Nonparametric Specification Testing of Conditional Asset Pricing Models

Francisco Peñaranda,*
Queens College CUNY, 65-30 Kissena Blvd., Flushing, NY 11367, US.
< francisco.penaranda@qc.cuny.edu >

Juan M. Rodríguez-Poo
Department of Economics, University of Cantabria
and IEE, University of Genève, Avda. de los Castros s/n, 39005 Santander, Spain.
< juan.rodriguez@unican.es >

Stefan Sperlich
IEE, University of Genève, Bd du Pont d’Arve 40, CH-1211 Geneva, Switzerland.
< stefan.sperlich@unige.ch >

June 1, 2018

Abstract

This paper presents an adaptive omnibus specification test of conditional asset pricing models. These models provide constraints that conditional moments of returns and pricing factors must satisfy, but most of them do not provide information on the functional form of those moments. Our test is robust to functional form misspecification, and also detects any relationship between pricing errors and conditioning variables. We give special emphasis to the test implementation and calibration. We find a conditional counterpart of a well-known empirical problem of unconditional models. The lack of rejection of consumption based conditional models seems to be due to the poor conditional correlation between consumption and the cross-section of stock returns.

Keywords: Adaptive testing, Bootstrap methods, Conditional Asset Pricing Model, Specification test.

JEL Codes: G12, C12, C14

*We would like to thank participants at the 10th Conference on Computational and Financial Econometrics (Sevilla) and the New Methods for the Empirical Analysis of Financial Markets (Comillas) for helpful comments and suggestions. Of course, the usual caveat applies. The authors gratefully acknowledge financial support from the Research Project APIE 1/2015-17: “New methods for the empirical analysis of financial markets” of the Santander Financial Institute (SANFI) of UCEIF Foundation resolved by the University of Cantabria and funded with sponsorship from Banco Santander.
1 Introduction

Asset pricing theory must consider that asset returns and pricing factors are predictable, in the sense of a significant time-variation in their joint conditional distribution. For this purpose, conditional asset pricing models provide constraints that conditional moments of returns and pricing factors must satisfy. On the other hand, most of these models do not provide much information on the functional form of those conditional moments.

Our paper presents an adaptive omnibus specification test for conditional asset pricing models. Importantly, our test is robust to functional form misspecification of both conditional moments and prices of risk, and has power against any relationship between pricing errors and conditioning variables.

There is a vast literature that uses parametric methods to evaluate conditional asset pricing models, and concludes that conditioning information substantially improves the empirical performance of asset pricing models. For instance, Jagannathan and Wang (1996) find that the conditional CAPM can explain the cross-section of stock returns even though the static CAPM cannot, while Lettau and Ludvigson (2001b) find that the value premium can be explained by a CCAPM with a time-varying price of risk. However, other authors suggest that this superior performance may be an illusion caused by low statistical power of standard asset pricing tests, see e.g. Lewellen and Nagel. (2006) and Nagel and Singleton (2011).

Along these lines, there have been significant contributions that avoid specifying the conditional distribution of returns and factors by using nonparametric techniques. Papers such as Nagel and Singleton (2011) estimate nonparametrically first and second conditional moments, but work with asset pricing models where the prices of risk are parametric. Nevertheless, Wang (2003) and Roussanov (2014) consider nonparametric prices of risk.

Wang (2003) uses nonparametric methods to evaluate conditional variants of the CAPM and Fama-French model. He uses the stochastic discount factor approach to estimate pricing errors nonparametrically, and test if pricing errors are independent of the set of conditioning variables. This methodology is easy to implement, and he develops the corresponding asymptotic theory, but unfortunately this test has zero power against pricing errors that exhibit nonlinear dependence with respect to the set of conditioning variables. On the contrary, our test is general enough to detect nonparametric alternatives from independence.

Roussanov (2014) estimates nonparametric prices of risk by minimizing a quadratic form of conditional pricing errors, with a focus on consumption based models. This estimation idea is very interesting and nests the prices of risk in Wang (2003). However, Roussanov (2014) does not develop a formal test of the asset pricing model. The models are evaluated by means of tests of a particular zero average pricing error, and figures of pricing errors with point-wise bootstrap bands. We develop the asymptotic theory of the test. Moreover, our test considers jointly all the pricing errors, and does not focus on unconditional means.

In Econometrics and Statistics, a plead of consistent specification tests are available to test a
fully parametric model against a broad set of alternatives. Following Hart (1997), mainly two fundamental approaches have been considered: statistics that appear as a weighted average of the residuals, and statistics that compare parametric and non-parametric fits. Our test statistic falls within this second class of tests although it exhibits a crucial difference with respect to the standard ones in the literature. The model under the null is also nonparametric and therefore some additional care needs to be taken into account with respect to the asymptotic bias of the test (see Rodriguez-Poo et al. (2015)). We obtain the asymptotic level of the test, and we also show that the test is consistent against any nonparametric form of dependence between pricing errors and conditioning variables.

We give special emphasis to the implementation and calibration of our test. We study practical issues such as bias reduction, adaptive bandwidth choice, and finite sample performance, including the resampling approximations. Our test works reasonably well even for small samples in a Monte Carlo exercise.

In our empirical application, we evaluate several asset pricing models against the cross-section of market, size, and value premia. We work with two prominent predictors, the default spread and the cointegrating residual of consumption and wealth. We find strong empirical evidence against the conditional CAPM, where the market return is the only relevant pricing factor. However, we do not find evidence against zero pricing errors when we test the CCAPM with consumption growth as the only pricing factor. In fact, in the case of a CCAPM with Epstein-Zin preferences, where both pricing factors are considered, the market price of risk seems to be zero.

Importantly, the low pricing errors with the consumption factor do not necessarily mean that this is an economically meaningful pricing factor. The lack of rejection seems to be due to the poor conditional correlation between consumption and the cross-section of stock returns. We find a conditional counterpart of a well-known problem with unconditional asset pricing models. Kan and Zhang (1999) among others argue that some empirical asset pricing models rely on useless factors with a zero unconditional covariance with returns.

The rest of the paper is organised as follows. Section 2 reviews conditional asset pricing models. Section 3 develops our omnibus specification test of conditional asset pricing models, and the corresponding asymptotic behaviour. Section 4 describes the implementation and calibration of the test, jointly with a Monte Carlo exercise, while Section 5 reports our empirical application. Finally, Section 6 concludes. Proofs and auxiliary results are relegated to appendices.

## 2 Conditional Asset Pricing Models

In this section we represent a conditional asset pricing model by its stochastic discount factor (SDF). The simplicity of the SDF to present a general theory of asset pricing is well recognised (see Cochrane (2001)).
2.1 SDF Pricing Conditions

The investment set is given by \( N \) assets, with a vector of excess returns \( r_{t+1} \) that is known at time \( t+1 \). The information set known at \( t \) is given by a \( d \times 1 \) vector \( z_t \) of conditioning variables. Standard arguments such as lack of arbitrage opportunities, or the first order conditions of a representative investor, imply that

\[
E (m_{t+1} r_{t+1} \mid z_t) = 0
\]

for some random variable \( m_{t+1} \) called SDF, which discounts uncertain payoffs in such a way that their expected discounted value equals their cost. The standard approach in empirical finance is to model the SDF as an affine transformation of some \( K < N \) observable risk factors \( f_{t+1} \). There are two common variants to express such an SDF, the uncentred SDF variant

\[
m_{t+1} = 1 - f_{t+1}' \delta (z_t),
\]

and the centred SDF variant,

\[
m_{t+1} = 1 - (f_{t+1} - \nu (z_t))' \tau (z_t),
\]

which demeans the factors with

\[
\nu (z_t) = E (f_{t+1} \mid z_t)
\]

to have a unit conditional mean. The vectors \( \delta (z_t) \) and \( \tau (z_t) \) represent the prices of risk.

For instance, Wang (2003) uses the uncentred SDF variant, while Roussanov (2014) uses the centred SDF variant. In the following, we focus on the uncentred SDF for its simplicity and robustness. Nevertheless, we develop the centred SDF in Appendix B, and show how to translate results from one variant to another. Peñaranda and Sentana (2015) provide a detailed explanation of these approaches for parametric unconditional asset pricing models.

2.2 Choice of Prices of Risk

Let us define the vector of risk premia

\[
\mu (z_t) = E (r_{t+1} \mid z_t)
\]

and the risks or factor sensitivities

\[
D (z_t) = E (r_{t+1} f_{t+1}' \mid z_t).
\]

\footnote{See Section 2.4 for an empirically relevant example where the uncentred SDF exists, but the centred SDF does not.}
We can express the model pricing errors as

\[ e (z_t; \delta (z_t)) = E \left[ (1 - f_{t+1}' \delta (z_t)) r_{t+1} \mid z_t \right] = \mu (z_t) - D (z_t) \delta (z_t), \]  

(7)

for some vector \( \delta (z_t) \) of prices of risk.

Roussanov (2014) pins down the prices of risk by minimising a quadratic form of the pricing errors. Similarly, the particular vector \( \delta (z_t) \) can be chosen by minimising the following quadratic form

\[ q (z_t; \delta (z_t)) = e (z_t; \delta (z_t))' W (z_t) e (z_t; \delta (z_t)) \]  

(8)

for some weighting matrix \( W (z_t) \) that may depend on \( z_t \), but not on \( \delta (z_t) \). We will comment two common choices below.

In fact, this choice of prices of risk could be interpreted as a two-pass procedure where first the conditional moments from time series data are estimated, and second the prices of risk are obtained from cross-sectional data. In particular, the prices of risk are obtained from a weighted least-squares projection of the vector of risk premia onto the span of the factor sensitivities.

Let us denote \( \delta^* (z_t) \) the prices of risk that minimise the criterion \( q (z_t; \delta (z_t)) \). The first order conditions that pin down these prices of risk are

\[ D (z_t)' W (z_t) e (z_t; \delta^* (z_t)) = 0, \]  

(9)

and we obtain

\[ \delta^* (z_t) = (D (z_t)' W (z_t) D (z_t))^{-1} D (z_t)' W (z_t) \mu (z_t). \]  

(10)

### 2.3 Common Weighting Matrices

The chosen weighting matrix is irrelevant under the null hypothesis, when there is a vector of risk prices that makes the pricing errors equal to zero. However, under the alternative hypothesis different weighting matrices pin down different prices of risk, and hence different pricing errors.

Two common choices of \( W (z_t) \) are the identity matrix and

\[ W (z_t) = E^{-1} (r_{t+1} r_{t+1}' \mid z_t). \]  

(11)

We use the latter in our empirical application because then the criterion (8) represents a conditional counterpart of the Hansen-Jagannathan distance (see Hansen and Jagannathan (1997)), which is widely used in empirical finance. Moreover, the first order conditions (9) can be interpreted as pricing the uncentred factor mimicking portfolios, and the prices of risk (10) as the inverse of their second moment matrix times their mean vector.

Importantly, if there are traded factors that are also part of the vector of returns, like the market portfolio return in the CAPM, then these particular returns will be priced exactly. In fact, Wang (2003) proposes the inverse of the second moment matrix of the traded factors times their mean
vector as the prices of risk for models where all the factors are traded. Given our previous comments, such prices of risk can be justified by the weighting matrix (11).

2.4 Special Case: Uncorrelated SDF

Finally, this section describes a special case that is empirically relevant. For this purpose, let us rewrite the pricing conditions (1) as

\[ E(r_{t+1} | z_t) E(m_{t+1} | z_t) + \text{Cov}(r_{t+1}, m_{t+1} | z_t) = 0, \]

or equivalently

\[ E(r_{t+1} | z_t) (1 - \nu(z_t)' \delta(z_t)) + \text{Cov}(r_{t+1}, f_{t+1} | z_t) \delta(z_t) = 0 \]

for an uncentred SDF (2).

If the matrix \( \text{Cov}(r_{t+1}, f_{t+1} | z_t) \) does not have full column rank, then there is a linear combination \( \delta(z_t) \) of the columns of this matrix such that \( \text{Cov}(r_{t+1}, f_{t+1} | z_t) \delta(z_t) = 0 \), or equivalently \( \text{Cov}(r_{t+1}, m_{t+1} | z_t) = 0 \). That linear combination can be chosen to satisfy \( 1 - \nu(z_t)' \delta(z_t) = 0 \), or equivalently \( E(m_{t+1} | z_t) = 0 \), and hence construct an SDF \( m_{t+1} = 1 - f_{t+1}' \delta(z_t) \) that satisfies the pricing conditions.\(^2\) For instance, if factor \( i \) has zero conditional correlation with the returns, then a vector \( \delta(z_t) \) with zero entries for all factors but factor \( i \), and \( \delta_i(z_t) = 1/\nu_i(z_t) \), satisfies the pricing conditions.

However, such an SDF is not meaningful from an economic perspective. It does not explain the cross-section of returns, in the sense of why some entries in \( E(r_{t+1} | z_t) \) are higher than others. Instead, it simply exploits a rank failure in \( \text{Cov}(r_{t+1}, f_{t+1} | z_t) \). Moreover, \( E(m_{t+1} | z_t) = 0 \) is neither appealing from an economic perspective because such an SDF would allow arbitrage opportunities. Therefore, in the empirical evaluation of an asset pricing model, it is important to check if an SDF that yields small pricing errors also yields a small SDF mean. That would mean that the SDF is actually uncorrelated with the cross-section of returns.

3 Omnibus Specification Test

In this section we introduce a new approach to test conditional asset pricing models. We discuss our test statistic and its asymptotic theory.

\(^2\)In this special case, a centred SDF does not exist. Therefore, the corresponding asymptotic theory of the estimators and test would break down. On the contrary, the asymptotic theory for an uncentred SDF of Section 3 is well behaved in this case.
3.1 Test Statistic

There are many types of theoretically possible statistics to test the null hypothesis (1). Note that, in terms of the pricing errors, it is equivalent to test for the null hypothesis

\[ H_0 : \mu (z) = D (z) \delta (z) \]  

(12)
or, by equation (7),

\[ H_0 : \ e (z; \delta (z)) = 0, \]  

(13)

for all \( z \in \mathcal{Z} \). The main problem in nonparametric testing is to (i) estimate the distribution of the test statistic under the null, and to (ii) calibrate this distribution estimate for a given significance level to guarantee that the practitioner can control the error of the first type, i.e. the rejection rate when the null hypothesis was correct. Finally, it has to be checked that (iii) this calibrated version exhibits some power toward all interesting alternatives. For more details see Sperlich (2014) and Rodriguez-Poo et al. (2015).

Out of a huge set of theoretically valid test statistics we find one for which we can provide solutions to these points (i) to (iii). The test distribution is estimated using wild bootstrap. Calibration and power are obtained by a proper combination of the bandwidth choices, but also depend on the way the bootstrap sample is generated.

Our test statistic directly checks whether the pricing errors are significantly different from zero, or any other given constant. That is, it checks whether (1) is fulfilled or not. Indeed, testing condition (1) in the uncentred SDF variant could be checked by a statistic of the form

\[ S = \int e (z_t; \delta (z_t))' \ W (z_t) e (z_t; \delta (z_t)) \pi_1 (z_t) \ dz_t, \]  

(14)

where \( \pi_1 (z_t) \) is a weighting function, typically used for trimming or e.g. the (squared) pdf, say \( p(\cdot) \), of \( z_t \). Under the null hypothesis that the conditional asset pricing model is correctly specified, i.e. (1) holds, we have that \( S = 0 \), while under the alternative, it follows that \( S > 0 \).

In order to obtain a sample analog of (14), let us start from the standard approach used in specification testing for regression functions, which is looking at the residuals under the Null model. These residuals are

\[ u_{t+1} (z_t) = r_{t+1} - \mu (z_t) + e (z_t; \delta (z_t)) = r_{t+1} - D (z_t) \delta (z_t). \]  

(15)

Note that under the Null, \( E (u_{t+1} (z_t) | z_t) = e (z_t; \delta (z_t)) = 0. \) Actually, we can define the residuals from the fitted Null model as

\[ \tilde{u}_{t+1} (z_t) = r_{t+1} - D_g (z_t) \delta_{g, h_1} (z_t), \]  

(16)

where \( \delta_{g, h_1} (z_t) \) is a nonparametric estimator of (10), i.e.

\[ \delta_{g, h_1} (z_t) = (D_g (z_t)' \ W (z_t) D_g (z_t))^{-1} D_g (z_t)' \ W (z_t) \tilde{u}_{h_1} (z_t), \]  

(17)
with $\hat{\mu}_{h_1}(z_t)$ being a local linear estimator of $\mu(z_t)$ (see e.g. Fan and Gijbels (1995)) with a bandwidth $h_1$. For simplicity, let us assume that we have already estimated $D(z)$ and $W(z)$ with some data-adaptively chosen bandwidth $g$, in our case cross validation. For the ease of notation we suppress their hats, and simply write $D_g(z)$ referring to their nonparametric estimates with bandwidth $g$. For bandwidth $h_2$ we obtain as a sample counterpart of (14)

$$S_{1T}(h_1, h_2) = \int \left[ \frac{1}{T} \sum_t K_{h_2}(z - z_t) \hat{u}_{t+1}(z_t) \right]' W(z) \left[ \frac{1}{T} \sum_t K_{h_2}(z - z_t) \hat{u}_{t+1}(z_t) \right] \pi_2(z) \, dz ,$$

(18)

where $\pi_2(\cdot) = \pi_1(\cdot)p^2(\cdot)$. This means that we consider the values of $\hat{u}_{t+1}(z_t)$ only at the values of $z_t$ within the support of $p(\cdot)$ and pay substantially less attention in areas where the data are sparse. Further, this modification not only simplifies the theoretical derivations, but also makes the statistic stable in practice – regardless of the choice of weight function $\pi_2(\cdot)$. The kernel function is

$$K_{h_2}(v) = \frac{1}{h_2^d} K\left(\frac{v}{h_2}\right)$$

with $K(\cdot)$ being a standard $d$-variate kernel function. This test statistic follows the maybe most common approach in nonparametric specification testing being based on the difference between a direct estimator of $\mu(z)$ and our estimator of the Null $D(z)\delta(z)$ convoluted with the same smoother as the raw output $r_{t+1}$.

Note that $h_1$ has to be chosen in a way that our test is calibrated and does not over- (or too much under-) reject under the Null. We can further maximise the power along bandwidth $h_2$ by looking at

$$\tilde{S}_{1T}(h_1) = \max_{h_2} \frac{S_{1T}(h_1, h_2) - E[S_{1T}(h_1, h_2) | H_0]}{\sqrt{V[S_{1T}(h_1, h_2) | H_0]}} .$$

(19)

We need, however, to estimate $E[S_{1T}(h_1, h_2) | H_0]$ and $V[S_{1T}(h_1, h_2) | H_0]$ for any bandwidth $h_2$ and given $h_1$.

In sum we have a test statistic which seems to be quite attractive, with a quite natural interpretation, an automatic procedure for maximising the power, and one parameter, namely bandwidth $h_1$, as calibration parameter. Certainly, bandwidth $g$ could also be used for calibration; but it is more natural to simply look for a bandwidth that estimates $D(z)$ and $W(z)$ quite well with the tendency to undersmooth in order to keep the smoothing bias small.

### 3.2 Asymptotic Behavior

As we remark in the previous section, the level and the power of our test are going to depend crucially on the theoretical properties of $D_g(z)$, estimated with bandwidth $g$, and $\hat{\mu}_{h_1}(z)$, estimated with bandwidth $h_1$. These estimators appear in the test through the estimator $\hat{\delta}_{g,h_1}(z)$ that is defined in equation (17).

Therefore, we first analyze the properties of this estimator and then the test. In order to do so, we introduce the following assumptions,
(C.1) Let $\mathcal{Z}$ be a compact subset of $\mathbb{R}^d$ and let $D_g (z)$ be an estimator of $D(z)$ at observation points $z$ such that $p(z) > 0$, then

$$\sup_{z \in \mathcal{Z}} |D_g (z) - D(z)| = O \left( \sqrt{\frac{\log T}{Tg^d}} \right) + O (g^{\alpha_1}),$$

for some $\alpha_1 > 0$, as $T$ tends to infinity.

(C.2) Let $\mathcal{Z}$ be a compact subset of $\mathbb{R}^d$ and let $\hat{\mu}_{h_1} (z)$ be an estimator of $\mu(z)$ at observation points $z$ such that $p(z) > 0$, then

$$\sup_{z \in \mathcal{Z}} |\hat{\mu}_{h_1} (z) - \mu(z)| = O \left( \sqrt{\frac{\log T}{T h_1^d}} \right) + O (h_1^{\alpha_2}),$$

for some $\alpha_2 > 0$, as $T$ tends to infinity.

We remark that these assumptions are just designed to allow for different possible estimators of $D(z)$ and $\mu(z)$. In fact, for $D(z)$ we propose the following,

$$D_g (z) = \frac{1}{T} \sum_{s} K_g (z - z_s) r_{s+1} f'_{s+1} \frac{1}{T} \sum_{s} K_g (z - z_s),$$

(20)

Under some primary conditions (see Technical Conditions (C1) to (C6) in Yin et al. (2010)) it is possible to show (C.1) as a result. For $\hat{\mu}_{h_1} (z)$ we propose a local linear estimator by noting that, as in the previous case, it is possible to show (C.2) as a result under some more primitive conditions stated for example in Masry (1996). However, for the sake of simplicity, and to allow for other possible estimators we prefer to impose conditions (C.1) and (C.2).

(C.3) The elements of $W(z)$, $D(z)$ and $\mu(z)$, at observation points $z$ such that $p(z) > 0$, must fulfill

$$\sup_{z \in \mathcal{Z}} |W(z) \mu(z)| < \infty, \quad \sup_{z \in \mathcal{Z}} |D(z)' W(z)| < \infty,$$

for fixed $N$, $K$ and $d$.

(C.4) At observation points $z$ such that $p(z) > 0$, assume that

$$\inf_{z \in \mathcal{Z}} D(z)' W(z) D(z) > \kappa_T.$$

(C.5) At observation points $z$ such that $p(z) > 0$ the weighting matrix $W(z)$ is bounded from above and below.

Conditions (C.3) and (C.4) are standard bounds in first and second order moment conditions. Note also that the matrix $W(z)$ needs to be bounded at values of $z$ such that $p(z) > 0$. The estimator of the prices of risk satisfies the following property.
Theorem 1. Let \( \mathcal{F} \) be a compact subset of \( \mathbb{R}^d \) and assume conditions (C.1) to (C.5) hold. Then as \( \kappa_T^{-1} g \) and \( \kappa_T^{-1} h_1 \) tend to zero in such a way that \( \kappa_T^2 T g^d \to \infty \) and \( \kappa_T^2 T h_1^d \to \infty \) we have that, under the null hypothesis,

\[
\sup_{z \in \mathcal{F}} |\hat{\delta}_{h_1}(z) - \delta(z)| = O \left( \kappa_T^{-1} \sqrt{\frac{\log T}{T g^d}} + \kappa_T^{-1} \sqrt{\frac{\log T}{T h_1^d}} \right) + O \left( \kappa_T^{-1} g^{\alpha_1} + \kappa_T^{-1} h_1^{\alpha_2} \right),
\]

as \( T \) tends to infinity.

Further, to establish the properties of \( S_{1T} (h_1, h_2) \) we include the following conditions

(A.1) The process \((z_t, r_t)\) is absolutely regular, i.e.

\[
\beta(s) \equiv \sup_{t \geq 1} E \left\{ \sup_{A \in \mathcal{F}_s} \left| P \left( A \mid \mathcal{F}_t \right) - P(A) \right| \right\} \to 0,
\]

as \( s \to \infty \),

where \( \mathcal{F}_t \) is the \( \sigma \)-field generated by \( \{ (z_k, r_k) : k = t, \cdots, s \} \). Further it is assumed that \( \beta(s) \) exhibits a geometric rate of decay.

(A.2) \( z_t \) has a bounded density function \( p(z_t) \). Further, the joint density of distinct elements of \((z_1, r_1, z_s, r_s, z_t, r_t)\), \( (t > s > 1) \), is continuous and bounded by a constant independent of \( s \) and \( t \). Furthermore, \( p(z_t) \) is two times continuously differentiable in all its arguments.

(A.3) Let \( \varepsilon_{t+1} = r_{t+1} - \mu(z_t) \). With probability one, \( E \left( \varepsilon_{i,t+1} \mid z_t, \mathcal{F}_i^{t-1} \right) = 0 \), for all \( t \) and \( i = 1, \cdots, N \). Let \( \sigma_t^2(z) = \text{Var} (r_{t,t+1} \mid z_t = z) = E \left( \varepsilon_{t,t+1}^2 \mid z_t = z \right) \). Then \( \sigma_t^2(z) \) satisfies some Lipschitz conditions:

\[
|\sigma_t^2(z + u) - \sigma_t^2(z)| \leq M(z) \|u\|, \quad \text{with } E \left( \|M(z)\|^{2+\eta} \right) < \infty,
\]

for all \( t \) and \( i = 1, \cdots, N \). Furthermore, for all \( i \neq j \), and \( t \neq s \), \( E \left( \varepsilon_{i,t+1} \varepsilon_{j,t+1} \mid z_t, z_s \right) = 0 \). Finally, \( E \left\{ \|\mu(z_t)\|^{16} + \|r_{t+1}\|^{16} \right\} < \infty \).

(A.4) \( K(\cdot) \) is a product kernel, i.e. \( K(z) = \prod_{d} k(z_d) \), and \( k(\cdot) \) is a symmetric density function with bounded support in \( \mathbb{R} \) and \( |k(z_1) - k(z_2)| \leq c |z_1 - z_2| \) for all \( z_1, z_2 \) in its support.

(A.5) \( h_2 \sim T^{-\frac{2(1-\varepsilon)}{3d-1}} \) for some \( 1 > \varepsilon > 0 \).

(A.6) \( \frac{h_{2,d}}{h_2^{d/2}} \kappa_T^{-2} \to 0 \) and \( \frac{h_{2,d}^2}{g^2} \kappa_T^{-2} \to 0 \) as \( T \) tends to infinity.

(A.7) \( T g^{d+\alpha_1} \to 0 \) and \( T h_1^{d+\alpha_2} \to 0 \) as \( T \) tends to infinity.

Assumption (A.1) imposes a structure that is necessary to apply a Central Limit Theorem for \( U \)-statistics under dependence. This will be used to obtain the asymptotic distribution of the test statistic under the null. Several CLT’s are available in the literature, see among others Hjellvik et al. (1996), Fan and Li (1999) and Gao and Hong (2008). Assumption (A.2) is a standard smoothness assumption. Assumption (A.3) imposes boundedness and smoothness conditions in moments of \( \varepsilon_t \). Note also that we need to make assumptions about the structure of the
matrix $E \left( \varepsilon_{t+1} \varepsilon'_{s+1} \big| z_t, z_s \right)$, for $t \neq s$. Assumption (A.4) on the kernel weight is also standard in nonparametric tests. Finally, assumptions (A.5) to (A.7) relate all bandwidths included in the testing procedure. In particular, assumption (A.5) is concerned with the rate of the bandwidth of the test. We discuss later in the paper how to choose empirically the bandwidth $h_2$ according to the rate assumed in (A.5).

The asymptotic behavior of the test statistic under the null hypothesis is given in the following result.

**Theorem 2.** Under conditions (C.1) to (C.5), (A.1) to (A.7), and if the null hypothesis $H_0$ is true, then

$$Th_2^{d/2} S_{1T} (h_1, h_2) - Bh_2^{-d/2} \rightarrow_d N(0, V),$$

where

$$B = \int_{\mathbb{R}^d} K^2(u) du \times \sum_i \sum_j \int_{\mathbb{R}^d} w_{ij}(z) E \left( \varepsilon_{i,t+1} \varepsilon_{j,t+1} \big| z \right) p(z) \pi_2(z) dz$$

and

$$V = 2 \int_{\mathbb{R}^d} K(u) K(v) K(u-x) K(v-x) dudvdx$$

$$\times \sum_i \sum_j \int_{\mathbb{R}^d} w_{ij}(z) E \left( \varepsilon_{i,t+1}^2 \big| z \right) E \left( \varepsilon_{j,t+1}^2 \big| z \right) p^2(z) \pi_2(z) dz$$

as $T$ tends to infinity.

The bias term is of factor $h_2^{-d/2}$, which makes necessary the use of a bootstrap procedure to approximate the asymptotic distribution of the test statistic. A proposal of bootstrap procedure designed to overcome this problem is shown in the next section.

Finally, we determine the power of our test against local alternatives to the null hypothesis. Let us define the sequence of local alternatives to the null hypothesis

$$H_1 : \mu(z) = D(z) \delta(z) + \gamma_T \Delta(z) \quad (23)$$

where $\gamma_T$ is a sequence that tends to zero such that

(H.1) $\sqrt{T}h^{d/2} \gamma_T \rightarrow \infty$, and $\gamma_T = o \left( \frac{1}{\sqrt{Th_2^d}} \right)$, as $T$ tends to infinity.

(H.2) At observation points $z$ such that $p(z) > 0$, we assume that

$$\sup_{z \in \mathbb{Z}} |W(z) \Delta(z)| < \infty.$$

Assumptions (H.1) is also related to the bias problem that has been already analyzed above. The power of our test is described in the following result.
Theorem 3. Under conditions (C.1) to (C.5) and (A.1) to (A.7), if \( H_1 \) verifies (H.1) and (H.2), then for all sequences of random variables \( \{ c_T \colon T \geq 1 \} \) with \( c_T = O_p(1) \) we have
\[
P\left( T h_d^{d/2} S_1 T > c_T \right) \to 1
\]
as \( T \) tends to infinity.

Theorem 3 indicates that our test has nontrivial power only against sequences of local alternatives for which \( \gamma_T \) tends to zero at a rate that is smaller than \( \sqrt{T} \). As seen in Andrews (1997), tests based on weighted parametric residuals have nontrivial power against local alternatives for which the rate is exactly \( \sqrt{T} \). Thus, at least in terms of the asymptotic local power these tests appear to dominate tests that require slower rates. However, as shown in Horowitz and Spokoiny (2001), at a exact rate of \( \sqrt{T} \) no test can have nontrivial power uniformly over reasonable classes of functions \( \Delta(\cdot) \) in (23).

3.3 Previous Estimators and Tests

In this section, we review previous papers that test conditional asset pricing models with nonparametric prices of risk. \(^3\)

In the case of traded factors, once the SDF has been estimated, Wang (2003) proposes to test the restrictions \( E (m_{t+1} r_{i,t+1} \mid z_t) = 0 \) by running the following regression
\[
\hat{e}_{i,t+1} = z_t' \hat{\gamma}_i + u_{i,t+1}
\]
(24)
to obtain
\[
\hat{\gamma}_i = \left( \frac{1}{T} \sum_{s=1}^{T} \hat{\omega}_s z_s z_s' \right)^{-1} \left( \frac{1}{T} \sum_{s=1}^{T} \hat{\omega}_s z_s \hat{e}_{i,s+1} \right)
\]
(25)
where \( \hat{\omega}_s \) is a weight function defined to cope with the random denominator in semiparametric regressions. If the linear model in (24) is correctly specified, then it is sufficient to test
\[
\gamma_1 = \cdots = \gamma_N = 0.
\]

Let us define \( \hat{\gamma}_T = (\hat{\gamma}_1, \cdots, \hat{\gamma}_N) \). Using standard GMM techniques, Wang (2003) shows, that under the null hypothesis of a valid SDF, the test statistic \( T \hat{\gamma}_T' \hat{\Omega}_T^{-1} \hat{\gamma}_T \) follows a limiting chi-square distribution with \( Nd \) degrees of freedom, where \( \hat{\Omega}_T^{-1} \) is the estimator of the inverse of the asymptotic variance-covariance matrix of \( \hat{\gamma}_T \).

\(^3\)Cai et al. (2015) analyze the estimation of \( \delta(z) \) for the particular case of the CAPM. Given a single return \( i \), they propose to estimate \( \hat{\delta}(z) \) through the ratio
\[
\hat{E}(r_{i,t+1} \mid z_t) / \hat{E}(r_{i,t+1} r_{p,t+1} \mid z_t),
\]
where \( r_{p,t+1} \) is the market return. They obtain the asymptotic distribution of \( \hat{\delta}(z) \). However, the authors do not obtain an estimator of the prices of risk for \( K > 1 \), and do not provide a formal testing device.
Unfortunately, if the pricing errors are actually nonlinear functions of $z$ then the previous test proposed in Wang (2003) exhibits zero power against those nonlinear alternatives. Our new test appears as a solution to this drawback.

Roussanov (2014) proposes some explorative but informal methods based on nonparametrically estimated pricing errors. He presents the estimated pricing errors with some point-wise bootstrap bands. However, the only test of the asset pricing model that is suggested is for a zero unconditional mean of a particular pricing error, and the corresponding asymptotic theory is not developed. We develop the asymptotic theory of our test, which considers jointly all the pricing errors, and does not focus on unconditional means.

4 Implementation and Calibration

4.1 Wild Bootstrap-type Resampling

Härdle et al. (1993) study three different bootstrap procedures and conclude that the wild bootstrap is the most pertinent method for testing the regression structure. Therefore we adopt a wild bootstrap scheme to estimate the distribution of our test statistic. First note that $S_{1T}$ can be written as

$$S_{1T}(h_1, h_2) = \frac{1}{T^2} \int \left[ \sum_t K_{h_2}(z - z_t) \hat{u}_{t+1}(z_t) \right]' W(z) \left[ \sum_t K_{h_2}(z - z_t) \hat{u}_{t+1}(z_t) \right] \pi_2(z) \, dz,$$

with $\hat{u}_{t+1} = r_{t+1} - D_g(D_g WD_g)^{-1} D_g WH_h r_{t+1} := (I - P_{g,h_1}) r_{t+1}$ where $I$ is the identity matrix, and $H_{h_1}$ is the so-called hat- or smoothing matrix of the kernel estimator used when estimating $\mu(\cdot)$. When following the procedure of Härdle et al. (1993), we would like to generate $B$ bootstrap analogues of $S_{1T}$ defined exactly as above but replacing $\hat{u}_{t+1}$ by $u^*_{t+1} = (I - P_{g,h_1}) r^*_{t+1}$ with $r^*_{t+1}$ being generated under the null hypothesis. We may then set $r^*_{t+1} = D_g(z_t) \hat{h}_b(z_t) + \varepsilon^*_{t+1}$ with bootstrap residuals $\varepsilon^*_{t+1}$ (see below for details) and bootstrap bandwidth $h_b$ which is supposed to converge slower to zero than $h_1$ does. However, while the proper choice of $h_b$ is crucial for the calibration\(^4\) of the test, in practise this can become a difficult task, see Sperlich (2014). An additional problem is that we may rather want to generate bootstrap returns $r^*_{t+1} = D^*_g(z_t) \tilde{h}_b(z_t) + \varepsilon^*_{t+1}$ because $D$ is $E[r_{t+1} f_{t+1} | z_t]$, but along the hypothetical pricing model $r^*_{t+1}$ should refer to $E[r^*_{t+1} f^*_{t+1} | z_t]$. That is, these bootstrap excess returns are defined only implicitly and therefore hard to obtain; but simulations with the simpler version based on $D_g$ give less satisfying results than the alternative we propose further below.

Kreiss et al. (2008) propose a bootstrap specification test that has important similarities with ours. For the resampling, they propose to use only $\varepsilon^*_{t+1}$ instead of $u^*_{t+1}$ when simulating the

\(^4\)In the sense that under $H_0$ it holds the fixed $\alpha$-levels, i.e. do not exceed the error of the first type.
distribution of the statistic under the Null. In fact, from (15) and (A.3) it is easy to see that

\[
Th_2^{d/2} S_{1T} (h_1, h_2) = Th_2^{d/2} S'_{1T} (h_1, h_2) + \frac{2h_2^{d/2}}{T} \int \left[ \sum_t v_h (z - z_t) \varepsilon_{t+1} \right] W(z) \]  

\[
\times \left[ \sum_t K_h (z - z_t) \left( e (z_t; \delta (z_t)) - \left( D_g (z_t) \hat{h}_1 (z_t) - D (z_t) \delta (z_t) \right) \right) \right] \pi_2 (z) dz \]  

\[
+ \frac{h_2^{d/2}}{T} \int \left[ \sum_t K_h (z - z_t) \left( e (z_t; \delta (z_t)) - \left( D_g (z_t) \hat{h}_1 (z_t) - D (z_t) \delta (z_t) \right) \right) \right] W(z) \]  

\[
\times \left[ \sum_t K_h (z - z_t) \left( e (z_t; \delta (z_t)) - D_g (z_t) \hat{h}_1 (z_t) - D (z_t) \delta (z_t) \right) \right] \pi_2 (z) dz, \]  

with innovations \( \varepsilon_{t+1} = r_{t+1} - \mu (z_t) \) being invariant under the null hypothesis, and

\[
S'_{1T} (h_1, h_2) = \frac{1}{T^2} \int \left[ \sum_t K_h (z - z_t) \varepsilon_{t+1} \right] W(z) \left[ \sum_t K_h (z - z_t) \varepsilon_{t+1} \right] \pi_2 (z) dz, \]  

a quadratic of them. It is not hard to show that under the null hypothesis, \( Th_2^{d/2} S_{1T} (h_1, h_2) \) is asymptotically normal, and more importantly, its asymptotic distribution is the same as that of \( Th_2^{d/2} S'_{1T} (h_1, h_2) \). This indicates that we could mimic the distribution of \( S_{1T} (h_1, h_2) \) by bootstrapping \( S'_{1T} (h_1, h_2) \). The latter does not depend on the validity of the null hypothesis, although \( S_{1T} (h_1, h_2) \) does, so it follows the guidelines set by Hall and Wilson (1991).

Kreiss et al. (2008) discuss various advantages of their bootstrap testing approach compared to that of Härdle et al. (1993). In their simulations, however, they only consider cases where the null hypothesis is a parametric model. That is, while their findings are useful for large samples, they miss the calibration problem we face under non- and semiparametric null hypotheses highlighted in Rodriguez-Poo et al. (2015). For our case, this means the impact of \( P_{g,h_1} \) on the test distribution matters in finite samples. This is confirmed by some simulation studies not shown here.

The obvious balance between them to generate bootstrap replicates of the test by substituting \( u_{t+1} = (I - P_{g,h_1}) \varepsilon_{t+1} \) for \( u_{t+1} \) in (26). This can either be seen as the refinement of Kreiss et al. (2008) by including \( P_{g,h_1} \), or as a simplification of Härdle et al. (1993) by reducing \( r_{t+1} \) to \( \varepsilon_{t+1} \). In all versions the bootstrap innovations \( \varepsilon_1^*, \cdots , \varepsilon_T^* \) have to be conditionally independent given the observed data \( \{(r_{t+1}, z_t) : 1 \leq t \leq T\} \). Furthermore,

\[
E^* \varepsilon_{t+1} = 0, \quad E^* [\varepsilon_{t+1} \varepsilon_{t+1}'] = \tilde{\varepsilon}_{t+1} \tilde{\varepsilon}_{t+1}' = \{r_{t+1} - \hat{\mu}_h (z_t)\} \{r_{t+1} - \hat{\mu}_h (z_t)\}', \]  

where \( E^* \) denotes the expectation under the bootstrap distribution and \( \hat{\mu}_h (\cdot) \) is a local linear estimator of \( \mu \). In practise, we can define \( \varepsilon_{t+1}^* = \tilde{\varepsilon}_{t+1} \eta_{t+1} \), where \( \eta_{t+1} \) is a sequence of i.i.d. random variables with zero mean and unit variance.
We reject the null hypothesis if

\[ S^*_{1T} (h_1, h_2) = \frac{1}{T^2} \int \left[ \sum_t K_{h_2} (z - z_t) u_{t+1}^* \right]' W(z) \left[ \sum_t K_{h_2} (z - z_t) u_{t+1}^* \right] \pi_2 (z) \, dz, \tag{29} \]

for which we can give the following result for the bootstrap distribution.

**Theorem 4.** Assume that the conditions of Theorem 2 hold. For the bootstrap statistic \( S^*_{1T} \) defined in (29), we have as \( T \) tends to infinity,

\[ Th_2^{d/2} \left[ S^*_{1T} - E^* (S^*_{1T}) \right] \to_d N (0, V) \tag{30} \]

conditionally on \( \{ (r_{t+1}, z_t) : 1 \leq t \leq T \} \), where \( V \) is the same as given in Theorem 2, and \( Th_2^{d/2} \left[ E (S_{1T}) - E^* (S^*_{1T}) \right] \to_p 0. \)

We reject the null hypothesis if \( S_{1T} > t^*_\alpha \), where \( t^*_\alpha \) is the upper \( \alpha \)-quantile of the conditional distribution of \( S^*_{1T} \) obtained via bootstrap, i.e. the \( p \)-value of the test is the relative frequency of the event \( \{ S^*_{1T} > S_{1T} \} \) in the bootstrap replications. Next we prove that the bootstrap is asymptotically correct in the sense that its significance level converges to \( \alpha \) as \( T \) tends to infinity.

**Theorem 5.** Assume that the conditions of Theorem 2 hold. Let \( t^*_\alpha \) be the upper \( \alpha \)-quantile of the conditional distribution of \( S^*_{1T} \) given \( \{ (r_{t+1}, z_t) : 1 \leq t \leq T \} \) and \( \alpha \in (0, 1) \). Then, as \( T \to \infty \), under the null hypothesis \( P \{ S_{1T} > t^*_\alpha \} \to \alpha. \)

Nonetheless, in finite samples the test performance depends on the bandwidths \((g, h_1, h_2)\). To understand the calibration problem in nonparametric testing, the following facts must be recognised. While bootstrap methods in nonparametric inference work quite well for the variance estimation, they do less so for the approximation of the smoothing bias. Moreover, even if the variance is well approximated and maybe also the bias problem reduced, the bootstrap distribution may differ from the true one in some higher moments so that the critical values (quantiles) are not well estimated. One may use bootstrap methods that asymptotically adapt the higher moments as e.g. the golden-cut method. However, there exists an important literature showing that these only improve performance for either huge samples or some particular cases.

In order to solve all these problems at once, we recommend to interpret the quantiles of the distribution of \( S_{1T}(h_1, h_2)|H_0 \), say \( t_\alpha \), as a function of the smoothing parameters, see Sperlich (2014). But before doing so, it is recommended to maximise the power of the test along \( h_2 \) by using the statistic \( \tilde{S}_{1T}(h_2) \), see (19). For this we also need estimates of \( E[S_{1T}(h_1, h_2)|H_0] \) and \( V[S_{1T}(h_1, h_2)|H_0] \) which can easily be obtained from our bootstrap replicates. Therefore, we are actually looking for bandwidth \( h_1 \) such that the corresponding quantile of \( \tilde{S}_{1T}(h_2)|H_0 \), say \( t^*_\alpha (h_2) \), is smaller or equal to the nominal level \( \alpha. \)

Here, \( \tilde{S}_{1T}(h_2) \) is defined as \( S_{1T}(h_2) \) but replacing \( S_{1T}(h_1, h_2) \) by \( S^*_{1T}(h_1, h_2) \). These bandwidths are obtained then from prior simulations as those outlined below. We do not provide here an extension of Theorem 2 to the test statistic

\[ ^5 \text{We actually want to get } t^*_\alpha (h_2) \approx \alpha, \text{ but for small samples this may be a numerically difficult task in practise.} \]
\[ \tilde{S}_{1T}(h_2). \] See Rodriguez-Poo et al. (2015) and references therein for details related to a possible proof of this result.

### 4.2 Some Simulation Results

Appendix C describes the DGP of the Monte Carlo experiment in detail. Importantly, we designed the experiment along our real data set, which is described in the next subsection. That is, the data generating functions in the model are estimated from the real data set. Therefore, the simulation study below reflects a realistic situation, and we can use its outcome for the calibration of the test when the test is applied to the real data.

We generated a series \( \{z_t\}_{t=1}^{T} \) with \( T = 251 \), as in the real data set, but repeated the simulations also with \( T = 500 \). The reason is that all theoretical results only reflect the first order approximation. But it is well known that in nonparametrics not only the convergence rate is slower than in the parametric world, but also that second order terms can still play an important role in finite samples. Note that for the testing problem we consider, \( T = 251 \) is actually a pretty small sample whereas \( T = 500 \) seems to be a reasonably large size.

Using the procedure described in Appendix C, we can simulate the factors and excess returns. In particular, we use a DGP calibrated along (a) a one-factor model with \( z \) being the log of the default spread, a traded factor \( f \) ’market’, and three excess returns \( r \) being ‘market’, ‘SMB’, and ‘HML’ (see section 5 for their definitions); (b) a two-factor model with the same ingredients as above but ‘consumption’ as a second factor. The asset pricing model in (a) represents the CAPM, while the model in (b) represents a CCAPM with Epstein-Zin preferences. The simulation was repeated 500 times under the null \( H_0 \) and 100 times under \( H_1 \). In these simulations we used 200 bootstrap samples for estimating the \( p \)-values.

For calculating our test, the integral in (18) was approximated by the Trapezoidal rule with an equidistant grid of hundred points, using \( \pi_2 \) only for trimming or setting \( \pi_2 = 1 \). Trimming is done in order to avoid that boundary effects of the nonparametric estimators distort the testing outcome. On the other hand, often the interesting things happen at the boundaries of the support. In our simulations we compare the case of no trimming or to cut off 5% (2.5% each side) of the most extreme values of \( z_t \) in the test statistic (not in the estimation).

As explained before, there are several parameters and estimators to be chosen. The conditional expectations of factors and excess returns are calculated with local linear estimators in order to eliminate biases in the linear direction, and keep boundary effects small. The matrices \( \mathbf{D} \) and \( \mathbf{W} \) are estimated by a Nadaraya-Watson estimator. We use the Quartic kernel throughout and ran our simulations with bandwidth \( g := c_g \cdot g^{CV} \) and \( h_1 := c_1 \cdot h_1^{CV} \), where \( g^{CV} \) and \( h_1^{CV} \) are the (typically somewhat undersmoothing) cross validation bandwidths. Not surprisingly, for the two-factor model containing many more nonparametric functions to be estimated, the calibration turns out to be harder than for the one-factor model.

In Tables 1 and 2 some simulation results are summarised for the test statistic \( \tilde{S}_{1T} \) with \( \mathbf{W}(z_t) \).
following (11) in the one-factor and two-factor models. Obviously, in Table 1 the trimming has no important effect. This is also true for the two factor model and is therefore skipped from Table 2. It can be seen there how calibration can be reached by means of the proper choice of \( c_1 \) and \( c_g \). However, in the two-factor model with only \( T = 251 \), the calibration does not work as nicely as in the one-factor model. In order to see whether this is a general problem of having more factors or simply a problem of sample size, we repeated the simulations for the two factor model with \( T = 500 \). The results in Table 2 show that it is indeed just a problem of sample size as for the larger sample we can perfectly calibrate the test. It is nonetheless important that already with \( T = 251 \) we find a calibration for the two-factor model that prevents the test from over-rejecting, while preserving some power.

Of course, this is just a small extract of all simulations we performed. For example, we also repeated the simulations with \( W \) being the identity matrix, other bootstrap procedures, or using local linear estimators also for the second moments, and even re-estimating the second order moments in each bootstrap sample. All this leads just to slightly different parameters \((c_g, c_1)\) for the calibration.

<table>
<thead>
<tr>
<th>( c_g )</th>
<th>( c_1 )</th>
<th>( H_0 )</th>
<th>( H_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \text{av.pv} )</td>
<td>0.10</td>
</tr>
<tr>
<td>without</td>
<td>.75</td>
<td>.75</td>
<td>.493</td>
</tr>
<tr>
<td>trimming</td>
<td>.75</td>
<td>1.0</td>
<td>.495</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>.491</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.5</td>
<td>.489</td>
</tr>
<tr>
<td>trimming</td>
<td>.75</td>
<td>.75</td>
<td>.494</td>
</tr>
<tr>
<td>5% of data</td>
<td>.75</td>
<td>1.0</td>
<td>.495</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>.493</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.5</td>
<td>.489</td>
</tr>
</tbody>
</table>

Table 1: One-factor model: Average \( p \)-values (av.pv) and percentages of rejections for rejection levels of \( \alpha = 0.1, 0.05 \) and 0.01 for \( \tilde{S}_{1T} \) under different calibration scenarios with and without trimming, \( T = 251 \). The numbers are obtained from 500 replications under the \( H_0 \) and 100 under \( H_1 \) using 200 bootstrap samples.

We also simulated the CCAPM, where there is a single non-traded pricing factor, and we obtained similar results. Generally, when using the same bandwidths as for the CAPM, the test becomes a bit more conservative, i.e. the same bandwidths give slightly larger \( p \)-values. Therefore it is recommended to use slightly larger \((c_g, c_1)\) for the CCAPM.

We can see from these tables that, for appropriately chosen parameters \((c_g, c_1)\), the test is calibrated to hold the rejection level under the Null. For the simulated alternatives \( H_1 \) described in Appendix C, the test exhibits some power. The function \( q^*(z_t) \) in Appendix C, which drives the distance to the Null, is estimated from the data of the following empirical section. It is away from zero and moves quite a bit in the one-factor model, but not so for the two-factor model. In some simulations that are not shown here, we either increased sample size to \( T = 500 \) or multiplied the estimated delta function by 2, and then the power went up by a factor of about 2 as expected. Putting all together, our test works reasonably well even for small samples.
<table>
<thead>
<tr>
<th>$c_0$</th>
<th>$c_1$</th>
<th>$H_0$ av.pv</th>
<th>$H_1$ av.pv</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.01 .05 .01</td>
<td>.01 .05 .01</td>
</tr>
<tr>
<td>$T = 251$</td>
<td>1.0</td>
<td>.568 .026 .006 .000</td>
<td>.397 .200 .120 .060</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>.491 .042 .012 .000</td>
<td>.348 .280 .170 .060</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>.475 .072 .036 .000</td>
<td>.311 .210 .130 .050</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>.464 .086 .036 .000</td>
<td>.305 .230 .140 .050</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>1.5</td>
<td>.474 .066 .020 .000</td>
<td>.174 .450 .360 .150</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>.430 .100 .038 .002</td>
<td>.157 .540 .380 .160</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>.428 .104 .042 .002</td>
<td>.153 .540 .380 .150</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>.405 .124 .054 .002</td>
<td>.153 .580 .390 .160</td>
</tr>
</tbody>
</table>

Table 2: Two-factor model: Average $p$-values (av.pv) and percentages of rejections for rejection levels of $\alpha = 0.1$, 0.05 and 0.01 for $S_{1T}$ under different calibration scenarios without trimming for $T = 251$ and $T = 500$. The numbers are obtained from 500 replications under the $H_0$ and 100 under $H_1$ using 200 bootstrap samples.

5 Empirical Application

We evaluate several asset pricing models with a vector $\mathbf{r}_{t+1}$ that has three excess returns. They are obtained from Ken French’s Data Library, and they are associated with three portfolios that summarize the relevant properties of the Fama and French size and book-to-market sorted portfolios. In particular, the first excess return is associated with the US market portfolio, while the other two excess returns are associated with portfolios that capture size and value effects, denoted SMB and HML, respectively. See his web page, as well as Fama and French (1993) for further details.

We work with two prominent predictors as components of $\mathbf{z}_t$: the default spread and the cointegrating residual of consumption and wealth. The former predictor is constructed from FRED data, using yields on AAA and BAA-rated bonds. We denote $lds$ the logarithm of the default spread. The latter predictor is taken from Martin Lettau’s web page, and was developed by Lettau and Ludvigson (2001a). This predictor is usually denoted $cay$.

We apply our test to three popular asset pricing models, or equivalently three choices of $\mathbf{f}_{t+1}$. The first model is the conditional CAPM, where the excess return on the market portfolio is the only pricing factor. This model extends the traditional CAPM of Sharpe (1964) and Lintner (1965) to conditional moments. See, e.g., Wang (2003) and the references therein. The second model is a linearised consumption CAPM (CCAPM) where the per capita consumption growth is the only pricing factor. See, e.g., Breeden (1979) for the canonical CCAPM. We also consider a third model that nests the previous two models, the linearized CCAPM with Epstein and Zin (1989) preferences (EZ), where the two previous pricing factors are considered jointly. In this way, our empirical application considers a model with a traded factor, a model with a nontraded factor, and a model with the two types of factors.

---

5.1 SDF Estimation

We use quarterly real data from 1951 to 2014 (T=251). Figure 1 shows the conditional means and covariances with the factors of the three excess returns. In this figure and the next ones, the conditioning variable is \( lds \). The performance of the models is similar if we use \( cay \) instead, and the corresponding figures are available upon request. All figures display smooth 90% point-wise confidence intervals obtained from wild bootstrap. The function estimates are calculated with the corresponding cross validation bandwidth.

Figure 1: Means and covariances conditioning on \( lds \) with 90% point-wise confidence intervals.

In the first column of Figure 1, we find that the market risk premium increases with \( lds \). This is expected as we can associate high values of this conditioning variable to "bad states", and low values to "good states". Regarding risks with respect to the market and consumption factors, the market variance also increases with \( lds \), but this is not always the case for the covariance between the market return and consumption growth \( cgp \). This covariance decreases at low values of the predictor, even though it finally increases at higher values.

The second and third columns of Figure 1 describe the risk premium and risks of SMB and HML. The size premium increases for most of the predictor values, but decreases for extreme values. The covariance between the SMB return and the market factor also increases for most values of the predictor, but the covariance between the SMB return and the consumption factor has a non-monotonic shape. The value premium increases for most of the predictor values, but it is much flatter than the other two risk premia. The covariances between the HML return and
the two factors are also flatter than the covariances of the other returns. Its covariance with the market factor is negative for most values of $lds$.

Figure 2 shows the prices of risk that are obtained from the previous means and covariances, and the weighting matrix in (11). Both the market and consumption factors have a positive and non-monotonic price of risk in the CAPM and CCAPM, respectively. Nevertheless, when we join the two factors in the EZ model, the market factor does not seem to be a relevant pricing factor in the SDF.

![Figure 2: Prices of risk conditioning on $lds$ with 90% point-wise confidence intervals.](image)

Figure 3 shows the pricing errors of the three models. In the first column, as Section 2.3 comments for the chosen weighting matrix, the market pricing error is zero in the two models that have the market as a pricing factor. The three models show a high estimated SMB pricing error in the second column of Figure 3, but not clearly significant from a statistical perspective. In the last column, the HML pricing error is clearly nonzero for the CAPM. However, we do not find evidence against zero HML pricing errors for the CCAPM and the EZ models.

The first row of Figure 4 shows the square root of the criterion function that the prices of risk minimise, and the second row shows the conditional mean of the associated SDFs. The minimised criterion is clearly lower for the CCAPM and EZ models than for the CAPM, being close to zero for the CCAPM and EZ models at many values of $lds$. Nevertheless, the SDF mean is also closer to zero for those two models in those regions, and hence those low criterion functions must be interpreted with care following Section 2.4. In fact, the confidence bands of the third row of Figure 1 do not show that the covariances of returns and consumption are clearly different from zero from a statistical point of view.

---

7Figure 3 in Roussanov (2014) shows the value premium conditioning on $cay$. He finds a decreasing premium, but an increasing covariance with the consumption factor. This is also the case in our study, and the corresponding figure is available upon request.
5.2 Asset Pricing Tests

We apply our test to these data, with 500 bootstrap replications and the parameter combinations \((c_g, c_1)\) giving the best calibration in Tables 1 and 2 for the one- and two-factor models, respectively. Table 3 reports the \(p\)-values of our test for the three models with and without
trimming. The first row of Table 3 reports the results for \( z := lds \), while the second row focuses on \( cay \). We also computed the \( p \)-values for several bandwidths, and the results were similar. All these results, which are commented below, are available upon request.

<table>
<thead>
<tr>
<th></th>
<th>CAPM No trim.</th>
<th>Trim.</th>
<th>CCAPM No trim.</th>
<th>Trim.</th>
<th>EZ No trim.</th>
<th>Trim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>lds</td>
<td>.000</td>
<td>.000</td>
<td>.510</td>
<td>.526</td>
<td>.514</td>
<td>.525</td>
</tr>
<tr>
<td>cay</td>
<td>.000</td>
<td>.000</td>
<td>.094</td>
<td>.096</td>
<td>.096</td>
<td>.120</td>
</tr>
</tbody>
</table>

Table 3: \( P \)-values of asset pricing tests with \((c_g, c_1) = (1, 1.5)\) for the CAPM and the CCAPM; \((c_g, c_1) = (3, 3)\) for the EZ.

In all cases we found that the CAPM is clearly rejected with a \( p \)-value smaller than 0.01. However, for the CCAPM, the \( p \)-value is around 0.5 when using \( lds \). If we use \( cay \) instead, then the \( p \)-value is slightly below 0.1 no matter whether looking at trimmed or not trimmed statistics, which indicates that deviations from the Null model were not due to the model behaviour at the boundaries of \( cay \). If we put both pricing factors together in the EZ model, then the \( p \)-value is again slightly above 0.5 when using \( lds \), while it is still around 0.1 when using \( cay \), with \( p \)-values for trimmed statistics always being larger than for non-trimmed statistics. As expected from the findings in the simulation study, larger \((c_g, c_1)\) tend to give smaller \( p \)-values.

The three returns that we use as test assets serve as a summary of the relevant properties of the Fama and French size and book-to-market sorted portfolios. As a robustness check, the conclusions are similar if we use the six Fama and French portfolios as our test assets. We find again a strong rejection of the CAPM, but this is not the case for the CCAPM and EZ models.

Roussanov (2014) does not perform joint tests of zero pricing errors as we do, but focuses on the pricing of particular portfolios instead. Using the centred SDF approach, he finds evidence against the correct pricing of some portfolios by the CCAPM and EZ models with similar data. We highlight two possible reasons.

First, to our knowledge, our test is the first consistent omnibus test for conditional asset pricing models. So other studies may reject or not, depending on their specific alternative or inconsistent inference; for example, looking at point-wise confidence intervals is not equivalent to testing, and bootstrap methods typically used for parametric regression analysis are often inconsistent in nonparametric analysis.

Second, Peña-Randa and Sentana (2015) find similar discrepancies between the centred and uncentred SDF approaches when testing unconditional asset pricing models. In their data, this fact seems to be due to a poor unconditional correlation between consumption growth and the cross-section of excess returns to be priced. Interestingly, we find a conditional counterpart of this problem. As shown in the third row of Figure 1, there is little evidence against the hypothesis of a zero conditional covariance of these excess returns with the consumption factor. Moreover, the estimated prices of risk in the second row of Figure 2 show that consumption growth rather than the market portfolio seems to be driving the SDF of the EZ model, and hence such a model is similar to the CCAPM.
Following Section 2.4, non-traded factors that are conditionally uncorrelated with excess returns will automatically price those returns with a SDF whose conditional mean is 0. The second row of Figure 4 shows SDF means for the CCAPM and EZ models that are close to zero in the regions where the pricing errors are low. Note that the figures display point-wise confidence bands, and uniform bands would be wider. We do not provide a formal test of a zero SDF mean, but we expect that such a test would not reject for the CCAPM and EZ model.

Finally, the centred SDF approach is not well defined in this case, and the corresponding asymptotic theory of the estimators and test would break down. In fact, if we implement the centred approach, the procedure often breaks down numerically simply because for some values of $z$ ($lds$ or $cay$) the nonparametric conditional covariances are very close to zero. Further inference based on its inverse is then meaningless, and actually corresponds to identification problems.\textsuperscript{8}

### 6 Conclusions

In this paper we present an adaptive omnibus specification test of conditional asset pricing models. These models provide constraints that conditional moments of returns and pricing factors must satisfy, but frequently do not provide information on the functional form of those conditional moments. The main interest of our test is that it is not only robust to functional form misspecification of conditional moments, but it also detects any relationship between pricing errors and conditioning variables. This last issue is of crucial interest for power in testing conditional models.

Our test statistic belongs to the class of consistent specification tests that compare parametric and non-parametric fits. However, our model under the null is also nonparametric, which is a crucial difference with respect to the standard tests in the literature. We develop the asymptotic theory of the test, and also make special emphasis on practical issues. The test distribution is estimated using wild bootstrap, and calibration and power are obtained by a proper combination of the bandwidth choices. Our test works reasonably well even for small samples in a Monte Carlo exercise.

In our empirical application, we evaluate several asset pricing models against the cross-section of stock returns. We find strong empirical evidence against the conditional CAPM, where the only relevant pricing factor is the market return. We do not find evidence against zero pricing errors with models that consider consumption growth, but the low pricing errors do not necessarily mean that consumption is an economically meaningful pricing factor. The lack of rejection seems to be due to the poor conditional correlation between consumption and the cross-section of stock returns.

Finally, there are several interesting avenues for future research. In particular, recently there has been interest in testing asset pricing models with individual stocks instead of portfolios.\textsuperscript{8}

\textsuperscript{8}In fact, for several bootstrap samples, we obtain conditional covariances equal to zero for some $z$-values, leading to numerical break-downs.
We plan to extend our econometric methodology to the case of a number of assets that grows without bound.

References


Horowitz, J. and V. Spokoiny (2001). An adaptive, rate-optimal test of a parametric mean-


Lettau, M. and S. Ludvigson (2001b). Resurrecting the (c)capm: A cross-sectional test when


Rodriguez-Poo, J. M., S. Sperlich, and P. Vieu (2015). Specification testing when the null is

Roussanov, N. (2014). Composition of wealth, conditioning information, and the cross-section

Sharpe, W. (1964). Capital asset prices: a theory of market equilibrium under conditions of


nance* 58, 161–196.

Sinica* 20, 469–479.
Appendices

A Proofs

In the following proofs, we denote by $a$ and $A$ a real valued vector and matrix, respectively. We denote by $\|a\|_W^2 = a' W a$ and $\|A\|_{W,F}^2 = \text{tr} (A' W A)$.

A.1 Proof of Theorem 1

From (17), premultiplying by $D(z)' W(z) D(z)$ and its inverse we can obtain

$$
\hat{\delta}_{g,h_1}(z) = \left( (D(z)' W(z) D(z))^{-1} D_g(z)' W(z) D_g(z) \right)^{-1} \times \\
\left( D(z)' W(z) D(z) \right)^{-1} D_g(z)' W(z) \hat{\mu}_{h_1}(z) \\
\equiv A_{1T}^{-1}(z) \times A_{2T}(z). 
$$

Under assumptions (C.1) and (C.4) and using a matrix inverse formulae such as those in page 33 of Rao (1973) we have that, if $g \to 0$ such that $Tg^d \to \infty$, as $T$ tends to infinity

$$
A_{1T}^{-1}(z) = I_N + O_p \left( \sqrt{\log T \over Tg^d} \right) + O_p (g^{\alpha_1}),
$$

uniformly for any $z \in \mathcal{Z}$. Furthermore, note that under assumptions (C.2) and (C.3) if both $g$ and $h_1$ tend to zero in such a way that $Tg^d \to \infty$ and $Th_1^d \to \infty$, as $T$ tends to infinity

$$
D_g(z)' W(z) \hat{\mu}_{h_1}(z) = D(z)' W(z) \mu(z) \\
+ O_p \left( \sqrt{\kappa_T^{-1} \log T \over Tg^d} + \kappa_T^{-1} \sqrt{\log T \over Th_1^d} \right) + O_p \left( \kappa_T^{-1} g^{\alpha_1} + \kappa_T^{-1} h_1^{\alpha_2} \right),
$$

uniformly in $z \in \mathcal{Z}$. Finally, noting that under the null hypothesis $\mu(z) = D(z) \delta(z)$ the proof is closed. 

A.2 Proof of Theorem 2

From (27),

$$
Th_2^{d/2} S_{1T}(h_1, h_2) = Th_2^{d/2} S_{1T}'(h_1, h_2) + Th_2^{d/2} S_{2T}(h_1, h_2) + Th_2^{d/2} S_{3T}'(h_1, h_2) 
$$

(35)
where

\[ S'_{1T} (h_1, h_2) = \frac{1}{T^2} \int \left[ \sum_t K'_{h_2} (z - z_t) \varepsilon_{t+1} \right]' W(z) \left[ \sum_t K'_{h_2} (z - z_t) \varepsilon_{t+1} \right] \pi_2 (z) \, dz, \]

\[ S'_{2T} (h_1, h_2) = \frac{2}{T^2} \int \left[ \sum_t K'_{h_2} (z - z_t) \varepsilon_{t+1} \right]' W(z) \]
\[ \times \left[ \sum_t K'_{h_2} (z - z_t) \left( e(z_t; \delta (z_t)) - \left( D_g (z_t) \delta_{h_1} (z_t) - D (z_t) \delta (z_t) \right) \right) \right] \pi_2 (z) \, dz, \]

\[ S'_{3T} (h_1, h_2) = \frac{1}{T^2} \int \left[ \sum_t K'_{h_2} (z - z_t) \left( e(z_t; \delta (z_t)) - \left( D_g (z_t) \delta_{h_1} (z_t) - D (z_t) \delta (z_t) \right) \right) \right]' W(z) \]
\[ \times \left[ \sum_t K'_{h_2} (z - z_t) \left( e(z_t; \delta (z_t)) - \left( D_g (z_t) \delta_{h_1} (z_t) - D (z_t) \delta (z_t) \right) \right) \right] \pi_2 (z) \, dz. \]

The proof of this result will be divided in two parts. First we show that, under the conditions established in the Theorem, as \( T \) tends to infinity,

\[ Th_2^{d/2} S'_{1T} (h_1, h_2) - Bh_2^{-d/2} \rightarrow_d N(0, V), \quad (36) \]

where

\[ B = \int_{\mathbb{R}^d} K^2 (u) \, du \times \sum_i \sum_j \int_{\mathbb{R}^d} w_{ij} (z) E (\varepsilon_{i,t+1} \varepsilon_{j,t+1} | z) p (z) \pi_2 (z) \, dz \]

and

\[ V = 2 \int_{\mathbb{R}^{3d}} K (u) K (v) K (u - x) K (v - x) \, du \, dv \, dx \]
\[ \times \sum_i \sum_j \int_{\mathbb{R}^d} w_{ij} (z) E (\varepsilon_{i,t+1}^2 | z) E (\varepsilon_{j,t+1}^2 | z) p^2 (z) \pi_2 (z) \, dz. \]

Second, we show that, under the null hypothesis, the other two remaining terms in the left hand side of (35) are of order \( o_p (1) \). In order to show (38) note that \( S'_{1T} (h_1, h_2) \) can be written as

\[ S'_{1T} (h_1, h_2) = \frac{1}{T^2} \sum_{i,j} \sum_t \int K_{h_2}^2 (z - z_t) w_{ij} (z) \varepsilon_{i,t+1} \varepsilon_{j,t+1} \pi_2 (z) \, dz \]
\[ + \frac{2}{T(T - 1)} \sum_{i,j} \sum_t \sum_{s < t} \int K_{h_2} (z - z_t) K_{h_2} (z - z_s) w_{ij} (z) \varepsilon_{i,t+1} \varepsilon_{j,s+1} \pi_2 (z) \, dz. \quad (37) \]

By the Ergodic Theorem, the first term of the right hand side in (37), multiplied by \( Th_2^{d/2} \), is equal to

\[ h_2^{-3d/2} \sum_{i,j} E \left\{ \int K_{h_2}^2 (z - z_t) w_{ij} (z) E (\varepsilon_{i,t+1} \varepsilon_{j,t+1} | Z_t) \pi_2 (z) \, dz \right\} + O_p \left( \frac{1}{\sqrt{Th_2^d}} \right). \]

The second term in the right hand side of (37) has the structure of a centered U-statistic (it has zero mean because of (A.3)) and therefore, under assumptions (A.1) to (A.5), by Theorem A of
Furthermore, applying (41) assumption (C.1) and Theorem 1 we obtain the following bound

$$Th_2^{d/2} \left( S_{ST}' (h_1, h_2) - E [S_{ST}' (h_1, h_2)] \right) \longrightarrow_d N(0, V), \quad (38)$$

as $T$ tends to infinity. Next, we show that, under the null hypothesis, i.e. $e (z_t; \delta (z_t)) = 0$,

$$Th_2^{d/2} S_{ST}' (h_1, h_2) = o_p(1), \quad (39)$$

as $T$ tends to infinity. In order to show this, note that under the null hypothesis,

$$S_{ST}' (h_1, h_2) = \int \left[ \sum_t K_{h_2} (z - z_t) \left( D_g (z_t) \hat{\delta}_{h_1} (z_t) - D (z_t) \delta (z_t) \right) \right] W(z) \quad (40)$$

then

$$\left| S_{ST}' (h_1, h_2) \right| \leq C \sup_{z \in \mathbb{R}} \left\| D_g (z_t) \hat{\delta}_{h_1} (z) - D (z_t) \delta (z) \right\| W(z) \int \left( \frac{1}{T} \sum_t \left| K_{h_2} (z - z_t) \right| \right)^2 \pi_2 (z) dz.$$ 

By (A.1), (A.2) and a strong law of large numbers we obtain

$$\left| S_{ST}' (h_1, h_2) \right| \leq C \sup_{z \in \mathbb{R}} \left\| D_g (z_t) \hat{\delta}_{h_1} (z) - D (z_t) \delta (z) \right\| W(z) \int p^2 (z) \pi_2 (z) dz + o_p(1). \quad (41)$$

Furthermore, applying (41) assumption (C.1) and Theorem 1 we obtain the following bound

$$\left| S_{ST}' (h_1, h_2) \right| = O \left( \kappa_T^{-2} \log T \frac{g^d}{T} + \kappa_T^{-2} \log T \frac{T}{Th_1^d} \right) + O \left( \kappa_T^{-2} g^{2\alpha_1} + \kappa_T^{-2} h_1^{2\alpha_2} \right).$$

Applying conditions (A.6) and (A.7) the proof of (39) is closed. Now, to close the proof of Theorem 2 we need to show that $Th_2^{d/2} S_{ST}' (h_1, h_2) = o(1)$. Under the null hypothesis, i.e. $e (z_t; \delta (z_t)) = 0$

$$S_{ST}' (h_1, h_2) = -2 \int \left[ \frac{1}{T} \sum_t K_{h_2} (z - z_t) \varepsilon_{t+1} \right] W(z) \quad (42)$$

$$\times \left[ \frac{1}{T} \sum_t K_{h_2} (z - z_t) \left( D_g (z_t) \hat{\delta}_{h_1} (z_t) - D (z_t) \delta (z_t) \right) \right] \pi_2 (z) dz.$$ 

Using assumption (C.1) and Theorem 1 we have that

$$\left| S_{ST}' (h_1, h_2) \right| \leq C \int \left[ \left\| \frac{1}{T} \sum_t K_{h_2} (z - z_t) \varepsilon_{t+1} \right\| \right]^2 W(z) \left( \frac{1}{T} \sum_t K_{h_2} (z - z_t) \right) \pi_2 (z) dz \times O \left( \kappa_T^{-1} \frac{\log T}{T g^d} + \sqrt{\kappa_T^{-1} \frac{\log T}{T h_1^d}} + \kappa_T^{-1} g^{\alpha_1} + \kappa_T^{-1} h_1^{\alpha_2} \right). \quad (43)$$

28
Now we split (43) in the following two terms

\[
\frac{C'}{T^2} \int \left[ \sum_t K_{h_2} (z - z_t) \varepsilon_{t+1} \right]' W(z) \\
\times \left( \sum_t \{ K_{h_2} (z - z_t) - E [K_{h_2} (z - z_t)] \} \right) \pi_2 (z) \, dz, \tag{44}
\]

and

\[
\frac{C''}{T} \int \left[ \sum_t K_{h_2} (z - z_t) \varepsilon_{t+1} \right]' W(z) E [K_{h_2} (z - z_t)] \pi_2 (z) \, dz, \tag{45}
\]

such that \( C, C', C'' > 0 \) are constants. By the Cauchy-Schwarz inequality, the expectation of the absolute value of (44) is bounded by

\[
\frac{C'}{T^2} \int \left\{ E \left[ \sum_t K_{h_2} (z - z_t) \varepsilon_{t+1} \right] \right\}^2 \pi_2 (z) \, dz \\
\times E \left( \sum_t \{ K_{h_2} (z - z_t) - E [K_{h_2} (z - z_t)] \} \right)^2 \frac{1}{2} \pi_2 (z) \, dz. \tag{46}
\]

Assumption (A.3) implies that

\[
E \left[ \sum_t K_{h_2} (z - z_t) \varepsilon_{t+1} \right]^2 \pi_2 (z) \, dz = O \left( T h_2^{-d} \right), \tag{47}
\]

and because of (A.1) and the fact that absolute regularity with geometrically decaying mixing coefficients implies strong mixing with mixing coefficients decaying at the same rate by applying the covariance inequality for strong mixing processes (Corollary 1.1 from Bosq (1996)), we have that

\[
E \left( \sum_t \{ K_{h_2} (z - z_t) - E [K_{h_2} (z - z_t)] \} \right)^2 \leq O (T) \times \left\{ E (K_{h_2} (z - z_t) \varepsilon_{t+1}) \right\}^{2/2+\delta} \\
= O \left( T h_2^{-d(1+\delta)} \right). \tag{48}
\]

Then (44) is of order \( O_p \left( T^{-1} h_2^{-d(2+\delta)/2} \right) \). It is also easy to check that the expectation of the square of (45) is of order \( O \left( T^{-1} h_2^{-d} \right) \). Under assumption (A.5) then (44) is \( o(1) \), as \( T \) tends to infinity. Assumption (A.5) again guarantees that (45) is also is \( o(1) \), as \( T \) tends to infinity. Therefore, under the assumptions of the Theorem we have proved that both (44) and (45) converge to zero in first and second moments and hence \( T h_2^{-d/2} |S_{ST}^T (h_1, h_2)| = o(1) \) as \( T \) tends to infinity. This closes the proof.
A.3 Proof of Theorem 3

Note that under the alternative hypothesis $H_1$, and under conditions of Theorem 1 we have

$$\sup_{x \in X} |\delta_{b_1}(x) - \delta(x)| = \gamma_T \left( \mathbf{D}(z)' \mathbf{W}(z) \mathbf{D}(z) \right)^{-1} \mathbf{D}(z)' \mathbf{W}(z) \Delta(z) + O \left( \kappa_T^{-1} \left( \frac{\log T}{Th_2^2} + \gamma_T^{-1} \sqrt{\frac{\log T}{Th_2^2}} \right) + O \left( \kappa_T^{-1} g^{a_1 + \kappa_T^{-1} h_1^2} \right). \quad (49)$$

Decomposing again $S_{1T}(h_1, h_2)$ as in (35), note that $S'_{1T}(h_1, h_2) = \frac{1}{Th_2^2} B + O_p \left( T^{-1/4} h_2^{-d/2} \right)$ because of (38). Now, under $H_1$ and assumptions (A.6) and (A.7)

$$e(z; \delta(z)) - \left( \mathbf{D}(z) \delta_{h_1}(z) - \mathbf{D}(z) \delta(z) \right) = \gamma_T \Delta(z) - \gamma_T \left\{ \mathbf{D}(z) \left( \mathbf{D}(z)' \mathbf{W}(z) \mathbf{D}(z) \right)^{-1} \times \mathbf{D}(z)' \mathbf{W}(z) \Delta(z) + o_p \left( T^{-1/2} h_2^{-d/2} \right) \right\}. \quad (50)$$

Therefore if we substitute (50) into (35) by the triangle inequality we have

$$\left| S'_{2T}(h_1, h_2) \right| \leq C \gamma_T^2 \int \Delta(z)' \mathbf{W}(z)' \left( \mathbf{W}(z)^{-1} - \mathbf{D}(z) \left( \mathbf{D}(z)' \mathbf{W}(z) \mathbf{D}(z) \right)^{-1} \mathbf{D}(z)' \right) \times \mathbf{W}(z) \Delta(z) p_2(z) \pi_2(z) dz + o_p \left( T^{-1/2} h_2^{-d/2} \right). \quad (51)$$

Then, assuming (C.3), (C.4), (C.5) and (H.2), we have that $|S'_{2T}(h_1, h_2)| = O_p \left( \gamma_T^2 \right)$, as $T$ tends to infinity.

Finally, applying the Cauchy-Schwarz inequality to $S'_{2T}(h_1, h_2)$, we obtain

$$\left| S'_{2T}(h_1, h_2) \right| \leq \left[ S_{2T}^2(h_1, h_2) \right]^{1/2} \left[ S'_{2T}(h_1, h_2) \right]^{1/2},$$

and by (38) and (51) we have $S'_{2T}(h_1, h_2) = O_p \left( \gamma_T / \sqrt{Th_2^2} \right)$. Now, note that assuming (H.1) $S_{1T}(h_1, h_2) = o \left( \left| S'_{2T}(h_1, h_2) \right| \right)$ and $S'_{1T}(h_1, h_2) = o \left( \left| S_{2T}(h_1, h_2) \right| \right)$. The proof is closed by noting that under assumptions (H.1) and (A.5), $Th_{2}^{d/2} |S_{2T}(h_1, h_2)| \rightarrow \infty$ because $\sqrt{T} \gamma_T \rightarrow \infty$.

A.4 Proof of Theorem 4

The proof of this result follows the same lines as in the proof of Theorem 2 of Kreiss et al. (2008), p. 379. Indeed, it easy to show that $Th_{2}^{d/2} \left[ S'_{1T} - E^* \left( S'_{1T} \right) \right] \rightarrow_d N(0, V^*)$ and $V^* = V + o_p(1)$. Unfortunately, the result of $Th_{2}^{d/2} \left[ E \left( S_{1T} \right) - E^* \left( S_{1T} \right) \right] \rightarrow_p 0$ is harder to show. Note that

$$E^* \left[ S_{1T}^*(h_1, h_2) \right] = \frac{1}{T^2} \int \left[ \sum_{t} K_{h_2} (z - z_t) (I - P_{g,h_1}(z_t)) \tilde{e}_{t+1} \right]' \mathbf{W}(z) \quad (52)$$

$$\times \left[ \sum_{t} K_{h_2} (z - z_t) (I - P_{g,h_1}(z_t)) \tilde{e}_{t+1} \right].$$
By standard properties of smoothing techniques, applying assumptions (C.2), (A.5), (A.6) and (A.7)

$$H_{h_1}(z) \hat{\varepsilon}_{t+1} = O_p \left( \frac{1}{T h_1^d} \right) + o_p \left( \frac{1}{T h_1^d} \right), \quad (53)$$

as $T$ tends to infinity, uniformly in $z \in \mathcal{Z}$. Furthermore, applying conditions (C.1)-(C.5) and (A.5)-(A.7), and following exactly the same lines as in the proof of Theorem 1 we obtain

$$\mathbf{P}_{g,h_1}(z) \hat{\varepsilon}_{t+1} := \mathbf{D}_g(z) \left( \mathbf{D}_g(z)^T \mathbf{W}(z) \mathbf{D}_g(z) \right)^{-1} \mathbf{D}_g(z) \mathbf{W}(z) \mathbf{H}(z)_{h_1} \hat{\varepsilon}_{t+1} = o_p \left( \frac{1}{T h_1^d} \right), \quad (54)$$

uniformly in $z \in \mathcal{Z}$. Now, substituting (54) into (52) and applying a triangle inequality we obtain

$$E^* \left[ S_{1T}^* (h_1, h_2) \right] = \frac{1}{T^2} \int \left[ \sum_t K_{h_2} (z - z_t) \hat{\varepsilon}_{t+1} \right]' \mathbf{W}(z) \times \left[ \sum_t K_{h_2} (z - z_t) \hat{\varepsilon}_{t+1} \right] + o_p \left( \frac{1}{T h_1^d} \right). \quad (55)$$

Making a standard decomposition

$$\hat{\varepsilon}_{t+1} = \varepsilon_{t+1} - \{ \hat{\mu}_{h_1} (z_t) - \mu (z_t) \}, \quad (56)$$

and substituting (56) into (55) we have

$$E^* \left[ S_{1T}^* (h_1, h_2) \right] = \frac{1}{T^2} \sum_i \sum_j \int K_{h_2}^2 (z - z_t) w_{ij} (z) \varepsilon_{i,t+1} \varepsilon_{j,t+1} \pi_2 (z) dz \quad (57)$$

$$+ \frac{1}{T^2} \sum_i \sum_j \int K_{h_2}^2 (z - z_t) w_{ij} (z) \{ \hat{\mu}_{i,h_1} (z_t) - \mu_i (z_t) \} \pi_2 (z) dz \quad (58)$$

$$+ \frac{1}{T^2} \sum_i \sum_j \int K_{h_2}^2 (z - z_t) w_{ij} (z) \varepsilon_{i,t+1} \pi_2 (z) dz + o_p \left( 1/T h_1^d \right) \quad (59)$$

By the ergodic theorem $Th_1^d \times (57)$ is equal to $h_2^{-d/2} B + O_p \left( 1/T h_1^d \right)$ where

$$B = \int_{\mathbb{R}^d} K^2 (u) d\mathbf{u} \times \sum_i \sum_j \int_{\mathbb{R}^d} w_{ij} (z) E \left( \varepsilon_{i,t+1} \varepsilon_{j,t+1} | z \right) p(z) \pi_2 (z) dz. \quad (60)$$

By assumption (C.2), $Th_2^{-d/2} \times (58)$ is of order $O_p \left( h_2^{-d/2} \left[ \frac{\log T}{Th_1^d} + h_1^{2\alpha_2} \right] \right)$. But this is $o_p(1)$ applying assumptions (A.6) and (A.7). Finally, applying assumptions (C.2) and (C.5) we have that (59) is bounded by

$$\left\{ \max_j \sup_{z \in \mathcal{Z}} | \hat{\mu}_{j,h_1} (z) - \mu_j (z) | \right\} \left\{ \max_i \sup_{z \in \mathcal{Z}} | w_{ij} (z) | \right\} \frac{N}{T^2} \sum_i \sum_j \int K_{h_2}^2 (z - z_t) \varepsilon_{i,t+1} \pi_2 (z) dz,$$
but, noting that \( N \) is fixed, by applying (C.2), (C.5) and a strong law of large numbers the previous bound becomes of order
\[
O \left( \sqrt{\frac{\log T}{Th_1^N}} + h_1^{\alpha_2} \right) \times O_p \left( T^{-1} h_2^{-d} \right).
\]
This is \( o_p(1/Th_2^d) \) just because of assumptions (A.5), (A.6) and (A.7).

\[\blacksquare\]

### A.5 Proof of Theorem 5

This Theorem follows immediately from Theorems 2 and 4.

\[\blacksquare\]

### B Centred SDF

Section 2.1 introduces the uncentred and centred SDF variants in equations (2) and (3), respectively, while Section 2.2 describes the choice of prices of risk for an uncentred SDF. In particular, we minimize a quadratic form \( q(z_t; \delta(z_t)) \) in pricing errors (8), obtaining the prices of risk \( \delta^*(z_t) \) in (10). The minimised criterion becomes
\[
q^*(z_t) = q(z_t; \delta^*(z_t))
\]
\[
= \mu(z_t)' W(z_t) \mu(z_t) - (\mu(z_t)' W(z_t) D(z_t)) (D(z_t)' W(z_t) D(z_t))^{-1} (D(z_t)' W(z_t) \mu(z_t)).
\]

In this Appendix, we relate \( q^*(z_t) \) and the prices of risk \( \delta^*(z_t) \) to their centred SDF counterparts.

### B.1 Choice of Prices of Risk

The centred variant of the SDF approach demean the factors in the SDF, or equivalently measures risks or factor sensitivities as
\[
C(z_t) = Cov(r_{t+1}, f_{t+1} | z_t).
\]
The corresponding pricing errors are
\[
\epsilon(z_t) = \mu(z_t) - C(z_t) \tau(z_t)
\]
for some vector \( \tau(z_t) \) of prices or risk. The common criterion to choose this vector is
\[
q(z_t; \tau(z_t)) = \epsilon(z_t; \tau(z_t))' W(z_t) \epsilon(z_t; \tau(z_t)),
\]

32
for some weighting matrix $\mathcal{W}(z_t)$.

Let us denote $\tau^*(z_t)$ the prices of risk that minimise $q(z_t; \tau(z_t))$. The first order conditions that pin down $\tau^*(z_t)$ are

$$C(z_t)' \mathcal{W}(z_t) \epsilon(z_t; \tau^*(z_t)) = 0,$$

which yield

$$\tau^*(z_t) = \left( C(z_t)' \mathcal{W}(z_t) C(z_t) \right)^{-1} \left( C(z_t)' \mathcal{W}(z_t) \mu(z_t) \right).$$

The minimised criterion becomes

$$q^*(z_t) = q(z_t; \tau^*(z_t))$$

$$= \mu(z_t)' \mathcal{W}(z_t) \mu(z_t) - \left( \mu(z_t)' \mathcal{W}(z_t) C(z_t) \right) \left( C(z_t)' \mathcal{W}(z_t) C(z_t) \right)^{-1} \left( C(z_t)' \mathcal{W}(z_t) \mu(z_t) \right).$$

Two common choices of $\mathcal{W}(z_t)$ are the identity matrix and $\mathcal{W}(z_t) = \text{Var}^{-1}(r_{t+1} | z_t)$.

The latter matrix is associated to a conditional counterpart of the Hansen-Jagannathan distance, and gives simple interpretations to the previous objects. The first order conditions price the centred factor mimicking portfolios, and the chosen prices of risk can be expressed as the inverse of the variance matrix of these portfolios times their mean vector. The minimised criterion becomes the difference between the maximum conditional Sharpe ratio that can be achieved from the span of $r_{t+1}$ and the span of the factor mimicking portfolios.

**B.2 From Uncentred to Centred SDF**

We develop our testing methodology for an uncentred SDF, but a researcher that wants to work with a centred SDF can translate our results into that setting.

The uncentred SDF variant

$$m_{t+1} = 1 - f_{t+1}' \delta(z_t)$$

has a conditional mean

$$c(z_t) = 1 - \nu(z_t)' \delta(z_t),$$

where $\nu(z_t)$ is the vector of factor means.

The centred SDF variant can be expressed as

$$m_{t+1} = a(z_t) - f_{t+1}' \tau(z_t),$$

where the conditional intercept keeps a unit SDF conditional mean

$$a(z_t) = 1 + \nu(z_t)' \tau(z_t).$$
Therefore, each uncentred SDF has an implied centred SDF, except in the special case $c(z_t) = 0$ studied in Section 2.4. If we divide the uncentred SDF by its mean $c(z_t)$, then we obtain the implied centred SDF. The relationship between the prices of risk is

\[ \tau(z_t) = \frac{1}{c(z_t)} \delta(z_t). \] (61)

However, the researcher may not be interested in the implied centred SDF, but in the centred SDF given by some chosen weighting matrix. In that case, the relationship may be more complex, but we can still relate both SDFs.

As an example, let us show their connection when both variants use the identity matrix to weight the pricing errors. For that choice, the uncentred SDF prices of risk in (10) become

\[ \delta^* (z_t) = (D(z_t)'D(z_t))^{-1} \left( D(z_t)' \mu(z_t) \right), \]

and the uncentred SDF criterion evaluated at those prices of risk is

\[ q^* (z_t) = \mu(z_t)' \mu(z_t) - \left( \mu(z_t)' D(z_t) \right) \left( D(z_t)' D(z_t) \right)^{-1} \left( D(z_t)' \mu(z_t) \right). \]

A researcher interested in the centred SDF variant with the same weighting matrix wants to compute the prices of risk

\[ \tau^* (z_t) = (C(z_t)' C(z_t))^{-1} \left( C(z_t)' \mu(z_t) \right), \]

and the criterion

\[ q^* (z_t) = \mu(z_t)' \mu(z_t) - \left( \mu(z_t)' C(z_t) \right) \left( C(z_t)' C(z_t) \right)^{-1} \left( C(z_t)' \mu(z_t) \right). \]

Let us define the following linear transformation of the factor means

\[ \eta(z_t) = (D(z_t)'D(z_t))^{-1} \nu(z_t), \]

and the associated scalar function

\[ d(z_t) = \nu(z_t)' \eta(z_t). \]

We can prove that the centred SDF prices of risk can be computed as a linear combination of the uncentred SDF prices of risk and the transformed factor means

\[ \tau^* (z_t) = a(z_t) \delta^* (z_t) - q^* (z_t) \eta(z_t), \]

where the coefficients are directly the centred SDF intercept and criterion. In addition, the
centred SDF intercept is related to the uncentred SDF mean by
\[
a(z_t) = \frac{c(z_t)}{c^2(z_t) + d(z_t) q^*(z_t)},
\]
and the centred SDF criterion is related to the uncentred SDF criterion by
\[
q^*(z_t) = \frac{q^*(z_t)}{c^2(z_t) + d(z_t) q^*(z_t)}.
\]
When the asset pricing model holds, both SDF criteria are zero
\[
q^*(z_t) = q^*(z_t) = 0,
\]
and the relationship between the prices of risk is simply
\[
\tau^*(z_t) = \frac{1}{c(z_t)} \delta^*(z_t).
\]
That is, this centred SDF is equal to the implied one that we defined in (61).

C Monte Carlo Design

C.1 Pricing Factors and Predictor

Following the empirical application, we think of the market return and consumption growth as our factors \(g_{t+1}\) in the data. However, to simplify the DGP, the vector \(f_{t+1}\) that we simulate represents the corresponding conditional Cholesky orthogonalisation, as we describe below.

We simulate the two factors from their conditional means plus the corresponding forecast error
\[
\begin{align*}
    f_{1t+1} &= \nu_1(z_t) + \epsilon_{1t+1}, \\
    f_{2t+1} &= \nu_2(z_t) + \epsilon_{2t+1},
\end{align*}
\]
where \((\epsilon_{1t+1}, \epsilon_{2t+1})\) are two uncorrelated standard normal shocks.

The main focus of the simulation study is to generate data from a DGP calibrated along our real data set. Therefore, if the original factors have conditional variance
\[
Var(g_{t+1} \mid z_t) = \begin{pmatrix}
    \sigma_{11}(z_t) & \sigma_{12}(z_t) \\
    \sigma_{12}(z_t) & \sigma_{22}(z_t)
\end{pmatrix},
\]
then the DGP factor means are
\[
\begin{pmatrix}
    \nu_1(z_t) \\
    \nu_2(z_t)
\end{pmatrix} = \begin{pmatrix}
    \frac{1}{\sqrt{\sigma_{11}(z_t)}} E(g_{1t+1} \mid z_t) \\
    \frac{1}{\sqrt{\sigma_{22}(z_t)}} \left[ E(g_{2t+1} \mid z_t) - \frac{\sigma_{12}(z_t)}{\sigma_{11}(z_t)} E(g_{1t+1} \mid z_t) \right]
\end{pmatrix},
\]
and
\[
\omega_{22}(z_t) = \sigma_{22}(z_t) - \frac{\sigma_{12}^2(z_t)}{\sigma_{11}(z_t)}.
\]
Note that, when the first factor is traded, that mean is the original Sharpe ratio of the factor. Finally, we think of the predictor as the logarithm of the default spread. We use the following AR(1) for the predictor

\[ z_{t+1} = v + \rho (z_t - v) + \sigma \left[ \rho_1 \epsilon_{1t+1} + \sqrt{1 - \rho_1^2} \epsilon_{2t+1} \right], \]

where \( \rho \) is the autocorrelation of the predictor, \( v \) is the steady state mean, and \( \sigma / \sqrt{1 - \rho^2} \) is the steady state standard deviation. The parameter \( \rho_1 \) is the conditional correlation between the predictor and the first factor shocks. We assume constant conditional second moments to simplify. We also focus on the correlation with the first factor because in the empirical application we use the market portfolio and predictors associated to that payoff.

C.2 Excess Returns

Let us think of SMB and HML as our original returns \( r_{o,t+1} \), in addition to the market return. Our total payoffs in the data are \( x_{t+1} = (r_{o,t+1}, g_{1t+1}) \). The vector \( r_{t+1} \) that we simulate represents some transformations of those returns that keep the same payoff space but simplify the DGP.

We simulate those returns from their conditional least squares projections onto the span of the factors and a constant, plus the corresponding projection errors. We apply two types of transformations. First, the two returns are uncorrelated to the market portfolio. Second, we can perform Cholesky and Householder transformations of those returns to obtain the simple structure

\[ r_{1t+1} = \mu (z_t) + \beta_1 (z_t) \epsilon_{2t+1} + u_{1t+1}, \]

\[ r_{2t+1} = \beta_2 (z_t) \epsilon_{2t+1} + u_{2t+1}, \]

where \( (u_{1t+1}, u_{2t+1}) \) are two uncorrelated standard normal shocks. If we want more excess returns in our Monte Carlo, then we simply need to add returns that are equal to the corresponding normal shocks.

To simulate the excess returns, we only need to estimate the three functions \( \mu (z_t), \beta_1 (z_t), \) and \( \beta_2 (z_t) \) from the real data. We will do so by means of three functions that are easy to obtain and interpret: the maximum Sharpe ratio from the three excess returns, the variance of the second factor mimicking portfolio, and the Hansen-Jagannathan distance.

We can obtain these functions from the original data as follows. As the span of \( (r_{t+1}, f_{1t+1}) \) is equal to the span of \( x_{t+1} \), we can estimate the maximum Sharpe ratio from the latter

\[ S^2 (z_t) = E (x_{t+1} \mid z_t)' Var^{-1} (x_{t+1} \mid z_t) E (x_{t+1} \mid z_t) . \] (64)
We can also estimate the variance of the second factor mimicking portfolio in the DGP from
\[
c_2(z_t) = \frac{1}{\omega_{22}(z_t)} \left[ \text{Var}(s_{2t+1} | z_t) - \sigma_{12}^2(z_t) \sigma_{11}(z_t) \right], \tag{65}
\]
where \( s_{2t+1} \) is the factor mimicking portfolio of the second factor in our data, and hence
\[
\text{Var}(s_{2t+1} | z_t) = \text{Cov}(x_{t+1}, g_{2t+1} | z_t) \text{Var}^{-1}(x_{t+1} | z_t) \text{Cov}(x_{t+1}, g_{2t+1} | z_t). \tag{66}
\]

The quadratic form of the centred SDF pricing errors with a weighting matrix equal to the inverse of the return variance, which can be interpreted as a squared Hansen-Jagannathan distance, is
\[
q^*(z_t) = E(x_{t+1} | z_t)' \text{Var}^{-1}(x_{t+1} | z_t) E(x_{t+1} | z_t) - \left[ E(x_{t+1} | z_t)' \text{Var}^{-1}(x_{t+1} | z_t) \text{Cov}(x_{t+1}, g_{t+1} | z_t) \right]^{-1}
\times \left[ \text{Cov}(x_{t+1}, g_{t+1} | z_t)' \text{Var}^{-1}(x_{t+1} | z_t) E(x_{t+1} | z_t) \right]. \tag{66}
\]

Of course, we can choose \( q^*(z_t) \) with the shape that we prefer to study power, leaving the previous empirical consideration as a reference.

Note that the three functions are based on centered second moments, even though our empirical methods are applied to an uncentred SDF. This is natural in our DGP because we simulate returns from their least squares projections onto the factors and, as it is common, a constant. Let us compute these three functions for our DGP.

We can show that the corresponding quadratic form of the pricing errors is
\[
q^*(z_t) = \mu^2(z_t) \frac{\beta_2^2(z_t)}{\beta_1^2(z_t) + \beta_2^2(z_t)},
\]
at the minimum. Therefore, if we assume \( \mu(z_t) \neq 0 \) in this setting, a proper 2-factor SDF can price the 3 assets iff
\[
\beta_2(z_t) = 0.
\]

Similarly, in our DGP, the maximum Sharpe ratio from the three excess returns, and the variance of the second factor mimicking portfolio are
\[
S^2(z_t) = \mu^2(z_t) \frac{1 + \beta_2^2(z_t)}{1 + \beta_1^2(z_t) + \beta_2^2(z_t)} + \nu_1^2(z_t),
\]
\[
c_2(z_t) = \frac{\beta_1^2(z_t) + \beta_2^2(z_t)}{1 + \beta_1^2(z_t) + \beta_2^2(z_t)}. \tag{66}
\]

Now we can obtain the three DGP functions of the excess returns by inverting the previous
equations

\[
\beta_2^2 (z_t) = \frac{q^* (z_t) c_2 (z_t)}{\theta (z_t) - q^* (z_t) c_2 (z_t)},
\]

\[
\beta_1^2 (z_t) = \frac{c_2 (z_t)}{1 - c_2 (z_t) \theta (z_t) - q^* (z_t) c_2 (z_t)},
\]

\[
\mu^2 (z_t) = \frac{\theta (z_t) - q^* (z_t) c_2 (z_t)}{1 - c_2 (z_t)},
\]

where

\[
\theta (z_t) = S^2 (z_t) - \nu^2 (z_t).
\]

When the model holds, the previous expressions simplify to

\[
\beta_2^2 (z_t) = 0,
\]

\[
\beta_1^2 (z_t) = \frac{c_2 (z_t)}{1 - c_2 (z_t)},
\]

\[
\mu^2 (z_t) = \frac{\theta (z_t)}{1 - c_2 (z_t)}.
\]

Note that the chosen functions must satisfy \( c_2 (z_t) \in (0, 1) \) and \( \theta (z_t) > 0 \) for these equations to be well-defined. In addition, if we use the same \( c_2 (z_t) \) and \( \theta (z_t) \) under the alternative as the ones under the null hypothesis, then the chosen \( q^* (z_t) \) must satisfy

\[
\theta (z_t) > q^* (z_t)
\]

for these equations to be well-defined.

Finally, we also simulate under a traded one-factor SDF. We can show that the corresponding quadratic form of the pricing errors is

\[
q^* (z_t) = \mu^2 (z_t) \frac{\beta_2^2 (z_t) + 1}{\beta_1^2 (z_t) + \beta_2^2 (z_t) + 1} = \theta (z_t),
\]

at the minimum. Therefore, this SDF prices the 3 returns iff \( \mu^2 (z_t) = 0 \). Moreover, if we choose \( \beta_1^2 (z_t) = 0 \), then we have the simple relationship \( q^* (z_t) = \mu^2 (z_t) \).

C.3 Parameter Values

The conditioning variable \( lds \) has steady state (i.e. historical) mean \( v = -0.114 \), and a steady state standard deviation of 0.41. Its autocorrelation \( \rho = 0.915 \), and the correlation between its shocks and the market return shocks is \( \rho_1 = -0.007 \).

The functions \( \nu_1 \) and \( \nu_2 \) can directly be calculated from the data along (62). Similarly, \( S \) and \( c_2 \) can be calculated from (64) and (65), respectively. We approximate our nonparametric estimates
with the following polynomials, which are fitted by least squares,

\begin{align*}
\nu_1(z_t) &= 0.2010 + 0.1182 \cdot z_t + 0.4594 \cdot z_t^2 - 0.0043 \cdot z_t^3 - 0.4218 \cdot z_t^4 \\
\nu_2(z_t) &= 0.9285 - 0.5239 \cdot z_t + 0.5106 \cdot z_t^2 - 0.4151 \cdot z_t^3 - 0.5268 \cdot z_t^4 \\
c_2(z_t) &= 0.1087 - 0.0687 \cdot z_t + 0.2435 \cdot z_t^2 - 0.0565 \cdot z_t^3 + 0.0717 \cdot z_t^4 \\
S(z_t) &= 0.3862 + 0.2282 \cdot z_t + 0.1908 \cdot z_t^2 - 0.1531 \cdot z_t^3 + 0.2930 \cdot z_t^4
\end{align*}

For all models we have either \( q^*(z_t) = 0 \) under the Null, or (66) – calculated from the data – under the alternative. To generate the alternatives, we approximate our nonparametric estimates with

\begin{align*}
q^*(z_t) &= 0.1849 + 0.1231 \cdot z_t + 0.3658 \cdot z_t^2 - 0.2742 \cdot z_t^3 - 0.0742 \cdot z_t^4 \\
\end{align*}

for the two-factor model, and

\begin{align*}
q^*(z_t) &= 0.3233 + 0.1793 \cdot z_t + 0.0150 \cdot z_t^2 - 0.1203 \cdot z_t^3 + 0.4642 \cdot z_t^4 \\
\end{align*}

for the one-factor model.