STOCKHOLM SCHOOL OF ECONOMICS
Department of Finance
MSc Thesis in Finance
Academic Year 2017-2018

Nonparametric Asset Pricing with Conditioning Information

Christoffer Sjöström*
Dominik Schmitz†

May 14, 2018

Abstract

This study sets out to be the very first in introducing the notion of a nonlinear pricing kernel in conditional asset pricing for the Swedish equity market. By implementing a flexible nonparametric methodology, we are able to conduct tests that are completely free from functional form specifications of time-varying betas, risk premia and the stochastic discount factor. The test provides a refined view about the empirical performance of asset pricing models with conditioning information. In this paper, we investigate the most general versions of the conditional Fama and French (1993) model and the CAPM ever examined in Sweden. We demonstrate that both these models significantly outperform their static counterparts, that they price stocks very well and that nonlinearity in the first and second moment of the market return is important.

Keywords: Conditional Asset Pricing, Nonparametric SDF, Nonlinear Pricing Kernel, Stochastic Discount Factor, Time-varying Betas, Dynamic Asset Pricing, Fama-French Three Factor Model, CAPM, Conditioning Information, Value, Size
JEL Classification codes: C14, C52, G12

Supervisor: Associate Professor Michael Halling
Examiner: Professor Mariassunta Giannetti
Acknowledgements: We would like to express our gratitude to our supervisor and teacher, Associate Professor Michael Halling, for his valuable support during the thesis. Access to data on the Fama and French (1993) three factors for Sweden from Professor Paolo Sodini and George Cristea at SHoF is greatfully acknowledged.

*22997@student.hhs.se
†40979@student.hhs.se

Contents

1 Introduction

The asset pricing field is in a state of turmoil with hungry practitioners constantly striving to find the optimal framework. In particular, the unconditional CAPM has for long been the workhorse of empirical finance. However, lots of studies have unveiled serious empirical deficiencies against this constant beta model that are so forceful that it has been widely argued that a replacement model is needed (see for instance Fama and French (1996a)). With the field under such strenuous attack, the common response has been either to improve already established models, propose alternative pricing frameworks or introduce multi-factor extensions. The Fama and French (1993) three factor model (hereafter FF-3F) represents a measure taken with regard to the latter category and is undisputedly regarded the most prominent asset pricing model to date. Despite numerous positive documentations, several empirical shortcomings have been shown to persist. For instance, the inability of the FF-3F model in its unconditional form to correctly price the momentum anomaly is widely documented and an unresolved issue in many cases (FF (1996b) and (2012)). Still, several authors uncover evidence that suggest that the static FF-3F fails miserably to capture the dynamic behaviour of asset returns.

Given the strong flood of negative empirical documentations on unconditional model performance, numerous studies have been targeted at exploring the roots and causes thereof. In particular, several studies argue that the ignorance of time-variability of betas and their associated asset pricing implications serve as important sources in explaining poor model performance. As such, conditional asset pricing models that allow for improved empirical efficiency by accounting for these crucially important aspects have become focal points of investigation. In line with this, Dybvig and Ingersoll (1982) among others have shown that the conditional CAPM can hold perfectly regardless of its unconditional counterpart exhibiting significant mispricing.

When modelling risk dynamics by incorporating conditioning information in the stochastic discount factor (hereafter SDF), the econometrician faces several technical challenges. To reduce the inherent complexity pertaining to the modeling procedure, it has become common practice to impose linearity restrictions whereby it is assumed that conditional expected returns on an asset is a linear function of one or more conditional betas. While the convenient linearity assumption still makes dynamic models intuitively appealing, its imposition might come at a significant cost as there is no theoretical guidance on how betas and risk premia comove with state variables that represent conditioning information. In other words, the empirical results can be shaped by inappropriate modelling assumptions imposed by the econometrician, which can lead to devastating empirical consequences. This can be highly important if betas are not well approximated by linear relationships and it is difficult to capture nonlinear beta functions. In line with this, Ghyssels (1998) among others demonstrate that several popular dynamic beta models are so severely misspecified that they are outperformed by their static counterparts because of the linearity assumption.

In this paper, we resurrect the hope for asset pricing with conditioning information by being the very first, to the best of our knowledge, to introduce the notion of a nonlinear SDF (or pricing kernel) in the Swedish equity market. Our study builds on the flexible nonparametric testing methodology of Wang (2002, 2003) that completely avoids functional form (mis)specifications of risk dynamics and the stochastic discount factor. This is achieved by deriving a nonparametric presentation of restrictions on the SDF implied by a conditional linear factor pricing model. Thanks to Wang (2002, 2003), we provide a novel view on the empirical performance of the conditional CAPM and FF-3F models for Sweden. In particular, the conditional CAPM and the fully non-parametric FF-3F model that we consider are the most general versions considered to date for the Swedish equity market.

By implementing sophisticated econometric techniques, we extensively show that convenient linear approximations of beta functions indeed are empirically inadequate. Rather, our findings reveal that nonlinearity in the first two conditional moments of the market return is highly present and that a nonparametric SDF that captures this feature is a solution that can bring significant value.

Our empirical results on the model performance of the conditional asset pricing models are striking. For the nonparametric version of the conditional CAPM, we find that it significantly outperforms the unconditional CAPM and that the magnitude of improvement is large. More specifically, the standard deviations of Jensen's alphas are lower across each and every of the 25 size-BE/ME test portfolios, and this largely holds for average pricing errors as well. While the conditional CAPM is statistically rejected at the 10% significance level, it seems to price risky assets very well: The intercepts, or average pricing errors, are economically small and statistically insignificant.

For the conditional FF-3F model, we find that the nonparametric version generates even more favourable empirical conclusions. It is statistically as well as economically significant and captures the most prominent pricing error deviations from the conditional CAPM. By jointly adding the size and book-to-market factors, pricing error measures are reduced with more than 80%! In sum, the model performs very well and conditional alphas are small.

Ghyssels (1998) and Wang (2003) argue that once dynamics of beta risk are correctly specified, one can expect time-varying beta models to beat static models. Consistent with this argument, we find that nonparametric versions of conditional asset pricing models outperform their constant beta counterparts significantly. These findings differ from several empirical results produced by conditional asset pricing models with functional form restrictions on betas (see for instance Hodrick and Zhang (2001), Ferson et al. (2009) as well as Petkova and Zhang (2005)). This contrast indeed reinforces the finding that nonlinearity in the first and second moment of the market return is an important consideration in the specification of dynamic asset pricing models. By being the first out-of-sample country check, we show that a nonparametric SDF that successfully incorporates the dynamic behaviour of returns generates significant payoffs in the study of asset pricing models with conditioning information in the Swedish equity market.

The remainder of this paper is structured as follows: Section 2 reviews previous research findings on the topic and identifies how our study uniquely contributes to the literature. Section 3 extensively describes the data employed to estimate and evaluate conditional asset pricing models. Section 4 introduces the nonparametric methodology to test for conditional mean-variance efficiency. Section 5 presents and thoroughly analyses empirical results. Finally, Section 6 concludes the paper and briefly aggregates the most important contributions as well as their implications for dynamic asset pricing in Sweden.

2 Previous Literature

This section reviews prior literature on asset pricing with conditioning information and identifies how our study uniquely contributes to previous findings. First, the topic is motivated in light of empirical manifestations providing harsh critique against static models. Then, several common estimation methods and their implications for the dynamics of conditional asset pricing are described. Finally, documentations on the pricing power ability pertaining to the conditional CAPM and the FF-3F are presented.

2.1 Motivation for Conditional Asset Pricing

The failure of the capital asset pricing model (CAPM) to explain not only the cross-sectional, but also the time-series of average equity returns is well demonstrated in the empirical asset pricing literature. The evidence against this static, i.e. constant beta model are so strong that practitioners argue that a replacement model is needed. FF(1992, 1996a), Banz (1981) and Basu (1983) only represent a few of the early authors who document that the univariate relationship between the single-factor and average returns is weak. More recently, FF (2004, 2006, 2012) verify and revisit trading strategies that are anomalous to the static CAPM, showing that the empirical results remain robust to a wide range of re-specifications and that they are not sample (i.e. country) specific.

A natural response to the criticism has been to propose either extensions or new models that lead to superior empirical performance (For example; FF (1993), Jagganathan and Wang (1996); Lettau and Ludvigsson (2001); Chen and Zhang (2011); Savov (2011); Koijen et al. (2017)). With the FF-3F model counting towards the most powerful asset pricing frameworks to date, the model cast in its unconditional form does not explain the momentum effect (FF(1996b), (2012)). Despite its history of having successfully addressed several shortcomings, the empirical record of the FF-3F model is still subject to a lot of controversy even when challenged against the static CAPM. For instance, Bartholdy and Peare (2005) bring into question the empirical adequacy of either models.

In light of the severe critique aimed at unconditional models and the related asset pricing anomalies identified against them, numerous studies have been aimed at investigating potential reasons for the poor performance. A prominent explanation is that time-variability of betas, risk premia and their asset pricing implications have been largely ignored. Such an explanation is based on several time-series predictability literature papers having evidenced the dynamic nature of risk premia (See Campbell and Shiller (1988); Ferson and Harvey (1991); Ferson and Korajczyk (1995); Li and Yang (2011); Engle (2016))

With the field under such strenuous attack, another reaction to the empirical deficiencies of constant models has given rise to a literature series directed at improving already established models by incorporating time-variability of risk premia. As a result, conditional versions of classical asset pricing models that allow for improved empirical efficiency by capturing beta-dynamics have become focal points of investigation. Dybvig and Ross (1985) as well as Hansen and Richard (1987) all show that conditional asset pricing models that account for these desirable features can hold perfectly, regardless of the unconditional CAPM exhibiting significant mispricing. As such, asset pricing models incorporating conditioning information have become popular objects at the forefront of empirical asset pricing (Cochrane, 2009). Given the suggested pervasive influence of these models, they remain interesting to study.

2.2 The SDF and Conditioning Information

Given the first-order importance of appropriately modelling betas once dynamic specifications have been introduced, a range of different statistical implementations are available for the econometrician. Our review mainly concerns, but is not limited to the stochastic discount factor (SDF) approach[1].

With the SDF representation of whichever asset pricing model, time-variability of betas can successfully be accounted for by incorporating conditioning information. This is achieved by expressing the parameters of the SDF as functions of lagged state variables (Cochrane (2009) extensively dis-

[1]Other methodologies have been suggested: For instance, time variation in conditional covariances has been modelled with M-GARCH (See Bollerslev et al. (1988); Ng (1991)) A disadvantage of this popular procedure is the strong assumption about the functional form of second moments. Furthermore, ARCH processes do not necessarily aggregate: In particular, if asset returns follow a specific ARCH process, it does not necessarily follow that a portfolio of assets follow that ARCH process (Harvey, 1989).

cusses this in his book). The most widely applied methodology makes use of a result that shows that the parameters of the SDF can be expressed as affine *linear* functions of instruments[2]. By imposing the linearity restriction on the functional form of conditional betas and the associated conditional moments, the econometrician can circumvent several technical modelling challenges. When the SDF is a fully specified parametric function of data, the model is usually estimated and tested by the generalized method of moments (GMM) of either Hansen and Jaggannathan (1997) or Hansen (1982). The former is usually implemented when the purpose is to run a horse race among different linear models, using the second moment matrix of returns as the weighting matrix[3]. Some of the first authors that pursue tests of conditional mean-variance efficiency in the linearized-GMM manner are Cochrane (1996), Harvey (1989) and He et al. (1996). More recently, Ferson et al. (2009) and Smith (2007) pursue such an application, where the latter relates to the SDF indirectly. Similarly, it has been popular to make use of conditional linear market regression models. Petkova and Zhang (2005), Ho and Hung (2009), Hyunh and Smith (2014) as well as Lewellen and Nagel (2006) count to this group.

While convenient linearity assumptions make dynamic models intuitively appealing, it has been demonstrated that it might come at a significant cost. These models imply that the conditional expected portfolio return is linearly related to conditional betas that measure the portfolio's sensitivity to systematic risk. However, there is no theoretical guidance on how betas and risk premia vary with state variables that reflect conditioning information. By imposing linear restrictions on the parameters of the SDF, empirical result can be shaped by the modelling assumptions imposed by the econometrician. An author strongly highlighting this issue is Ghysels (1998). The author demonstrates that several popular dynamic beta models are so severely misspecified to the extent that they are outperformed by their static counterparts[4].

[2]This approach is sometimes referred to as multiplicative, as it amounts to forming dynamic models whose factors and returns are multiplied/scaled by the lagged instruments.

[3]For strong critique against the Hansen and Jaggannathan (HJ) version and the associated testing procedure, see Ferson and Siegel (2003), who detect striking finite-sample biases in the HJ bounds.

[4]The inherent problem pertaining to the linearity assumption is also acknowledged by for instance He et al. (1996) as well as Nagel and Singleton (2001). In contrast, Harvey (2001) provides a rationale for linearity in several conditional asset pricing applications.

Given the debatable empirical adequacy of assuming linear relationships between conditional expected returns and betas, Wang (2002, 2003) proposes a flexible nonparametric methodology constructed within the SDF framework that is completely free from functional form misspecification about beta dynamics, risk premia and the SDF. With Ghyssels well-motivated critique (1998) as the source of inspiration for a nonparametric SDF, the author resurrects the hope for asset pricing models with conditioning information by providing a novel view about their empirical performance.

Other authors that may have been influenced by Wang (2003) are Li and Yang (2011) as well as Ang and Kristensen (2012). On the one hand, they are related to Wang (2003) by the application of nonparametric tests of conditional mean-variance efficiency. On the other hand, these authors differ significantly in several aspects: They use local information to approximate betas and conditional alphas rather than using conditioning information, whereby it is assumed that information is stable within short windows. Li and Yang (2011) consider high frequency tests designed to optimally select the right amount of local data in their GLS regression windows, whereas Ang and Kristensen (2012) consider kernel-weighted OLS regressions. The latter provide that their conditional alpha estimators in general are inconsistent.

The concept of a nonparametric SDF has previously been suggested: Bansal et al. (1993) as well as Bansal and Viswanathan (1993) are perhaps the pioneers to advocate the usage of a flexible SDF in empirical asset pricing[5]. Unlike Wang (2003), these authors focus on nonlinear APT models and propose a series (polynomial) expansion approach combined with the standard GMM for estimating and testing. The drawback with this methodology is that it is problematic to derive a distribution theory and an effective assessment of finite sample performance. However, nor are fully nonparametric econometric methodologies in general free from drawbacks. Yet, Wang (2003) successfully circumvents many of the disadvantages aimed at such econometric methods by carefully designing the asset pricing tests accordingly (further discussed in the Methodology section).

By building upon the empirical success of Wang (2002, 2003), our paper is the very first, to the best of our knowledge, to introduce the notion of a

[5]See Lustig and Van Nieuwerburgh (2005) for a compromise; they specify a semiparametric SDF in conditional asset pricing.

nonparametric SDF in asset pricing with conditioning information in Sweden. The restrictions implied by this nonlinear pricing kernel give rise to the most general versions of the CAPM and the FF-3F models ever considered in the Swedish equity market.

2.3 Empirical Manifestations

Although asset pricing with conditioning information provides the foundation for the reestablishment of the backbone CAPM, the literature has not reached full consensus regarding whether the conditional model explains average returns. Strong modelling assumptions and functional form restrictions on the first and second moment of returns do not necessarily ensure empirical success all the way: When allowing the parameters of the SDF to fluctuate linearly with the business cycle, Hodrick and Zhang (2001) lend support in favour of some conditional CAPM specifications. However, some of the models fail to pass Ghyssel's (1998) parameter stability test, strongly suggesting the occurrence of potential misspecification of betas. Similarly, Harvey (1989) and He et al. (1996) demonstrate that time-varying conditional covariances scaled by instrumental variables in the CAPM do not suffice in explaining the dynamic behaviour of returns. These documentations are further reinforced by Petkova and Zhang (2005), who further investigate the conditional CAPM and conclude that the magnitude of the value premium remains mostly significant even after controlling for time-varying risk. The empirical evidence of Ang and Kristensen (2012) as well as Li and Yang (2011) point in the same direction: Their optimal window regression methods based on OLS and GLS nonparametric estimators indicate that neither the conditional CAPM, nor the conditional FF-3F model is able to correctly price the momentum and value premium. This failure could be due to the assumption that short windows provide local information that is sufficient for accurate pricing power.

On the contrary, Lettau and Ludvigsson (2001) consider the log-consumption to wealth-ratio as a conditioning variable, whereby they demonstrate clearly improved empirical performance of the conditional consumption-CAPM. Similarly, Santos and Veronesi (2006) shows that the fraction of total income funded by total labor income improves the CAPM. Other authors that produce positive conclusions on the CAPM with conditioning information are

Bali and Engle (2014), Ho and Hung (2009) as well as Duffee (2005)[6].

Turning to the FF-3F model, it is possible to highlight that albeit its largely undisputable superiority relative to the CAPM, authors still uncover empirical evidence at odds with the framework once the model is set in its conditional form. Harvey and Ferson (1999) demonstrate that even the conditional version of the FF-3F does not offer an improvement large enough to justify significant pricing error patterns[7]. Similarly, Ferson et al. (2009) produce negative empirical conclusions on the FF-3F performance, although the authors make the claim that their test uses conditioning information efficiently. Nonetheless, these authors reject conditional mean-variance efficiency of both all static or time-varying combinations of the FF-3F model. However, the authors do not abandon linear restrictions on the SDF, which might have an important bearing on their empirical conclusions. For documentations that indicate improved efficiency on the FF-3F, see Hong and Hu (2009) as well as Huynh and Smith (2014).

Moving to the nonparametric tests of conditional mean-variance efficiency used in this study, Wang (2002, 2003) is able to demonstrate significant improvements relative to previous research. He attributes the empirical success to his flexible approach which relaxes linearity assumptions and functional restrictions on the parameters of the SDF. For the conditional CAPM, the author finds that the nonparametric version offers substantially improved performance compared to its unconditional counterpart. Although the model is statistically rejected, the author demonstrates that the patterns of pricing errors are interesting: They exhibit a prominent size pattern in volatility but not in time-series averages. At the same time, the pricing errors have a strong book-to-market pattern in time-series averages but not in volatility. Put differently, Wang (2003) suggests that the size and book-to-market effects tend to occur across different channels. Furthermore, the author evidences that

[6]Lewellen and Nagel (2006) assert that several of the cross-sectional specifications don't provide a full test of the CAPM, arguing that the results should be subject to further reinvestigation. However, Boguth et al. (2010) argue that the regression method of Lewellen and Nagel (2006), which leads to rejection of the conditional CAPM, potentially suffers from biased estimates of alphas.

[7]Petkova (2006) claims that the FF-3F model is not successful at capturing the effect of conditioning information. Yet, the author argues that the parts of HML and SMB that are important for pricing risky assets proxy for innovations in state variables that predict the excess market return and the yield curve.

the conditional FF-3F model performs well in contrast to several previous documentations. More specifically, the model is able to capture the most significant features of pricing error deviations from the conditional CAPM. Once challenging the models with momentum portfolios, the author is able to demonstrate that the conditional FF-3F successfully captures the perhaps most debated asset pricing anomaly - the momentum effect.

Even though asset pricing with conditioning information is acknowledged to have pervasive influence on empirical performance, previous research on the topic remains strikingly scarce and out-dated for the Swedish equity market. This seems to apply for the whole conditional asset pricing field as well. Hansson and Hördahl (1998), who are among the few to test the conditional CAPM, consider the time-period 1977-1990 by using a multivariate GARCH-M process for the conditional covariance matrix of asset returns. The authors provide strong support for the conditional CAPM. Given the significant shortage of literature covering conditional asset pricing in Sweden, our study seeks to not only fill the research room, but also provide an important link to other related asset pricing fields.

Besides from being the pioneers in introducing a nonparametric SDF in asset pricing with conditioning information, we provide a refined view of the empirical performance of conditional asset pricing models in the Swedish equity market.

3 Data

This section aims to give the reader an overview of the data used to evaluate and estimate conditional asset pricing models. We first provide further details on the construction of both benchmark portfolios as well as risk factors and then discuss the conditioning instruments that are considered in empirical tests.

3.1 Test Portfolios and Fama-French Three Factors

This study considers 25 Swedish stock portfolios: These are the size and book-to-market sorted portfolios of FF (1993). The usage of portfolios as test assets as opposed to individual stocks follows from a couple of natural considerations. First and foremost, we would like to comply with established common practice within the conditional empirical asset pricing literature. Second, we would like to maximize comparability across previous research documentations. Third, we would like to avoid the errors-in-variables problem as discussed by Jagannathan et al. (2010). Additionally, by including the test portfolios we would like to circumvent any potential methodological shortcomings in the empirical testing procedure, such as those brought forward by Lewellen et al. (2010).

Since currently the 25 FF-3F portfolios are not available for Sweden, we construct them ourselves. We adopt the methodology proposed in the seminal papers by FF (1992, 1993). The firm market and book data is collected on a monthly frequency spanning the period July 1994 to December 2016 from Finbas, which is free from firm survivorship bias[8].

To begin with, we remove all firms with negative BE/ME (book-to-market equity divided by market equity) values to mitigate any concerns with erroneous interpretation. Furthermore, all financial firms are exempt from the analysis since high leverage for these firms in general does not have the same meaning as for nonfinancial firms, for which high leverage more likely is linked to distress. In addition to removing any duplicate observations, we also mold

[8]Another popular choice is the Compustat database. However, due to its prominent exclusion bias of firms and several incorrect firm observations, we use Finbas. Similarly, another alternative would be to use Datastream, but it contains a significant portion of errors.

our sample by excluding firms for which no consecutive trading data for the last 24 months is reported. Last but not least, for a firm to be included in the tests we require it to have a stock price for December of year $t-1$, June of t and book common equity for year $t-1$. The full sample contains monthly observations of 996 unique firms, stretching over a time period of 270 trading months, yielding a total number of 88529 observations. The sample period is chosen as to maximize the data quality while ensuring a satisfactory amount of observations in each test portfolio. Table 1 reports some statistics for the monthly number of stocks included in the sample.

Table 1:
Summary Statistics of the Monthly Number of Stocks

Full Sample	
Total	996
Min	147
Max	471
Mean	307
Median	320

In June of each year t from 1994 to 2016, all stocks are ranked on size (market equity or ME). The median of ME is used to split stocks into two groups; referred to as small and big (S and B). The stocks are also divided into three book-to-market equity groups (BE/ME) based on three different breakpoints: The bottom 30 % (Low or L), middle 40% (Middle or M) and top 30 % (High or H). In order to establish the correct lead-lag relationship between accounting variables and returns, BE/ME is calculated based on book common equity for the fiscal year ending in calendar year $t-1$ divided by ME at the end of December of $t-1$. Thus, it is assumed that the accounting variables are known before the returns they are used to explain.

Next, we apply a 2×3 sorting schedule and construct six intersecting portfolios based on the two ME and the three BE/ME groups: S/L, S/M, S/H, B/L, B/M and B/H[9]. For instance, the B/H contains the big-ME stocks that have high BE/ME. Having obtained the portfolios, monthly value-weighted returns are calculated from July of year t to June of $t+1$. The portfolios

[9]According to FF (1993), the usage of three BE/ME and only two size groups is motivated by a stronger role of the former in average stock returns.

are rebalanced every June of $t+1$. We commence with return calculations in July of year t to ensure that book values for the year $t-1$ is known. We ensure that the portfolio returns are stationary by applying a unit root (ADF) test, the results of which are not reported for brevity.

The return on the size portfolio SMB (small minus big) is meant to mimic the risk factor in returns related to size. It is defined as the difference each month between the equally-weighted return on the three small-stock portfolios (S/L, S/M and S/H) and the three big-stock portfolios (B/L, B/M and B/H):

$$SMB_t = \frac{R_{S/L,t} + R_{S/M,t} + R_{S/H,t}}{3} - \frac{R_{B/L,t} + R_{B/M,t} + R_{B/H,t}}{3} \tag{1}$$

Similarly, the return on the HML (high minus low) portfolio aims to mimic the risk factor in returns related to book-to-market equity. It is defined as the difference each month between the equally-weighted return on the two high BE/ME portfolios (S/H, B/H) and the return on the two low BE/ME portfolios (S/L, B/L)

$$HML_t = \frac{R_{S/H,t} + R_{B/H,t}}{2} - \frac{R_{S/L,t} + R_{B/L,t}}{2} \tag{2}$$

The market factor ($MRKT$) is derived by subtracting the monthly return on 1-month Swedish treasury bills from the monthly excess market return:

$$MRKT_t = R_{m,t} - r_{f,t} \tag{3}$$

The treasury bill data is obtained from the Swedish Riksbank's website and is deannualized prior to usage[10]. The return on the market portfolio is calculated as the value-weighted returns of the stocks in the six size-BE/ME portfolios, including any firms with negative BE/ME.

In a similar fashion to what has been stated previously, we form a larger sample of test portfolios for the purpose of empirical model evaluation by conducting further independent sortings. In the original study, Fama and French (1993) construct 25 portfolios (5x5 sorting scheme). In several ap-

[10]The deannualized rate is computed as $r_{f,t} = (1 + \text{T-bill rate})^{1/n} - 1$ assuming 252 trading days and 12 trading months.

plications, we use this number of portfolios. However, in line with Wang (2003), it is worthwhile to reduce the number of test portfolios in some applications as the the test otherwise becomes computationally demanding[11]. Thus, reducing the number of portfolios while keeping a good representation of the size/BE-ME patterns is worthwile. As such, we consider the following five size-BE/ME quintile combinations in some empirical tests: SZ1/BM1, SZ1/BM5, SZ3/BM3, SZ5/BM1 and SZ5/BM5. Table 2 presents summary statistics of the FF-3F test portfolios and the factor returns.

Table 2:
Summary Statistics of the Fama and French Size-BE/ME Portfolios and Factors

Panel A presents means and standard deviations of average monthly excess returns of the 25 Fama and French (1993) Size-BE/ME portfolios. The return measures are arithmetic rates of return, measured in percent (multiplied by 100). The sample period is from July 1994 to December 2016. The variable SZ1 refers to the bottom quintile (20%) firms with regard to size (market capitalization). The same logic applies to book-to-market (book-value of equity divided by market value of equity), where BM1 through BM5 stands for the five quintiles (from low to high). Tabulated in Panel B are summary statistics for the Fama and French (1993) three factors. For more details on the factors and the factor mimicking portfolios, see Fama and French (1992, 1993).

Panel A: The Fama and French Size-BE/ME Portfolios

	Means					Standard Deviations				
	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5
SZ1	0.90	1.42	1.98	0.29	0.81	12.86	11.33	15.69	9.84	8.91
SZ2	0.25	−0.30	0.27	0.45	0.19	9.04	9.01	7.50	8.43	6.60
SZ3	0.12	0.98	0.61	0.52	0.22	9.66	7.38	6.44	5.83	5.55
SZ4	−0.04	0.90	0.63	1.05	0.84	8.54	7.09	7.19	6.17	6.38
SZ5	0.77	0.78	0.88	0.94	1.40	8.57	6.51	5.65	6.34	6.95

Panel B: The Fama and French Three Factors

Variable	Mean	Std. dev.	First auto.	Cross Correlations	
MKT	0.87	5.56	0.12		
SMB	-0.04	5.42	-0.02	-0.01	
HML	0.14	5.13	0.36	-0.23	-0.42

In line with the findings of Fama and French (1992,1993), both the MKT and HML factors earn positive returns (see Panel B of Table 2). However,

[11] If using 25 portfolios, the test would depend on the inverse of a 125×125 matrix in several applications.

the SMB factor is slightly negative. At odds with FF, it seems that there are no pronounced size and BE/ME patterns in the Swedish equity market (see Panel A of Table 2)[12]. This holds for the underlying six portfolios that are used to create the factors as well (not reported for brevity). Yet, these results are not unexpected due to a several number of reasons. First and foremost, even the averages of the monthly U.S and European factor returns provided by Kenneth Frenchs's website are also negative during the corresponding and similarly defined periods of our data sample. Second of all, various authors demonstrate that several developed capital markets do not exhibit any size patterns: Schrimpf et al. (2006) and Ziegler et al. (2007) even find that there is a reversed size pattern in the German stock market. Similarly, Malin and Veeraghavan (2004) as well as Eraslan (2013) document that the SMB factors in the UK and the Turkish markets earn negative average returns. Thirdly, it seems that the global economic recession that took place during the sample period has a large bearing on the results: An exclusion of the tech bubble (year 2000) and the financial crisis (2008) alters the factor returns a lot. As a testimony to form Size-BE/ME portfolios that are largely free of the return behaviour influence of each other, FF (1992) reports low correlations between the factors. While this is successfully the case for the MKT factor, the correlation between SMB and HML is noteworthy. Last but not least, the spread between average excess returns for different portfolios are quite large: the smallest excess return amounts to -0.30% while the largest is roughly 2%.

3.2 Conditioning Variables

In this paper we select a number of conditioning instruments from a larger set of five variables that represent conditioning information. All the state variables are standard in the conditional asset pricing literature and are collected on a monthly frequency over the sample period from July 1994 to December 2016. The first conditioning variable considered is $OMXS30$, the OMX Stockholm 30 index. It is a price return index constructed with the objective of creating a measure, which develops in close correlation with the stocks listed on NASDAQ Stockholm. The index comprises the 30 shares that have the largest volume of trading on NASDAQ Stockholm during a

[12]However, there are indications of SMB patterns and reversed value patterns in return volatilities.

certain period. The inclusion of an index aimed at mirroring the average development of stocks as a conditioning variable is common practice (see for instance Wang (2003) and Harvey (1989)). We collect the data from Datastream, a Thomson Reuters Software Application.

The second conditioning variable is the Swedish total government debt, TGD. Its usage is motivated by several authors having demonstrated that sovereign budget deficits bear a close relationship with the development of the economy (see for instance Abdullah et al. (1993)). Similarly, the trade deficit, i.e. the balance of trade or net exports, is another instrument that comoves with the aggregate economy (see for instance Flannery and Protopapadakis (2002)). Therefore, we have included exports, EXP, in the pool of conditioning instruments. We retrieve the data on exports and total government debt from Statistics Sweden.

This paper also considers the industrial production index, IPI, which is perhaps one of the most widely included variables in conditional asset pricing models that aim to capture beta dynamics. It serves as a proxy for aggregate business conditions in the Swedish economy and is designed to mirror the development of Swedish industrial production. Last but not least, the 1 month short-term interest rate, TB, represents an instrumental variable that has gained increased attention in the literature (see for instance Ferson and Harvey (1999)). The IPI and short-term interest rate data are downloaded from Statistics Sweden and the Swedish Riksbank, respectively. Summary statistics of the conditioning variables employed in this study are tabulated in Table 3.

To test for the predictive power of our set of conditioning variables and narrow down the number of instruments used to the largest possible extent, we run the following regression

$$R_{i,t+1} = a + b' z_t + \varepsilon_{t+1} \tag{4}$$

where $z_t = (TGD_t, OMXS30_t, EXP_t, IPI_t, TB_t)$ is the vector of conditioning variables. Gross returns of the market portfolio proxy are denoted $R_{i,t+1}$. To test the hypothesis that the coefficients on the predictive variables are jointly equal to zero ($H_0 : b = 0$), the Wald statistics are computed. The results show that the joint use of the three popular forecasters TGD, $OMXS30$ and EXP drive out the other conditioning variables in predicting the market. These results are excluded for brevity, but are available upon further request. We limit the choice to these variables in applications, letting other variables be used in robustness checks.

Table 3:
Summary Statistics of Conditioning Variables

This table presents summary statistics of monthly observations from July 1994 to December 2016 of changes in five conditioning variables: Swedish Total Government Debt (TGD), The return on the OMXS30 index (OMXS30), Swedish aggregate exports (EXP), the industrial production index (IPI) and the annualized return on 1-month treasury bills (TB). All the measures are percentages.

Conditioning variables								
Variable	Mean	Std. dev.	First auto.	Cross correlations				
TGD	0.07	2.25	0.004	1.000				
OMXS30	0.79	5.85	0.099	−0.012	1.00			
EXP	0.96	10.91	−0.206	−0.108	−0.058	1.00		
IPI	2.78	2.24	0.984	0.012	−0.038	0.012	1.00	
TB	0.13	2.03	−0.268	0.011	0.047	0.213	0.029	1.00

4 Methodology

In the following section, we in detail explain the different methodologies employed in estimating as well as evaluating both single and multi-factor conditional asset pricing models. More specifically, we begin with presenting a nonparametric test of conditional mean-variance efficiency for the CAPM, whose inference methodology is constructed within the SDF framework. To further set the stage for our methodology, we introduce a wide range of structural change tests to thoroughly examine the degree of nonlinearity. Next, we introduce the procedure according to which statistical inference about mispricing is made. Finally, we extend the testing procedure by considering the conditional FF-3F model.

4.1 The Stochastic Discount Factor

All asset pricing models can be conveniently expressed through the SDF representation. The flexibility of the SDF to present a general theory of asset pricing is now widely recognized, due to its universality and intuitive interpretation. Given the law of one price and the absence of any arbitrage opportunities, each and every asset pricing model delivers a simple pricing equation

$$E_t[m_{t+1}R_{i,t+1}] = 1 \tag{5}$$

that holds for all assets i in the economy ($i = 1, ..., n$). In particular, $R_{i,t+1}$ denotes the gross return of asset i at time $t+1$. The SDF, also known as the pricing kernel, is denoted by m_{t+1} and takes a particular form depending on different asset pricing model specifications. Given an investable riskless asset paying a return of R_f, equation (5) can be expressed in terms of excess returns

$$E_t[m_{t+1}r_{i,t+1}] = 0 \tag{6}$$

where $r_{i,t+1}$ is the excess return (i.e the risk-free rate subtracted from the gross return) on the ith of n assets.

4.2 The Nonparametric Conditional CAPM

In what follows, we will demonstrate how the relationship between returns and beta can be reduced into a non-parametric restriction on the stochastic discount factor m_{t+1} in accordance with equation (6). With a nonparametric expression of the SDF implied by the conditional CAPM, we have the fundamental base for establishing the statistical inference approach. As such, the procedure is designed to test the model and its restriction on the SDF, without assuming any auxiliary functional form.

In a similar fashion to the unconditional version, the conditional CAPM states that the market portfolio is conditionally mean-variance efficient by satisfying the following equation for $t = 1, ..., N$,

$$E[r_{i,t+1}|I_t] = E[r_{p,t+1}|I_t]\frac{\text{cov}[r_{i,t+1}, r_{p,t+1}|I_t]}{\text{var}[r_{p,t+1}|I_t]} \tag{7}$$

where $r_{i,t+1}$ is the excess return on the ith asset in excess of the risk-free rate $r_{f,t}$. Similarly, $r_{p,t+1}$ denotes the excess return of the market portfolio and $E_t[\cdot|I_t]$ represents the conditional expectation given the agent's information set I_t at t. With the use of simple algebra, the well-known covariance representation (7) can be translated into a cross-moment expression[13]:

$$E[r_{i,t+1}|I_t] = E[r_{p,t+1}|I_t]\frac{E[r_{i,t+1}r_{p,t+1}|I_t]}{E[r_{p,t+1}^2|I_t]} \tag{8}$$

This formulation of the CAPM rather than the former serves as the basis for our econometric tests[14]. In equation (8), the moment conditions are based on the agent's entire information set I_t, which is not directly available in a data-set for the econometrician. As a consequence, he has to consider x_t, which is a $k \times 1$ vector of strictly stationary conditioning or instrumental

[13]By multiplying both sides of the covariance representation (7) with $\text{var}[r_{p,t+1}|I_t]$, we end up with $E[r_{i,t+1}|I_t]\text{var}[r_{p,t+1}|I_t] = E[r_{p,t+1}]\text{cov}[r_{i,t+1}, r_{p,t+1}|I_t]$. Next, exchange both $\text{cov}[r_{i,t+1}, r_{p,t+1}|I_t]$ and $\text{var}[r_{p,t+1}|I_t]$ with $E[r_{i,t+1}r_{p,t+1}|I_t] - E[r_{i,t+1}|I_t]E[r_{p,t+1}|I_t]$ and $E[r^2_{p,t+1}|I_t] - (E[r_{p,t+1}|I_t])^2$, respectively. After removing the common term on both sides of the equation, it is easy to see that the covariance representation implies the cross-moment representation and vice versa.

[14]Wang (2003) shows that the nonparametric test based on the cross-moment representation (8) is simpler and performs better in Monte Carlo simulation than (7).

variables such that[15]

$$E[r_{p,t+1}|I_t] = E[r_{p,t+1}|x_t] \tag{9}$$

$$E[r^2_{p,t+1}|I_t] = E[r^2_{p,t+1}|x_t] \tag{10}$$

Next, we define

$$g_p(x_t) = E[r_{p,t+1}|x_t] \tag{11}$$

$$g_{pp}(x_t) = E[r^2_{p,t+1}|x_t] \tag{12}$$

$$b(x_t) = \frac{g_p(x_t)}{g_{pp}(x_t)} \tag{13}$$

Assuming that equations (9) and (10) hold, Jensen's alphas (i.e. conditional pricing errors) can be conveniently formulated as

$$E[r_{i,t+1}|I_t] - E[r_{p,t+1}|I_t]\frac{E[r_{i,t+1}r_{p,t+1}|I_t]}{E[r^2_{p,t+1}|I_t]} = E[m_{t+1}r_{i,t+1}|I_t] \tag{14}$$

Given the cross-moment representation equation (8), it is possible to solve for the implied SDF as

$$m_{t+1} = 1 - b(x_t)r_{p,t+1} \tag{15}$$

As such, equation (8) is congruent with the SDF representation in (6)

$$E_t[m_{t+1}r_{i,t+1}|I_t] = 0 \tag{16}$$

Hence it is possible to verify that the stochastic discount factor m_{t+1} implied by the conditional version of the single-factor CAPM is a function of the first two conditional moments of the market portfolio.

[15]Using x_t for the conditional asset pricing tests is weaker than using the full characterization of the information set I_t. Furthermore, note that the eqs. (9) and (10) pertain to the market portfolio p. These equations are sufficient for the development of nonparametric tests.

4.2.1 Non-parametric Estimation of the Implied SDF

With a non-parametric representation of the stochastic discount factor, one can replace it with an estimate

$$\hat{m}_{t+1} = 1 - \hat{b}(x_t) r_{p,t+1} \tag{17}$$

where

$$\hat{b}(x_t) = \frac{\hat{g}_p(x_t)}{\hat{g}_{pp}(x_t)} \tag{18}$$

For this task, we use a range of sophisticated non-parametric tools at our disposal

$$\hat{f}(x) = \frac{1}{Nh^k} \sum_{s=1}^{N} K\left(\frac{x - X_s}{h}\right) \tag{19}$$

$$\hat{g}_p(x) = \frac{1}{Nh^k} \hat{f}(x)^{-1} \sum_{s=1}^{N} K\left(\frac{x - X_s}{h}\right) r_{p,s+1} \tag{20}$$

$$\hat{g}_{pp}(x) = \frac{1}{Nh^k} \hat{f}(x)^{-1} \sum_{s=1}^{N} K\left(\frac{x - X_s}{h}\right) r_{p,s+1}^2 \tag{21}$$

All of the nonparametric estimators above (19), (20) and (21) are widely popular and standard. The first estimator $\hat{f}(x)$ is a multivariate version of the kernel density estimator (KDE) and represents a nonparametric approach to estimate the probability density function of a random variable. It is also known as the Rosenblatt-Parzen estimator (1962) and has kernel function $K(\cdot)$ and bandwidth parameter h. In general, but not in our case (as explained in Appendix D), the choice of h is the most important factor affecting the accuracy of the KDE, since it governs the orientation and degree of smoothing induced by its user. The reader should be aware of the wide range of different bandwidth operators available and that optimal bandwidth selection still remains a topic subject to debate in the econometric literature.

The estimator $\hat{g}_p(x)$ and $\hat{g}_{pp}(x)$ is a multivariate kernel regression estimator, known as the Nadaraya-Watson kernel estimator (1964). It makes use of the KDE (19) as a weighting function. While (19) and (20) are linear smoothers, the estimator does not imply a linear regression. Rather, the objective is to let the data speak for itself by potentially revealing any non-linear relationships. This helps us bring to light the underlying structure of the regression data when estimating conditional expected returns. As such, nonparametric methods not imposing any stringent parametric assumptions on the underlying probability of the model associated with the data generating process have the ability to expose the dynamic structure of data, which would be completely missed by usual parametric methods. Put differently: Given that an assumed parametric model is not appropriate, the erroneous statistical inference associated with it can lead to seriously misleading interpretations of empirical results. However, nonparametric methods are computationally complex. Moreover, it is crucial to ensure that samples are of satisfactory size, since the motivation of nonparametric methods typically rely on asymptotic distribution theory. As such, ensuring good statistical power in finite sample applications is essential. This can be challenging, since the perhaps most prominent issue faced by nonparametric estimators is the rate at which they converge; typically, the convergence rate is significantly slower compared to parametric estimators. This problem is called the "curse of dimensionality" and is in depth discussed by Silvermann (1986), Pritsker (1998) as well as Chapman and Pearson (2000). In the following sections, we demonstrate how the design of Wang's nonparametric (2003) test successfully addresses and largely eliminates any of these issues.

While the conditional CAPM does not imply a fully specified parametric SDF, it is congruent with an SDF of particular structure: The SDF here is not necessarily positive. This does neither imply any arbitrage opportunities, nor the violation of the law of one price, because there is at least one strictly positive SDF of the CAPM in a discrete time finite asset setting, as discussed by Dybvig and Ingersoll (1982). Furthermore, statisticians have raised the issue that a positive SDF is troublesome in the design of empirical tests, simultaneously as the imposition of such a restriction does not necessarily lead to any distinct advantages in econometric analysis. Hence, this strict specification is often relaxed in tests of asset pricing models (see for example Cochrane (1996)).

4.2.2 Nonlinearity Tests

As a rationale for our nonparametric tests, we begin with investigating the degree of significance of nonlinearity extensively. First, we turn to the suggestiveness of nonparametric conditional beta plots. Then, we inspect the presence of nonlinearity by conducting parameter instability and structural change tests by applying the supF test suggested by Andrews (1993) as well as Andrews and Ploberger (1994). To further set the stage for our tests, we compare cross-sectional forecasts generated by the nonparametric SDF versus those implied from an SDF based on linear conditional moments.

A. SupF Test for Structural Change and Parameter Stability

The supF test is described in Appendix B. The testing approach is based on the following standard linear regression model for all test assets $(i = 1, ..., n)$ across each period $(t = 1, ..., N)$

$$r_{i,t} = r_{p,t}\beta_{i,t} + u_{i,t} \tag{22}$$

where $r_{p,t}$ denotes the time-series vector of excess returns on the market portfolio, $\beta_{i,t}$ represents the regression coefficient and $r_{i,t}$ is the excess return on the test portfolios. To test for any structural changes across any potential change points, one can compute a series of F statistic estimates based on the linear regression model fit. As a next step, these are aggregated into a single test statistic, the sup$F$, which can reveal any nonlinear relationships between conditional betas and excess returns on the assets.

B. Cross-sectional Forecasts

First, the sample is split into two parts of equal lengths, where τ denotes the time points in the second half of the sample. Making use of data up to $\tau - 1$, we estimate the SDF

$$m_t = 1 - \frac{g_p(x_{t-1})}{g_{pp}(x_{t-1})} r_{p,t} \tag{23}$$

where $g_p(x_{t-1}) = E(r_{p,t}|x_{t-1})$, $g_{pp}(x_{t-1}) = E(r_{p,t}^2|x_{t-1})$ and r_p is the excess market return. We first use linear and then nonparametric regressions to estimate the conditional moments g_p and g_{pp}. After computing a parametric and nonparametric SDF, we compare performances, i.e. average pricing errors, of the one-step ahead forecasts generated by the different SDFs. These are computed as in Cochrane (2009). The average forecast error of an SDF $\hat{m}_\tau$ for the ith asset is

$$\frac{\frac{1}{N-N_0}\sum_{\tau=N_0+1}^{N}\hat{m}_\tau r_{i,\tau}}{\frac{1}{N-N_0}\sum_{\tau=N_0+1}^{N}\hat{m}_\tau} \tag{24}$$

where N_0 is the last observation in the first half of the data sample.

4.3 Empirical Evaluation of Mispricing

The statistical inference about the degree of mispricing builds upon the notion that conditionally discounted excess returns implied by the nonparametric SDF follow

$$E[e_{i,t+1}|I_t] = 0 \tag{25}$$

where $e_{i,t+1} = m_{t+1}r_{i,t+1}$. Since $e_{i,t+1}$ is not predictable by conditioning information I_t, a simple way of testing is by running a weighted least squares regression (hereafter WLS regression)

$$e_{i,t+1} = z_t'\delta_i + u_{i,t+1} \tag{26}$$

where z_t is a $q \times 1$ vector of instrumental variables observed in the information set I_t[16]. As standard practice and an implication of (25), $E[u_{i,t+1}|I_t] = 0$ for all $i = 1, 2..., n$. Intuitively, the regression coefficients are expected to be zero as a manifestation of the moment condition (16). Put differently: $\delta = 0$, where $\delta_N = (\delta_1'\delta_2'...\delta_n')'$.

For econometric evaluation of pricing errors, we nonparametically estimate the SDF $\hat{m}_{t+1}$ according to (17) and then derive the parameter vector of coefficients δ_i for all $i = 1, ..., n$.

$$\hat{\delta}_i = \left(\frac{1}{N}\sum_{t=1}^{N}\hat{w}_t z_t z_t'\right)^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\hat{w}_t z_t \hat{e}_{i,t+1}\right) \tag{27}$$

The weighting function is $\hat{w}_t = \hat{f}(x_t)\hat{g}_{pp}(x_t)$, the choice of which is motivated purely by the technicality to establish the large sample theory. In

[16] The reader should be made aware that with a one period lag behind m_{t+1}, z_t can be whichever vector included in the information set I_t. The choice of $z_t = x_t$ is a natural choice in several testing approaches.

other words, $\frac{1}{N}\sum_{t=1}^{N}\hat{w}_t z_t z_t'$ and $\frac{1}{N}\sum_{t=1}^{N}\hat{w}_t z_t \hat{e}_{i,t+1}$ can be expressed as secondorder generalized U-statistics, enabling empirical investigation of large sample properties of $\hat{\delta}_i$. If the weighting functions do not resemble U-statistics, the econometrician faces complex issues in developing the distribution theory[17]. The weighting function employed in (27) represents the easiest of those implying U-statistics structures. Based on the comfortability of this particular weighting function, Robinson (1989) and Powell et al. (1989) among others apply density weighting in their empirical studies as well.

The test, according to which we make statistical inference, is based on the WLS estimator $\hat{\delta}_N$ in equation (27) above

$$\hat{\delta}_N = (\delta_1'\delta_2'...\delta_n')' \tag{28}$$

The WLS estimator gives the test a simple and nice interpretation: In case of conditional mean-variance efficiency of the market portfolio p, the estimator $\hat{\delta}_N$ converges to zero. If mean-variance efficiency is violated, the estimator rather converges to a non-zero limit (unless $e_{i,t+1}$ is orthogonal to the components in z_t for all i). Naturally, one can test whether the distance of the estimator is zero and thus quantify the degree of mispricing by using asymptotic distribution theory to look into sampling errors. Furthermore, the test is designed to spot any time series variation in pricing errors: since a significant component of the estimator δ_N indicates that returns are correlated with the corresponding predictive variables, the test implies a view upon the time-variation of mispricing. This is clearly demonstrated if the user applies (26) to model pricing errors, letting $z_t'\delta_i$ proxy for $E[e_{i,t}|I_t]$, the conditional pricing errors from equation (8).

The test developed by Wang (2002, 2003) has very appealing statistical properties that successfully address and circumvent the otherwise prominent drawbacks of nonparametric methods. The estimator $\hat{\delta}_N$ demonstrates characteristics of its parametric counterpart by exhibiting the fast parametric convergence rate ($\sqrt{N}$ convergence rate), no matter how many conditioning variables are used. This is implied by a statistical fact that the average of slowly converging nonparametric kernel density estimates can lead to parametric convergence rates (More on this fact in Hardle and Stoker (1989),

[17]We choose not to set $\hat{w}_t = 1$, since it does not imply U-statistic structures.

Powell et al. (1989), Robinson (1989) and Lee (1992))[18]. Given this desirable feature, the test may result in good power in statistical settings. Furthermore, if $h \to 0$, $Nh^{2k} \to \infty$ and $Nh^{2k+2} \to 0$, the WLS estimator $\hat{\delta}_N$ is such that $\sqrt{N}(\hat{\delta}_N - \delta)$ has limiting multivariate normal distribution with mean 0 and variance-covariance matrix Ω. Using a consistent estimator $\hat{\Omega}_N$ of the variance-covariance matrix, the following test statistic is proposed by Wang (2002)

$$T_\delta = N\hat{\delta}_N'\hat{\Omega}_N^{-1}\hat{\delta}_N \tag{29}$$

The test statistic can be shown to have a limiting chi-squared distribution with qn degrees of freedom given that the conditional CAPM holds. Understanding the desirable features of the test is of crucial importance in motivating the methodology employed in this study. Therefore, derivations of proofs and detailed test mechanics as brought forward in Wang (2002) are presented in Appendix A. In particular, the reader can find details on the asymptotic variance-covariance matrix Ω and the test statistic calculation under the Appendix A.3 Section given the notation in A.2.

4.3.1 Estimation and Measurement of Pricing Errors

In applications, we have to estimate the conditional Jensen's alphas (i.e. pricing errors) related to $E_t(r_{i,t+1})$ for all $t = 1, ..., N$ across all test assets $i = 1, ..., n$. Letting $\varepsilon_{i,t}$ represent the pricing error series, we run

$$\hat{\varepsilon}_{i,t} = z_t'\hat{\delta}_i \tag{30}$$

where the vector of regressors is $z_t = (1_t x_t')'$ and the WLS $\hat{\delta}_i$ is defined in (27). Then, we use three summary measures to evaluate the degree of mispricing. The first is the average absolute bias (AAB), which simply measures the average pricing error or bias of the model. The second is the average standard deviation (ASD) and is aimed at the volatility of the pricing errors. The third is the average root mean squared error (ARMSE), which represents a measure that accounts for both bias and volatility of the conditional alphas. These measures are used to evaluate the cross-sectional degree of mispricing of the

[18] Wang(2002) even finds that the nonparametric tests have better power than the standard GMM approach.

n pricing errors series as follows

$$\text{AAB} = \frac{1}{n}\sum_{i=1}^{n} |B_i| \tag{31}$$

$$\text{ASD} = \frac{1}{n}\sum_{i=1}^{n} \text{SD}_i \tag{32}$$

$$\text{ARMSE} = \frac{1}{n}\sum_{i=1}^{n} \text{RMSE}_i \tag{33}$$

where B_i, SD_i and $RMSE_i$ are the sample mean, standard deviation and root mean squared error of the pricing error series $\varepsilon_{i,1}, \varepsilon_{i,2}, ..., \varepsilon_{i,N}$, respectively.

4.4 The Nonparametric Conditional FF-3F Model

This subsection aims to describe an application of multi-factor extensions, which is used to estimate and evaluate the conditional FF-3F model. Specifically, no parametric assumptions about excess returns on the benchmark portfolio p are made. Given a parameter vector θ, we investigate the following case:

The Conditional FF-3F: $\theta_{k,t}$ are nonparametric for k factors

To the best of our knowledge, this version of the FF-3F model is the most general ever considered in the Swedish asset pricing literature[19]. The hypothesis of mean-variance efficiency implied by the conditional FF-3F model with nonparametric excess returns can be expressed

$$r_{p,t+1} = MKT_{t+1} + \theta_{1,t}SMB_{t+1} + \theta_{2,t}HML_{t+1} \tag{34}$$

where the proportions of the size and book-to-market portfolios in the benchmark are not fixed constants, but instead are time-varying. This approach allows us to test the hypothesis in absence of any functional form restrictions. Furthermore, both the $l \times 1$ parameter vectors $\theta_{1,t}$ and $\theta_{2,t}$ are fully

[19]By publication of his article, Wang (2003) made the claim that "The three-factor model in such a general form has never been examined in the literature".

nonparametric for time $(t + 1)$ excess returns. This is clearly emphasised by the definition (34), where the dependent variable is formulated as $r_{p,t+1}$ rather than $r_{p,t+1}(\theta)$.

Given that the benchmark portfolio is conditionally mean-variance efficient, the conditional FF-3F is congruent with a SDF of the following type[20]

$$m_{t+1} = 1 - b_{0,t}MKT_{t+1} - b_{1,t}SMB_{t+1} - b_{2,t}HML_{t+1} \tag{35}$$

As such, the econometrician can solve for the coefficients $b_{0,t}$, $b_{1,t}$ and $b_{2,t}$ by imposing the following system of pricing equations, where it is implicitly assumed that the SDF prices the three factors correctly.

$$E_t[m_{t+1}MKT_{t+1}] = 0 \tag{36}$$

$$E_t[m_{t+1}SMB_{t+1}] = 0 \tag{37}$$

$$E_t[m_{t+1}HML_{t+1}] = 0 \tag{38}$$

This SDF fully conforms with that of the beta-pricing equation. The above equations can conveniently be expressed in the set of equations with matrix notation

$$A_t b_t = c_t \tag{39}$$

where

$$A_t = \begin{bmatrix} E_t[MKT_{t+1}^2] & E_t[MKT_{t+1}SMB_{t+1}] & E_t[MKT_{t+1}HML_{t+1}] \\ E_t[MKT_{t+1}SMB_{t+1}] & E_t[SMB_{t+1}^2] & E_t[SMB_{t+1}HML_{t+1}] \\ E_t[MKT_{t+1}HML_{t+1}] & E_t[SMB_{t+1}HML_{t+1}] & E_t[HML_{t+1}^2] \end{bmatrix} \tag{40}$$

and $b_t = [b_{0,t}, b_{1,t}, b_{2,t}]'$ and $c_t = [E_t[MKT_{t+1}], E_t[SMB_{t+1}], E_t[HML_{t+1}]]'$. Next, we replace the SDF in (35) with

$$\hat{m}_{t+1} = 1 - \hat{b}_{0,t}MKT_{t+1} - \hat{b}_{1,t}SMB_{t+1} - \hat{b}_{2,t}HML_{t+1} \tag{41}$$

[20] To demonstrate this, one can start with the basic idea of the SDF (i.e. $E_t[m_{t+1}r_{i,t+1}] = 0$) and then show that $E_t(r_{i,t+1})$ is a linear function of the conditional betas with the three factors. The reader is directed to Cochrane (2001) for more information on general equivalence pertaining to beta representations and the SDF.

where $\hat{b}_t = \hat{A}_t^{-1}\hat{c}_t$ and $\hat{A}_t$ as well as $\hat{c}_t$ are derived by replacing A_t and c_t with each element using the kernel density estimate in the previously outlined equations (20) and (21) above. As an example: to obtain the kernel estimate for $E_t[MKT_{t+1}HML_{t+1}]$, just replace $r_{p,s+1}$ with $MKT_{t+1}HML_{t+1}$ in (20).

Given the estimates of the SDF, we reiterate the aforementioned process in the section above to derive the test statistic

$$T_\delta = N\hat{\delta}_N'\hat{\Omega}_N^{-1}\hat{\delta}_N \tag{42}$$

To investigate the pricing power of the conditional multifactor FF-3F model, we implement the Politis and Romano (1994) stationary bootstrap, the technicalities of which are more in detail explained in Appendix C. In general, there are several block bootstrap methods available for the econometrician. The specific methodology by Politis and Romano (1994) makes use of overlapping blocks with lengths that are randomly sampled from the geometric distribution. The procedure is well suited for stationary and weakly dependent data. In brief, the stationary bootstrap is characterised by dividing the data into different blocks with random resampling of the blocks with replacement. A significant advantage with randomized block lengths is the implied stationarity, which is not necessarily obtained when using blocks with fixed lengths. In our case, we resample the data and get a number of N_b bootstrapped values of $\hat{\delta}_N$, which are referred to as $\hat{\delta}^*_{N,j}$ with j denoting the indexes pertaining to N_b. Then, we derive the following test statistic

$$T^*_{\delta,j} = N(\hat{\delta}^*_{N,j} - \hat{\delta}_N)'(\hat{\Omega}^*_{N,j})^{-1}(\hat{\delta}^*_{N,j} - \hat{\delta}_N) \tag{43}$$

Intuitively, the statistical inference is based on comparing the $T^*_{\delta,j}$ quantiles with T_δ. More specifically, making use of $T^*_{\delta,j}$ for $j = 1, ..., N_b$, we approximate the distribution of $N(\hat{\delta}_N - \delta)'\hat{\Omega}_N^{-1}(\hat{\delta}_N - \delta)$, which is the statistic T_δ under the H_0 that the weighted least squares estimator δ is zero. This approach avoids any auxiliary functional form assumptions[21]. Furthermore, by using a time series bootstrap method it is possible to avoid the limiting distribution theory.

[21] Given this methodology, the econometrician circumvents several technicalities pertaining to asymptotic results for the FF-3F model. More specifically, the econometrician has to cope with generalized U-statistics of order $k + 1$ for a conditional k factor model.

5 Empirical Results

In this section, we present and thoroughly analyse the empirical results. First, we describe the evidence on nonlinearity and its implications for a nonparametric SDF. As a next step, we turn to the investigation of the conditional CAPM and discuss its empirical performance on both a stand-alone basis and when compared to its unconditional counterpart. In a similar fashion, we proceed with the dynamic version of the FF-3F. In Appendix E, we elaborate on the robustness of the empirical results under several methodological re-specifications.

5.1 Evidence for Nonlinearity

We first present some graphical evidence on nonlinearity to motivate the usage of a nonparametric SDF in empirical asset pricing. To do so, we estimate conditional betas of the Fama and French (1993) size-BE/ME portfolios with respect to the MKT, SMB and HML factors, respectively. Figure 1 plots the betas of the SZ5/BM5 portfolio. That is to say, the portfolio that contains the top 5 quintile book-to-market value and size firms. For conciseness, we merely report the results for this portfolio, as patterns and empirical conclusions about nonlinearity remain qualitatively robust to the choice of portfolio. Conditional betas are computed by the aforementioned standard kernel regression estimates (19), (20) and (21) of the first two conditional moments of returns. The bounds represent the 95% confidence intervals. Furthermore, to cope with any shortcomings of nonparametric methodologies, we focus on univariate functions, i.e. the relationships between one instrumental variable and conditional betas, while keeping the other conditioning variables constant at their averages. By this practice, we successfully focus on areas with more data points and also avoid the statistical curse of dimensionality.

Several cases in Figure 1 strongly suggest the inappropriateness of assuming linear relationships, implying that a nonparametric SDF is fruitful in empirical analysis. In particular, the MKT factor beta seems to be highly nonlinear in TGD, while the SMB and HML factors betas demonstrate strong deviations from linearity with respect to OMXS30. On the other hand, one could argue that the factor beta functions can be well approximated with a linear function when examining EXP as a conditioning variable. In sum, the conditional beta plots reflect the presence of nonlinearity. These findings reinforce the well established critique against assuming convenient linear specifications without any prior investigation of its empirical adequacy. The results in Figure 1 are expected, as there is only little theoretical guidance about how betas should vary with conditioning variables. By investigating the Swedish equity market, we augment previous findings by providing a first out-of-country sample check.

Even though the results reported in Figure 1 suffice in motivating a nonlinear pricing kernel in Swedish asset pricing, some beta estimates are too noisy in light of their respective confidence intervals. To complement and verify the results obtained this far, we turn to more sophisticated statistical tools.

A. SupF Test for Parameter Stability and Structural Change

Panel A in Table 4 reports the results from performing the SupF test for parameter stability and structural change. We find that the test leads to a rejection of linearity for most of the portfolios with a varying degree of significance. Most rejections either occur at the 5% or 1% significance level. In line with Wang (2003), our tests lead to strong rejections of linearity between conditional betas and variables that represent conditioning information. These findings lend further support to the notion that linear modelling assumptions do not suffice in explaining the dynamic behaviour of returns in the Swedish market.

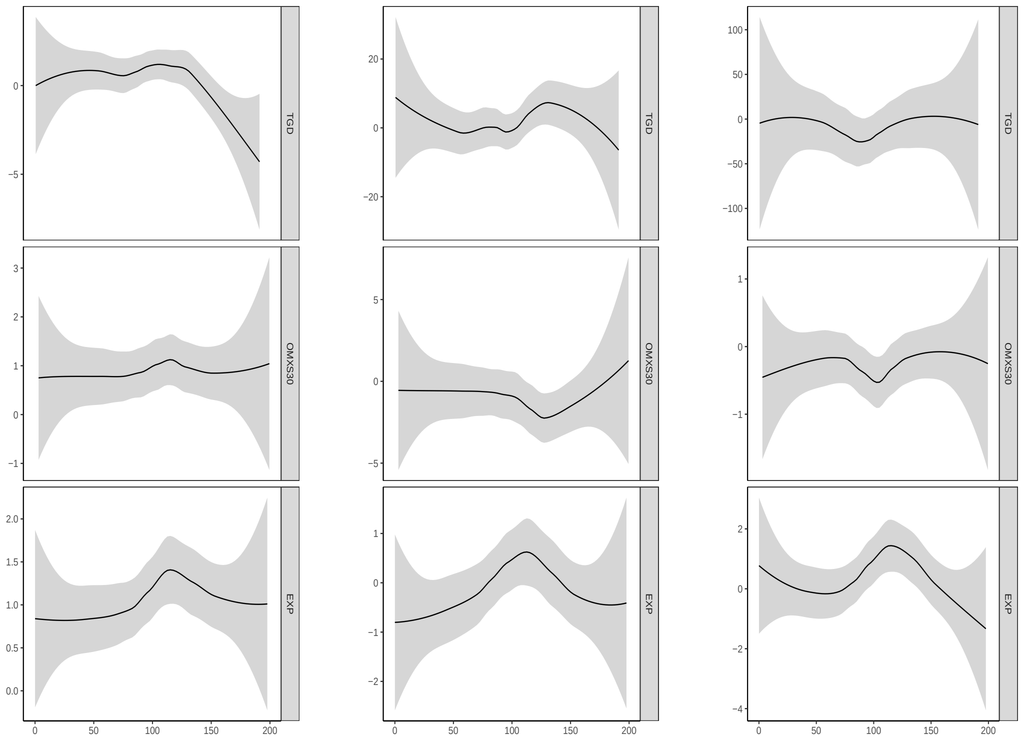
B. Cross-sectional Forecasts

Given the strong evidence that a nonparametric SDF can yield significant benefits in empirical analysis, we proceed our investigation by challenging the cross-sectional forecasts generated by a parametric SDF with those generated by a SDF based on nonlinear moments. Panel B in Table 4 presents the average pricing errors of the one-step ahead forecasts of the two SDFs for each portfolio. For the 25 size-BE/ME portfolios, the nonparametric SDF proves to have superior forecasting power than its parametric counterpart in slightly more than half of the cases with the margin of outperformance being significant for many portfolios. In particular, it proves to be significantly

Figure 1: Plots of Conditional Betas

This figure reports the estimated beta functions. The solid lines represent univariate beta function estimates for the relation between the beta and a single instrumental variable, holding other instrumental variables fixed at their respective averages. Highlighted areas represent 95% confidence intervals. The multivariate beta function is obtained from the standard kernel regression estimates (19) (20) and (21) of the first two conditional moments of the market return. The plot is constructed as follows: The horizontal axis corresponds to the interval for the conditioning instrument and ranges from two standard deviation shocks below and above the mean, respectively. The intervals are rescaled to a range from 0 to 200, where the value 100 serves as the location of the mean.

Panel A: MKT Beta | Panel B: SMB Beta | Panel C: HML Beta



better for all the SZ5/BM combinations, and almost across all SZ/BM2 combinations. Furthermore, when comparing the cross-sectional forecasts, it is interesting to conclude that the SDFs predict pricing errors with different signs for about 10 of the portfolios. In sum, the results brought forward this far strongly evidence that nonlinearity in the first two conditional moments of the market return is important and that a nonlinear pricing kernel that captures this feature is valuable.

Table 4:
Tests for Nonlinearity

This table presents the results from two approaches for examining the significance of nonlinearity. Panel A tabulates the result from the supF method suggested by Andrews (1993) as well as Andrews and Ploberger (1994) to test for structural breaks. The testing procedure is more in detail described in Appendix B. To examine the suitability of linearity, the test is aimed at the following simple linear regression model:

$$r_{i,t} = r_{p,t}\beta_{i,t} + u_{i,t} \tag{44}$$

where $r_{p,t}$ is the time-series vector of excess returns on the market portfolio. Furthermore, $\beta_{i,t}$ represents the regression coefficient and $r_{i,t}$ is the excess returns on the test portfolios. We use the five Fama and French (1993) size-BE/ME portfolios SZ1/BM1, SZ1/BM5, SZ3/BM3, SZ5/BM1 and SZ5/BM5 as tests assets. The sampling period is from July 1994 to December 2016. Significance levels for statistical rejections of linearity between conditional betas and market returns at the ten percent level appear in the table with a single $*$, at five percent with$**$, and at one percent with $***$. Panel B shows the output related to the cross-sectional forecasts generated by a nonparametric and parametric SDF. Reported are the average forecast errors, i.e. pricing errors, of the one-step ahead forecasts pertaining to the SDFs. Here, all the 25 size-BE/ME portfolios serve as test assets.

Panel A: Suprenum F Test Results for Structural Change

	BM1	BM2	BM3	BM4	BM5
SZ1	10.12	18.50***	33.78***	10.14	6.52
SZ2	7.84	7.59	29.63**	14.76**	28.59***
SZ3	8.90	20.87**	14.32**	17.48**	13.04*
SZ4	22.34***	12.41*	16.23**	20.31***	6.86
SZ5	52.39***	14.67**	22.10***	9.15	21.84***

Panel B: Cross-Sectional Forecasts for Fama and French (1993) Test Portfolios

	Nonparametric SDF					Parametric SDF				
	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5
SZ1	−1.28	0.49	−1.65	−3.40	−1.81	0.29	1.08	2.19	−0.81	0.42
SZ2	0.42	0.16	−0.31	0.55	−0.53	−0.29	−0.80	0.00	0.05	-0.49
SZ3	−0.08	1.01	−1.01	0.39	−0.39	−0.73	0.20	−0.20	−0.26	-0.84
SZ4	0.92	−0.27	0.87	0.12	−1.71	−0.33	0.58	0.26	0.39	0.08
SZ5	0.27	0.14	−0.53	0.60	0.21	0.74	1.03	1.18	1.28	1.35

5.2 The Nonparametric Conditional CAPM

Having motivated the usage of a nonparametric SDF in asset pricing with conditioning information for the Swedish equity market, we turn to investigating the conditional CAPM. The results reported are based on the previously motivated conditioning variables: TGD, OMXS30 and EXP. To verify the robustness of the empirical results, we have conducted a series of sensitivity analyses. The measures taken are further described in Appendix E. In Appendix E, any limitations pertaining to the study are also acknowledged and discussed. To a large extent, the empirical findings remain qualitatively robust to a wide range of model re-specifications. The output from the robustness checks are excluded in this paper for brevity but are available upon further request.

Table 5 reports the output for the conditional CAPM. Panel A shows that conditional mean-variance efficiency is statistically rejected at the 10% significance level (p-value 0.081). However, all individual regressors including the intercept are statistically insignificant when viewed separately. In particular, OMXS30 produces a low test statistic with a corresponding high 0.94 p-value.

Moving to the output from running WLS regressions, it is possible to look further into the time-varying mispricing. Because of the specification that the three regressors are the demeaned conditioning variables, the intercepts reflect average pricing errors[22]. In view of this, Panel B shows that the conditional CAPM seems to price average returns very well: The intercepts or pricing errors are statistically insignificant. Furthermore, average pricing errors are in the range of -0.25 to -0.01% and are small in economic terms.

When examining the factor loadings in more detail, it is possible to conclude that the magnitudes are very small. In particular, OMXS30 exhibits the smallest coefficients. In absolute terms, the loadings on OMXS30 and EXP are always smaller than those on TGD. The same pattern holds for the standard errors of the regression coefficients as well. Interestingly, TGD

[22]By letting the regressors be the conditioning variables minus their corresponding averages, neither is the estimation and inference about slopes on time-varying regressors affected, nor is the joint statistic and pricing errors. With regard to the intercepts, statistical inference is affected. Statistically we do not pay attention to any potential estimation noise in the sample means of the regressors. This is statistically equivalent to assuming that the regressors are computed in deviation form.

Table 5:
Investigating the Conditional CAPM,

This table provides results from testing the conditional CAPM. Panel A provides test statistics for both individual regressors and a joint test, in which all regressors are considered. The regressor vector is $z_t(1x_t')'$ in the tests and the vector of conditioning variables is $x_t = (TGD_t, OMXS30_t, EXP_t)$. In order to investigate whether $H\delta = 0$, where H denotes a $n \times qn$ matrix, the test statistic is constructed as $N(H\hat{\delta}_N)'(H\hat{\Omega}_N H')^{-1}(H\hat{\delta}_N)$. It has a limiting $\chi^2(n)$ distribution. The p-value represents the probability that a draw from the chi-squared distribution exceeds the test statistic. Panel B tabulates the coefficient estimates obtained from the WLS regression method explained in Section 4.3. Standard errors appear in parentheses. In both panels, the variable regressors are the demeaned conditioning variables. The following five size/BE-ME portfolios are considered as test assets: SZ1/BM1, SZ1/BM5, SZ3/BM3, SZ5/BM1, SZ5/BM5. The sample period is from July 1994 to December 2016.

Panel A: Significance Tests of Regressors

Output	Joint Test	Individual Regressors			
		Intercept	TGD	OMXS30	EXP
Test stat	9.81	5.49	3.46	1.28	3.35
p-value	0.081	0.36	0.63	0.94	0.65

Panel B: The Weighted Least Squares Regression Estimates

Test Portfolios	Intercept	TGD	OMXS30	EXP
SZ1/BM1	-0.247 (0.135)	-0.129 (0.122)	-0.001 (0.019)	-0.018 (0.012)
SZ1/BM5	-0.021 (0.049)	-0.028 (0.045)	-0.002 (0.007)	-0.003 (0.004)
SZ3/BM3	-0.149 (0.077)	-0.036 (0.070)	0.002 (0.011)	0.000 (0.007)
SZ5/BM1	-0.160 (0.069)	0.033 (0.063)	0.000 (0.001)	0.001 (0.006)
SZ5/BM5	-0.011 (0.068)	-0.050 (0.061)	0.001 (0.001)	-0.001 (0.006)

enters with a negative sign in most cases, suggesting a negative correlation pattern with pricing errors.

Next, we challenge the nonparametric version of the conditional CAPM against the parametric CAPM to evaluate whether any distinct payoffs are to be made by studying asset pricing models with conditioning information. Table 6 reports conditional pricing errors and the associated standard deviations.

The empirical results are striking! The conditional CAPM strongly outperforms its unconditional counterpart and the margin is substantial. More specifically, the volatility of the pricing errors are significantly lower for each and every test portfolio. Furthermore, the nonparametric CAPM outperforms the unconditional version for 20 out of the 25 size-BE/ME portfolios when investigating average pricing errors. The results are largely consistent with those of Wang (2003)[23]. When examining conditional alphas more in detail, it is interesting to note two things. First and foremost, they are economically small. In detail, the nonparametric CAPM generates standard errors and pricing errors who are in the range between 0.34 to 0.02% and -0.33 to 0.16 %, respectively. Second, all the pricing errors are negative except for two portfolios, implying that the conditional CAPM seems to systematically underestimate returns to some small extent.

Ghyssels (1998) claims that if the econometrician correctly captures the time-varying behaviour of risk premia, dynamic beta models are much likely to be superior than static models. However, the author points out that beta risk can be severely misspecified by the imposition of functional form restrictions, which can lead to devastating empirical consequences when investigating conditional asset pricing models. In line with this, our empirical findings evidence that a nonparametric version of the conditional CAPM works well and offers a large improvement over the unconditional CAPM. In other words, a nonparametric SDF can bring significant value in the study of conditional asset pricing models.

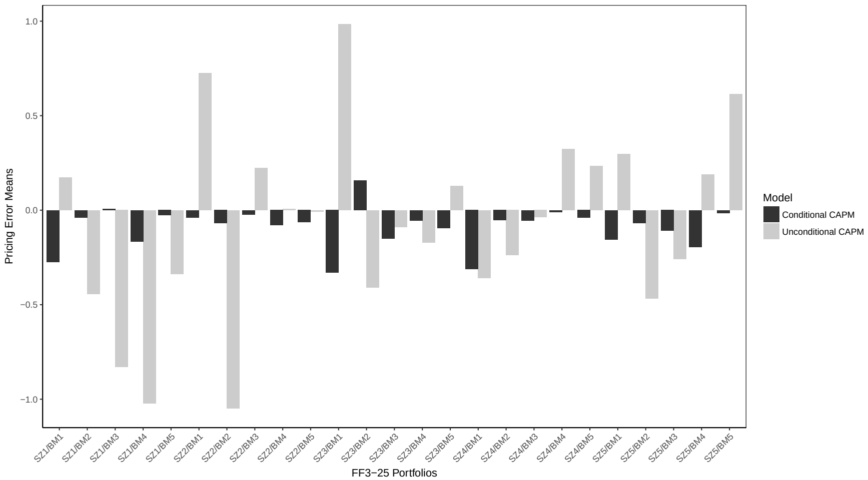
[23] Wang (2003) statistically rejects the conditional CAPM at the 1 % significance level.

Table 6:

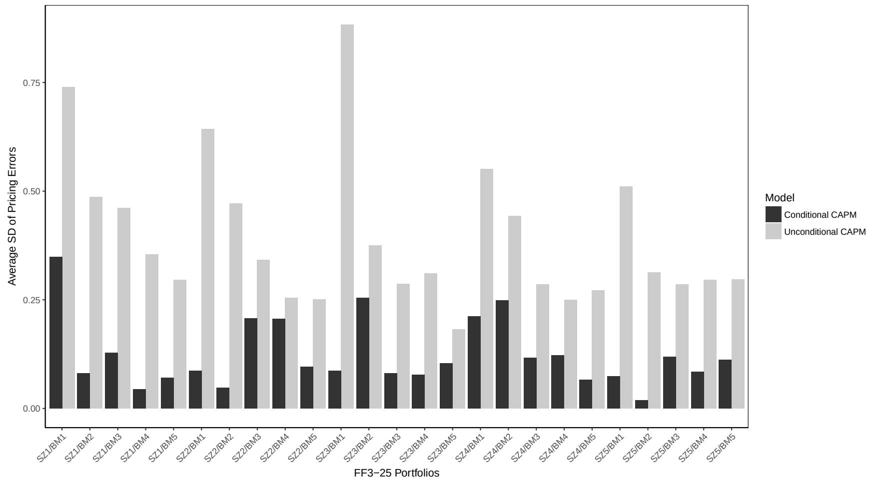
The Conditional CAPM versus the Unconditional CAPM

This table presents and plots means and standard deviations of estimated pricing errors of the conditional and unconditional CAPM for the 25 Fama and French (1993) size-BE/ME test portfolios. The pricing errors for the conditional CAPM are estimated in accordance with the WLS regression approach described in Section 4.3. The same method applies for the unconditional CAPM as well, where the SDF is $m_t = 1 - b_0 MKT_t$. Panel A tabulates the estimates, while Panel B and Panel C plots the estimates.

Panel A: Pricing Errors																				
	The Conditional CAPM										The Unconditional CAPM									
	Average					Std. Dev					Average					Std. Dev				
	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5
SZ1	-0.27	-0.04	-0.33	-0.31	-0.16	0.35	0.09	0.09	0.21	0.07	0.17	0.72	0.98	-0.36	0.30	0.74	0.64	0.88	0.55	0.51
SZ2	-0.04	-0.07	0.16	-0.05	-0.07	0.08	0.05	0.25	0.25	0.02	-0.44	-1.05	-0.41	-0.24	-0.47	0.49	0.47	0.38	0.44	0.31
SZ3	0.01	-0.02	-0.15	-0.05	-0.11	0.13	0.21	0.08	0.12	0.12	-0.83	0.22	-0.09	-0.04	-0.26	0.46	0.34	0.29	0.29	0.29
SZ4	-0.17	-0.08	-0.05	-0.01	-0.19	0.04	0.21	0.08	0.12	0.08	-1.02	0.01	-0.17	0.32	0.19	0.36	0.25	0.31	0.25	0.30
SZ5	-0.03	-0.06	-0.09	-0.04	-0.01	0.07	0.10	0.10	0.07	0.11	-0.34	-0.01	0.13	0.23	0.61	0.30	0.25	0.18	0.27	0.30



Panel B: Pricing Error Means



Panel C: Pricing Error Standard Deviations

5.3 The Nonparametric Conditional FF-3F Model

Next, we consider the most general version of the FF-3F Model examined in the Swedish equity market. As aforementioned, we investigate conditional mean-variance efficiency of the benchmark portfolio p with period $(t+1)$ excess returns for the following specification

$$r_{p,t+1} = MKT_{t+1} + \theta_{1,t} SMB_{t+1} + \theta_{2,t} HML_{t+1}$$

where the three factors reflect the excess returns on the market portfolio and the factor mimicking portfolios for the size and value portfolios, respectively. In this case, the model is cast in its most general form, without the imposition of any parametric assumptions about the parameter vectors $\theta_{1,t}$ and $\theta_{2,t}$. Table 8 reports the FF-3F test output.

As a testimony of the empirical success, the nonparametric FF-3F is not statistically rejected by the stationary bootstrap, which generates a p-value of 34%. Even more importantly, the conditional FF-3F proves to capture the most dispersion in the pricing errors of the conditional CAPM. Across all the three measures, the conditional FF-3F reduces the pricing errors with more than 80 %! More specifically, AAB and ARMSE are reduced by 83%, followed by 82% for ASD. Furthermore, all the pricing error measures are economically small. The empirical findings are in line with those of Wang (2002). Yet, the pricing error measures are significantly lower for the Swedish equity market, in some cases about half the magnitude when compared to Wang's (2003) investigation of the U.S equity market.

The empirical findings indeed show that the SMB and HML factors significantly can lead to improved model performance. To summarize, the fully nonparametric version of the FF-3F model serves as a prominent contribution to the Swedish asset pricing literature.

To wrap up, our empirical findings strongly and consistently suggest that nonlinearity in the first and second moment of the market return is crucially important. If accounted for correctly, significant payoffs in the study of conditional asset pricing models can be generated. For this purpose, a nonparametric SDF is a solution that provides significant value. Furthermore, the results that the nonparametric versions of the conditional CAPM and FF-3F exhibit strong performance contrast several conclusions generated by models that assume linear relationships. Taken together, this implies that functional form mispecifications of betas, risk premia and the SDF can drastically shape empirical results.

Table 8:
Investigating the Conditional Fama and French (1993) Three Factor Model

This table shows the results obtained when testing the fully nonparametric conditional Fama and French (1993) model. The following model is hypothesised to be conditionally mean-variance efficient for $(t+1)$ excess returns

$$MKT_{t+1} + \theta_{1,t}SMB_{t+1} + \theta_{2,t}HML_{t+1}$$

In other words, $\theta_{1,t}$ and $\theta_{2,t}$ are fully non-parametric. MKT_{t+1} is the excess returns of the market portfolio and SMB_{t+1} as well as HML_{t+1} are the returns on the factor mimicking portfolios of the size and value factors, respectively. The regressor vector amounts to $z_t(1x_t')'$ in the tests and the vector of conditioning variables is $x_t = (TGD_t, OMXS30_t, EXP_t)$. The test portfolios are the following 5 Fama and French (1993) size-BE/ME portfolios combinations: SZ1/BM1, SZ1/BM5, SZ3/BM3, SZ5/BM1 and SZ5/BM5. The sample period is from January 1994 to December 2016. The p-value is the probability that a draw from the $\chi^2(n)$-distribution exceeds the test statistic. Additionally, the table tabulates the degree of mispricing as indicated by the three pricing error measures described in Section 4.3. The pricing errors are calculated according to the WLS regression method for all the 25 Fama and French (1993) size-BE/ME portfolios. AAB represents the average absolute bias, whereas ASD is the average standard deviation and ARMSE is the average root mean squared error.

Test output and Pricing Error Measures					
	Test statistic	p-value	AAB	ASD	ARMSE
Conditional FF-3F Model					
$\theta_{1,t}$ & $\theta_{2,t}$ nonparametric	8.51	0.340	0.018	0.022	0.030
Conditional CAPM					
$\theta_{1,t} = \theta_{2,t} = 0$	9.81	0.081	0.104	0.124	0.175

6 Conclusion and Future Research

The main objective of this study is to be the very first in introducing a nonparametric SDF in asset pricing with conditioning information for the Swedish equity market. For this purpose, we conduct asset pricing tests based on the nonparametric methodology proposed by Wang (2002, 2003), which completely avoids functional form mispecifications of beta dynamics, risk premia and the SDF. As a result, this paper provides a unique view about the empirical performance of conditional asset pricing models in Sweden. In particular, this paper examines the most general versions of the conditional FF-3F and CAPM in Sweden to date.

With a strong flood of negative empirical documentations on unconditional model performance, several studies have argued that the ignorance of time-variability of betas and their associated asset pricing implications serve as important sources in explaining poor model performance. Ghyssels (1998) and Wang (2003) among others have emphasized that once beta risk dynamics are correctly specified, conditional asset pricing models are sure to outperform their unconditional counterparts. Yet, several dynamic specifications of asset pricing models have failed miserably with this task. As such, many authors argue that this failure of conditional asset pricing models to capture the dynamic nature of returns is due to convenient but econometrically harmful modeling assumptions. In line with this, we use sophisticated econometric techniques and extensively show that linear approximations of beta functions are empirically inappropriate. Rather, our empirical results strongly reflect the presence of nonlinearity in the first two conditional moments of the market return. This is expected, as there is no theoretical reason to expect betas to be linear.

The investigation of a nonparametric version of the conditional CAPM yields strongly positive conclusions. More specifically, the empirical analysis shows that intercepts, or average pricing errors, are insignificant and economically quite small. In other words, the conditional CAPM seems to price stocks well. When challenged against its unconditional counterpart, the dynamic version based on nonlinear moments offers clearly significant improvement: For all size-BE/ME test portfolios, the standard deviations of conditional alphas are distinctly lower for the nonparametric CAPM. Similarly, average pricing errors are significantly reduced across almost all test portfolios. While the conditional CAPM proves to be economically significant, it is statistically rejected at the 10 % level.

The nonparametric version of the conditional FF-3F serves as the most general asset pricing model of its kind ever considered in the Swedish equity market. It proves to be very successful in explaining the dynamic behaviour of equity returns in Sweden. More specifically, it captures the most prominent deviations of the conditional CAPM, with pricing error reductions in the range of 80 %! Furthermore, conditional alphas are economically small. This shows that the joint usage of the size and book-to-market factor is fruitful in empirical analysis.

Taken together, this study strongly reinforces the findings that nonlinearity in the first and second moment of the market return is a highly important consideration to account for in dynamic asset pricing model specifications for the Swedish equity market. In particular, we demonstrate that a nonparametric SDF is a solution that can bring valuable benefits. Last but not least, by being the first out-of-country sample test, we extend previous findings showing that functional form specifications of beta dynamics can shape and have a bearing impact on empirical results.

With these results, we hope to attract more practitioners to engage in the conditional asset pricing field in Sweden. Dynamic model specifications achieved through flexible nonparametric methodologies could provide an interesting link to other asset pricing areas by bringing further insights to the market anomalies literature, for instance through the examination of trading strategies that are not priced by current models. Wang (2003) for instance demonstrates that conditional models are far better in explaining short-term return persistence than unconditional models. Furthermore, performance evaluation represents a potentially interesting research area where nonparametric methodologies can bring value. Given the wide documentations of the time-varying behaviour of expected returns, conditional measures have become increasingly adopted in evaluating mutual fund and managed portfolio performance. In several cases, performance measures that are constructed by linear functions of state variables have been suggested[24]. Since perfor-

[24]For instance, Ferson and Schadt (1996) suggest conditional measures in performance evaluation.

mance evaluation outcomes can be highly susceptible to the choice of measure, further studies with nonparametric methods can be giving. Moreover, a detailed analysis on optimal variable selection in conditional asset pricing for the Swedish market could successfully address some of the limitations brought forward in Appendix E[25]. Finally, the consideration of other asset classes beyond stocks seems to be a very interesting research project for the Swedish market.

[25] Wang (2004) presents a methodology concerning optimal variable selection in conditional asset pricing.

7 References

J. H. . A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, 104(3):572–621, 1996.

D. A. Abdullah and S. C. Hayworth. Macroeconometrics of stock price fluctuations. *Quarterly Journal of Business and Economics*, pages 50–67, 1993.

Y. Ait-Sahalia. Nonparametric pricing of interest rate derivative securities. Technical report, National bureau of economic research, 1995.

D. W. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

D. W. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

A. Ang and D. Kristensen. Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156, 2012.

T. G. Bali and R. F. Engle. The conditional capm explains the value premium. 2014.

R. Bansal, D. A. Hsieh, and S. Viswanathan. A new approach to international arbitrage pricing. *The Journal of Finance*, 48(5):1719–1747, 1993.

R. Bansal and S. Viswanathan. No arbitrage and arbitrage pricing: A new approach. *The Journal of Finance*, 48(4):1231–1262, 1993.

J. Bartholdy and P. Peare. Estimation of expected return: Capm vs. fama and french. *International Review of Financial Analysis*, 14(4):407–427, 2005.

S. Basu. The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence. *Journal of financial economics*, 12(1):129–156, 1983.

O. Boguth, M. Carlson, A. Fisher, and M. Simutin. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics*, 102(2):363–389, 2011.

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

J. Y. Campbell and R. J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228, 1988.

D. A. Chapman and N. D. Pearson. Is the short rate drift actually nonlinear? *The Journal of Finance*, 55(1):355–388, 2000.

L. Chen, R. Novy-Marx, and L. Zhang. An alternative three-factor model. 2011.

J. H. Cochrane. *Asset Pricing:(Revised Edition)*. Princeton university press, 2009.

P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. In *Annales de l'IHP Probabilités et statistiques*, volume 30, pages 63–82. Gauthier-Villars, 1994.

P. H. Dybvig and J. E. Ingersoll Jr. Mean-variance theory in complete markets. *Journal of Business*, pages 233–251, 1982.

P. H. Dybvig and S. A. Ross. Differential information and performance measurement using a security market line. *The Journal of finance*, 40(2):383–399, 1985.

R. F. Engle. Dynamic conditional beta. *Journal of Financial Econometrics*, 14(4):643–667, 2016.

V. Eraslan. Fama and french three-factor model: evidence from istanbul stock exchange. *Business and Economics Research Journal*, 4(2):11, 2013.

E. F. Fama and K. R. French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

E. F. Fama and K. R. French. The capm is wanted, dead or alive. *The Journal of Finance*, 51(5):1947–1958, 1996.

E. F. Fama and K. R. French. Multifactor explanations of asset pricing anomalies. *The journal of finance*, 51(1):55–84, 1996.

E. F. Fama and K. R. French. Size, value, and momentum in international stock returns. *Journal of financial economics*, 105(3):457–472, 2012.

W. E. Ferson and C. R. Harvey. The variation of economic risk premiums. *Journal of Political Economy*, 99(2):385–415, 1991.

W. E. Ferson and C. R. Harvey. Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360, 1999.

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

W. E. Ferson and A. F. Siegel. Stochastic discount factor bounds with conditioning information. *The Review of Financial Studies*, 16(2):567–595, 2003.

W. E. Ferson and A. F. Siegel. Testing portfolio efficiency with conditioning information. *The Review of Financial Studies*, 22(7):2735–2758, 2009.

M. J. Flannery and A. A. Protopapadakis. Macroeconomic factors do influence aggregate stock returns. *The review of financial studies*, 15(3):751–782, 2002.

E. Ghysels. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573, 1998.

B. E. Hansen. Approximate asymptotic p values for structuras-change tests. *Journal of Business & Economic Statistics*, 15(1):60–67, 1997.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

L. P. Hansen and R. Jagannathan. Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590, 1997.

L. P. Hansen and S. F. Richard. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*, pages 587–613, 1987.

B. Hansson and P. Hordahl. Testing the conditional capm using multivariate garch-m. *Applied Financial Economics*, 8(4):377–388, 1998.

W. Härdle and T. M. Stoker. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408):986–995, 1989.

C. R. Harvey. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics*, 24(2):289–317, 1989.

C. R. Harvey. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics*, 24(2):289–317, 1989.

C. R. Harvey. The specification of conditional expectations. *Journal of Empirical Finance*, 8(5):573–637, 2001.

J. He, R. Kan, L. Ng, and C. Zhang. Tests of the relations among marketwide factors, firm-specific variables, and stock returns using a conditional asset pricing model. *The Journal of Finance*, 51(5):1891–1908, 1996.

C. Ho and C.-H. Hung. Investor sentiment as conditioning information in asset pricing. *Journal of Banking & Finance*, 33(5):892–903, 2009.

R. J. Hodrick and X. Zhang. Evaluating the specification errors of asset pricing models. *Journal of Financial Economics*, 62(2):327–376, 2001.

T. D. Huynh and D. R. Smith. Conditional asset pricing and momentum. 2014.

I. Ibragimov. Linnik yu. v. *Independent and stationary sequences of random variables*, 1971.

R. Jagannathan, G. Skoulakis, and Z. Wang. The analysis of the cross section of security returns. *Handbook of financial econometrics*, 2:73–134, 2010.

R. Jagannathan and Z. Wang. The conditional capm and the cross-section of expected returns. *The Journal of finance*, 51(1):3–53, 1996.

R. S. Koijen, H. Lustig, and S. Van Nieuwerburgh. The cross-section and time series of stock and bond returns. *Journal of Monetary Economics*, 2017.

B.-J. Lee. A heteroskedasticity test robust to conditional mean misspecification. *Econometrica: Journal of the Econometric Society*, pages 159–171, 1992.

M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849, 2001.

M. Lettau and S. Ludvigson. Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, 109(6):1238–1287, 2001.

J. Lewellen and S. Nagel. The conditional capm does not explain asset-pricing anomalies. *Journal of financial economics*, 82(2):289–314, 2006.

J. Lewellen, S. Nagel, and J. Shanken. A skeptical appraisal of asset pricing tests. *Journal of Financial economics*, 96(2):175–194, 2010.

Y. Li and L. Yang. Testing conditional factor models: A nonparametric approach. *Journal of Empirical Finance*, 18(5):972–992, 2011.

H. N. Lustig and S. G. Van Nieuwerburgh. Housing collateral, consumption insurance, and risk premia: An empirical perspective. *The Journal of Finance*, 60(3):1167–1219, 2005.

M. Malin and M. Veeraraghavan. On the robustness of the fama and french multifactor model: evidence from france, germany, and the united kingdom. *International Journal of Business and Economics*, 3(2):155, 2004.

D. L. McLeish et al. A maximal inequality and dependent strong laws. *The Annals of probability*, 3(5):829–839, 1975.

E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

L. Ng. Tests of the capm with time-varying covariances: A multivariate garch approach. *The Journal of Finance*, 46(4):1507–1521, 1991.

A. R. Pagan and G. W. Schwert. Alternative models for conditional stock volatility. *Journal of econometrics*, 45(1):267–290, 1990.

E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

R. Petkova. Do the fama–french factors proxy for innovations in predictive variables? *The Journal of Finance*, 61(2):581–612, 2006.

R. Petkova and L. Zhang. Is value riskier than growth? *Journal of Financial Economics*, 78(1):187–202, 2005.

D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.

J. L. Powell, J. H. Stock, and T. M. Stoker. Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pages 1403–1430, 1989.

M. Pritsker. Nonparametric density estimation and tests of continuous time interest rate models. *The Review of Financial Studies*, 11(3):449–487, 1998.

P. M. Robinson. Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer, 1989.

T. Santos and P. Veronesi. Vlabor income and predictable stock returnsv. *V Review of Financial Studies*, 19(1):44, 2006.

A. Savov. Asset pricing with garbage. *The Journal of Finance*, 66(1):177–201, 2011.

A. Schrimpf, M. Schröder, and R. Stehle. Evaluating conditional asset pricing models for the german stock market. 2006.

B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

D. R. Smith. Conditional coskewness and asset pricing. *Journal of Empirical Finance*, 14(1):91–119, 2007.

K. Q. Wang. Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio. *Journal of Empirical Finance*, 9(2):133–169, 2002.

K. Q. Wang. Asset pricing with conditioning information: A new test. *The journal of finance*, 58(1):161–196, 2003.

H. White and I. Domowitz. Nonlinear regression with dependent observations. *Econometrica: Journal of the Econometric Society*, pages 143–161, 1984.

A. Ziegler, M. Schröder, A. Schulz, and R. Stehle. Multifaktormodelle zur erklärung deutscher aktienrenditen: eine empirische analyse. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 59(3):355–389, 2007.

A Asymptotic results and proofs

This section in detail derives and discusses the properties of the test as put forward in Wang (2002). The reader is assumed to be familiar with high-level statistics and non-parametric econometrics. We have structured the section so as to maximize and facilitate understanding for the reader.

A.1 Assumptions

Under the following set of assumptions, we provide a large sample justification for the WLS regression approach as a testing procedure.

For a matrix of random variables $X \equiv (x_{ij})$, and any positive ρ, denote $||X||_\rho \equiv (||x_{ij}||_\rho)$ and $||x_{ij}||_\rho \equiv [E|x_{ij}|^\rho]^{\frac{1}{\rho}}$. The matrix $||X||_\rho$ is the ρ-norm of X. Denote $|X|^\rho$ with $(|x_{ij|}|^\rho)$. The notations stated are also used for the vectors, with $||x||$ as the Euclidean norm of a vector x. As aforementioned, the information set I_t stands for a σ-field which includes the σ-field generated by a data sequence $\{y_t, y_{t-1}, ...\}$. Additionally, denote

$$W_1(t,s) \equiv (r_{p,s+1}^2 - r_{p,s+1}r_{p,t+1})r_{t+1} + (r_{p,t+1}^2 - r_{p,s+1}r_{p,t+1})r_{s+1}$$

$$W_2(t,s) \equiv r_{p,s+1}^2 z_t z_t' + r_{p,t+1}^2 z_s z_s'$$

A function $\phi(x)$ is said to satisfy local Lipschitz conditions for some function $m(x)$ if

$$|\phi(x+v) - \phi(x)| < m(x)||v||$$

Denote $\Delta_{l_1,\ldots,l_j}\phi(x)$ and let it stand for the partial derivative $\frac{\delta^j \phi(x)}{\delta x_{l_1} \ldots \delta x_{l_j}}$ where x_l is the lth element of x.

Assumption 1

The first assumption pertains to the data sequence $\{y_{t+1}\}$: It is a strictly stationary β mixing process and the subvector x_t has absolutely continuous distribution with density $f(x_t)$. For $\rho > 2$ the mixing numbers β_n, $n = 1, 2, ...$, satisfy $\sum_{n=1}^{\infty} n\beta_n^{\frac{\rho-2}{\rho}} < \infty$.

Assumption 2

(a) $r^2_{p,t+1}, r_{t+1}$, and $r_{p,t+1}r_{t+1}$ have finite first moment; (b) $||W_1(t,s)||_\rho < \infty$ and $||W_2(t,s)||_\rho < \infty$ for all $t < s$; (c) $||\eta(y_{t+1})||_\rho < \infty$ and $||a(y_{t+1})||_\rho < \infty$

Assumption 3

fg_p, fg_{pp}, fg_r, fg_{pr} and fg_{zz} satisfy the local Lipschitz condition for some $m(x)$, where $m(x_t)r_{t+1}$, $m(x_t)r_{t+1}r_{p,t+1}$, $m(x_t)r_{p,t+1}$, $m(x_t)r^2_{p,t+1}$ and $m(x_t)z_tz_t'$ have finite ρ-norm.

Assumption 4

$K(u)$ is a bounded symmetric kernel function with

$$\int K(u)du = 1$$

$$\int |u|^j|K(u)|du < \infty \text{ if } 0 \leq j \leq k+1$$

$$\int u_1^{l_1}...u_k^{l_k}K(u)du = 0 \text{ if } 0 < l_1 + ... + l_k < k+1$$

where u_j denotes the element j in the vector u. In other words, $K(u)$ is a bounded symmetric kernel function with order $k+1$.

Assumption 5

(a) The j number of partial derivatives pertaining to the functions $fg_p, fg_{pp}, fg_r, fg_{pr}$ and fg_{zz} are given for all $j \leq k+1$
(b) The same applies for expectations $E[g_{pr}\nabla_{l_1,...,l_j}(fg_p)]$,$E[g_r\nabla_{l_1,...l_j}(fg_{pp})]$,$E[g_{pp}\nabla_{l_1,...l_j}(fg_r)]$, $E[g_p\nabla_{l_1,...l_j}(fg_{pr})]$,$E[g_{pp}\nabla_{l_1,...l_j}(fg_{zz})]$ where the evaluation point for the functions and their partial derivatives is x_t.

Assumption 6

Both matrices A and Γ_0 are nonsingular and are thus invertible.
The postulated condition given by Assumption 1 represents a limitation on the degree of dependence in the data and is crucially important as it allows the application of the central limit theorem. The reader should note that

processes alike the one stated in Assumption 1 that require the parameter (i.e. β_n) to decrease at the power-of-n-rate is not generally restrictive for economic data and is widely implemented (See Ait-Sahaila (1995) for an example). Furthermore, the 1st Assumption interacts with the 2nd Assumption in terms of letting a larger ρ (stronger moment restrictions) allow more dependence in the data series. This represents a trade-off and is a widely adopted basis for the establishment of asymptotic results for data that exhibits serial correlations, as discussed by White and Domowitz (1984). The Lipschitz conditions and kernel assumptions stated in both Assumption 3 and 4 are tools that have been applied in a vast amount of literature, for instance in Hardle and Stoker (1989). Assumption 5 represents the regular condition for asymptotic bias correction by using a higher order kernel.

A.2 Notations

The notations presented in this section are used to present the propositions in the following section: Define r_{t+1} as a vector of scaled excess returns $(r_{1,t+1}, ..., r_{n,t+1})' \otimes z_t$, where $\otimes$ denotes the standard Kronecker product. Let $y_{t+1} = (x_t' z_t' r_{p,t+1} r_{t+1}')'$, $w_t = f(x_t) g_{pp}(x_t)$, $A = i_n \otimes E[w_t z_t z_t']$ and $\hat{A}_N = i_n \otimes N^{-1} \sum_{t=1}^{N} \hat{w}_t z_t z_t'$, where i_n is the $n \times n$ identity matrix. Denote $\delta = (\delta_1', ..., \delta_n')'$ with $\delta_i = [E(w_t z_t z_t')]^{-1} E[w_t z_t e_{i,t+1}]$ and then define the following equations:

$$\gamma(y_{t+1}) = \eta(y_{t+1}) - [i_n \otimes a(y_{t+1})]\delta$$

$$\eta(y_{t+1}) = f(x_t)[g_{pp}(x_t) r_{t+1} - g_p(x_t) r_{p,t+1} r_{t+1} + g_r(x_t) r_{p,t+1}^2 - g_{pr}(x_t) r_{p,t+1}]$$

$$\text{where } g_r(x_t) = E(r_{t+1}|x_t) \text{ and } g_{pr}(x_t) = E(r_{p,t+1} r_{t+1}|x_t)$$

$$a(y_{t+1}) = f(x_t)[g_{pp}(x_t) z_t z_t' + r_{p,t+1}^2 g_{zz}(x_t)]$$

$$\text{where } g_{zz}(x_t) = E(z_t z_t'|x_t).$$

Both $g_p(x_t)$ and $g_{pp}(x_t)$ are defined in equation (11) and (12) in the methodology section above. As aforementioned, they are estimated with the standard Nadaraya-Watson kernel estimator (1964) as in (20) and (21), with the usage of $\hat{f}(x)$ in (19) as the weighting function. In a similar fashion:

$$\hat{g}_r(x) = \frac{1}{Nh^d}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - X_s}{h}\right) r_{i,s+1}$$

$$\hat{g}_{pr}(x) = \frac{1}{Nh^d}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - X_s}{h}\right) r_{p,s+1}r_{i,s+1}$$

$$\hat{g}_{zz}(x) = \frac{1}{Nh^d}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - X_s}{h}\right) z_s z_s'$$

In the subsection of this Appendix, a couple of interesting results necessary for stating a couple of propositions are derived: We demonstrate that the estimator $\hat{A}_N$ asymptotically converges in terms of probability to A and that the limiting distribution of $\sqrt{N}\hat{A}_N(\hat{\delta}_N - \delta)$ equals the same for that of $\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\gamma(y_{t+1})$. Last but not least, it is shown that $E\gamma(y_{t+1})$ is equal to zero. With the usage of the central limit theorem (which is feasible since $\frac{1}{N}\sum_{t=1}^{N}\gamma(y_{t+1})$ is a simple arithmetic mean of stationary random vectors), the desirable features of the test can be derived in the Propositions section.

A.3 Propositions

Wang (2002) makes two propositions that build upon an application of a central limit theorem allowed by the assumptions stated above.

Proposition 1

Given the assumptions previously stated above: If $h \to 0$, $Nh^{2k} \to \infty$ and $Nh^{2k+2} \to 0$, the weighted least squares estimator $\hat{\delta}_N$ is of the type that $\sqrt{N}(\hat{\delta_N} - \delta)$ has a limiting multivariate normal distribution with mean 0 and variance-covariance matrix Ω, where $\Omega = A^{-1}\Gamma A^{-1}, \Gamma = \sum_{-\infty}^{\infty}\Gamma_j$ and $\Gamma_j = E[\gamma(y_{t+1})\gamma(y_{t+j+1})']$.

Proposition 2

Given that Proposition 1 holds, with the equations (9) and (10): (a) if the market portfolio p demonstrates mean-variance efficiency in conditional terms, the test statistic $\hat{T}_\delta$ has a limiting chi-squared distribution with $q \times n$ degrees of freedom. (b) Moreover, $\hat{\Gamma}_j$ is a consistent estimator of Γ_j for any fixed j. The second proposition thus gives the limiting distribution of the test statistic T_δ.

Proposition 1 demonstrates that $\hat{\delta}_N$ is congruent with the standard limiting properties,$\sqrt{N}$-consistency and asymptotic normality to those of parametric estimators. The convergence conditions $Nh^{2k} \to \infty$ and $Nh^{2k+2} \to 0$ imply lower and upper bounds on the convergence rate of the bandwidth parameter h to zero and are imposed for the WLS estimator $\hat{\delta}_n$ to establish its appealing asymptotic statistical properties. Furthermore, the latter convergence condition is motivated by the usage of the higher order kernel (of order $k+1$) in accordance with Assumption 4. As such, the admissible range pertaining to the convergence rate can be successfully relaxed when allowing the usage of a kernel of order higher than the order $k+1$.

Proposition 2 contributes in the following way: The regression test is designed with a simple estimator of the variance-covariance matrix Ω. The second proposition also gives inputs (Γ_j) to obtain the consistent variance-covariance matrix estimator $\hat{\Omega}_N$ given general circumstances. The properties listed have implications for bandwidth selection. We discuss this in further detail in Appendix.

Estimation of Ω

In order to calculate an estimate of Ω, we replicate the results in Wang (2002). One should first consider the estimation of $\gamma(y_{t+1})$. As an initial step, substitute; $f(x), g_p(x), g_{pp}(x), g_r(x), g_{pr}(x)$ and $g_{zz}(x)$ by their estimates implied by the Nadaraya-Watson kernel estimators. Next, substitute δ by its weighted least squares estimator counterpart by $\hat{\delta}_N$. Taken together, an approximation for $\gamma(y_{t+1})$ is given by

$$\hat{\gamma}_N(y_{t+1}) = \hat{\eta}_N(y_{t+1}) - [i_n \otimes \hat{a}_N(y_{t+1})]\hat{\delta}_N$$

$$\hat{\eta}_N(y_{t+1}) = \hat{f}(x_t)[\hat{g}_{pp}(x_t)r_{t+1} - \hat{g}_p(x_t)r_{p,t+1}r_{t+1} + \hat{g}_r(x_t)r^2_{p,t+1} - \hat{g}_{pr}(x_t)r_{p,t+1}]$$

$$\hat{a}_N(y_{t+1}) = \hat{f}(x_t)[\hat{g}_{pp}(x_t)z_t z_t' + r_{p,t+1}^2 \hat{g}_{zz}(x_t)]$$

The reader should note that $g_{zz}(x_t) = z_t z_t'$ when z_t is a fixed transformation of x_t. For instance, in the case when $z_t = (1x_t')'$. Then, kernel estimation is not needed; By replacing $g_{zz}(x_t)$ with $z_t z_t'$, this gives $\hat{a}_N(y_{t+1}) = \hat{f}(x_t)[\hat{g}_{pp}(x_t) + r_{p,t+1}^2]z_t z_t'$. Furthermore, a consistent estimator for Γ_j is

$$\hat{\Gamma}_j = \frac{1}{N}\sum_{t=1}^{N-j} \hat{\gamma}_N(y_{t+1})\hat{\gamma}_N(y_{t+j+1})'$$

This result is further derived in the proofs section. Furthermore $\Gamma_j = 0$ for any j different from zero given eqs. (9) and (10), provided that the specification of the WLS regression equation in 26 hold. The covariance matrix estimator that we use in the test T_δ is calculated as follows

$$\hat{\Omega}_N = \hat{A}_N^{-1}\hat{\Gamma}_0\hat{A}_N^{-1}$$

A.4 Lemmas and Proofs

To prove Proposition 1 and Proposition 2, Wang (2002) applies three lemmas, which are explained below.

Yoshihara's fundamental lemma: Lemma 1

Denote $\{y_t\}$, which is a strictly stationary β mixing process with numbers $\beta_n, n = 1, 2, ...$ For any given j , $1 \leq j \leq m-1$ and $t_1 < ... < t_m$, denote $\xi_{j+1}, ..., \xi_m$ as $m-j$ random vectors that are identical with regard to the joint distribution $y_{t_{j+1}}, ..., y_{t_m}$ yet independent from $y_{t_1}, ..., y_{t_j}$. Moreover, denote $\phi(y_{t_1}, ..., y_{t_m})$ as the function

$$E[\phi(y_{t_1}, ..., y_{t_j}\xi_{j+1}, ..., \xi_m)] = 0$$

and

$$sup_{1\leq t_1<...<t_m<\infty}||\phi(y_{t_1}..., y_{t_m})||_{\rho 1} \leq M$$

for ρ_1 *and* $M > 0$*.* Then, specify that for $n_j = t_{j+1} - t_j$ that,

$$|E\phi(y_{t_1}, ..., y_{t_m})| \leq 4M\beta_{n_j}^{\frac{\rho_1 - 1}{\rho_1}}$$

As a next step, write two generalized second order U-statistics

$$U_{1N} \equiv \frac{1}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} q_N(y_{t+1}, y_{s+1})$$

$$U_{2N} \equiv \frac{1}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} \kappa_N(y_{t+1}, y_{s+1})$$

with

$$q_N(y_{t+1}, y_{s+1}) \equiv \frac{1}{h^k} K\left(\frac{x_t - x_s}{h}\right) W_1(t,s) - [i_n \otimes W_2(t,s)]\delta$$

$$\kappa_N(y_{t+1}, y_{s+1}) \equiv \frac{1}{h^k} K\left(\frac{x_t - x_s}{h}\right) W_2(t,s)$$

Both κ_N and q_N vary with N through the parameter h and exhibit symmetry. Then, in order to use the Hoeffding projection technique, the following equations below are written (where $F(y_{s+1})$ is the distribution of y_{s+1})

$$\gamma_N(y) \equiv \int q_N(y, y_{s+1}) dF(y_{s+1})$$

$$\hat{H}_{1N} \equiv E\gamma_N(y_{t+1}) + \frac{1}{N}\sum_{t=1}^{N} [\gamma_N(y_{t+1}) - E\gamma_N(y_{t+1})]$$

$$a_N(y) \equiv \int \kappa_N(y, y_{s+1}) dF(y_{s+1})$$

$$\hat{H}_{2N} \equiv \frac{1}{2} Ea_N(y_{t+1}) + \frac{1}{N}\sum_{t=1}^{N} [a_N(y_{t+1}) - Ea_N(y_{t+1})]$$

Hoeffding Composition: Lemma 2

Denote $\{y_t\}$, which is a strictly stationary β mixing process with numbers $\beta_n, n = 1, 2, ...$ that satisfies

$$\sum_{n=1}^{\infty} n\beta_n^{(\rho-2)/\rho} < \infty$$

for $\rho > 2$. Given that $B_N = o(\sqrt{N})$ so that $||\gamma_N(y_{t+1})||_\rho \leq B_N$ and

$$sup_{1\leq t<s<\infty}||q_N(y_{t+1}, y_{s+1})||_\rho \leq B_N$$

then

$$NE[(U_{1N} - H_{1N})(U_{1N} - H_{1N})'] = o(1)$$

Lemma 3

Given that the aforementioned assumptions 1-5 hold. Denote

$$\varepsilon_{1N}(y_{t+1}) \equiv \gamma_N(y_{t+1}) - \gamma(y_{t+1})$$

$$\varepsilon_{2N}(y_{t+1}) \equiv a_N(y_{t+1}) - a(y_{t+1})$$

(a) Given the condition $h \to 0$, then $||\varepsilon_{1N}(y_{t+1})||_\rho \leq b_N$ for $b_N = o(1)$ and additionally

$$\frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\varepsilon_{1N}(y_{t+1}) - E\varepsilon_{1N}(y_{t+1})] = o_p(1)$$

This same result also is the same for $\varepsilon_{2N}(y_{t+1})$
(b) Given the convergence condition $Nh^{2k+2} \to 0$, then $\sqrt{N}E(H_{1N}) = o(1)$

Proof 1: Proposition 1

In the weighted least squares (WLS) estimator definition, it is implicit that

$$\hat{A}_N(\hat{\delta}_N - \delta) \equiv \left(1 - \frac{1}{N}\right) U_{1N}$$

$$\hat{A}_N = \left(1 - \frac{1}{N}\right) i_n \otimes U_{2N}$$

From the assumptions 2 and 4, it follows that an upper bound of M is in place so that

$$E\left|K\left(\frac{x_t - x_s}{h}\right) W(t,s)\right|^{\rho} \leq M$$

with $W(t,s) = W_1(t,s) - [i_n \otimes W_2(t,s)]\delta$. Thus, it can be written that

$$E|q_N(y_{t+1}, y_{s+1})|^{\rho} = \frac{1}{h^{pk}} E\left|K\left(\frac{x_t - x_s}{h}\right) W(t,s)\right|^{\rho} \leq \frac{M}{h^{pk}}$$

Given this equation, it is possible to derive the equation below (because $Nh^{2k} \to \infty$)

$$||q_N(y_{t+1}, y_{s+1})||_{\rho}^2 \leq h^{-2k} M^{2/\rho} = O(h^{-2k}) = O\left(\frac{N}{Nh^{2k}}\right) = o(N)$$

From the 2nd assumption stated in the section above combined with the triangular inequality and the 3rd Lemma

$$||\gamma_N(y_{t+1})||_{\rho} \leq ||\gamma(y_{t+1})||_{\rho} + ||e_N(y_{t+1})||_{\rho}$$

It is possible to see that $||\gamma_N(y_{t+1})||_{\rho}$ is consistent with an upper bound that is constant. Taken together, it is possible to note that the conditions implied and pertaining to Lemma 2 are satisfied. Namely, Lemma 2 suggests that

$$\sqrt{N}(U_{1N} - EH_{1N}) = \sqrt{N}(H_{1N} - EH_{1N}) + o_p(1) \tag{45}$$

Reiterating the above arguments for both κ_N and a_N to pursue an application of the second Lemma to U_{2N}, one gets the following result

$$U_{2N} = H_{2N} + o_p(1)$$

Then by the presented "Lemma 3 (a)", it can be shown that:

$$H_{2N} = \frac{1}{2} Ea(y_{t+1}) + \frac{1}{N}\sum_{t=1}^{N}[a(y_{t+1}) - Ea(y_{t+1})] + o_p(1) = \frac{1}{2} Ea(y_{t+1})) + o_p(1)$$

Since $Ea(y_{t+1})$ equals $2E(w_t z_t z_t')$, it follows that $\hat{A}_N \to^p A$

The same Lemma 3 (a) can also be used to demonstrate that

$$\sqrt{N}(H_{1N}-EH_{1N}) \equiv \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\gamma_N(y_{t+1})-E\gamma_N(y_{t+1})] = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\gamma(y_{t+1})-E\gamma(y_{t+1})]+o_p(1) \tag{46}$$

It follows from "Lemma 3(b)" as well as eqs. (45) and (46)

$$\sqrt{N}U_{1N} = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\gamma(z_{t+1}) - E\gamma(z_{t+1})] + o_p(1)$$

The reader can now note that $E\gamma(y_{t+1}) = 0$ since a simple applications from the Law of Iterated Expectations implies

$$E\eta(y_{t+1}) = 2E(w_t e_{t+1})$$

$$Ea(y_{t+1}) = 2E(w_t z_t z_t')$$

where the series $e_{t+1} = (e_{1,t+1}, ..., e_{n,t+1})' \otimes z_t$ and $\otimes$ denotes the Kronecker product.

Hence, it can be demonstrated that

$$E\eta(y_{t+1}) - [i_n \otimes Ea(y_{t+1})]\delta = 0$$

Then, from improvements of the standard central limit theorems of Ibragimov and Linnik (1971) pertaining to Theorem 18.5.3, Doukhan et al. (1994) have provided evidence that if $2 < p < \infty$, $E|X|^p < \infty$ and $\sum_{n=1}^{\infty} \alpha_n n^{2/(p-2)} < \infty$, then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - EX_i)$ converges to a centered normal random vector. Because: $\sum_{n=1}^{\infty} n\beta_n^{\rho-2/\rho} < \infty$ implies $\sum_{n=1}^{\infty} n\beta_n^{\rho/\rho-2} < \infty$,both assumptions 1 and 2 imply central limit theorem conditions that are also sufficient for Γ to be finite. Using this, it can be demonstrated that

$$\frac{1}{\sqrt{N}}\sum_{t+1}^{N}\gamma(y_{t+1}) \to^d \mathcal{N}(0, \Gamma)$$

where

$$\Gamma = \sum_{-\infty}^{\infty} E\gamma(y_t)\gamma(y_{t+j})'.$$

Hence, with these equations the proof is demonstrated. Q.E.D

Proof 2: Proposition 2

Proof of part (b)
To begin with, denote

$$\tilde{\gamma}_N(y_{t+1}) = \hat{\eta}_N(y_{t+1}) - [i_n \otimes \hat{a}_N(y_{t+1})]\delta$$

Because of the convergence condition pertaining to the WLS $\hat{\delta}_N \to^p \delta$, it is sufficient to demonstrate

$$\frac{1}{N}\sum_{1}^{N-j} \tilde{\gamma}_N(y_{t+1})\tilde{\gamma}_N(y_{t+j+1})' \to^p \Gamma_j$$

where $\tilde{\gamma}_N(y_{t+1})$ can be written as

$$\tilde{\gamma}_N(y_{t+1}) = \frac{1}{N}\sum_{s=1}^{N} q_N(y_{t+1}, y_{s+1})$$

Hence,

$$\tilde{\gamma}_N(y_{t+1}) - \gamma_N(y_{t+1}) = \frac{1}{N}\sum_{s=1}^{N} J_N(t,s)$$

where

$$J_N(t,s) \equiv q_N(y_{t+1}, y_{s+1}) - \gamma_N(y_{t+1})$$

As has been previously demonstrated: $||q_N(y_{t+1}, y_{s+1})||_\rho$ has an upper bound in place of the order $o(\sqrt{N})$. Also: $||\gamma_N(y_{t+1})||_\rho$ has a constant upper bound. It follows from "Minkowski's inequality" that $||J_N(t,s)||_\rho$ also has an upper bound of the same order $o(\sqrt{N})$. As such, $M_N = o(N)$ is such that

$$||J_N(t,s_1)J_N(t,s_2)'||_{\rho/2} \leq ||J_N(t,s_1)||_\rho ||J_N(t,s_2)'||_\rho \leq M_N$$

Additonally, note that for: $s_1 < s_2$

$$\int J_N(t, s_1) J_N(t, s_2)' dF(z_{j+1}) = 0$$

with $j = s_1$ if $s_1 < t$, else $j = s_2$.

With the usage of the first Lemma (Yoshihara's Lemma), one can derive that:

$$\begin{aligned} N^2 E\{[\tilde{\gamma}_N(y_{t+1}) - \gamma_N(y_{t+1})][\tilde{\gamma}_N(y_{t+1})\gamma_N(y_{t+1}]'\} \\ \leq \sum_{s_1=1}^{N} |EJ_N(t, s_1) J_N(t, s_1)'| + 2 \sum_{1 \leq s_1 < s_2 \leq N} |EJ_N(t, s_1) J_N(t, s_2)'| \\ \leq NM_N + 8NM_N \sum_{n=1}^{N-1} \beta_n^{\rho-2/\rho} = o(N^2) \end{aligned}$$

So $b_{1,N} = o(1)$ such that $E|\tilde{\gamma}_N(y_{t+1}) - \gamma_N(y_{t+1})|^2 \leq b_{1,N}$. Taking together with the 3rd Lemma, it is possible to show that $E|\tilde{\gamma}_N(y_{t+1}) - \gamma_N(y_{t+1})|^2 \leq b_{2,N}$ for $b_{2,N} = o(1)$. Because of constant upper bounds for $E|\tilde{\gamma}_N(y_{t+1})|^2$ and $\mathrm{E}|\gamma_N(y_{t+1})|^2$, it is implied that for any fixed j, the $b_{3,N} = o(1)$ is such that

$$E|\tilde{\gamma}_N(y_{t+1})\hat{\gamma}_N(y_{t+j+1})' - \gamma(y_{t+1})\gamma(y_{t+j+1})'| \leq b_{3,N}$$

Using the Chebyshev's inequality in representation, it gives rise to

$$\frac{1}{N}\sum_{t+1}^{N} \tilde{\gamma}_N(y_{t+1})\tilde{\gamma}_N(y_{t+j+1})' - \frac{1}{N}\sum_{t=1}^{N-j} \gamma(y_{t+1})\gamma(y_{t+j+1})' = o_p(1)$$

Noting the following: $\beta_n = o(n^{-2\rho/(\rho-2)})$. This is because:

$$\sum_{j=n}^{2n} j\beta_j^{(\rho-2)/\rho} \geq \beta_{2n}^{(\rho-2)/\rho} \sum_{j=n}^{2n} j = \beta_{2n}^{(\rho-2)/\rho} O(n^2)$$

and

$$\sum_{j=n}^{2n} j\beta_j^{(\rho-2)/\rho} = o(1)$$

by Assumption 1 above. Hence, large mixing coefficients of the data satisfies

$\alpha_n = o(n^{-r/(r-2)})$ for $2 < r < \rho$. Moreover, from the second assumption, $||\gamma(y_{t+1})\gamma(y_{t+j+1})'||_{\rho/2} < \infty$. From Lemma 2.1, Theorem 2.3 and 2.1 by White and Domowitz (or the theorem 2.10 of McLeish (1975)), thus gives rise to

$$\frac{1}{N}\sum_{t=1} N - j\gamma(y_{t+1})\gamma(y_{t+j+1})' = E[\gamma(y_{t+1})\gamma(y_{t+j+1})'] + o_p(1)$$

The above derived equation establishes the theorem that the estimator $\hat{\Gamma}_j$ is consistent for Γ, i.e. part (b). Q.E.D

Proof of part (a):

Given the equations (9) and (10)

$$E[\eta(y_{t+1})|I_t] = E(w_t e_{t+1}|I_t) + E(w_t e_{t+1}|x_t)$$

$$E[a(y_{t+1})|I_t] = w_t z_t z_t' + w_t g_{zz}(x_t)$$

where (as before) the series $e_{t+1} = (e_{1,t+1}, ..., e_{n,t+1})' \otimes z_t$. As aforementioned, the equation (26) leads to

$$E(w_t e_{t+1}|I_t) - [i_n \otimes (w_t z_t z_t')]\delta = 0$$

$$E(w_t e_{t+1}|x_t) - [i_n \otimes (w_t g_{zz}(x_t)]\delta = 0$$

Hence, given the regression specification (21)

$$E[\gamma(y_{t+1})|I_t] = 0$$

Then, $\Gamma_j = 0$ for $j \neq 0$. Given the conditional expectations (9) and (10), the null hypothesis of conditional mean variance efficiency implies (26).

So the convergence: $\hat{\Omega}_N \to^p \Omega$. With $\delta = 0$: Proposition 1 thus delivers the following distributional asymptotic result:

$$N\hat{\delta}_N'\hat{\Omega}_N^{-1}\hat{\delta}_N \to^d \chi^2(qn)$$

Q.E.D

B SupF Test for Structural Change

This section gives a brief overview of the supF test for parameter instability and structural change[26]. The standard linear regression model upon which the statistical test is based can be formulated as follows:

$$y_i = x_i'\beta_i + u_i \quad (i = 1, ..., n)$$

where at time i, $x_i = (1, x_{i2}, ..., x_{ik})'$ denotes a $k \times 1$ vector of independent variable observations, with the first component equal to unity, and the terms u_i are i.i.d $(0, \sigma^2)$. Furthermore, β_i is a $k \times 1$ regression coefficient vector and y_i is the dependent variable observation. Tests on structural change are concerned with testing the null hypothesis of "no structural change"

$$H_0 : \beta_i = \beta_0 \text{ for } (i = 1, ..., n)$$

The alternative is that the vector of regression coefficients varies over time

$$\begin{aligned} H_1 : \beta_i = \beta_A & \text{ for } (1 \leq i \leq i_0) \\ \beta_B & \text{ for } (i_0 < i \leq n) \end{aligned}$$

where i_0 is some change point in the interval $(k, n-k)$. The testing idea is to fit two separate regressions; a full model and a restricted model. This is done for two subsamples as defined by i_0. Then the econometrician rejects whenever the following statistic is too large:

$$F_{i0} = \frac{\hat{u}'\hat{u} - \hat{e}'\hat{e}}{\hat{e}'\hat{e}/(n-2k)}$$

Here, $\hat{e} = (\hat{u}_A, \hat{u}_B)'$ denotes the residuals from the full model, where the regression coefficients in the subsamples are estimated separately. Similarly, $\hat{u}$ are the residuals from the restricted model, where the parameters are just fitted once for all observations. The statistic F_{i0} has an asymptotic chi-squared distribution with k degrees of freedom. Furthermore, given normality

[26]As background information, the most important classes of tests of structural change are generally either tests from the generalized fluctuation test framework of Kuan and Hornik (1995) as well as tests on based on F statistics.

F_{i0}/k has an exact F distribution with k and $n-2k$ degrees of freedom. In order to test for structural change across every possible change point i_0, the econometrician has to calculate F statistics for all potential change points within a specified interval $[\underline{i}, \bar{i}]$[27]. As such, the procedure starts out with calculating F statistics F_i for $k < \underline{i} \leq i \leq \bar{i} < n-k$. The next step is to aggregate the series of F statistics into a single test statistic. Andrews (1993) as well as Andrews and Ploberger (1994) propose to use[28]

$$\sup F = \sup_{\underline{i} \leq i \leq \bar{i}} F_i$$

The p-values, upon which statistical inference about linearity is made, are calculated according to Hansen (1997).

C The Stationary Bootstrap

In this section, the application of the stationary bootstrap pertaining to the fully non-parametric FF-3F model is described.

Denote the data $\{y_{t+1}\}$ for $t = 1, ..., N$. The methodology makes use of the resampled time series $\{y^*_{t+1,j}\}$ for $t = 1, ..., N$ where j indexes one of the N_b bootstrapped samples. With the resampled data, one can obtain resampled counterparts of both $\hat{\delta}_N$ and $\hat{\Omega}_N$ through $\hat{\delta}^*_{N,j}$ as well as $\hat{\Omega}^*_{N,j}$, respectively. The following test statistics are used as the tools for statistical inference related to T_δ.

$$T^*_{\delta,j} = N(\hat{\delta}^*_{N,j} - \hat{\delta}_N)'(\hat{\Omega}^*_{N,j})^{-1}(\hat{\delta}^*_{N,j} - \hat{\delta}_N)' \tag{47}$$

Across each bootstrapped sample $j = 1..., N_b$, the data $\{y^*_{t+1,j}\}$ follows

$$y^*_{t+1,j} = y_{\xi_j(t)+1} \tag{48}$$

[27] Andrews (1993) proposes several appropriate interval specifications depending on several contextual factors. In our applications, we use the following interval; [0.15;0.85], which is in line with Andrews(1993).

[28] The authors also provide two other alternative test statistics: ave$F = \frac{1}{\bar{i}-\underline{i}+1} \sum_{i=\underline{i}}^{\bar{i}} F_i$ and exp$F = \frac{1}{\bar{i}-\underline{i}+1} \sum_{i=\underline{i}}^{\bar{i}} \exp(0.5 \cdot F_i)$, respectively. The test statistics have certain optimality properties.

for each $t = 1, ..., N$ with $\xi_j(t)$ denoting a randomly chosen index. This index is chosen by the stationary bootstrap algorithm of Politis and Romano (1994). When facing this implementation, the user needs to set a "smoothing parameter" $q = q_N$, such that $0 < q_N \leq 1$, $q_N \to 0$, and $Nq_N \to \infty$ as $N \to \infty$. As a general rule of thumb, a small value of the parameter q is particularly suitable when the data exhibits strong dependence. On the other hand, a large value of q is appropriate given limited amount of dependence demonstrated in the data. We set the parameter $q = 0.10$ as in Wang (2003), which corresponds to an average block length of 10.

As a next step, the user proceeds as follows:

1. Draw an independently distributed random variable $\xi_j(1)$ for $t = 1$, which is uniformly distributed over $\{1, ..., N\}$
2. As a next step, increase t incrementally by 1. If $t > N$ stop the procedure. Else, draw a standard uniform and independently uniform random variable u.
   - If $u < q$, draw $\xi_j(t)$ as a random variable that is independently distributed as well as uniformly distributed over $\{1, ..., N\}$
   - If $u \geq q$ set $\xi_j(t) = \xi_j(t-1) + 1$. If $\xi_j(t) > N$, set $\xi_j(t) = 1$
3. Repeat step 2.

D Bandwidth and Kernel Selection

In this appendix section, we describe and motivate our choice of kernel function and smoothing parameter based on their respective properties. We also explain the bandwidth selection problem and the inherent trade-offs pertaining to it. With regard to the choice of kernel function, we limit our choice to an independent multivariate normal density function (hereafter multivariate Gaussian Kernel)

$$K(x) = \prod_{i=1}^{k} \phi_i(x_i) \tag{49}$$

The choice of this kernel is due mathematical convenience and its wide popularity: ϕ_i is the density of a univariate normal with mean zero and variance σ_i^2. Furthermore, we focus on the independent multivariate normal density function because we want to maximize comparability with the empirical results of Wang (2003). In computation, σ^2 is the standard deviation of the ith state variable and is replaced with the sample estimate.[29].

As has been quickly mentioned before, optimal bandwidth selection remains a controversial topic within the nonparametric econometric literature. Yet, there are in general some selection approaches that are more well-suited given certain circumstances. We set the bandwidth as

$$h = cN^{\frac{-1}{(2k+1)}} \tag{50}$$

As such, a few remarks should me made: To begin with, the limiting distribution of the WLS $\hat{\delta}_N$ is not dependent on the scaling constant c. Such a desirable feature implies that in contrast to widely used kernel density regressions and KDEs, the estimate $\hat{\delta}_N$ becomes less susceptible to the choice of the scaling constant as the number of observations increases. That is to say, the bandwidth choice does not become as crucially important as in kernel density and kernel regression estimation. This property further suggests that bandwidth issues due to persistence of conditioning variables may not serve as a prominent challenge to the testing procedure, as discussed by Pritsker (1998) as well as Chapman and Pearson (2000). However, this attractive trait comes at a cost; the trade-off is that justification of cross-validation procedures becomes difficult.

In similarity to Wang (2003), we set the scaling constant c to 1. This is a practical choice and is widely regarded as an objective starting point previously implemented by many authors (see Silvermann (1986), Pagan and Schwert (1990) and Harvey (2001) etc.)

Moreover, regardless of how many conditioning variables are used, the WLS $\hat{\delta}_N$ still has parametric convergence rate and standard limiting distri-

[29]Wang (2003) also makes use of a bias-corrected higher order kernel $K^*(x) = \frac{K(x) - \sum_{j=1}^{k} a_j b_j^{-k} K(x/b_j)}{1 - \sum_{j=1}^{k} a_j}$ in accordance with Powell et al. (1989), where $a = B^{-1}e$ for $a = (a_i, ..., a_k)'$ and B is a $k \times k$ matrix. Each component $B_{ij} = b_j^i, e$ is a $k \times 1$ vector of ones and $b_j = k + j$ for $j = 1..., k$. However, Wang (2003) finds that the higher order kernel produces more variable estimates than the Gaussian kernel and that with regard to reducing the bias, it is not satisfactory in a finite sample context. Therefore, we follow Wang (2003) and only focus on the Gaussian kernel.

bution. On the other hand, the reader should be aware of that the user has to pay a price for this appealing property that is related to the variance-covariance matrix of the WLS $\hat{\delta}_N$: In finite samples contexts, it still is advantageous to exclude redundant conditioning variables.

E Robustness and Limitations

We consider a range of different tests to verify the robustness of our results. To begin with, the econometrician can choose among a wide range of different kernels in applications. A very popular choice is the Epanechnikov kernel, which often is the most efficient in a MSE (mean square error) sense. We reconsider our results with this kernel. Furthermore, we change the bandwidth slightly. The results are not reported for brevity, but are available upon request. Similarly, we respecify the sample by dividing it in two parts. However, since the motivation for nonparametric methods are asymptotic, we take any such measurements with careful consideration. Additionally, we challenge the conditional asset pricing models by equally-weighted size/BE-ME portfolios in addition to the value-weighted versions. Li and Yang (2011) suggest that this might have a bearing on empirical outcomes. Last but not least, we conduct the tests with the original six size/BE-ME of FF to mitigate any concerns that the Swedish equity market has a smaller amount of firm observations than in the US.

As this study aims to introduce the notion of nonparametric SDF in asset pricing with conditioning information in the Swedish equity market, a brief elaboration on any limitations is worthwhile. In this study, we consider five different conditioning variables, whereof three serve as the main inputs in our empirical tests and the rest are considered from a robustness perspective. It could be argued that another set of conditioning instruments would lead to significantly different conclusions. Furthermore, there are several limitations pertaining to the choice of the sampling period: Whereas Wang (2002, 2003) considers the U.S equity market over a longer sample horizon, we consider the significantly smaller equity market of Sweden. This implies that we have a much lower number of observations included in each test portfolio, which can have an impact on the empirical results. Additionally, as the motivation of nonparametric methods rely on asymptotic distribution theory to ensure good statistical power, the collection of a large number of observations is crucial. Last but not least, empirical findings on the performance of asset pricing models have been shown to be highly susceptible to the choice of sample period. As we investigate the Swedish market from July 1994 to December 2016, it could be argued that a more comprehensively specified horizon would be fruitful.