Hierarchical Time Varying Estimation of a Multi Factor Asset Pricing Model

Richard T. Baillie, Fabio Calonaci & George Kapetanios

Working paper No. 879 January 2019 ISSN1473-0278

School of Economics and Finance

Queen Mary University of London

Hierarchical Time Varying Estimation of a Multi Factor Asset Pricing Model

Richard T. Baillie*
Michigan State University, USA
King's College London, UK
Rimini Center for Economic Analysis, Italy
Fabio Calonaci†
Queen Mary University of London, UK
George Kapetanios‡
King's College London, UK

Abstract

This paper presents a new hierarchical methodology for estimating multi factor dynamic asset pricing models. The approach is loosely based on the sequential approach of Fama and MacBeth (1973). However, the hierarchical method uses very flexible bandwidth selection methods in kernel weighted regressions which can emphasize local, or recent data and information to derive the most appropriate estimates of risk premia and factor loadings at each point of time. The choice of bandwidths and weighting schemes, are achieved by cross validation. This leads to consistent estimators of the risk premia and factor loadings. Also, out of sample forecasting for stocks and two large portfolios indicates that the hierarchical method leads to statistically significant improvement in forecast $RMSE$.

JEL Classification: C22; F31; G01; G15

Keywords: Asset pricing model, Fama MacBeth model, estimation of beta, kernel weighted regressions, cross validation, time-varying parameter regressions.

*Address correspondence: King's Business School, 30 Aldwych, London, WC2B 4BG, Room (N)4.23, or email: richard.baillie@kcl.ac.uk, baillie@msu.edu.

†Address correspondence: School of Economics and Finance, Queen Mary University of London, Mile End Road London E1 4NS, Room GC 5.26, or email: f.calonaci@qmul.ac.uk.

‡Address correspondence: King's Business School, 30 Aldwych, London, WC2B 4BG, Room (N)4.14, or email: g.kapetanios@kcl.ac.uk.

1 Introduction

The concept of a time varying risk premium is a standard idea in financial economics and there are many articles approaching the subject in different ways; e.g. Campbell and Shiller (1988), Ferson and Harvey (1991) and Lewellen and Nagel (2006). There is also the cornerstone of asset pricing as exemplified by Fama and MacBeth (1973), who estimate equity risk premium by a now standard two stage, cross sectional regression method. The method assumes a linear multi factor setting where the pricing of different types of risk are constant.

The Fama and MacBeth (1973) approach has great intuitive appeal as well as being based on natural financial concepts. Hence this paper builds on their approach in the sense that a sequential hierarchical structure is developed which has several components. Our method maintains the Fama MacBeth (1973) stages of estimating risk factors, (or betas) from a time series regression and then the importance of factor loadings (or gammas) from cross sectional regressions. However, we also include intermediate multiple hierarchical stages of selecting optimal bandwidths by cross validation.in the estimation of the betas and also separately for the factor loadings and gammas. Hence a main contributions of this paper is to allow for very general time variation in the risk premium terms, or betas, and corresponding variation in time varying factor loadings coefficients, within an affine pricing kernel specification. The method that we use to calculate the time varying betas and factor loadings depends on a kernel weighted regression, which is an extension of the least squares rolling window regression approach, which has frequently been used in empirical finance; e.g. Jagannathan and Wang (1996), Ghysels (1998) and Lewellen and Nagel (2006), etc.

One great advance of our methodology is to employ a flexible method for the bandwidth selection, which essentially determines the lag of recent updates of the betas (risk factors) and also of the factor loadings. This avoids imposing any priori structure and allows a natural data orientated way for incorporating economic and financial news that is relevant for the pricing of assets under investigation. Our empirical results overwhelmingly indicate the importance of removing the restriction of constant betas and the full superiority of our methodology becomes apparent in terms of prediction of out of sample returns for a wide range of assets. Hence our results show that time variation of risk associated with stocks and portfolios must be captured with an estimation procedure that on one hand avoids imposing excessive a priori structure and on the other hand takes into account the specific features of each assets. The results also indicate the economic importance of the factor loadings, and how they relate to the risk premium. One highlight of our approach is that it leads to between 4% to 7% reduction in one step ahead forecasting performance of excess returns in comparison with a plethora of alternative methods and independently of the data employed.

The remainder of this paper is organized as follows: section 2 provides a brief review of previous literature, while section 3 discussion of the contribution of this paper relative to the existing literature. In section 4, we present the empirical evidence of our results and some robustness checks.

2 Background Literature

The Capital Asset Pricing Model, or ($CAPM$) of Sharpe (1964), Lintner (1965) and Markowitz (196?) is a benchmark of asset pricing models. The model 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 premium. Traditionally this is considered invariant over time and "beta" represents the covariance between the return of the asset and the return on the market portfolio. The basic model has been criticised by Black, Jensen and Scholes (1972), Fama and French (1992), Fama and MacBeth (1973) among others on the grounds that only one factor, the market beta, is inadequate to describe the systemic risk. Hence researchers many have attempted to improve the basic $CAPM$ by the introduction of other factors. Most notably there is the three factor model of Fama and French (1993) which introduced the size, or SMB where positive returns are related to small size, and the high minus low, or HML,factors, high book-to-market ratios are associated with higher returns. While Carhart (1997) introduced a fourth momentum factor, MOM, which describes the tendency of a stock price to continue recent trends.

Other developments with extending the basic $CAPM$ have centred on implementing more flexible estimation strategies where the the beta coefficient(s) are not necessarily assumed to be constant across time or space. For example, see Harvey (1989), Ferson and Harvey (1991, 1993), Bollerslev, Engle and Wooldridge (1988), Fama and French (1997, 2006), etc.

Adrian and Franzoni (2005) have argued 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 policy, and exchange rates. Many researchers, such as Zolotoy (2011), show that variation in betas are more evident around important news announcements.

Jagannathan and Wang (1996), Lettau and Ludvigson (2001), 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, which is the embryonic version of the method used in this paper, 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). This will be discussed more in section 4 of this paper.

Other research has directly exploited the covariation between the market and other assets; see Engle (2002) and Bali and Engle (2010) who estimate time varying betas using multivariate dynamic conditional correlation methods to exploit correlations between cross sectional average returns of various factor portfolios. Another approach by Anderson, Bollerslev, Diebold and Wu (2006) estimates dynamic betas through using the realized beta from high frequency intra day returns. However, the main framework in our analysis is that of Fama and MacBeth (1973), which is outlined below. It should also be noted that early work on the idea of estimating equity risk premia in a linear multi factor setting were also in Black, Jensen, and Scholes (1972)[1].

2.1 Fama and MacBeth formulation

The seminal paper of Fama and MacBeth (1973) advocates a two step procedure to estimate risk premia in the multi factor $CAPM$ setting and also provides a test the explanatory power of the various chosen factors. The usual model assumes the coefficients are constant and estimates them by OLS. 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 and hence estimate the beta parameters. The second step consists of a cross section 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 premia coefficients, or gammas, for each factor. The great insight of Fama and MacBeth (1973) is to average these coefficients to obtain expected for a unit exposure to each

[1]There is also a literature on the statistical theory of properties for estimated linear factor models with constant coefficients see e.g. Shanken (1985, 1992), Jagannathan and Wang (1998), Shanken and Zhou (2007), Kan, Robotti, and Shanken (2013).

risk factor over time and to test if these are adequately priced by the market.

To be specific, we consider N assets and m factors; and in the first step the factor exposures, or betas are computed from the regression considered for all the N assets:

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

with $i \in [1:N]$ and $t \in [1:T]$ and where $R_{i,t}$ is defined as the nominal return on the i'th asset between period t and $t-1$, and $R_{f,t}$ denotes the risk free rate. Then $F_{j,t}$, with $j \in [1:m]$, is a 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 section 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 factor loadings over time. Hence,

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

where the γ_js measure the risk premia associated with each F_j. Hence the method determines $m+1$ series of the γs, 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 cross sectional differences, $\hat{\bar{\gamma}}_{0,.} = 0$ and $\hat{\bar{\gamma}}_{j,.}$ represent the average risk premium associated with each risk factor.

3 Hierarchical Methodology

The key point in our contribution of a hierarchical methodology is to develop a flexible methodology to easily allow time variation in both the betas and gammas of the baseline Fama and MacBeth (1973) approach. The main tool for achieving this is to have a flexible bandwidth parameter which essentially controls the weight given to local information for updating the beta and gamma coefficients. The really important innovation is to optimize the bandwidth selection based on out of sample cross validation methods. Before outlining our method in detail we briefly mention some of the previous literature which has attempted to model more flexible coefficients in the Fama MacBeth $CAPM$.

As outlined below, there has already been considerable awareness of the importance of incorporating time variation in the estimation of the $CAPM$. In particular, rolling window estimates of betas have been considered by Chen, Roll, and Ross (1986), Ferson and Harvey (1991) and Petkova and Zhang (2005) among others.

Related to our approach, Adrian, Crump and Moench (2015) used the weighted kernel estimator proposed by Ang and Kristensen (2012) for the purpose of being robust to misspecification. However, one major innovation approach is the use of optimal bandwidth selection for kernel estimation for deriving improved local estimates of the betas and this improves on out of sample prediction of returns. The first part of our Hierarchical approach is to calculate the Time Varying Parameters (TVP) associated with the coefficients of the factors (βs). The method we use is based on a kernel weighted regression, hence

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

with $i \in [1 : N]$ number of assets, $t \in [1 : T]$ period of time and $k \in [1 : K]$ number factors and h is the bandwidth to be discussed later. The new hierarchical approach is subsequently tested against different combination of factors that are consistent with previous research in this area. It is generally assumed throughout that $u_{n,t+1}$ is $i.i.d.(0, \sigma^2)$; although as discussed in the appendix, corrections for autocorrelation or heteroskedasticity are available if required. But in general there appeared from diagnostic tests for there to be no reason to implement them. The β parameters are estimated by an extension of the methodology of Giraitis, Kapetanios and Yates (2014) and is summarized in the appendix of this paper. Hence the beta for the $k\prime th$ factor is estimated by

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

where $K(\frac{t-j}{H})$ is assumed to be a Gaussian kernel function. Intuitively, the kernel function places more weight on local observations and the rate of decay is governed by the bandwidth, h. Giraitis, Kapetanios and Yates (2014, 2018) show that under very mild conditions, kernel based estimates of random coefficient processes have very desirable properties such as consistency and asymptotic normality and in addition provide valid standard errors. They address also the delicate issue of the choice of bandwidth and the resulting bias and variance of the estimator[2]. The use of a cross validation approach implies that the choice of bandwidth parameter, $h,$ is selected from an out of sample, one-step ahead forecasting comparison over a grid search of h which incorporates 18 different values of h, for the grid of $h \in [0.05; 0.9]$ with an interval of 0.05 for each grid.

The end of the first stage of the Hierarchical process generates for each asset, i, a time series of beta estimates for different values of the bandwidth parameter h. These estimated betas allow the identification of the price of risk factor loadings for different values of h,γ_h; using the following equation:

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

where $\varepsilon_{n,t,h}$ is assumed to be $i.i.d.(0, \sigma^2_\varepsilon)$. This process generates $k + 1$ series of γs (including the constant) for every factor and for every h. The cross validation procedure then compares the forecasting performance of the competing models, via the computation of the forecast errors $e_{i,t+1,h}$ and is implemented over the initial T_0 observations in a *training period*, which also allows for out of sample forecast analysis to be based on the remaining $(T - T_0)$ periods. The training period is fixed at 60 observations, or 5 years of data and we also performed robustness tests with different values of T_0 of 120 and 180. The one step ahead forecast for each asset is then obtained from the following regression

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

The forecast errors, $e_{i,t+1,h}$ are computed for each period and for each of the eighteen different values of h, which leads to a forecast error matrix of dimension $18 \times (T - T_0)$ for each asset.

Essentially five alternative tests for robustness were calculated; namely (i) to try different training periods of $60, 120$ and 180 observations; (ii) to optimize the choice of the bandwidth on both the first step with the estimation of the betas and also with the second step estimations of the gamma; (iii) to try eighteen different values for h the bandwidth; (iv) to examine both periods of before and after the Great Financial Crisis and (v) to use a time varying lasso method with

[2]Ang and Kristensen (2012) have investigated the asymptotic distributions for conditional and long run estimates of the alpha and betas; and suggest the use of different bandwidth mainly for the purpose of reducing finite-sample biases and variances.

variation in the penalty function for the identification of the best model at each time period.

The time varying $RMSE$ is calculated at each point in time and for each asset; and the value of h selected is chosen to minimize the $RMSE$ loss function. Several different criteria and approaches were investigated; including rolling regressions, and non parametric kernel smoothed regressions. The former method refers to the classical rolling window approach with different window, or horizons, 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 computed instead:

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

with $H^{(i)} = T^{h'}$ and $h' \in [0.05; 0.9]$. Clearly when $W(H) = 1$ the formula reduces to the regular $RMSE_t$ formula in equation (4). 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. This matrix of $RMSE_t$ are used to determine the optimal values of h for each asset,$h_{i,t}^{opt}$.

Once the multi step hierarchical procedure has found the "optimum" beta coefficient estimates, then the Hierarchical method performs cross section regressions to identify the price of risk actor loadings for different values of h and γ_h; using the following equation:

$$(R_{i,t} - R_{f,t})_h = \gamma_{0,h} + \widehat{\beta}_{1,t,i,h}\gamma_{1,t,h} + \widehat{\beta}_{2,t,i,h}\gamma_{2,t,h} + \widehat{\beta}_{3,t,i,h}\gamma_{3,t,h} + \varepsilon_{n,t,h} \tag{7}$$

where $\varepsilon_{n,t,h}$ is assumed to be $i.i.d.(0, \sigma_\varepsilon^2)$. .

3.1 Estimation of factor risk loadings - β_t

The cross validation procedure is used turns out to be very important in searching for the most appropriate value of the bandwidth parameter, h, in the kernel function that sets the degree of smoothness of the estimates. This parameter is critical in providing the appropriate degree of persistence in determining the "memory" of the window used in the TVP estimation of the different stages of the Hierarchical modelling procedure. classical three FF model: market factor, $MRKT$, size factor, SML, and book-to-market factor, HML. The following list other possible ways for determining the factor risk loadings β_t, using the optimal parameter $h_{i,t}^{opt}$; they are:

(i) A kernel weighted approach with $h = 0.5$. As showed by Giraitis et al. (2014), this bandwidth allows to get smooth estimates with desirable properties such as consistency and asymptotic normality and in addition provides asymptotically valid standard errors. This model is then used as a benchmark.

(ii) Alternative kernel approach, where h is fixed for each assets and time and is determined from a pooled average of the optimal bandwidth parameters, $h_{i,t}^{opt}$, as follow:

$$\bar{h}^{Pooled} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} h_{t,i} \tag{8}$$

(iii) 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:

$$\bar{h}_t^{Average} = (N)^{-1} \sum_{i=1}^{N} h_{t,i} \tag{9}$$

(iv) The optimal h, which is specific for each asset to give $h_{i,t}^{opt}$.

All these 5 approaches are implemented in the three factor Fama and French (1993) model:

$$R_{i,t} - R_{f,t} = \beta_{1,t,i,m} F_{MRKT,t} + \beta_{2,i,t,m} F_{SMB,t} + \beta_{3,i,t,m} F_{HML,t} + u_{i,t,m} \tag{10}$$

where $m \in [1 : 5]$ represents the different approach used for the computation of the factor loadings, βs.

3.1.1 Hierarchical estimation of Risk Premium

The TVP estimates of beta, $\hat{\beta}$, are then used in the third of our procedure, which facilitates computation of risk premia associated with the factors under investigations,γs. Hence this stage of our procedure follows has some similarities with Fama and MacBeth (1973) article, except that at each point in time, we consider multiple TVP estimates of beta instead of them fixed as a constant. Our hierarchical methodology then replaces the assets' excess returns by their corresponding time varying kernel weighted average, $(R_{n,t} - R_{f,t})$, which are computed using the bandwidth, h that had been selected in the previous step for the computation of the βs. The kernel weighted averages for the excess returns are then

$$(R_{i,}\widehat{- R_{f,t+1}}) = \sum_{k=1}^{T} K\left(\frac{t-k}{H}\right)(R_{i,k} - R_{f,k}) \tag{11}$$

where $K(\frac{t-k}{H})$ is the same continuously bounded kernel function and $H^* = T^{h_m}$ with $m \in [2:5]$. The cross section of these smoothed excess returns are then used for the OLS regressions, to identify the risk premia,

$$(R_{i,t}\widehat{-R}_{f,t}) = \gamma_{0,m} + \widehat{\beta}_{1,t,i,m}\gamma_{1,t,m} + \widehat{\beta}_{2,t,i,m}\gamma_{2,t,m} + \widehat{\beta}_{3,t,i,m}\gamma_{3,t,m} + \varepsilon_{i,t,m} \tag{12}$$

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 βs; each one of which is of length $T-T_0$. The last stage of the hierarchical approach is to select the best bandwidth in terms of $RMSE$ minimization in out of sample forecasting. One method is to forecast the average excess return across all assets using the average of the estimated betas, which realizes the time series of forecasts of the average,

$$\overline{R_{t+1} - R_{f,t+1}} = \hat{\gamma}_{0,m} + \hat{\gamma}_{1,t,m}\hat{\bar{\beta}}_{1,t,m} + \hat{\gamma}_{2,t,m}\hat{\bar{\beta}}_{2,t,m} + \hat{\gamma}_{3,t,m}\hat{\bar{\beta}}_{3,t,m} \tag{13}$$

where

$$\overline{R_{t+1} - R_{f,t+1}} = \frac{1}{N}\sum_{i=1}^{N}(R_{t+1} - R_{f,t+1}) \tag{14}$$

and

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

The $RMSE$ are then computed for each method and compared across estimation strategies using different bandwidths. Another method is to use the excess return for each asset using its betas but assuming that the gammas are identical across the portfolios. This allows investigation of how the different models perform for each asset and appreciate better the contribution of the different approaches for the estimation of the betas. The forecasts are computed as follow:

$$R_{i,t+1} - R_{f,t+1} = \hat{\gamma}_{0,m} + \hat{\gamma}_{1,t,m}\hat{\beta}_{1,t,i,m} + \hat{\gamma}_{2,t,m}\hat{\beta}_{2,t,i,m} + \hat{\gamma}_{3,t,m}\hat{\beta}_{3,t,i,m} \tag{15}$$

Then $RMSE$ for each of these quantities are calculated their significance analysed by using the Diebold and Mariano (1995) test.

3.2 Data

The new hierarchical methodology is now applied to three different financial returns data sets. The first is an $N = 25$ sized portfolio which is sorted by size and book-to-market; while the second is an $N = 30$ sized portfolios sorted by industry. We further use 200 individual stock prices from the Center for Research in Securities Prices (CRSP), so that $N = 200$. We consider excess return over the 30-day Treasury bill yield, with the total series covering the period from August 1973 through April 2016, for a total of $T = 514$ observations. All of these portfolios are constructed from Kenneth French's on-line web site.

We use the following set of factors in our subsequent Hierarchical Analysis; namely the 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) factors are derived in the same was as in Fama and French (1993). We further considered the momentum factor, MOM, of 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. Again, all of them are available from Ken French's on-line data library. Full details of the results including the momentum factor are available on line; as are analysis of thee five factor model from 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. Full details of the analysis form the five factor model is again available on-line, but is suppressed in this version of the paper in the interests of conserving space.

3.3 Empirical Results of the Hierarchical Analysis

Following the details of the above methodological framework, Table 1 presents the descriptive statistics for the three different portfolios. The results are categorized in terms of the bandwidth being used. The optimal bandwidth parameters is denoted by h^{opt}, while the other values of h are for two conventional rolling windows and then for the Gaussian kernel that is one of the features of the Hierarchical method. The results in Table 1 are presented in the form of the h^{opt} filtering, or kernel windows, across all the different assets in each portfolio.

Interestingly, it can be seen that the value of h selected by the cross validation procedure for the two portfolios of 25 and 55 assets are quite close to 0.50 across the assets and the standard deviation of these estimates is relatively small. 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, Kapetanios and Yates (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 average for h^{opt} for the individual stocks data, are higher than for the two portfolios and are in the 0.65 range. This may indicate the need for more weight on observations at higher lags since possibly the degree of smoothness is reduced due to a higher degree of heterogeneity of the data.

Further, the analysis of across the different method for computing the $RMSE_t$ shows that the non parametric approach, $Kernel$, provides the highest values for the standard deviations for each portfolio.

[TABLE 1 ABOUT HERE]

Figure 1 plots the selected h_t^{opt} at each point of time for the three different methods described in Table 1 and also for the different factors. Hence panel 1 of figure 1 plots the MKT beta for the three different portfolios; while panels 2 and 3 show the h_t^{opt} for the SMB and HML factors respectively. While the h_t^{opt} trend to be centred around 0.50, they also indicate some erratic, yet mean reverting paths. In general, the non parametric approach appears to be the most volatile; and is the only one that increases in the Global Financial Crisis, GFC.

[FIGURE 1 ABOUT HERE]

Table 2 provides details of the estimated beta coefficients for all three competing filtering or kernel approaches and for all three portfolios. Hence estimates are derived of $\beta_{MRKT,t}$, $\beta_{SMB,t}$ and $\beta_{HML,t}$ for each portfolio and for all three methods of filtering, or kernel weighting the data. It should be noted that the $\beta_{MRKT,t}$ for the stock portfolio is approximately 0.7, which is in accord with previous literature. The estimated market beta, $\hat{\beta}_{MRKT}$, is close to the unity for all the portfolio combinations. The standard errors provided by the $Specific$ approach are the smallest compared with the other methods and are very important for subsequent efficient estimation of risk premia.

[TABLE 2 HERE]

Figure 2 presents the factor risk loadings estimates,for the three factors along the columns and the three portfolios down the rows. Each one of the nine separate panels shows five TVP estimates of the respective beta derived from different methods. In particular, the rolling window is a green line, the kernel smoothed with constant bandwidth parameter of $h = 0.5$ is the black line, the constant bandwidth parameter from pooling average, $Pooled$, is the purple line; the time varying h set equal for all the asset, $Average$, is the blue line; and finally the time varying h optimised for each asset, $Specific$, is the red line. The last three methods all use the Gaussian kernel.

We note the TVP betas from the Gaussian kernel are similar to the five years rolling window estimates.

There is little difference in the beta estimates for the kernel approaches that uses optimal bandwidth parameter optimised on 25 or 55 portfolios.

In accord with Adrian, Crump and Moench (2015) there is greater time variation in the estimates produced with the five year rolling approach than those with constant h. Furthermore, the former estimates show higher variation also than the $Pooled$ and $Average$ approaches.

The time varying bandwidth parameter optimised for each asset, $Specific$, is the one with the highest time variation. All these methods produce estimates of the betas which are characterised by numerous sudden changes.

As expected, the beta estimates for portfolio case exhibit a lower degree of variation with respect to those that employ stock indexes, which is presumably due to noise using stock data and the loss of information induced by grouping stocks to build a portfolio. In general, the betas on the MKT and HML factors are the one that most often switch sign, while the SMB appear to be the most stable factor.

[FIGURE 2 ABOUT HERE]

Table 3 provides estimates of the market price of risk parameters γ_i with $i \in [1, K+1]$ where K is the number of factors (including also the constant term) provided by the different beta estimates. The average price of risk for each factor is in the second column, and its standard error in the penultimate column. The Newey West standard errors are also displayed in the last column.

Table 3 presents results for h^{opt} and the computed using $RMSE_t$ with kernel averaging approach.

The average prices of risk appear to be very similar across the different methods and within each dataset. The $Specific$ method is the only one that produces small difference both in terms of signs and magnitude; and has the smallest standard errors.

The sample size appears to matter and affects the significance of the price of all the factors. In particular, SMB is 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, Crump and Moench (2015) and Lettau and Livingston (2001).

Despite the fact most of the factors are not statistically different from zero on average, hence not priced, they exhibit statistically significant time variation and fluctuate a lot between positive and negative values. Such time variation of the price of risk is well documented by the set of figures 3 to 6. Figure 3 plots by columns the γs for the three different samples; with the top panel relating to 25 portfolios, the central panel to the 55 portfolios and bottom panel individual stocks.

As before, the value of $h = 0.50$ and the $Pooled$ methods describe a form of background path for the evolution of the price of risk while $Specific$ approach exhibits the highest volatility. From the analysis of these figures is clear how much of the information about the price of risk is lost using approaches such $h = 0.50$, where we do not consider the specificity of each asset.

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 to 6 instead, we produce an analysis of the significance for the different estimates across time: Figure 4 contains the results for 25 portfolios, Figure 5 for 55 portfolios while Figure 6 the ones for individual stocks. All the figures are structured as follow: in column are reported the different γs while in each row there is a different method for the computation of the $beta$ as in Section 2. For what concern the market risk premia, it show a significant positive sign at the beginning of the sample until early 2000, when it becomes significantly negative. Such change it has been captured by all the methods, despite it is more clear for the stock asset context.

An important aspect of the Hierarchical method is whether it is successful in terms of out of sample forecasting ability. Table 4 reports the $RMSE$ of one step ahead out of sample forecasts for the five Hierarchical methods and shows that the Specific method out performs the four other methods and for all three portfolios. The relative gains in forecast efficiency are greatest for the portfolios of size 25 and 55. The $Specific$ approach produces an improvements in the forecasting performance of around 6% with respect to the basic Fama and MacBeth (1973) two pass estimator based on a five year rolling window; and it improves forecast by 4.5% with respect to the Gaussian kernel approach with optimal bandwidth parameter set to 0.5. Since the optimal Hierarchical approach outperforms the other two TVP methods based on the optimal h, $Pooled$ and $Average$, it is clear that the crucial advance of our method is allowing the time variation in the bandwidth parameter h, and also to optimize it for each asset.

Also, the five year window regressions perform substantially less well than all the Gaussian kernel approaches for all the samples and for all the techniques for the choice of the bandwidth parameter; which confirms the findings of Adrian, Crump and Moench (2015).

[TABLE 4 HERE]

Table 5 presents pairwise analysis using the Diebold and Mariano (1995) test, henceforth DM. The p-$values$ of the DM test are calculated under the null hypothesis that two competing models have the same predictive accuracy; and the analysis is conducted for all the sample and methods. The results are very striking and indicate that the DM test for the Specific method are statistically significant at the 0.01 level. Indeed, the specific method appears to dominate all other methods.

[TABLE 5 HERE]

The key role of the time variation in the bandwidth parameter is also emphasized by the results of the method labeled $Average$ with respect to the $h = 0.5$ and $Pooled$ approaches. In these cases the null hypothesis of no difference in terms of performance cannot be rejected. The Fama and MacBeth (1973) five year rolling window approach is never preferred to the Gaussian kernel regression method with $h = 0.5$.

[TABLE 6 HERE]

Some further results on model comparisons are in Table 6 and concern the correlations between the beta estimates generated by each different asset. It is clear that the $Specific$ approach is the method that is least correlated with the other estimates of each beta. This is an important finding and shows that the Specific method is providing and using information not in the other methods.

The $Specific$ approach provides the smallest standard errors, allow such approach to get better estimates of risk premia and overcome the competing approaches in the forecasting exercise.

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

3.3.1 Robustness checks

A sustantial number of robustness checks were performed to check the previously reported empirical findings. Some of these results for the 55 sized stock portfolio are reported in Appendix. All the tables report the loss function of $RMSE$ in terms of deviation from the benchmark approach of $FFMcB$. Full details are available from our website. First, we investigate the sensitivity of the results to the choice of bandwidth, and we analyze the three alternatives of $[0.35; 0.9]$, $[0.05; 0.6]$ and $[0.25; 0.75]$. The results are reported in Table 7, and indicate the $Specific$ approach leads to a reduction of $RMSE$ ranging between 2.2% and 8% for the 55 sized stock Portfolio. One exception is for the $h = 0.5$ and the *Specific*: $h \in [0.35; 0.9]$ cases.

The specification of the TVP method was also investigated through using a $LASSO$ penalization in the cross sectional procedure for each value of h.

Finally, table 9 shows how changes in the size of sample size affects the results and we compare results from analyzing ten years of $T = 120$ obervations and fifteen years of $T = 180$ observations. Table 9 confirms that the *Specific* outperforms the benchmark model and is the one with the highest reduction of the $RMSE$. The use of a momentum factor was also reported but was not found to be a statistically significant factor.

3.3.2 Conclusion

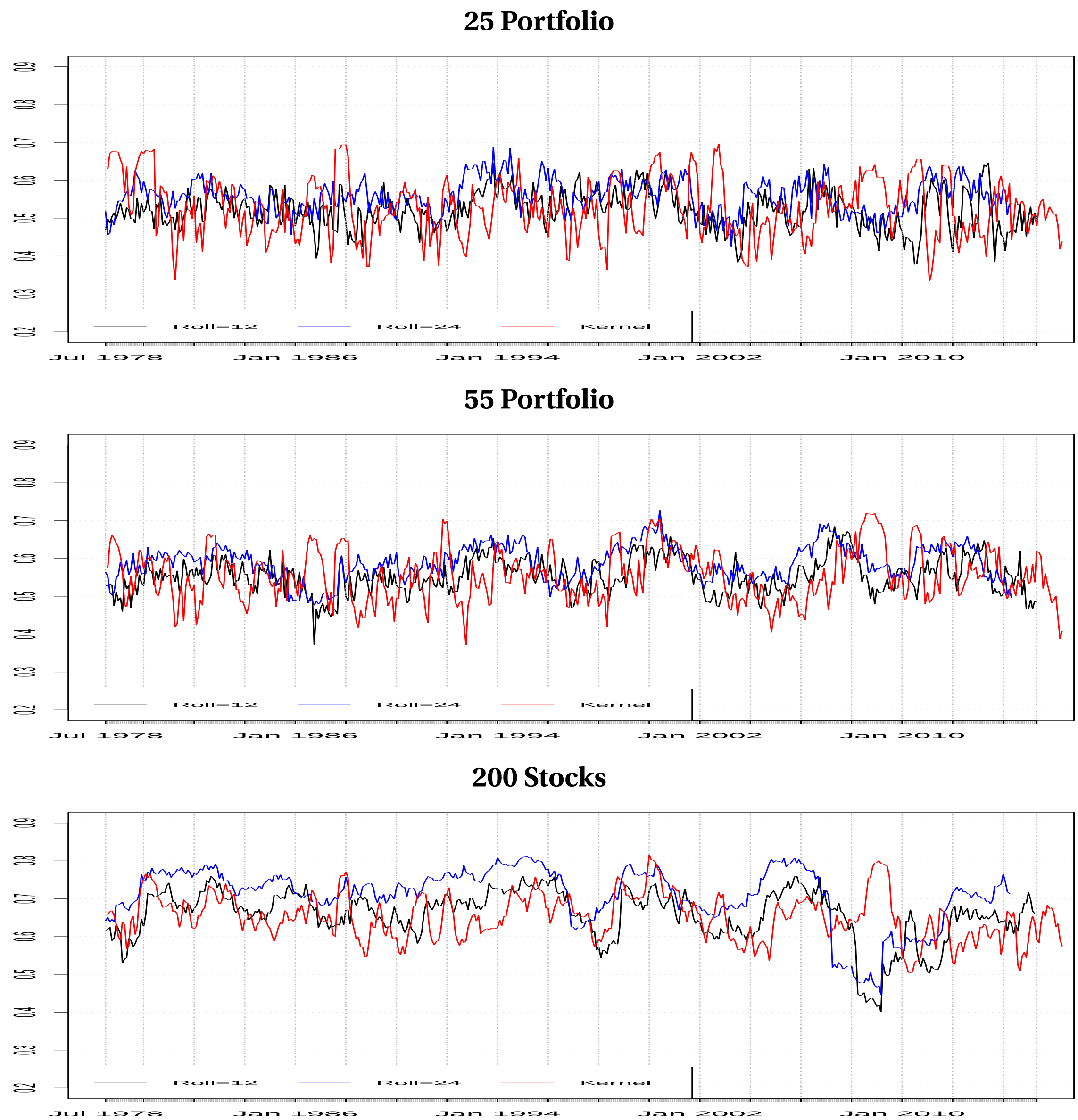
This paper has developed a new framework for the estimation of beta for a generic dynamic asset pricing model that imposes little a priori structure and generalizes the classic two step Fama and MacBeth (1973) procedure. Time variation in the beta estimates are found from a kernel weighted regression that significantly improves on conventional results in a $RMSE$ sense. We use a cross validation procedure which allows us to identify the optimal time bandwidth for each asset for each point in time. This very flexible approach without imposing an extensive a priori structure improves on the Fama and MacBeth (1973) approach. Further, it allows one to remain agnostic on the choice of data between portfolio and individual stock and to achieve significant improvements in the estimation of the risk premia proved by the reduction in the loss function.

Table 1: Descriptive Statistics for the optimal bandwidth parameter

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

Note: Descriptive statistics are given for the optimal bandwidth parameters of the three different portfolios. The $RMSE_t$ is calculated by a simple rolling window approach and also by kernel weighted averages.

Figure 1: Time Varying optimal bandwidth parameter



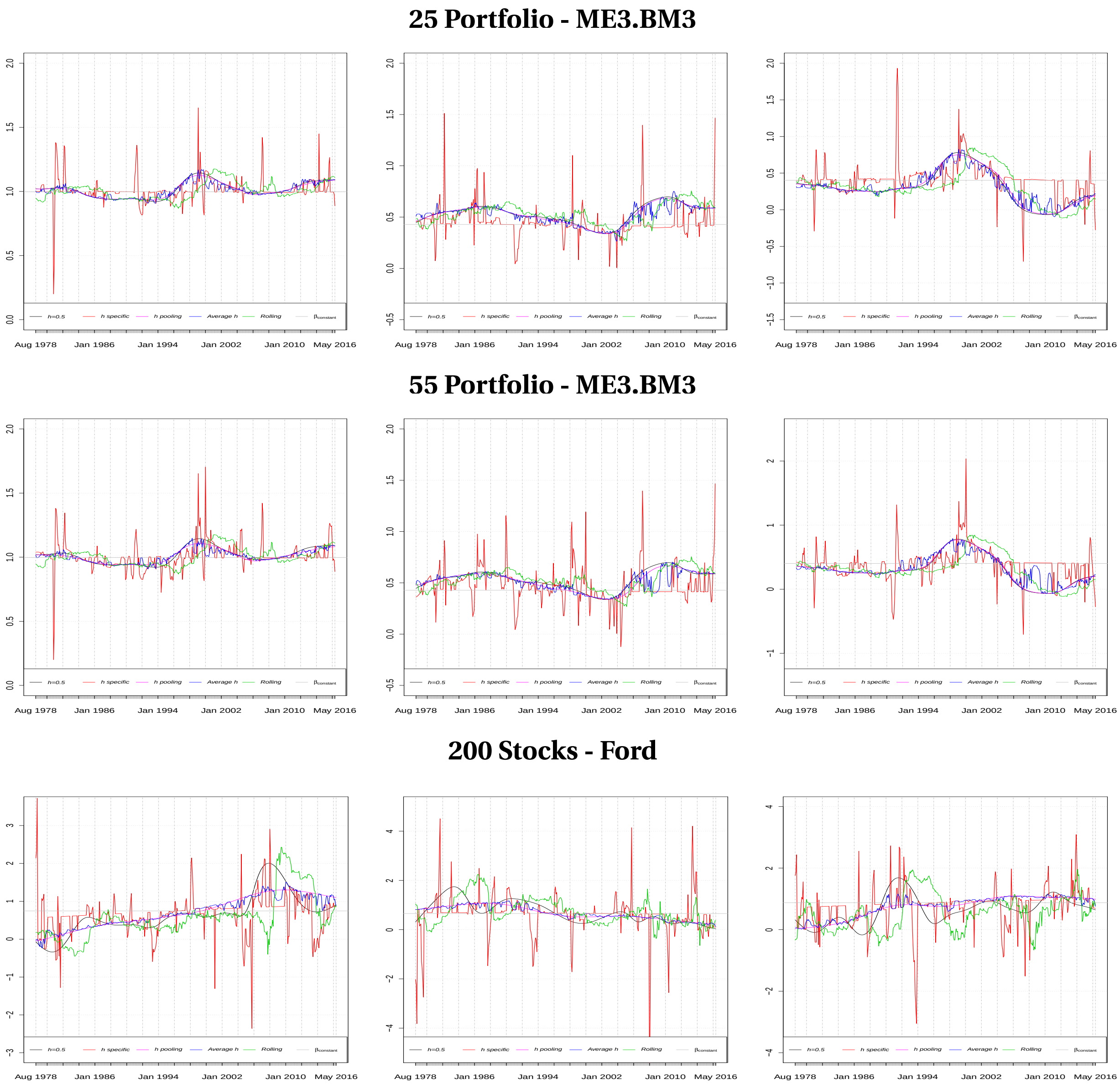
Note: The Figure reports the plots of the optimal bandwidth parameters considering different datasets and obtained using different methods for computing the $RMSE_t$: parametric and non parametric. The former method refers to the classical rolling window approach with different w such that $w \in [12; 24]$ as in equation 15, while the latter involves a kernel average method, as in equation 16. The portfolios are provided by the K.R. French's website while the individual stocks from the Center for Research in Security Prices ($CRSP$). The sample spans from August 1973 to April 216, for a total of 514 observation.

Table 2: Factor risk loading estimates

	Constant βs			*Rolling*			$H = 0.5$			H *pooling*			*Average* H			H *specific*		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
25 Portfolios																		
ME3.BM3																		
Roll w=12	0.9943	0.4265	0.4017	1.0034	0.5275	0.3101	1.0142	0.5363	0.3035	1.0132	0.5322	0.3072	1.0114	0.5281	0.3085	1.0123	0.4670	0.3830
	(0.0175)	(0.0259)	(0.0262)	(0.0413)	(0.0619)	(0.0651)	(0.0119)	(0.0282)	(0.0303)	(0.0104)	(0.0249)	(0.0265)	(0.0109)	(0.0267)	(0.0289)	(0.0048)	(0.0103)	(0.0158)
Roll w=24	0.9935	0.4211	0.4059	1.0022	0.5230	0.3165	1.0138	0.5342	0.2962	1.0099	0.5204	0.3101	1.0080	0.5200	0.3101	1.0041	0.4684	0.3792
	(0.0175)	(0.0259)	(0.0262)	(0.0384)	(0.0571)	(0.0596)	(0.0122)	(0.0291)	(0.0311)	(0.0083)	(0.0203)	(0.0208)	(0.0086)	(0.0217)	(0.0221)	(0.0053)	(0.0080)	(0.0167)
Kernel	0.9969	0.4282	0.4029	1.0052	0.5296	0.3036	1.0135	0.5304	0.3024	1.0116	0.5241	0.3074	1.0089	0.5199	0.3067	1.0126	0.4461	0.4163
	(0.0175)	(0.0259)	(0.0262)	(0.0453)	(0.0681)	(0.0721)	(0.0120)	(0.0279)	(0.0303)	(0.0099)	(0.0233)	(0.0250)	(0.0115)	(0.0268)	(0.0304)	(0.0067)	(0.0165)	(0.0176)
55 Portfolios																		
ME3.BM3																		
Roll w=12	0.9943	0.4265	0.4017	1.0034	0.5275	0.3101	1.0142	0.5363	0.3035	1.0105	0.5234	0.3145	1.0093	0.5192	0.3221	1.0078	0.4694	0.3856
	(0.0175)	(0.0259)	(0.0262)	(0.0413)	(0.0619)	(0.0651)	(0.0119)	(0.0282)	(0.0303)	(0.0084)	(0.0203)	(0.0212)	(0.0088)	(0.0214)	(0.0231)	(0.0078)	(0.0132)	(0.0204)
Roll w=24	0.9935	0.4211	0.4059	1.0022	0.5230	0.3165	1.0138	0.5342	0.2962	1.0068	0.5104	0.3203	1.0051	0.5069	0.3277	1.0071	0.4741	0.3925
	(0.0175)	(0.0259)	(0.0262)	(0.0384)	(0.0571)	(0.0596)	(0.0122)	(0.0291)	(0.0311)	(0.0069)	(0.0169)	(0.0171)	(0.0072)	(0.0179)	(0.0190)	(0.0071)	(0.0125)	(0.0219)
Kernel	0.9969	0.4282	0.4029	1.0052	0.5296	0.3036	1.0135	0.5304	0.3024	1.0091	0.5164	0.3133	1.0071	0.5105	0.3203	1.0058	0.4703	0.3864
	(0.0175)	(0.0259)	(0.0262)	(0.0453)	(0.0681)	(0.0721)	(0.0120)	(0.0279)	(0.0303)	(0.0083)	(0.0197)	(0.0209)	(0.0095)	(0.0217)	(0.0242)	(0.0107)	(0.0200)	(0.0237)
200 Stocks																		
FORD																		
Roll w=12	0.7331	0.6690	0.8870	0.6008	0.7378	0.5933	0.6404	0.7413	0.1441	0.7267	0.7116	0.7683	0.7375	0.7060	0.7757	0.6333	0.7759	0.8200
	(0.1164)	(0.1728)	(0.1744)	(0.2993)	(0.4719)	(0.4822)	(0.1148)	(0.2719)	(0.2926)	(0.0438)	(0.1041)	(0.1067)	(0.0466)	(0.1217)	(0.1176)	(0.0482)	(0.0896)	(0.0916)
Roll w=24	0.7379	0.6988	0.8895	0.6249	0.7443	0.6284	0.6435	0.7615	0.1341	0.7410	0.7150	0.7986	0.7804	0.6880	0.8053	0.6197	0.7704	0.8294
	(0.1164)	(0.1728)	(0.1744)	(0.2778)	(0.4361)	(0.4428)	(0.1199)	(0.2864)	(0.3057)	(0.0343)	(0.0742)	(0.0782)	(0.0376)	(0.0964)	(0.0921)	(0.0401)	(0.0679)	(0.0828)
Kernel	0.7449	0.6571	0.8763	0.5719	0.7405	0.5524	0.6460	0.7135	0.6401	0.7214	0.6726	0.7392	0.7105	0.6767	0.7621	0.6441	0.6066	0.7476
	(0.1164)	(0.1728)	(0.1744)	(0.3260)	(0.5146)	(0.5292)	(0.1129)	(0.2622)	(0.2848)	(0.0438)	(0.1020)	(0.1067)	(0.0471)	(0.1163)	(0.1212)	(0.0721)	(0.1281)	(0.1505)

Average estimates of factor risk loadings for the portfolio ME3.BM3 in the first 2 Panels while those for Ford stock in the bottom one, as described in text.

Figure 2: A dynamic comparison of the factor loading estimates



(a) β_{MRKT} (b) β_{SMB} (c) β_{HML}

Note: Estimates of factor risk loadings computed as described in text.

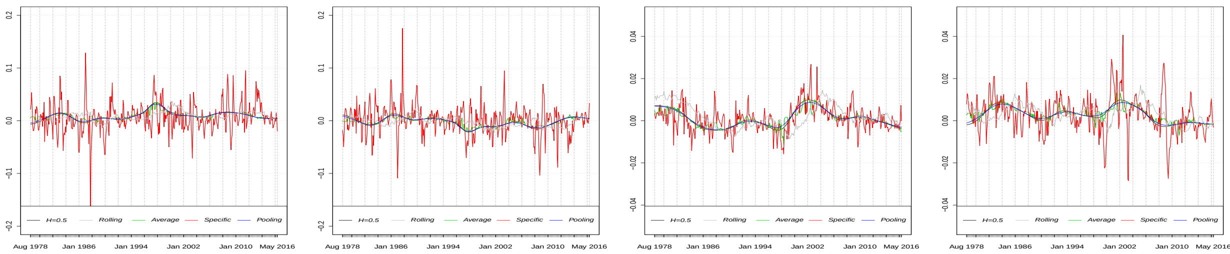
Table 3: Descriptive Statistics of risk premia estimates for different methods

	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
Pooling									
$\hat{\gamma}_0$	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
Pooling									
$\hat{\gamma}_0$	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
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
Pooling									
$\hat{\gamma}_0$	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

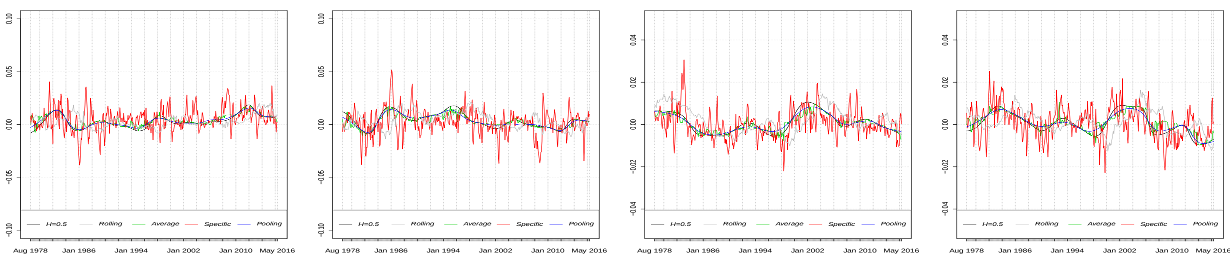
Note: Descriptive statistics of the estimated risk premia, computed for classical $FMcB$ approach and 4 different bandwidth as described in text.

Figure 3: Dynamic comparison of risk premia estimates for different approaches

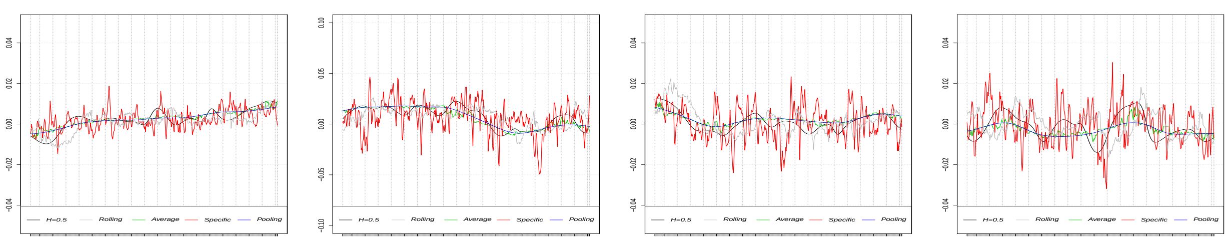
25 Portfolio



55 Portfolio



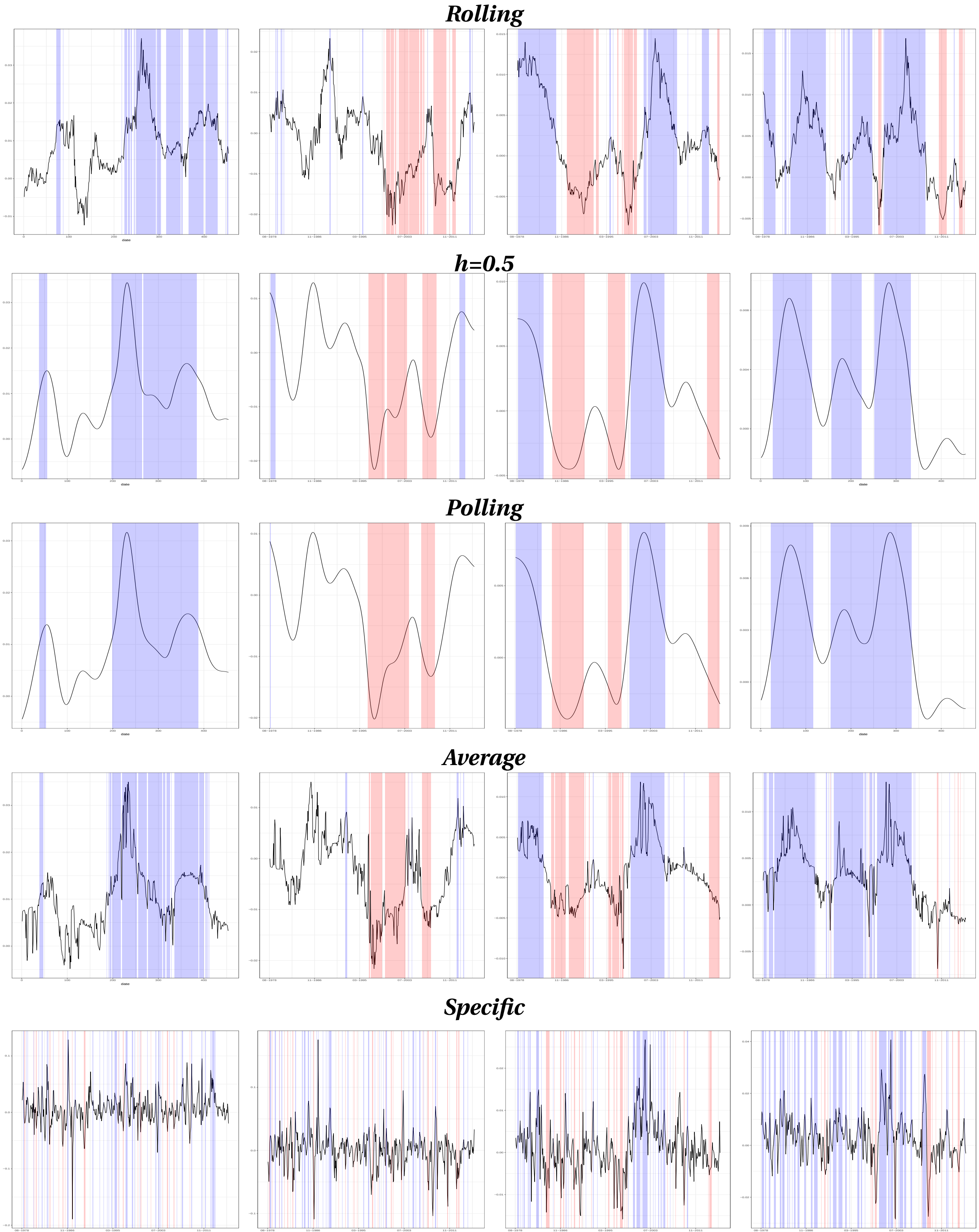
200 Stocks



(a) γ_0 (b) γ_1 (c) γ_2 (d) γ_3

Note: The figure provides the estimates of risk premia estimates computed using factor risk loadings calculated with Rolling window, with 60 estimation period ($FFMcB$ approach, green line) and kernel weighted regressions using 4 different optimal bandwidth; *h=0.5* (black line); *Pooled* a single value of h coming from the pooling average the cross asset and time (the values used correspond to the first column of Table 1; purple line); *Average*, a unique time varying bandwidth coming from the average of h across asset (blue line); *Specific*, multiple time varying bandwidths, one for each asset and time, h^{opt} (red line).

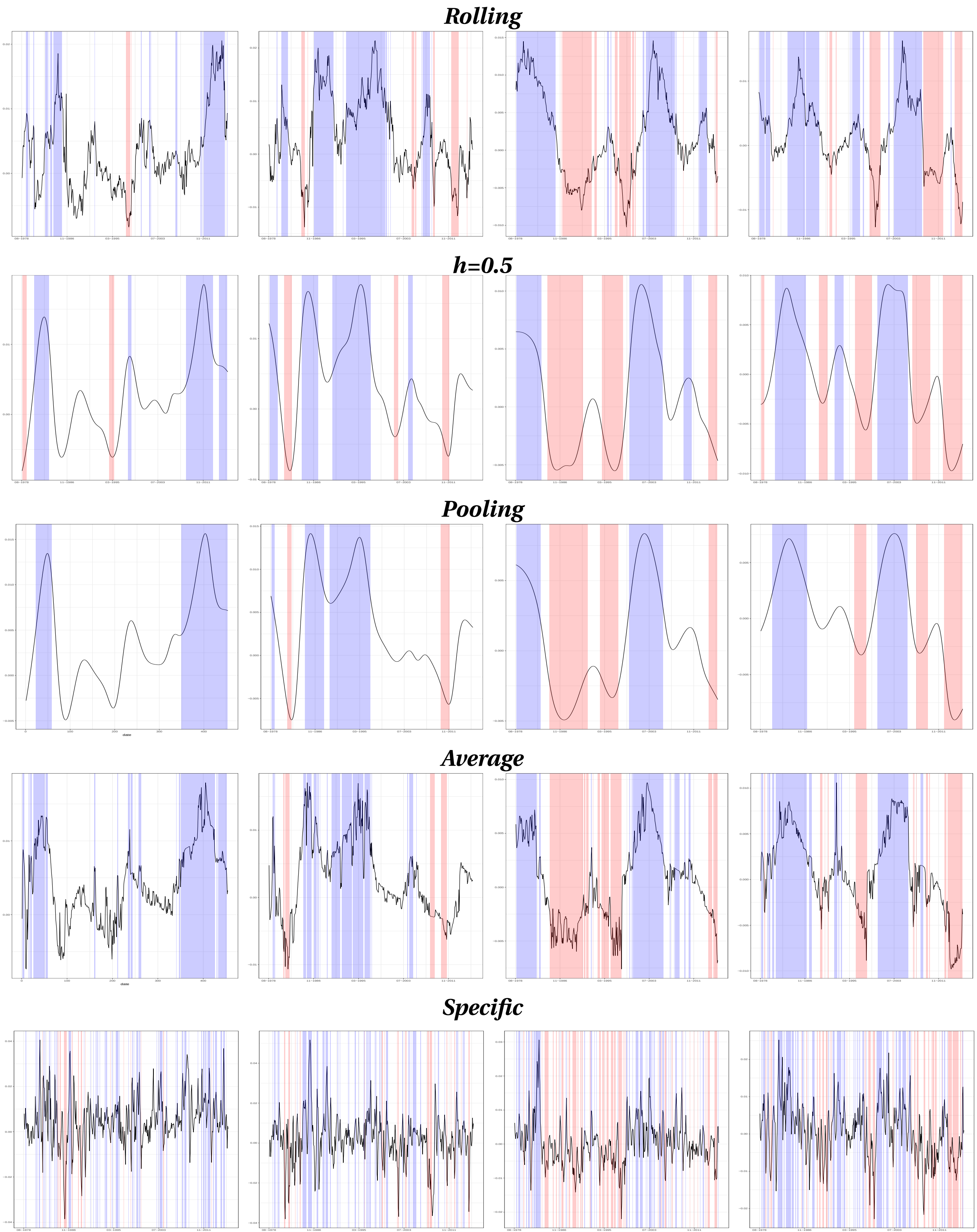
Figure 4: Comparison of $gamma$s significance of different approaches - 25 Portfolios



(a) γ_0 (b) γ_1 (c) γ_2 (d) γ_3

Note: The figure provides significance analysis of the estimates of risk premia estimates computed using factor risk loadings calculated with different approaches: Rolling window ($FFMcB$ approach), *h=0.5*; *Pooled*, *Average* and *Specific*.

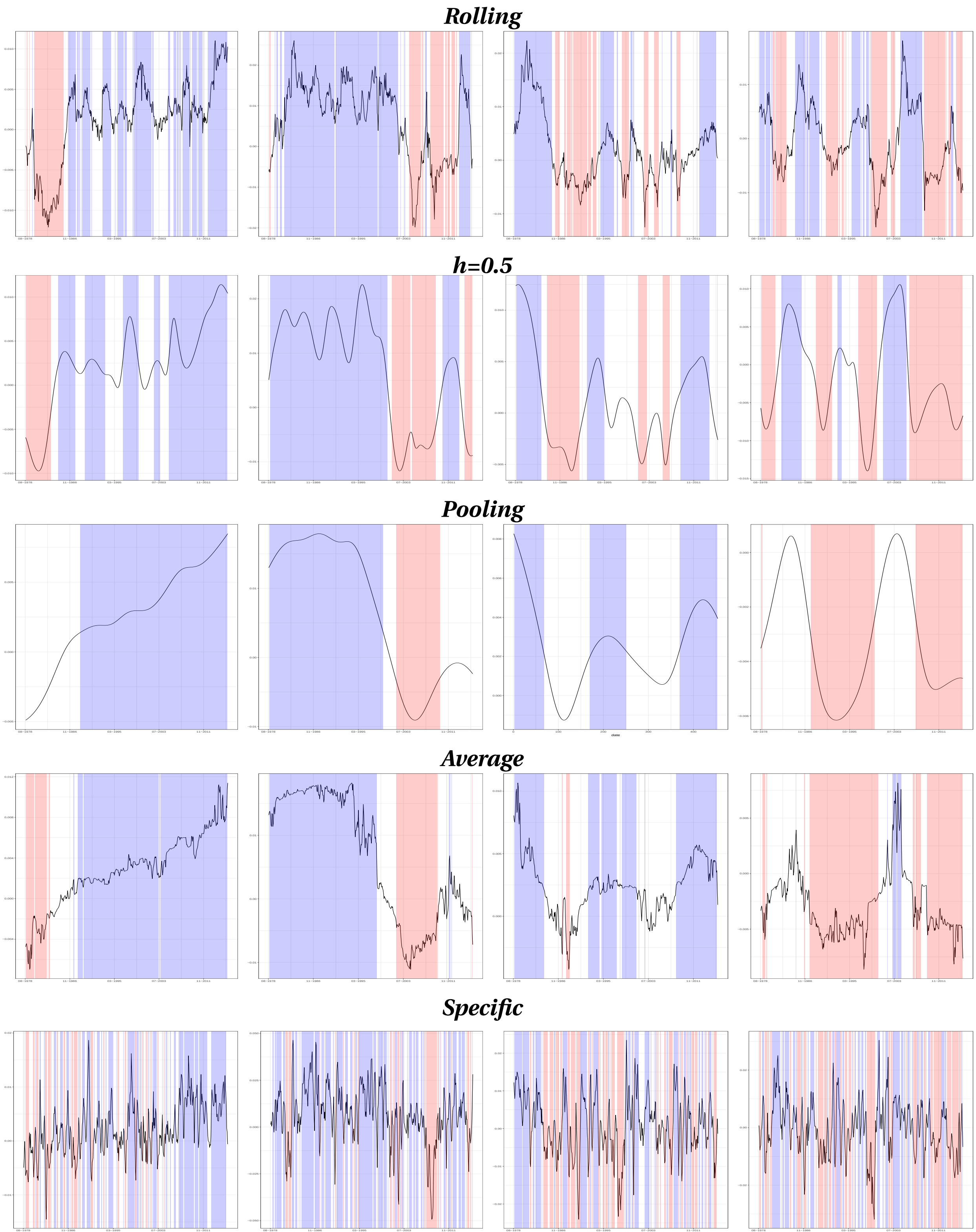
Figure 5: Comparison of $gamma$s significance of different approaches - 55 Portfolios



(a) γ_0 (b) γ_1 (c) γ_2 (d) γ_3

Note: The figure provides significance analysis of the estimates of risk premia estimates computed using factor risk loadings calculated with different approaches: Rolling window ($FFMcB$ approach), *h=0.5*; *Pooled*, *Average* and *Specific*.

Figure 6: Comparison of $gamma$s significance of different approaches - Individual stocks



(a) γ_0 (b) γ_1 (c) γ_2 (d) γ_3

Note: The figure provides significance analysis of the estimates of risk premia estimates computed using factor risk loadings calculated with different approaches: Rolling window ($FFMcB$ approach), *h=0.5*; *Pooled*, *Average* and *Specific*.

Table 4: $RMSE$ for different model and data

	Bandwidth choice: $RMSE$		
	Roll w=12	**Roll w=24**	**Kernel**
25 Portfolio			
h =0.5	-0.522	-0.483	-2.302
Pooled	-0.396	-0.168	-2.140
Average	-0.424	-0.444	-2.048
Specific	-5.800	-5.383	-6.738
55 Portfolio			
h =0.5	-0.528	-0.529	-2.235
Pooled	-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
Pooled	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 $Rolling$.

Table 5: Diebold and Mariano

	25 Portfolio				***55 Portfolio***				***200 Stocks***			
	Rolling	**h**$= 0.5$	**H Pooled**	**Average**	**Rolling**	**h**$= 0.5$	**H Pooled**	**Average**	**Rolling**	**h**$= 0.5$	**H Pooled**	**Average**
Roll w=12												
h $= 0.5$	0.0179				0.0144				0.0193			
Pooled	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
Roll w=24												
h $= 0.5$	0.0202				0.0134				0.0113			
Pooled	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
Kernel												
h $= 0.5$	0.0000				0.0001				0.0007			
Pooled	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 on the results of the Table 4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy, while the alternative is that the two method have significant different level of accuracy for the out of sample one step ahead forecasting exercise.

Table 6: Correlation matrix among betas

	25 Portfolio			*55 Portfolio*			*200 Stocks*		
	Roll w=12	**Roll w=24**	**Kernel**	**Roll w=12**	**Roll w=24**	**Kernel**	**Roll w=12**	**Roll w=24**	**Kernel**
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:

References

[1] Adrian, T., Crump, R.K. and Moench, E. , 2015. Regression based estimation of dynamic asset pricing models", Journal of Financial Economics, 118, 211-244.

[2] Andersen, T. G., Bollerslev, T., Diebold, F. X., & Wu, G. (2006). Realized beta: Persistence and predictability. In Econometric Analysis of Financial and Economic Time Series (pp. 1-39). Emerald Group Publishing Limited.

[3] Ang, A., & Kristensen, D. (2012). Testing conditional factor models. Journal of Financial Economics, 106(1), 132-156.

[4] Bali, T. G., & Engle, R. F. (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. Journal of Monetary Economics, 57(4), 377-390.

[5] Bauer, R., Cosemans, M., & Schotman, P. C. (2010). Conditional asset pricing and stock market anomalies in Europe. European Financial Management, 16(2), 165-190.

[6] Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. Journal of political Economy, 96(1), 116-131.

[7] Beach, S. L. (2011). Semivariance decomposition of country-level returns. International Review of Economics & Finance, 20(4), 607-623.

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

[9] Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. Journal of business, 383-403.

[10] Diebold, F.X. and R. Mariano (1995), "Comparing predictive accuracy", Journal of Business and Economic Statistics, 13, 253-263.

[11] Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics, 20(3), 339−350.

[12] Fama, E F & J MacBeth (1973), "Risk, return and equilibrium: empirical tests", Journal of Political Economy, 71, 607- 636.

[13] Fama, E F & K R French (1992), ”The cross section of expected stock returns”, Journal of Finance, 47, 427-465.

[14] Fama, E F & K R French (2016), ”Dissecting anomalies with a five factor model”, Review of Financial Studies, 29, 69-103.

[15] Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. Journal of Political Economy, 99(2), 385-415.

[16] Ferson, W. E., Kandel, S., & Stambaugh, R. F. (1987). Tests of Asset Pricing with TimeVarying Expected Risk Premiums and Market Betas. The Journal of Finance, 42(2), 201-220.

[17] Franzoni, F., & Adrian, T. (2005). Learning about Beta: Time-varying factor loadings, expected returns, and the Conditional CAPM (No. hal-00587579).

[18] Gagliardini, P., Ossola, E., & Scaillet, O. (2016). TimeVarying Risk Premium in Large Cross-Sectional Equity Data Sets. Econometrica, 84(3), 985-1046.

[19] Ghysels, E. (1998). On Stable Factor Structures in the Pricing of Risk: Do TimeVarying Betas Help or Hurt?. The Journal of Finance, 53(2), 549-573.

[20] Giraitis, L., Kapetanios, G., & Yates, T., 2014. Inference on stochastic time varying coefficient models. Journal of Econometrics 179, 46–65.

[21] Giraitis, L., Kapetanios, G., & Yates, T. (2018). Inference on multivariate heteroscedastic time varying random coefficient models. Journal of Time Series Analysis, 39(2), 129-149.

[22] Hansen, P. R., Lunde, A., & Voev, V. (2014). Realized beta garch: A multivariate garch model with realized measures of volatility. Journal of Applied Econometrics, 29(5), 774-799.

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

[24] Harvey, C.R. (1991), The world price of covariance risk, Journal of Finance, 46, 111-157.

[25] Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-€section of expected returns. The Journal of finance, 51(1), 3-53.

[26] Jensen, M. C., Black, F., & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests.

[27] Kan, R., Robotti, C., & Shanken, J. (2013). Pricing Model Performance and the TwoPass CrossSectional Regression Methodology. The Journal of Finance, 68(6), 2617-2649.

[28] Lettau, M., & Ludvigson, S. (2001a). Consumption, aggregate wealth, and expected stock returns. the Journal of Finance, 56(3), 815-849.

[29] Lettau, M., & Ludvigson, S. (2001b). Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy, 109(6), 1238-1287.

[30] Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. Journal of financial economics, 82(2), 289-314.

[31] Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. The journal of finance, 20(4), 587-615.

[32] Petkova, R., & Zhang, L. (2005). Is value riskier than growth?. Journal of Financial Economics, 78(1), 187-202.

[33] Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. Journal of financial economics, 4(2), 129−176.

[34] Shanken, J. (1985). Multi factor CAPM or Equilibrium?: A Reply. The Journal of Finance, 40(4), 1189-1196.

[35] Shanken, J. (1992). The current state of the arbitrage pricing theory. The Journal of Finance, 47(4), 1569-1574.

[36] Shanken, J., & Zhou, G. (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. Journal of Financial Economics, 84(1), 40-86.

[37] Sharpe, W. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

[38] Zolotoy, L. (2011). Earnings News and Market Risk: Is the Magnitude of the Post earnings Announcement Drift Underestimated?. Journal of Financial Research, 34(3), 523-535.

Appendix A: Kernel weighted regression

In order to give the above RV models the maximum opportunity to represent the realized volatility process we implement a non parametric approach for computing the time variation in the regression coefficients that requires minimal theoretical restriction on the functional form. We extend the work of Giraitis, Kapetanios and Yates (2014) work on autoregressive processes to that of a kernel smoothing regression. Giraitis, Kapetanios and Yates (2014) consider the $AR(1)$ process

$$y_t = \phi_{t-1} y_{t-1} + u_t \tag{16}$$

where u_t is $iid\ (0, \sigma_u^2)$ and there is some initialization of the process y_0 and ϕ_{t-1} is a random coefficient and $u_t|\Omega_{t-1} = 0$ and $\phi_t|\Omega_{t-1} = \phi$. The stability of the model depends on the TVP nature of the AR parameters satisfying various smoothness classes. 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

$$\phi_t = \phi \frac{a_t}{\max_{0 \leq k \leq t} |a_k|} ....t > 0 \tag{17}$$

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. Also, $\phi \in (0, 1)$ and ϕ_{t-1} is then 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

$$\phi_{t,i} = \phi \frac{a_{t,i}}{\max_{0 \leq k \leq t} |a_{k,i}|} ....t > 1 \tag{18}$$

where $0 < \phi < 1$ and each $a_{t,i}$ are independent versions 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, Kapetanios and Yates (2014) show that the coefficient process $\{\phi_t; t = 1, ..., T\}$ converges in distribution as T increases to the limit

$$\{\phi_t; 0 \leq \tau \leq 1\} \to_D \{\phi \tilde{W}_\tau; 0 \leq \tau \leq 1\}.$$

The approach for estimating the time varying parameter, ϕ_t is to use the moving window estimator for the $AR(1)$ RC model

$$\widehat{\phi}_t = \frac{\sum_{t=1}^{H} K\left(\frac{t-k}{H}\right) y_t y_{t-1}}{\sum_{t=1}^{H} K\left(\frac{t-k}{H}\right) y_{t-1}^2} \tag{19}$$

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. On generalizing a generic regression which can be expressed as

$$y_t = x_t' \beta_t + u_t \tag{20}$$

with $\beta_t = (\beta_{1,t}, \beta_{2,t}, ...\beta_{k,t})$ and 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. In general the kernel weighted regression estimator for $\beta_{j,t}$ is

$$\hat{\beta}_{j,t} = \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \left(\sum_{j=1} w_{jt} x_j y_j\right). \tag{21}$$

where $w_{jt} = K\left(\frac{t-k}{H}\right)$. From Giraitis, Kapetanios and Yates (2014) it follows that

$$H^{1/2}(1-\hat{\beta}_{j,t}^2)^{-1/2}(\hat{\beta}_{j,t}-\beta_{j,t}) \sim N(0,1) \tag{22}$$

The authors prove that if the bandwidth is $o_p\left(T^h\right)$ with $h = 1/2$, and given homoskedasticity of the error process, then

$$Var\left(\hat{\beta}_t\right) = \hat{\sigma}_u^2 \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \sum_{j=1} w_{jt}^2 x_j x_j' \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \tag{23}$$

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

$$Var\left(\hat{\beta}_{j,t}\right) = \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \left(\sum_{j=1} w_{jt}^2 x_j x_j' \hat{u}_t^2\right) \left(\sum_{j=1} w_{jt} x_j x_j'\right)^{-1} \tag{24}$$

which can be used for inference. One appealing characteristic of this approach is that they nest rolling window estimates of the regression betas and are 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 bandwidth and for a given kernel function, $K\left(\frac{t-k}{H}\right)$, the bandwidth, H, represents the degree of smoothness of the estimates. Giraitis, Kapetanios and Yates (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 provide valid standard error.

Another appealing characteristics of such approach is that they nest, as a special case, rolling window estimates of betas (for example, Chen, Roll, and Ross, 1986; Ferson and Harvey, 1991; Petkova and Zhang, 2005; among many others). Rolling beta estimates are equivalent to kernel smoothing estimators 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 approach a key role is played by the decision about 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 will be under smoothed, with high variability, otherwise if the value of H is big, the resulting estimators will be over smooth and farther 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 shows obtain desirable properties such as consistency and asymptotic normality and in addition provide valid standard errors.

Appendix B: Robustness checks

Table 7: RMSEs for different bandwidth parameters intervals as deviation from $FMcB$

25 Port	**Roll w=12**	**Roll w=24**	**Kernel**	**55 Port**	**Roll w=12**	**Roll w=24**	**Kernel**
$h \in [0.05; 0.9]$							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.004	-0.002	-0.021	**Pooling**	-0.002	-0.002	-0.019
Average	-0.004	-0.004	-0.020	**Average**	-0.001	0.000	-0.019
Specific	-0.058	-0.054	-0.067	**Specific**	-0.053	-0.044	-0.063
$h \in [0.05; 0.6]$							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.021	-0.016	-0.038	**Pooling**	-0.019	-0.010	-0.036
Average	-0.020	-0.019	-0.036	**Average**	-0.024	-0.018	-0.034
Specific	-0.070	-0.066	-0.083	**Specific**	-0.074	-0.065	-0.083
$h \in [0.35; 0.9]$							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	0.000	0.001	-0.018	**Pooling**	0.000	0.000	-0.017
Average	0.000	0.001	-0.018	**Average**	0.001	0.001	-0.017
Specific	-0.009	-0.008	-0.023	**Specific**	-0.007	-0.005	-0.022
$h \in [0.25; 0.75]$							
h=0.5	-0.005	-0.005	-0.011	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.004	-0.002	-0.008	**Pooling**	-0.003	-0.002	-0.007
Average	-0.003	-0.003	-0.007	**Average**	-0.002	-0.001	-0.007
Specific	-0.020	-0.019	-0.020	**Specific**	-0.019	-0.015	-0.018

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 8: RMSEs for different penalization parameters as deviation from $FMcB$ in LASSO context

25 Port	Roll w=12	Roll w=24	Kernel	55 Port	Roll w=12	Roll w=24	Kernel
NO LASSO							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.004	-0.002	-0.021	**Pooling**	-0.002	-0.002	-0.019
Average	-0.004	-0.004	-0.020	**Average**	-0.001	0.000	-0.019
Specific	-0.058	-0.054	-0.067	**Specific**	-0.053	-0.044	-0.063
LASSO 0.0001							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.006	-0.004	-0.025	**Pooling**	-0.003	-0.001	-0.020
Average	-0.013	-0.014	-0.050	**Average**	-0.002	0.000	-0.018
Specific	-0.068	-0.063	-0.102	**Specific**	-0.052	-0.046	-0.061
LASSO 0.00005							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.005	-0.002	-0.022	**Pooling**	-0.003	-0.001	-0.020
Average	-0.005	-0.004	-0.022	**Average**	-0.001	0.000	-0.018
Specific	-0.064	-0.057	-0.067	**Specific**	-0.052	-0.046	-0.060

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, λ: 0.0001 and 0.00005. The results are expressed as a deviation from the $RMSE$ produced by the benchmark model, $FMcB$.

Table 9: RMSEs for different sample size of the trading period

25 Port	Roll w=12	Roll w=24	Kernel	55 Port	Roll w=12	Roll w=24	Kernel
$T = 60$							
h=0.5	-0.005	-0.005	-0.023	**h=0.5**	-0.005	-0.005	-0.022
Pooling	-0.004	-0.002	-0.021	**Pooling**	-0.002	-0.002	-0.019
Average	-0.004	-0.004	-0.020	**Average**	-0.001	0.000	-0.019
Specific	-0.058	-0.054	-0.067	**Specific**	-0.053	-0.044	-0.063
$T = 120$							
h=0.5	0.005	0.012	-0.017	**h=0.5**	-0.015	-0.010	-0.036
Pooling	0.006	0.015	-0.015	**Pooling**	-0.011	-0.004	-0.033
Average	0.008	0.015	-0.015	**Average**	-0.010	-0.002	-0.031
Specific	-0.043	-0.030	-0.063	**Specific**	-0.052	-0.038	-0.075
$T = 180$							
h=0.5	0.016	0.030	-0.005	**h=0.5**	0.026	0.043	0.004
Pooling	0.017	0.035	-0.004	**Pooling**	0.031	0.051	0.008
Average	0.020	0.036	-0.002	**Average**	0.036	0.055	0.014
Specific	-0.035	-0.013	-0.058	**Specific**	-0.010	0.018	-0.038

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

School of Economics and Finance

Queen Mary
University of London

This working paper has been produced by the School of Economics and Finance at Queen Mary University of London

School of Economics and Finance
Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/