = 0.5$ and the *Polling* methods described a form of background path for the evolution of the price of risk, while the *Specific* approach exhibited the highest volatility. From the analysis of these graphs, it is clear how much of the information about the price of risk was lost using approaches such $h = 0.50$, where we did not consider the specificity of each asset.

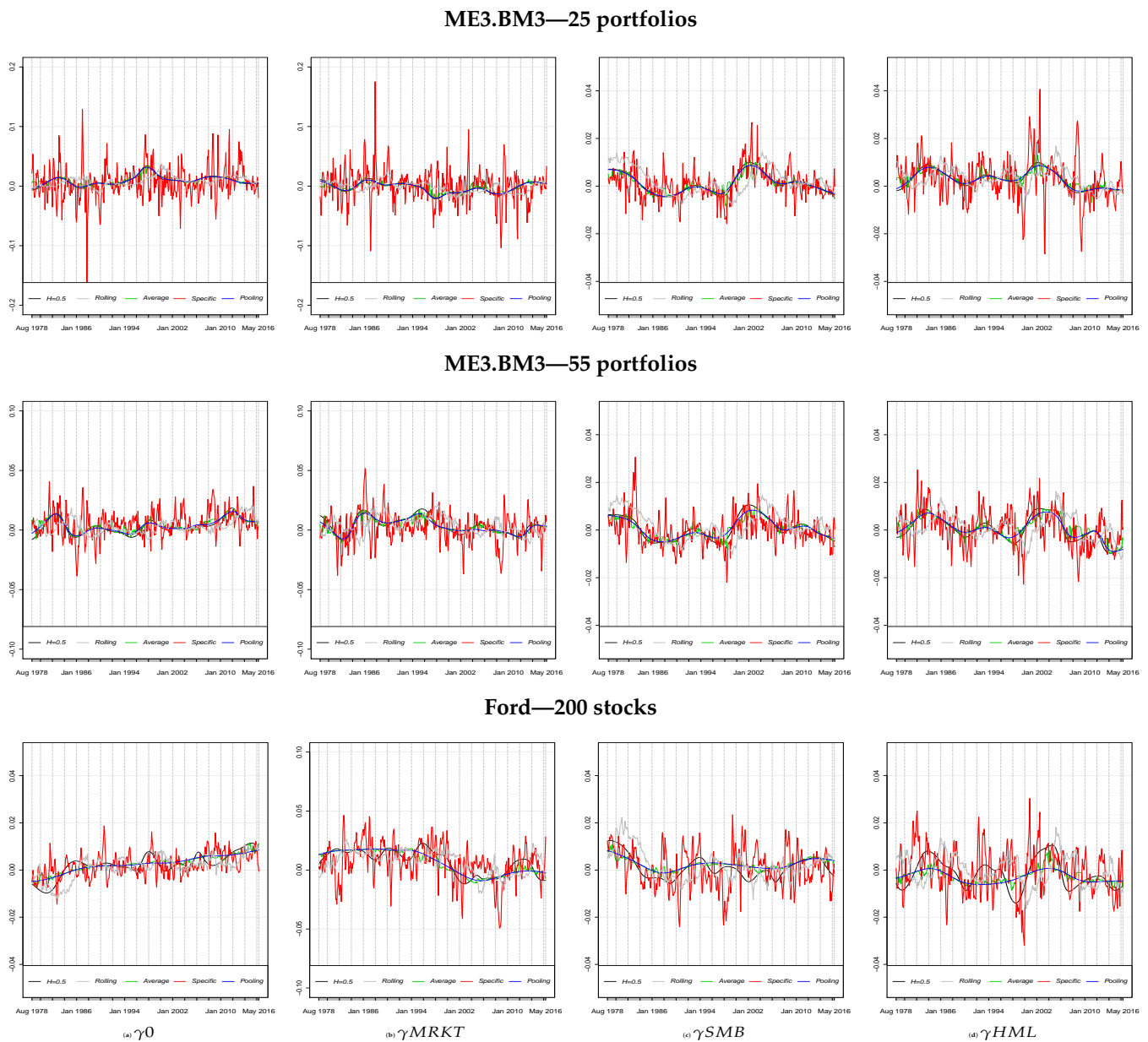


Figure 3. Dynamic comparison of risk premium estimates for different approaches. **Note:** The figure provides the estimates of risk premium estimates computed using factor risk loadings calculated with Rolling window, with a 5-year estimation period (*FMcB* approach, green line) and kernel-weighted regressions using 4 different optimal bandwidths; $h = 0.5$ (black line); *Pooling* a single value of h coming from the poll average across assets and time, as shown in Equation (7) (purple line); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8); blue line); *Specific*, multiple time-varying bandwidths, one for each asset and time (red line). The choice of the optimal bandwidth parameter h_t^{opt} was made using the kernel approach for the computation of the time-varying *RMSE*, as discussed in Equation (6).

In particular, only the *Specific* approach seems able to capture the *GFC*, where the drop in the estimates of the price of the factors is clearly evident. In Figures 4–6, instead, we reproduce an analysis of the significance for the different estimates across time—Figure 4 contains the results for 25 portfolios and Figure 5 for 55 portfolios, while Figure 6 the ones for individual stocks. All the Figures are structured as follows: in the columns, the different γ s are reported, while, in each row, there is a different method for the computation of β , as in Section 2. For what concerns the market risk premia, they show a significant

positive sign at the beginning of the sample until early 2000, when it becomes significantly negative. Such change was captured by all the methods, despite it being clearer for the stock asset context.

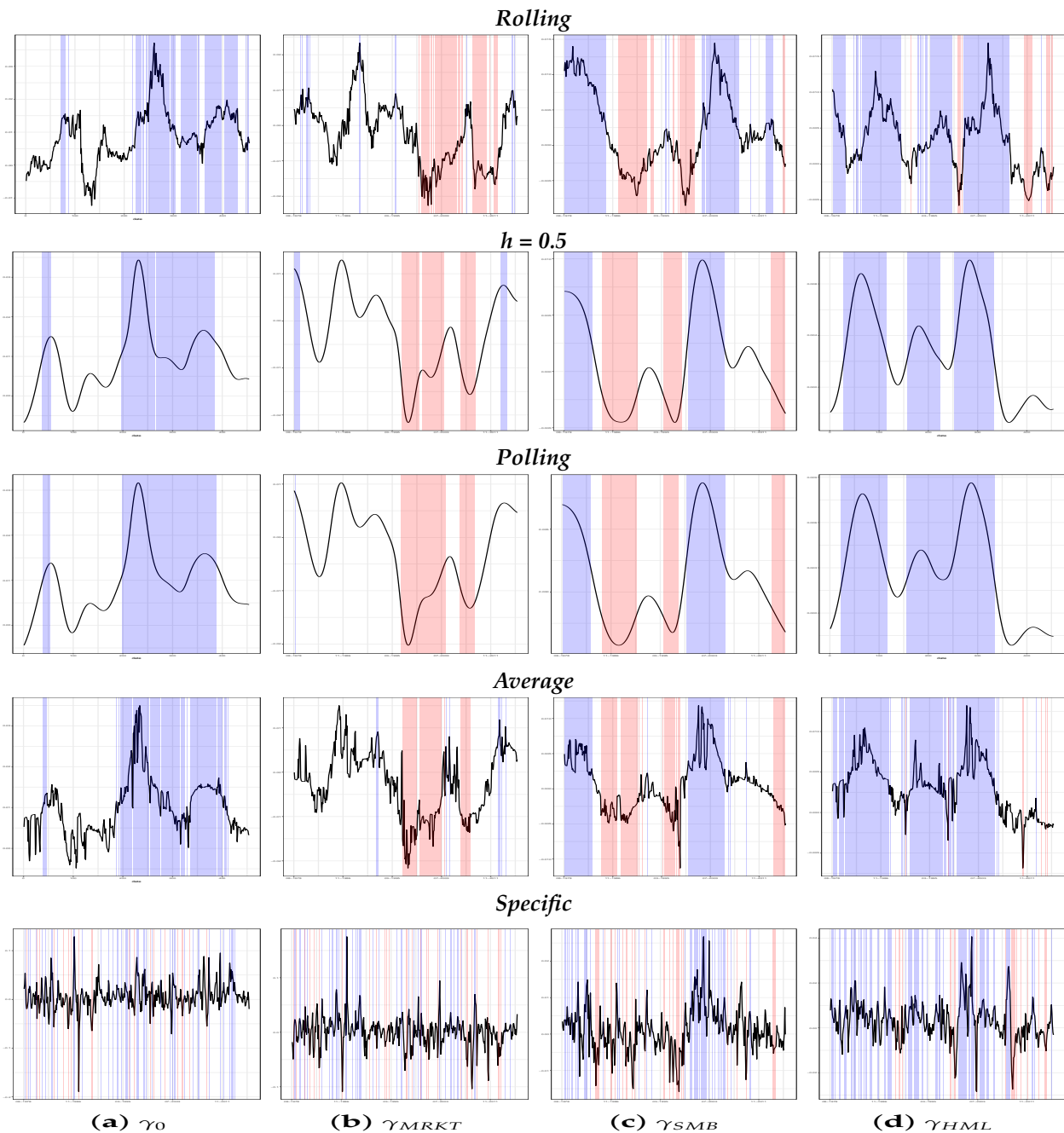


Figure 4. Comparison of γ s significance of different approaches—25 portfolios. **Note:** The figure provides a significance analysis of the estimates of risk premium estimates. The blue areas are periods in which the estimates are statistically positive at a 5% level of significance, while the red ones identify periods in which the estimates are negative. The series were computed using different approaches: Rolling window (*FMcB* approach), $h = 0.5$; *Polling* a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time. The choice of the optimal bandwidth parameter h_t^{opt} has been made using the kernel approach as discussed in Equation (6).

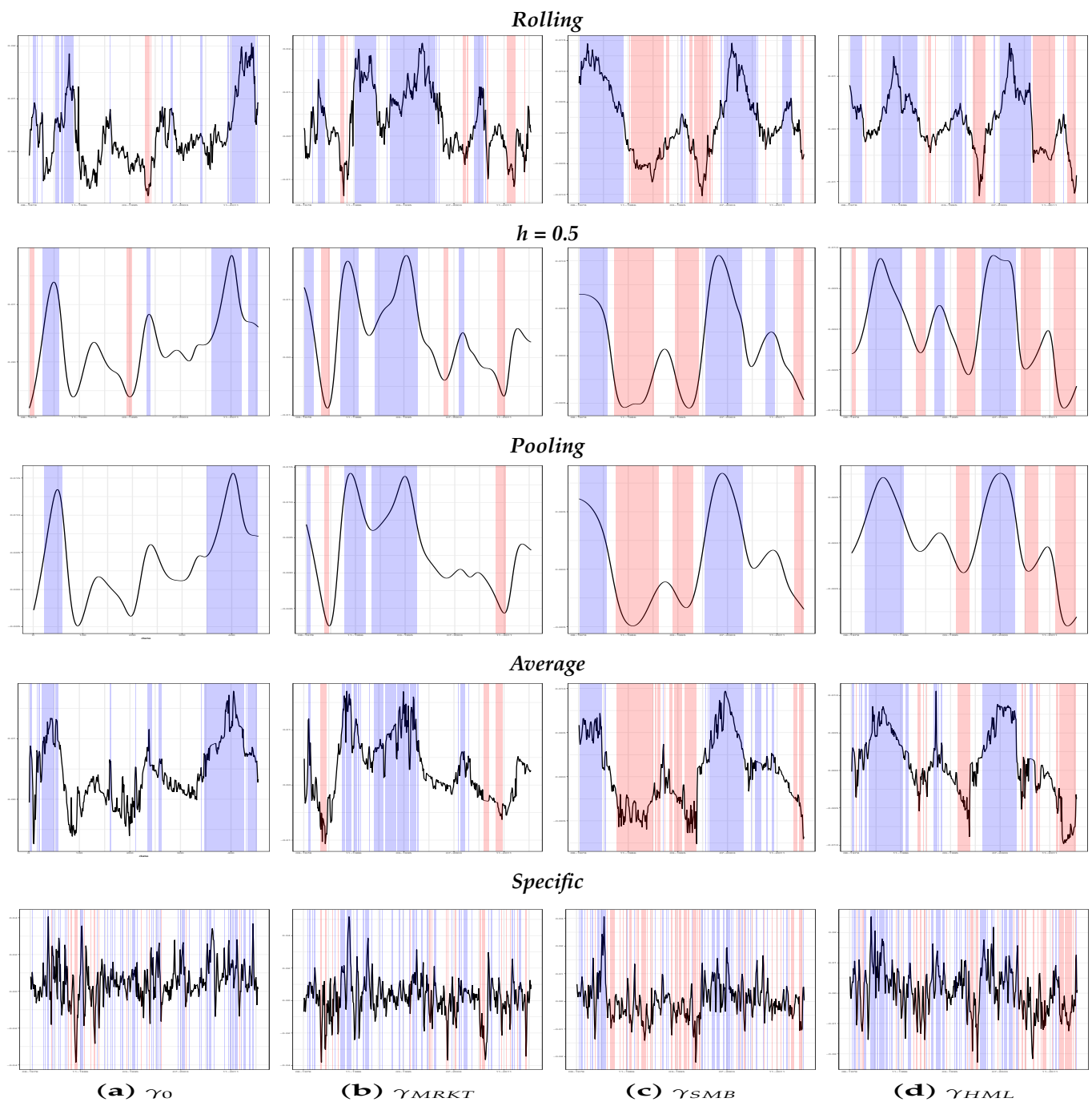


Figure 5. Comparison of γ s significance of different approaches—55 portfolios. **Note:** The figure provides a significance analysis of the estimates of risk premium estimates. The blue areas are periods in which the estimates are statistically positive at a 5% level of significance, while the red ones identify periods in which the estimates are negative. The series were computed using different approaches: Rolling window (*FMcB* approach), $h = 0.5$; *Pooling* a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time. The choice of the optimal bandwidth parameter h_t^{opt} was made using the kernel approach, as discussed in Equation (6).

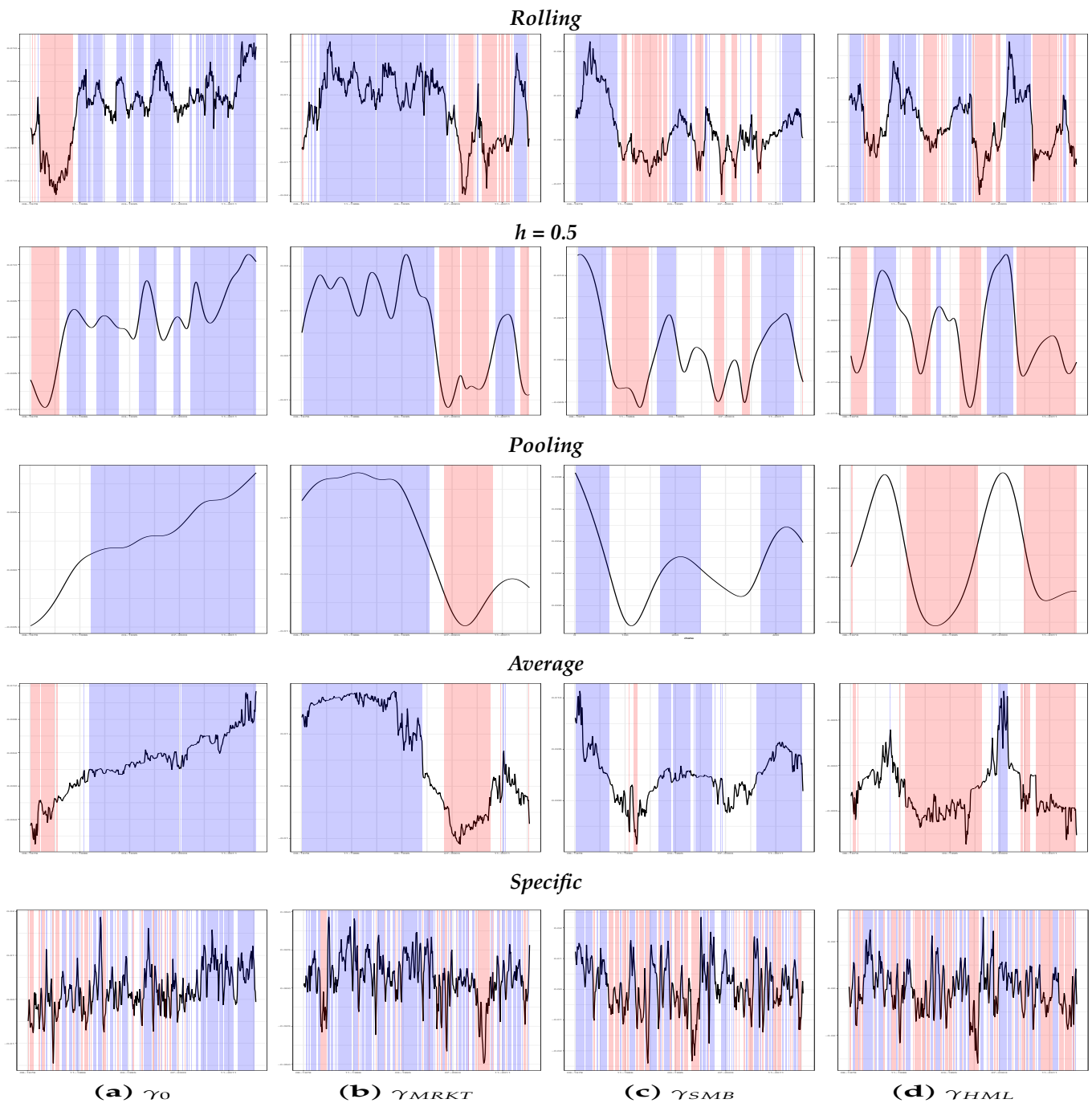


Figure 6. Comparison of γ s significance of different approaches—200 stocks. **Note:** The figure provides a significance analysis of the estimates of risk premium estimates. The blue areas are periods in which the estimates are statistically positive at a 5% level of significance, while the red ones identify periods in which the estimates are negative. The series were computed using different approaches: Rolling window (*FMcB* approach), $h = 0.5$; *Pooling* a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time. The choice of the optimal bandwidth parameter h_t^{opt} has been made using the kernel approach as discussed in Equation (6).

An important aspect of the hierarchical method is the improvement in the forecasting precision. Table 5 displays the *RMSE* of an out-of-sample forecast exercise, reported as deviation from the *RMSE* produced by the benchmark [Fama and MacBeth \(1973\)](#) approach

(Rolling). The analysis of the tables identifies the *Specific* approach as the best method, since it produced a remarkable reduction in the loss function, hence more precise forecast. The overall gains were greater for the portfolios of sizes 25 and 55. This approach produced improvements of around 6% with respect to the basic Fama and MacBeth (1973) and of 4.5% with respect to the kernel approach with optimal bandwidth parameter set to 0.5. Further, since the *Specific* approach also outperformed the other two kernel methods, *Polling* and *Average*, the importance of the time variation in the bandwidth parameters h and its optimisation for each asset is clear. Further, in line with Adrian et al. (2015), we observed that the classical rolling window approach was always outperformed by the kernel ones.

Table 5. Percentage reduction of RMSE in respect to the benchmark model.

	Bandwidth Choice: RMSE		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}
25 Portfolios			
$h = 0.5$	−0.522	−0.483	−2.302
Polled	−0.396	−0.168	−2.140
Average	−0.424	−0.444	−2.048
Specific	−5.800	−5.383	−6.738
55 Portfolios			
$h = 0.5$	−0.528	−0.529	−2.235
Polled	−0.219	−0.196	−1.943
Average	−0.147	0.019	−1.871
Specific	−5.349	−4.444	−6.339
200 Stocks			
$h = 0.5$	−0.535	−0.664	−1.902
Polled	0.033	−0.014	−1.382
Average	0.102	−0.107	−1.413
Specific	−2.393	−2.119	−3.310

Note: The table provides the RMSE for the out-of-sample one-step ahead forecasting exercise as a percentage deviation from the benchmark model of Fama and MacBeth—*Rolling*. The competing models are the following: $h = 0.5$; *Polling*, a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time.

Finally, increasing the sample size did not help to reduce the RMSE; the smallest values were reached performing the analysis for the 55-portfolio sample, while the largest ones for the constituents of *SandP500*.

Table 6 presents a pairwise analysis using the Diebold and Mariano (1995) test, henceforth *DM*, performed to certify the significance of the superior forecasting performance of the Gaussian kernel approach with a time-varying bandwidth. The p -values of the *DM* test were calculated under the null hypothesis that two competing models had the same predictive accuracy, while the alternative was that the two methods had significantly different levels of accuracy. The analysis was conducted for all the samples and methods. The results are very striking and indicate that the *DM* test for the *Specific* method were statistically significant at the 0.01 level, confirming the aforementioned results.

The key role of the time variation in the bandwidth parameter is also emphasized by the results of the method labelled *Average* with respect to the $h = 0.5$ and *Polling* approaches. Here, the null hypothesis of no difference in terms of performance could not be rejected. In line with the literature, the Fama and MacBeth (1973) five-year rolling window approach was never preferred to the kernel regression method with $h = 0.5$, *Polling*, or *Average*. Instead, ambiguous results were produced for the relations between $h = 0.5$ and *Polling*, where the former was preferred only in those cases where *Polling* had h s lower than 0.5, such as the 25 portfolios, with $RMSE_{w=12}$. Such evidence was not unexpected given the nature and the characteristics of the *Polling* approach. Indeed, as Table 1 shows, by increasing the sample size, the degree of smoothness increased, producing flatter estimates that performed well in a forecasting exercise.

Some further results on model comparisons and explanation of results are presented in Table 7; it displays the correlations between the beta estimates generated by each different asset. The *Specific* approach is the method that produced less correlated estimates; the

difference with respect the other methods is between 60% and 70%. This finding is robust in terms of changing sample and methods relatively to the choice of the bandwidth. Such results, together with the fact that the *Specific* approach provided small standard errors, let us solve two of the main critiques of the [Fama and MacBeth \(1973\)](#) method (error in variable problem and cross-sectional correlation), remaining agnostic on the choice of data between portfolios and individual stocks (see [Shanken \(1992\)](#) and [Adrian and Franzoni \(2005\)](#), among the others).

Table 6. Diebold and Mariano test results.

	25 Portfolios				55 Portfolios				200 Stocks			
	Rolling	$h = 0.5$	Polling	Average	Rolling	$h = 0.5$	Polling	Average	Rolling	$h = 0.5$	Polling	Average
$h_{w=12}$												
$h = 0.5$	0.0179				0.0144				0.0193			
Rolling	0.2688	0.0407		0.4797	0.0426		0.9222	0.1062				
Average	0.2401	0.4732	0.8207		0.6468	0.0767	0.4621		0.7755	0.0235	0.8493	
Specific	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0018	0.0061	0.0043	0.0005
$h_{w=24}$												
$h = 0.5$	0.0202				0.0134				0.0113			
Rolling	0.5511	0.0569			0.3998	0.0542			0.9527	0.1051		
Average	0.1953	0.7605	0.0466		0.9368	0.0392	0.1147		0.7303	0.0183	0.7439	
Specific	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019	0.0044	0.0025	0.0007
h_{Kern}												
$h = 0.5$	0.0000				0.0001				0.0007			
Rolling	0.0001	0.0521			0.0001	0.0542			0.0088	0.1181		
Average	0.0001	0.2588	0.5973		0.0004	0.1260	0.5895		0.0098	0.1721	0.5460	
Specific	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: The table provides the p -values of the DM test applied to the results of Section 5. The null hypothesis is that the two competing forecasting models have the same predictive accuracy, while the alternative is that the two methods have a significantly different level of accuracy for the out-of-sample one-step ahead forecasting exercise.

Table 7. Correlation matrix among factor risk loadings.

	25 Portfolios			55 Portfolios			200 Stocks		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
β_{MRKT}									
Rolling	0.2969	0.3216	0.2747	0.3118	0.3359	0.2888	0.3074	0.3183	0.2936
$h = 0.5$	0.3159	0.3185	0.3141	0.3334	0.3381	0.3317	0.3293	0.3335	0.3301
Pooling	0.3284	0.3608	0.3303	0.3785	0.4235	0.3782	0.5030	0.5966	0.4945
Average	0.3106	0.3389	0.2894	0.3663	0.4114	0.3491	0.4525	0.5235	0.4451
Specific	0.0887	0.0937	0.0892	0.0985	0.1108	0.0967	0.1026	0.1220	0.0755
β_{SMB}									
Rolling	0.3155	0.3379	0.2995	0.2984	0.3204	0.2797	0.4223	0.4314	0.4187
$h = 0.5$	0.3335	0.3356	0.3267	0.3266	0.3257	0.3293	0.5252	0.5368	0.4829
Pooling	0.3491	0.3864	0.3471	0.3721	0.4076	0.3740	0.6785	0.7183	0.6254
Average	0.3406	0.3774	0.3165	0.3549	0.3838	0.3403	0.6179	0.6618	0.5942
Specific	0.0940	0.0930	0.0884	0.0967	0.0978	0.0900	0.1546	0.1895	0.1151
β_{HML}									
Rolling	0.4920	0.5310	0.4525	0.4417	0.4831	0.4007	0.3418	0.3729	0.3160
$h = 0.5$	0.5263	0.5310	0.5214	0.4689	0.4744	0.4643	0.3625	0.3624	0.3575
Pooling	0.5467	0.5808	0.5488	0.5219	0.5650	0.4744	0.5373	0.6080	0.5309
Average	0.5082	0.5576	0.4826	0.4917	0.5335	0.4702	0.4952	0.5339	0.4774
Specific	0.1150	0.1135	0.1065	0.1094	0.1167	0.0998	0.1024	0.1271	0.0838

Note: The table provides the average correlation among the 3 factor loadings for all the approaches under analysis: Rolling window, with 5-year estimation period and kernel-weighted regressions using 4 different optimal bandwidths; $h = 0.5$; *Pooling* a single value of h coming from the poll average across assets and time (Equation (7)); *Average*, a unique time-varying bandwidth coming from the average of h across assets (Equation (8)); *Specific*, multiple time-varying bandwidths, one for each asset and time, h^{opt} .

These findings extend the results obtained by [Adrian et al. \(2015\)](#) and are consistent with those obtained by [Ferson and Harvey \(1991\)](#), highlighting the importance of using not only a dynamic framework but also a dynamic estimation approach with minimal theoretical restriction.

5.1. Robustness Checks

A substantial number of robustness checks were performed to test the aforementioned findings. Full details are available in Appendix C, where we report the *RMSE* for each approach in terms of deviation from the benchmark *FMcB*.

Firstly, we investigated the sensitivity of the results to the choice of bandwidth parameter range, originally set as [0.05;0.95]. Three alternative intervals were analysed: [0.35;0.95], [0.05;0.6] and [0.25;0.75]. The results, reported in Table A1, confirm that the *Specific* approach led to a reduction in *RMSE* that oscillated between 0.6% and 8.2% for the 25- and 55-portfolio samples and between 0.5% and 6.4% for the constituents of *SandP500*.

Further, the awareness of possible overfitting issues due to the combination of sample size of the training period and the small value for the bandwidth parameters led us to also investigate the specification of the bandwidth parameter using a time-varying *LASSO* approach inside the hierarchical methodology. After an accurate analysis for the choice of the penalization parameter, we decided to use values of λ that allowed us to maintain the model unchanged⁶ ($\lambda \in [0.00005;0.00001]$). The results, displayed in Table A3, show that, by increasing the penalisation, we significantly increased the gain of the *Specific* technique, which resulted to be between 2.8% and 10.2%.

To avoid the possible presence of overfitting concerns, Table A4 shows how changes in the size of training period affect the results. We investigated the results using ten years of data, $T = 120$ observations, and fifteen years of data, $T = 180$ observations. Table A4 confirms that the *Specific* outperformed the benchmark model and was the one with the highest reduction in *RMSE*. In Table A2, the analysis relatively to changing the sample period in order to exclude the global financial crisis is reported. The new sample tested was 1973–2007. The results confirm again that the *Specific* approach outperformed its competitors.

Finally, since the goal of this paper is to propose a new estimation method to increase the forecasting performances of any asset pricing model, here, we consider different model specifications, namely, the momentum factor by Carhart (1997) and the five-factor model by Fama and French (2015).

6. Concluding Remarks

In this paper, we developed a new framework for the estimation of beta coefficients for a generic dynamic asset pricing model that imposes little a priori structure and generalizes the classic two-step Fama and MacBeth (1973) procedure. The time variation in the beta estimates is found from a kernel-weighted regression that significantly improves on conventional results in terms of *RMSE*. The cross-validation procedure allows us to optimise the choice of the time bandwidth parameter for each asset at each point in time. This very flexible approach, without imposing an extensive a priori structure, improves the estimation of the risk premia. The empirical results overwhelmingly show that the time variation of risk associated with stocks and portfolios must be captured with an estimation procedure that, on one hand, avoids imposing an excessive a priori structure and, on the other hand, takes into account the specific features of each asset and the time variation of its generating mechanism. The methodology is able to produce an increase in the forecasting performance greater (between 4% and 7%) than the alternative methods and independently for any type of model and asset.

Despite the empirical nature of this work, further development and application of this methodology is possible. Especially from a modelling point of view, further studies could focus on re-assessing the existing financial factors in a time-varying context both in the American and European markets. For instance, accounting for the time component of the different factors might reveal the importance of financial ratios in revealing corporate financial soundness and helping the competitive position of an enterprise Valášková (2020).

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Formulation Details

In [Fama and MacBeth \(1973\)](#) seminal paper, they consider N assets and m factors; firstly, the factor exposures, or betas, are computed from the following time-series regression produced for all the N assets:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{1,i}F_{1,t} + \beta_{2,i}F_{2,t} + \dots + \beta_{m,i}F_{m,t} + u_{i,t},$$

where $i \in [1 : N]$; $t \in [1 : T]$; $R_{i,t}$ is defined as the nominal return on the i th asset between period t and $t - 1$; $R_{f,t}$ denotes the risk-free rate. Then, $F_{j,t}$, where $j \in [1 : m]$, is a potential explanatory factor, while $\beta_{j,i}$ represents the factor loading, that describes the degree of exposure of each asset to the factor, and $u_{i,t}$ is assumed to be $iid(0, \sigma_u^2)$.

The second step of the [Fama and MacBeth \(1973\)](#) method is to compute T cross-sectional regressions of the excess return of the assets on the m estimated betas, $\hat{\beta}$, computed in the previous step. All these regressions use the same $\hat{\beta}$, since the objective of the [Fama and MacBeth \(1973\)](#) approach is to estimate the exposure of the N returns to the m factors loadings over time. Hence,

$$(R_{i,t} - R_{f,t}) = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{1,i} + \gamma_{2,t}\hat{\beta}_{2,i} + \dots + \gamma_{m,t}\hat{\beta}_{m,i} + \varepsilon_{i,t},$$

where γ_j s measure the risk premia associated with each F_j . Hence, the method determines $m + 1$ series of the γ s, which are also generally considered to be constant. If the model is well specified and all the factors considered are significant, then the risk loadings explain the cross-sectional differences, $\hat{\gamma}_0 = 0$, and $\hat{\gamma}_j$ represent the average risk premia for each factor.

Appendix B. Kernel-Weighted Regression

In order to account for the time variation in the coefficients in our models, we implement a non-parametric kernel approach that has the main advantage of requiring minimal theoretical restriction on the functional form. Specifically, we extend the work by [Giraitis et al. \(2014\)](#) on autoregressive processes to a kernel smoothing regression. [Giraitis et al. \(2014\)](#) consider the $AR(1)$ process

$$y_t = \phi_{t-1}y_{t-1} + u_t, \tag{A1}$$

where u_t is $iid(0, \sigma_u^2)$ and there is some initialization of the process y_0 , whereas ϕ_{t-1} is a random coefficient, $u_t | \Omega_{t-1} = 0$ and $\phi_t | \Omega_{t-1} = \phi_t$. The stability of the model depends on the TVP nature of the AR parameters satisfying various smoothness conditions. [Giraitis, Kapetanios and Yates \(2014\)](#) model the TVP parameter, denoted by ϕ_t , for an $AR(1)$ as a rescaled random walk, where $\{a_t\}$ is a non-stationary process which defines the random drift and $-1 < \phi < 1$. In this context, ϕ_t is a standardized version of a_t , so that

$$\phi_t = \phi \frac{a_t}{\max_{0 \leq k \leq t} |a_k|}, \dots, t > 0, \tag{A2}$$

where the stochastic process a_t is assumed to be a drift-less random walk, so that $a_t = a_{t-1} + w_t$, and where w_t is a stationary process with zero mean. In addition, $\phi \in (0, 1)$ and

ϕ_{t-1} is bounded away from the boundary points of -1 and 1 . The above framework can be extended to the time-varying $AR(p)$ model

$$y_t = \sum_{i=1}^p \phi_{t-1,i} y_{t-i} + u_t$$

and can be used with the boundary conditions

$$\phi_{t,i} = \phi \frac{a_{t,i}}{\max_{0 \leq k \leq t} |a_{k,i}|}, \dots, t > 1, \tag{A3}$$

where $0 < \phi < 1$ and each $a_{t,i}$ in an independent version of the a_t process defined above. Under these assumptions, the maximum absolute eigenvalues of the matrix

$$A_t = \begin{bmatrix} \phi_{t,1} & \phi_{t,2} & \dots & \dots & \dots & \phi_{t,p} \\ 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & 0 \\ \dots & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & 1 & 0 \end{bmatrix}$$

are bounded above by unity for all t . Giraitis et al. (2014) show that the coefficient process $\{\phi_t; t = 1, \dots, T\}$ converges in distribution as T increases to the limit

$$\{\phi_t; 0 \leq \tau \leq 1\} \rightarrow_D \{\phi \tilde{W}_\tau; 0 \leq \tau \leq 1\},$$

where $\tilde{W}_{(\cdot)}$ is the standard Brownian motion. The approach for estimating the time-varying parameter, ϕ_t is to use the moving window estimator for the $AR(1)$ RC model

$$\hat{\phi}_t = \frac{\sum_{i=1}^H K\left(\frac{t-k}{H}\right) y_t y_{t-1}}{\sum_{i=1}^H K\left(\frac{t-k}{H}\right) y_{t-1}^2}, \tag{A4}$$

where $K\left(\frac{t-k}{H}\right)$ is a kernel and continuously bounded function, such as the Epanechnikov kernel with finite support, or the familiar Gaussian kernel with infinite support. Generalising this estimation method, a regression can be expressed as

$$y_t = x'_t \beta_t + u_t, \tag{A5}$$

where $\beta_t = (\beta_{1,t}, \beta_{2,t}, \dots, \beta_{k,t})$; it is assumed that each $\beta_{j,t}$ follows a bounded random walk. x'_t is the matrix $(m \times T)$ containing the time series of the factors. Therefore, the kernel-weighted regression estimator for $\beta_{j,t}$ is

$$\hat{\beta}_{j,t} = \left(\sum_{j=1}^T w_{jt} x_j x'_j \right)^{-1} \left(\sum_{j=1}^T w_{jt} x_j y_j \right), \tag{A6}$$

where $w_{jt} = K\left(\frac{t-k}{H}\right)$. The authors prove that, if the bandwidth is $o_p(T^h)$ with $h = 1/2$ and given homoskedasticity of the error process, then

$$Var(\hat{\beta}_t) = \hat{\sigma}_u^2 \left(\sum_{j=1}^T w_{jt} x_j x'_j \right)^{-1} \sum_{j=1}^T w_{jt}^2 x_j x'_j \left(\sum_{j=1}^T w_{jt} x_j x'_j \right)^{-1}, \tag{A7}$$

where $\hat{\sigma}_u^2 = \frac{1}{T} \sum_{i=1}^T (y_t - x'_t \beta_t)^2$. While, if u_t is heteroscedastic, then the covariance matrix of the TVP parameter estimates is given by

$$Var(\hat{\beta}_{j,t}) = \left(\sum_{j=1}^T w_{jt} x_j x'_j \right)^{-1} \left(\sum_{j=1}^T w_{jt}^2 x_j x'_j \hat{u}_t^2 \right) \left(\sum_{j=1}^T w_{jt} x_j x'_j \right)^{-1}, \tag{A8}$$

which can be used for inference. One appealing characteristic of this approach is that it nests rolling window estimates of the regression betas and is equivalent to kernel-smoothing estimators using a uniform one-sided kernel instead of a Gaussian two-sided kernel. A key role is played by the decision about the selection of the bandwidth; for a given kernel function $K\left(\frac{t-k}{H}\right)$, the bandwidth H represents the degree of smoothness of the estimates. Giraitis et al. (2014) proved that a bandwidth of $H = T^h$, with $h = 0.5$, provides an estimator with desirable properties such as consistency and asymptotic normality and, in addition, provides valid standard errors.

Another appealing characteristic of such approach is that it nests, as a special case, rolling window estimates of betas (see, for example, Chen et al. 1986; Ferson and Harvey 1991 and Petkova and Zhang 2005, among many others). Rolling beta estimates are equivalent to kernel-smoothing estimators obtained using a uniform one-sided kernel instead of a Gaussian two-sided kernel and it has been proved that the order of the smoothing bias of the estimator for the betas and the price of risk parameters is larger for one-sided kernels.

In the kernel estimation approach, a key role is played by the selection of the bandwidth. For a given kernel function $K\left(\frac{t-k}{H}\right)$, the bandwidth H represents and controls the degree of smoothness of the estimates. In other terms, if the bandwidth is small, the estimates are under-smoothed, with high variability, otherwise, if the value of H is big, the resulting estimators are over-smoothed and further from the real function. Different approaches have been proposed to handle the choice of the bandwidth. Ang and Kristensen (2012) suggest to optimise the choice of the bandwidth for conditional and long estimates in order to reduce any finite-sample biases and variances. Giraitis et al. (2014), instead, proved that, if the bandwidth is $H = T^h$, with $h = 0.5$, the estimator has desirable properties such as consistency and asymptotic normality and, in addition, provides valid standard errors.

Appendix C. Robustness Checks

Table A1. Percentage reduction in RMSEs for different bandwidth parameter intervals.

	25 Portfolios			55 Portfolios			200 Stocks		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
$h \in [0.05; 0.95]$									
$h = 0.5$	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
$h \in [0.05; 0.6]$									
$h = 0.5$	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.412	-0.413	-1.743
Polling	-2.086	-1.640	-3.800	-1.921	-0.963	-3.585	-1.499	-0.751	-2.796
Average	-1.994	-1.863	-3.625	-2.428	-1.786	-3.446	-1.894	-1.393	-2.688
Specific	-6.970	-6.638	-8.256	-7.374	-6.522	-8.315	-5.752	-5.087	-6.486
$h \in [0.35; 0.95]$									
$h = 0.5$	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.412	-0.413	-1.743
Polling	0.014	0.087	-1.758	0.028	0.047	-1.714	0.022	0.036	-1.337
Average	0.030	0.109	-1.754	0.075	0.100	-1.671	0.058	0.078	-1.304
Specific	-0.891	-0.766	-2.306	-0.654	-0.517	-2.210	-0.510	-0.403	-1.724
$h \in [0.25; 0.75]$									
$h = 0.5$	-0.522	-0.483	-1.056	-0.528	-0.529	-1.119	-0.412	-0.413	-0.872
Polling	-0.382	-0.245	-0.766	-0.279	-0.200	-0.740	-0.218	-0.156	-0.577
Average	-0.339	-0.288	-0.738	-0.155	-0.087	-0.714	-0.121	-0.068	-0.557
Specific	-2.014	-1.855	-2.029	-1.900	-1.533	-1.842	-1.482	-1.196	-1.436

Note: The table provides the RMSE for the out-of-sample one-step ahead forecasting exercise comparing 4 different intervals for the identification of the optimal bandwidth parameter. The results are expressed as a deviation from the RMSE produced by the benchmark model, FMcB.

Table A2. Percentage reduction in RMSEs for different sample.

	25 Portfolios			55 Portfolios			200 Stocks		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
08/1973–01/2020									
$h = 0.5$	−0.522	−0.483	−2.302	−0.528	−0.529	−2.235	−0.535	−0.664	−1.902
Polling	−0.396	−0.168	−2.140	−0.219	−0.196	−1.943	0.033	−0.014	−1.382
Average	−0.424	−0.444	−2.048	−0.147	0.019	−1.871	0.102	−0.107	−1.413
Specific	−5.800	−5.383	−6.738	−5.349	−4.444	−6.339	−2.393	−2.119	−3.310
08/1973 – 08/2007									
$h = 0.5$	0.334	0.719	−1.033	−0.947	−0.608	−2.263	−0.794	−0.510	−1.896
Polling	0.401	0.956	−0.959	−0.717	−0.238	−2.076	−0.601	−0.199	−1.740
Average	0.471	0.921	−0.913	−0.648	−0.151	−1.926	−0.543	−0.127	−1.614
Specific	−2.688	−1.904	−3.925	−3.233	−2.369	−4.696	−2.709	−1.986	−3.935

Note: The table provides the *RMSE* for the out-of-sample one-step ahead forecasting exercise comparing different sub samples identified around the global financial crisis (08/1973–08/2007). The results are expressed as a deviation from the *RMSE* produced by the benchmark model, *FMcB*.

Table A3. Percentage reduction in RMSEs for different penalization parameters in LASSO.

	25 Portfolio			55 Portfolio			200 Stocks		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
No Lasso									
$h = 0.5$	−0.522	−0.483	−2.302	−0.528	−0.529	−2.235	−0.535	−0.664	−1.902
Polling	−0.396	−0.168	−2.140	−0.219	−0.196	−1.943	0.033	−0.014	−1.382
Average	−0.424	−0.444	−2.048	−0.147	0.019	−1.871	0.102	−0.107	−1.413
Specific	−5.800	−5.383	−6.738	−5.349	−4.444	−6.339	−2.393	−2.119	−3.310
$\lambda = 0.0001$									
$h = 0.5$	−0.522	−0.483	−2.302	−0.528	−0.529	−2.235	−0.328	−0.328	−1.386
Polling	−0.556	−0.369	−2.496	−0.311	−0.142	−1.983	−0.193	−0.088	−1.229
Average	−1.294	−1.352	−4.975	−0.158	−0.019	−1.810	−0.098	−0.012	−1.122
Specific	−6.821	−6.271	−10.188	−5.229	−4.625	−6.066	−3.242	−2.868	−3.761
$\lambda = 0.00005$									
$h = 0.5$	−0.522	−0.483	−2.302	−0.528	−0.529	−2.235	−0.328	−0.328	−1.386
Polling	−0.469	−0.229	−2.196	−0.311	−0.142	−1.983	−0.193	−0.088	−1.229
Average	−0.514	−0.410	−2.202	−0.134	−0.026	−1.806	−0.083	−0.016	−1.120
Specific	−6.417	−5.661	−6.676	−5.197	−4.617	−6.015	−3.222	−2.862	−3.729

Note: The table provides the *RMSE* for the out-of-sample one-step ahead forecasting exercise when we consider the LASSO procedure inside our mechanism for the identification of the optimal bandwidth. Here, we report the results for 2 values of the penalty function, $\lambda = 0.0001$ and 0.00005 . The results are expressed as a deviation from the *RMSE* produced by the benchmark model, *FMcB*.

Table A4. Percentage reduction in RMSEs for different sample size of the trading period.

	25 Portfolios			55 Portfolios			200 Stocks		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
$T = 60$									
$h = 0.5$	−0.522	−0.483	−2.302	−0.528	−0.529	−2.235	−0.535	−0.664	−1.902
Polling	−0.396	−0.168	−2.140	−0.219	−0.196	−1.943	0.033	−0.014	−1.382
Average	−0.424	−0.444	−2.048	−0.147	0.019	−1.871	0.102	−0.107	−1.413
Specific	−5.800	−5.383	−6.738	−5.349	−4.444	−6.339	−2.393	−2.119	−3.310
$T = 180$									
$h = 0.5$	1.629	2.964	−0.490	2.641	4.327	0.435	2.213	3.626	0.365
Polling	1.724	3.522	−0.430	3.055	5.114	0.809	2.560	4.286	0.678
Average	1.986	3.558	−0.160	3.613	5.488	1.377	3.027	4.599	1.154
Specific	−3.454	−1.310	−5.815	−0.952	1.820	−3.835	−0.798	1.525	−3.214
$T = 120$									
$h = 0.5$	0.534	1.152	−1.654	−1.517	−0.974	−3.623	−1.271	−0.816	−3.036
Polling	0.642	1.531	−1.535	−1.148	−0.381	−3.324	−0.962	−0.319	−2.786
Average	0.753	1.474	−1.461	−1.037	−0.242	−3.084	−0.869	−0.203	−2.585
Specific	−4.303	−3.049	−6.284	−5.176	−3.794	−7.518	−4.338	−3.179	−6.300

Note: The table provides the *RMSE* for the out-of-sample one-step ahead forecasting exercise comparing 3 different trading periods T for the identification of the optimal bandwidth parameter. The results are expressed as a deviation from the *RMSE* produced by the benchmark model, *FMcB*.

Table A5. Percentage reduction in RMSEs for different asset pricing models.

	25 Portfolios			55 Portfolios			200 Stocks		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
3 Factors									
$h = 0.5$	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
MOM Factor									
$h = 0.5$	-0.569	-0.526	-2.508	-0.576	-0.577	-2.435	-0.357	-0.358	-1.510
Polling	-0.606	-0.402	-2.720	-0.339	-0.155	-2.160	-0.210	-0.096	-1.339
Average	-1.409	-1.473	-5.420	-0.172	-0.021	-1.972	-0.106	-0.013	-1.223
Specific	-7.432	-6.832	-11.099	-5.697	-5.039	-6.608	-3.532	-3.124	-4.097
5 Factors									
$h = 0.5$	-0.655	-0.605	-2.889	-0.663	-0.664	-2.805	-0.411	-0.412	-1.739
Polling	-0.589	-0.287	-2.756	-0.390	-0.179	-2.488	-0.242	-0.111	-1.543
Average	-0.645	-0.514	-2.763	-0.168	-0.032	-2.266	-0.104	-0.020	-1.405
Specific	-8.053	-7.103	-8.378	-6.522	-5.793	-7.548	-4.044	-3.592	-4.680

Note: The table provides the RMSE for the out-of-sample one-step ahead forecasting exercise comparing 3 different asset pricing models: 3-factor Fama and French (1992) model, momentum factor by Carhart (1997) and the 5-factor model by Fama and French (2015) for the identification of the optimal bandwidth parameter. The results are expressed as a deviation from the RMSE produced by the benchmark model, FMcB.

Notes

- 1 The aim of the paper is to produce evidence in support of the importance of the bandwidth liberalisation in a factor model. Any analysis about the importance of the factor is beyond the scope of the paper. In Appendix C, the methodology is applied to different well-established factor models.
- 2 The only Standard and Poor’s constituents which are used are for data available for the entire sample
- 3 See Section 5.1 and Appendix C for robustness checks against other well know financial factors—the momentum factor, MOM, by Carhart (1997), which is computed as the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. Finally, the methodology is tested on the five-factor model by Fama and French (2015). These additional factors represent the *robust minus weak*, RMW, and the *conservative minus aggressive*, CMA, factors. The RMW factor is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolio, while CMA represents the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios.
- 4 The remainder of the paper reports the results concerning the kernel approach methodology for the computation of the time-varying $RMSE_t$, as in Equation (6). The results for the other two approaches are available upon request to the authors
- 5 The aim of the paper is to produce evidence in support of the importance of the bandwidth liberalisation in a factor model. Any analysis about the importance of the factor is beyond the scope of the paper.
- 6 The purpose of the paper is not to identify the best factors but the identification of the best methodology. Therefore, to guarantee an identical setting, the factors in the model are fixed.

References

Adrian, Tobias, and Francesco Franzoni. 2005. Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance* 16: 537–56. [CrossRef]

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. 2015. Regression based estimation of dynamic asset pricing models. *Journal of Financial Economics* 118: 211–44. [CrossRef]

Ang, Andrew, and Dennis Kristensen. 2012. Testing conditional factor models. *Journal of Financial Economics* 106: 132–56. [CrossRef]

Bali, Turan G., and Robert F. Engle. 2010. The inter-temporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics* 57: 377–90. [CrossRef]

Beach, Steven L. 2011. Semi variance decomposition of country-level returns. *International Review of Economics and Finance* 20: 607–23. [CrossRef]

Black, Fischer, Michael C. Jensen, and Myron Scholes. 1972. The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets* 81: 79–121.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. A capital asset pricing model with time-varying covariances. *Journal of political Economy* 96: 116–31. [CrossRef]

Campbell, John Y., and Robert J. Shiller. 1988. Stock prices, earnings, and expected dividends. *Journal of Finance* 43: 661–76. [CrossRef]

Carhart, Mark M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52: 57–85. [CrossRef]

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross. 1986. Economic forces and the stock market. *Journal of Business* 59: 383–403. [CrossRef]

Diebold, Francis X., and Robert S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253–63.

- Engle, Robert. 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20: 339–50. [CrossRef]
- Fama, Eugene F., and James MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 71: 607–36. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 1992. The cross section of expected stock returns. *Journal of Finance* 47: 427–65. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 1995. Size and book-to-market factors in earnings and returns. *The Journal of Finance* 50: 131–55. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 1997. Industry costs of equity. *Journal of Financial Economics* 43: 153–93. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2006. Profitability, investment and average returns. *Journal of Financial Economics* 82: 491–518. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2020. Comparing cross-section and time-series factor models. *The Review of Financial Studies* 33: 1891–26. [CrossRef]
- Ferson, Wayne E., and Campbell R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99: 385–415. [CrossRef]
- Ferson, Wayne E., and Campbell R. Harvey. 1999. Conditioning variables and the cross section of stock returns. *The Journal of Finance* 54: 1325–60. [CrossRef]
- Ghysels, Eric. 1998. On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt? *Journal of Finance* 53: 549–73. [CrossRef]
- Giraitis, Liudas, George Kapetanios, and Tony Yates. 2014. Inference on stochastic time varying coefficient models. *Journal of Econometrics* 179: 46–65. [CrossRef]
- Giraitis, Liudas, George Kapetanios, Konstantinos Theodoridis, and Tony Yates. 2015. Estimating Time-Varying DSGE Models Using Minimum Distance Methods. No. 768. Working Paper. Available online: <https://www.qmul.ac.uk/sef/research/workingpapers/2015/items/768.html> (accessed on 17 November 2021).
- Giraitis, Liudas, George Kapetanios, and Tony Yates. 2018. Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis* 39: 129–49. [CrossRef]
- Markowitz, Harry M. 1968. *Portfolio Selection*. New Haven: Yale University Press.
- Harvey, Campbell R. 1989. Time varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24: 289–317. [CrossRef]
- Harvey, Campbell R. 2001. Asset pricing in emerging markets. *International Encyclopedia of the Social and Behavioral Sciences* 840–45. [CrossRef]
- Harvey, Campbell R., and Yan Liu. 2019. A Census of the Factor Zoo. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3341728 (accessed on 10 September 2021).
- Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. ... and the cross-section of expected returns. *The Review of Financial Studies* 29: 5–68. [CrossRef]
- Jagannathan, Ravi, and Zhenyu Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51: 3–53. [CrossRef]
- Lettau, Martin, and Sydney Ludvigson. 2001a. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56: 815–49. [CrossRef]
- Lettau, Martin, and Sydney Ludvigson. 2001b. Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109: 1238–87. [CrossRef]
- Lewellen, Jonathan, and Stefan Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82: 289–314. [CrossRef]
- Lintner, John. 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20: 587–615.
- Lo, Andrew W., and A. Craig MacKinlay. 1990. Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies* 3: 431–67. [CrossRef]
- Petkova, Ralitsa, and Lu Zhang. 2005. Is value riskier than growth? *Journal of Financial Economics* 78: 187–202. [CrossRef]
- Shanken, Jay. 1992. The current state of the arbitrage pricing theory. *Journal of Finance* 47: 1569–74. [CrossRef]
- Sharpe, William. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–42.
- Valášková, Katarína, Beáta Gavurová, Pavol Ďurana, and Mária Kováčová. 2020. Alter ego only Four times? the Case study of business profits in the visegrad group. *E & M Ekonomie a Management* 23: 101–19. [CrossRef]
- Zolotoy, Leon. 2011. Earnings News and Market Risk: Is the Magnitude of the Post earnings Announcement Drift Underestimated? *Journal of Financial Research* 34: 523–35. [CrossRef]