I N S T I T U T   D E   S T A T I S T I Q U E
B I O S T A T I S T I Q U E   E T
S C I E N C E S   A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/35

Semiparametric Estimation of
Risk-return Relationships

ESCANCIANO, JC., PARDO-FERNANDEZ, JC. and I. VAN KEILEGOM

SEMIPARAMETRIC ESTIMATION OF RISK-RETURN RELATIONSHIPS

Juan Carlos Escanciano*
Indiana University

Juan Carlos Pardo-Fernández
Universidade de Vigo

Ingrid Van Keilegom
Université catholique de Louvain

August 18, 2013

Abstract

This article proposes semiparametric least squares estimation of parametric risk-return relationships, i.e. parametric restrictions between the conditional mean and the conditional variance of excess returns given a set of unobservable parametric factors. A distinctive feature of our estimator is that it does not require a parametric model for the conditional mean and variance. We establish consistency and asymptotic normality of the estimates. The theory is non-standard due to the presence of estimated factors. We provide simple sufficient conditions for the estimated factors not to have an impact in the asymptotic standard error of estimators. A simulation study investigates the finite sample performance of the estimates. Finally, an application to the CRSP value-weighted excess returns highlights the merits of our approach. In contrast to most previous studies using non-parametric estimates, we find a positive and significant price of risk in our semiparametric setting.

---
*Corresponding address: Indiana University, Department of Economics, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA, e-mail: jescanci@indiana.edu.

1 Introduction

The relation between expected excess return of the aggregate stock market (the so-called *equity premium*) and its conditional variance is one of the fundamental problems of finance and has been the subject of an extensive theoretical and empirical research. Such relationship is often suggested by economic and financial theories of optimal portfolio choice and asset pricing; see e.g. Cochrane (2005). These theories, however, do not suggest any parametric functional forms for the two conditional moments involved, which has hampered their empirical evaluation. The standard approach in the literature has been to assume parametric functional forms for the conditional mean and variance (e.g. Generalized Autoregressive Conditional Heteroskedastic, GARCH, models; see e.g. Bollerslev, Engle and Wooldridge, 1988, or Glosten, Jaganathan and Runkle, 1993) and to proceed estimating the risk-return relationships by the Quasi-Maximum Likelihood Estimator (QMLE). This popular approach, although simple, is not satisfactory, as it is not immune to misspecification of the conditional mean and variance, and it leads to inconsistent estimation if correct specification fails. In stark contrast, this article proposes semiparametric least squares estimation for risk-return restrictions, without the need to specify parametric models for the conditional mean and variance. Our methodology is flexible enough to cover many different specifications of the link between risk and return, as suggested by competing asset pricing and general equilibrium theories. [1]

Formally, let $\mu(I_{t-1}) = E(Y_t \mid I_{t-1})$ and $\sigma^2(I_{t-1}) = Var(Y_t \mid I_{t-1})$ be the conditional mean and variance, respectively, of the the time series of excess returns Y_t, given the information set at time $t-1$, say I_{t-1}.[2] The most prominent example of risk-return restriction has been the linear specification between expected excess returns and the conditional variance, with a positive constant slope, i.e.

$$\mu(I_{t-1}) = \theta_0 \sigma^2(I_{t-1}) \tag{1}$$

with $\theta_0 > 0$; see Merton (1973, 1980). Popular examples of this specification include some of the currently most successful asset pricing models, namely, the external habit model of Campbell and Cochrane (1999) and the long-run risk model proposed by Bansal and Yaron (2004). Much of the existing literature has been focussed on empirically assessing the sign of the slope coefficient θ_0, which under some conditions has a structural interpretation as the coefficient of relative risk aversion.[3]

Alternative asset pricing theories imply different restrictions between mean and variance. For instance, Gennotte and Marsh (1993) constructed a general equilibrium model of asset returns and derive the equilibrium condition $\mu(I_{t-1}) = \theta_{00} + \theta_{01}\sigma^2(I_{t-1}) + g(\sigma^2(I_{t-1}))$, where the functional form of

---

[1]See e.g. Merton (1973, 1980), Abel (1988, 1999), Cox, Ingersoll and Ross (1985), Gennotte and Marsh (1993) and Whitelaw (2000), to mention but a few.

[2]We drop almost surely (a.s.) in equalities involving conditional moments for notational simplicity.

[3]The extensive empirical evidence on the sign of θ_0 within the linear specification is mixed. Ghysels, Santa-Clara and Valkanov (2005), Lundblad (2007), Pastor, Sinha and Swaminathan (2008) and Ludvigson and Ng (2007) find a positive risk-return relation, while Campbell (1987), Pagan and Hong (1991), Glosten, Jagannathan and Runkle (1993), Harvey (2001) and Brandt and Kang (2004) find a negative relation. Still many others find inconclusive evidence, see e.g. French, Schwert and Stambaugh (1987), Nelson (1991), Campbell and Hentschel (1992), Linton and Perron (2003), Whitelaw (1994) and Ghosh and Linton (2009), to mention but a few. The conflicting empirical findings have been often attributed to the different approaches to modelling conditional variance. See Lettau and Ludvigson (2010) for a recent survey.

g depends on the representative agent's preferences and on the parameters of the distribution of asset returns. For instance, if the representative agent has logarithmic utility, then $g \equiv 0$, and the model reduces to Merton's (1973) model. Further evidence of other functional forms for the link between $\mu(I_{t-1})$ and $\sigma^2(I_{t-1})$ are given in Backus and Gregory (1993), Whitelaw (2000) and Veronesi (2001), among others. More recently, Yogo (2008) has suggested an asset pricing model with habit formation and reference-dependent preferences that implies $\mu(I_{t-1}) = \theta_0 \sigma(I_{t-1})$. The latter specification yields a constant Sharpe Ratio (SR) $S_{t-1} = \mu(I_{t-1})/\sigma(I_{t-1})$. Engle, Lilien and Robins (1987) examined several models and conclude that $\mu(I_{t-1}) = \theta_{00} + \theta_{01} \log \sigma(I_{t-1})$, with $\sigma(I_{t-1})$ following an ARCH model, provided the best fit accross several parametric specifications. Others have found evidence of time varying "price of risk" θ_0 or have included a hedging component $h(I_{t-1})$ modeling changes in the investment opportunity set, so that $\mu(I_{t-1}) = \theta_0 \sigma^2(I_{t-1}) + h(I_{t-1})$, see Guo and Whitelaw (2006) and Brandt and Wang (2010), among others. In these applications, the component h is often parametrically specified as a function of some factors. Since several competing models between $\mu(I_{t-1})$ and $\sigma^2(I_{t-1})$ are available, it seems that a general estimation method for parametric restrictions between risk and return is most welcome.

Arguably, the major difficulty in estimating risk-return relationships is that neither expected return nor conditional variances are observable. The vast majority of existing results rely on parametric assumptions for the first and second conditional moments.[4] There seems to be a consensus in the literature that conclusions are quite sensitive to parametric assumptions on $\mu(I_{t-1})$ and $\sigma^2(I_{t-1})$, see e.g. French, Schwert and Stambaugh (1987), Nelson (1991), Glosten, Jagannathan and Runkle (1993) and the extensive Monte Carlo analysis in Harvey (2001) and Scruggs and Glabadanidis (2003), among many others. Parametric assumptions are practically convenient, but they are likely to be affected by misspecification errors and lead to erroneous conclusions (i.e. inconsistent estimators).

Motivated by the strong limitations of existing approaches, this article proposes estimators for general parametric specifications for risk-return relationships that do not need to specify a parametric model for the conditional mean and variance. A fully nonparametric approach is, however, not feasible given the potential infinite-dimensional information set I_{t-1} and small sample sizes. We use and justify inferences under an index structure to avoid the "curse of dimensionality" typically present in nonparametric analysis. We show how this index structure can also be used as a devise to account for volatility persistence in a flexible way. Our framework is general enough to include as a special case parametric specifications of possibly time-varying SR. In this case, and to the best our knowledge, our estimators are the first of their kind in the literature.

We propose a two-step Semiparametric Generalized Least Squares (SGLS) estimator for risk-return restrictions based on local-polynomial estimators of the first two conditional moments of excess returns. We obtain consistency and asymptotic normality of our estimator under weak conditions. In the general case, estimation of the driving factors has an impact in the asymptotic variance of estimates. We provide some simple sufficient conditions under which there is no estimation effect, and inference can be carried out as if the factors were known. Our asymptotic results can be used to propose valid tests

---

[4]There are few exceptions, such as, e.g., Pagan and Ullah (1988), Pagan and Hong (1991) and Harvey (2001).

for parameter restrictions, such as testing for time-varying price of risk.

Using monthly data from the CRSP value-weighted index (including dividends), we find a significant and positive market price of risk for the subperiod 1947-2008. Our results stand in stark contrast with previous existing literature using monthly data and nonparametric or semiparametric methods, that find negative (or positive but insignificant) coefficients; see e.g. Linton and Perron (2003) and references therein. The discrepancy between our findings and existing ones may be explained by the relative efficiency of our methods.

The remainder of this article is organized as follows. In Section 2 we introduce the model and the estimator. In Section 3 we develop the asymptotic distribution for the estimator. We apply our results to the linear model (1) with estimated factors in Section 4. Section 5 presents a simulation study and Section 6 contains an empirical application to excess returns of the CRSP value-weighted. Finally, we conclude in Section 7. Mathematical proofs are gathered in an Appendix.

2   Model and Estimator

In this section, we introduce the model, with some examples that illustrate the general applicability of our procedure, and the SGLS estimator. In order to handle flexible functional forms, while avoiding the "curse of dimensionality" problem inherent in nonparametric analysis, we assume that there exists a d-dimensional vector X_{t-1}, with possibly unobserved components, satisfying $\mu(I_{t-1}) = \mu(X_{t-1})$ and $\sigma^2(I_{t-1}) = \sigma^2(X_{t-1})$. Moreover, we assume that Y_t is generated from the index semiparametric regression model

$$Y_t = g(\theta_0, \sigma(X_{t-1}), X_{t-1}) + \sigma(X_{t-1})\varepsilon_t, \tag{2}$$

where g is a completely specified function up to the unknown parameter $\theta_0 \in \Theta \subset \mathbb{R}^q$, which we aim to estimate, and ε_t is an error term, independent of X_{t-1}. The vector X_{t-1} can contain lagged values of Y_t as well as other exogenous variables Z_t in I_{t-1}, and it is parametrically generated from the available information at time $t-1$, that is, $X_{t-1} \equiv X_{t-1}(\beta_0) = \gamma(I_{t-1}, \beta_0)$, where β_0 is an unknown parameter in $B \subset \mathbb{R}^k$ (to be estimated in practice) and γ is a known function. Intuitively, this means that all the relevant information contained in the set I_{t-1} is captured by the vector X_{t-1} through a certain parametric model. We shall assume that a $\sqrt{T}$-consistent estimator of β_0 is available. Obviously, the assumption on β_0 is superfluous when X_{t-1} is completely observed, as for example when $X_{t-1} = (Y_{t-1}, \ldots, Y_{t-q}, Z_t')'$ (A' denotes the transpose for any vector or matrix A). As we show below, allowing for unobserved X_{t-1} can be used as a device to allow for persistence in the nonparametric volatility in model (2) without requiring complicated estimation techniques, see the examples below. It is also worth noticing that our general set-up is consistent with an extensive empirical literature that uses estimated factors driving the first two conditional moments of excess returns, see e.g. Fama and French (1992, 1993), Guo and Whitelaw (2006), Ludvigson and Ng (2007), Tang and Whitelaw (2011), to mention just a few. From the inference point of view, estimating X_{t-1} creates in general a "generated regressors problem", in which standard errors and tests statistics need to be corrected by the estimation of the factors. Inference that does not account for the uncertainty in estimating X_{t-1} may lead to erroneous conclusions. We shall investigate this issue in detail below.

We illustrate the general applicability of our setting with some examples. The first example is motivated by an extensive literature documenting a time-varying SR; see e.g. Harvey (1989), Ferson (1989), Ferson, Foerster and Keim (1993) and references below.

Example 1. Models with parametric time-varying SR. Our setting is general enough to include as a special case models under which the SR, i.e $S_{t-1} = \mu(X_{t-1})/\sigma(X_{t-1})$, follows a specific parametric model. For instance, if the specified model is of a multiplicative form $g(\theta, \sigma(x), x) = \sigma(x)g_1(\theta, x)$, then the SR will have the parametric structure specified in g_1, i.e. $S_{t-1} = g_1(\theta_0, X_{t-1})$. Examples are the exponential models in De Santis and Gerard (1997) and Bekaert and Harvey (1995), where $g_1(\theta_0, X_{t-1}) = \exp(\theta_0' X_{t-1})$. To the best of our knowledge, general estimators for time-varying parametric SR are not available in the literature. ▲

Our second example shows that our approach can accommodate hedging components, see e.g. Guo and Whitelaw (2006) and Brandt and Wang (2010).

Example 2. Models with hedging component. The Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973) also includes a hedging component in the relation between $\mu(X_{t-1})$ and $\sigma(X_{t-1})$. Our model can accomodate this additional component with specifications such as $g(\theta_0, \sigma(x), x) = \theta_1\sigma^2(x) + g_1(\theta_2, x)$, $\theta_0 = (\theta_1', \theta_2')'$. This is, for instance, the specification used in Ghysels et al. (2005), Guo and Whitelaw (2006) and Brandt and Wang (2010), among others. ▲

The vast majority of inferences for risk-return relationships have been restricted to models where the variance is parametrically specified, as e.g. GARCH model. These parametric models are nested in our general set-up by an appropriate definition of $X_{t-1}(\beta_0)$, as shown in the next example. Hence, this example shows the flexibility of our semiparametric set-up by accounting for persistence in volatility in a flexible way, while addressing the curse of dimensionality problem. This example also serves to illustrate how a generic estimator for β_0 can be constructed by reference to the estimator developed in Yang (2006).

Example 3. Semiparametric extensions of parametric models. Parametric models have been traditionally used in the literature. Among them, the leading example is the GARCH(1,1) model and the GJR model, used in Glosten, Jaganathan and Runkle (1993). These are special cases of the semiparametric model studied in Yang (2006), which specifies

$$\sigma^2(I_{t-1}) = h(V_{t-1}(\beta_0)),$$

for an unknown function $h(\cdot)$ and $V_{t-1}(\beta_0) = \sum_{j=1}^{t} \beta_{01}^{j-1} v(Y_{t-j}, \beta_{02})$, $\beta_0 = (\beta_{01}, \beta_{02}')'$, with $v(\cdot)$ known up to the parameter β_{02}. For h equals the identity function and $v(y, \beta_{02}) = b_1 y^2 + b_2$, $\beta_{02} = (b_1, b_2)'$, and $v(y, \beta_{02}) = b_1(y^2 + b_3 y^2 1(y < 0)) + b_2$, $\beta_{02} = (b_1, b_2, b_3)'$, the model (2) leads to the GARCH(1,1) and GJR models, respectively. To extend this to our context and allow for covariates, we consider the semiparametric model in (2) where $X_{t-1} = (V_{t-1}(\beta_0), Z_t)$. In this semiparametric model the parameter β_0 can be estimated with a simple extension of the semiparametric least squares estimator proposed in Yang (2006). ▲

To define the new SGLS estimator we need to introduce estimators for $\mu(x)$ and $\sigma(x)$ as follows. First, set $\widehat{X}_{t-1} = \gamma(I_{t-1}, \widehat{\beta})$, where $\widehat{\beta}$ is a suitable consistent estimator of β_0; see Section 4 for an example. Then, we consider local polynomial estimators $\widehat{\mu}(x)$ and $\widehat{\sigma}(x)$ for $\mu(x)$ and $\sigma(x)$, respectively. That is, $\widehat{\mu}(x) = \widehat{\alpha}_0(x)$, where $\widehat{\alpha}_0(x)$ is the first component of the vector $\widehat{\alpha}(x)$, which is the solution of the local minimization problem

$$\min_{\alpha} \sum_{t=1}^{T} \left\{ Y_t - P_t(\alpha, x, p) \right\}^2 K_h(\widehat{X}_{t-1} - x),$$

where $P_t(\alpha, x, p)$ is a polynomial of order p built up with all $0 \leq i \leq p$ products of factors of the form $X_{t-1j} - x_j$, $j = 1, \ldots, d$, where d is the dimension of x. The vector α consists of all coefficients of this polynomial. Here, for $u = (u_1, \ldots, u_d) \in \mathbb{R}^d$, $K(u) = \prod_{j=1}^{d} k(u_j)$ is a d-dimensional product kernel, k is a univariate kernel function, $h = (h_1, \ldots, h_d)$ is a d-dimensional bandwidth vector converging to zero when n tends to infinity, and $K_h(u) = \prod_{j=1}^{d} k(u_j/h_j)/h_j$. To estimate $\sigma^2(x)$, define

$$\widehat{\sigma}^2(x) = \widehat{\gamma}_0(x) - \widehat{\alpha}_0^2(x),$$

where $\widehat{\gamma}_0$ is defined in the same way as $\widehat{\alpha}_0$, but with Y_t replaced by Y_t^2 in (2) ($t = 1, \ldots, T$).

With the estimator of $\widehat{\sigma}(x)$ in place, we introduce our SGLS estimator, say $\widehat{\theta}_{ls}$, defined as any minimizer of the criterium function

$$\theta \to \sum_{t=1}^{T} w(\widehat{X}_{t-1}) \left(\frac{Y_t - g(\theta, \widehat{\sigma}(\widehat{X}_{t-1}), X_{t-1})}{\widehat{\sigma}(\widehat{X}_{t-1})} \right)^2,$$

where w is a positive weight function, which is introduced as a technical device to allow for covariates with non-compact support. This is important because in applications X_{t-1} may contain highly persistent variables with large supports. The next section investigates the asymptotic theory for $\widehat{\theta}_{ls}$.

3 Asymptotic Theory

We introduce the following regularity conditions and notations. Let $F_X(x) = P(X_t \leq x)$ and let $F(x, y) = P(X_{t-1} \leq x, Y_t \leq y)$ (which under assumption A1 below do not depend on t). Lowercase letters will be used to denote the corresponding density functions. Henceforth, C is a generic constant that may change from expression to expression.

ASSUMPTION A1: The process (X_{t-1}, Y_t), $t = 0, \pm 1, \pm 2, \ldots$, satisfies (2) and is strictly stationary and absolutely regular (β-mixing), with mixing coefficients of order $O(t^{-b})$, for some $b > 2$.

ASSUMPTION A2:

(i) θ_0 belongs to the interior of a compact subset Θ of $\mathbb{R}^q$, and β_0 belongs to the interior of a compact subset B of $\mathbb{R}^k$.

(ii) The weight function w has compact support R_w in $\mathbb{R}^d$ and satisfies $w(x) > 0$ for all $x \in R_w$, $\sup_{x \in R_w} w(x) \leq C$ and $w(\gamma(I_{t-1}, \beta))$ is continuous in β a.s.

(iii) All partial derivatives of F_X up to order $2d+1$ exist on the interior of R_w, they are uniformly continuous and $\inf_{x \in R_w} f_X(x) > 0$.

(iv) All partial derivatives of μ and σ up to order $p+2$ exist on the interior of R_w, they are uniformly continuous and $\inf_{x \in R_w} \sigma(x) > 0$.

(v) The function $\gamma(I_{t-1}, \beta)$ is continuously differentiable with respect to β.

(vi) The function $g(\theta, u, x)$ is continuously differentiable with respect to the components of θ, u and x.

ASSUMPTION A3:

(i) $E(|Y_1|^s) < \infty$ and $\sup_{x \in R_X} E(|Y_1|^s \mid X_0 = x) < \infty$ for some $s > 2 + 2/(b-2)$, where b is as in Assumption A1.

(ii) There exists some j' such that for all $j \geq j'$,

$$\sup_{x_0, x_j \in R_X} E(|Y_1 Y_{j+1}|^2 \mid X_0 = x_0, X_j = x_j) f_j(x_0, x_j) < \infty,$$

where $f_j(x_0, x_j)$ denotes the joint density of (X_0, X_j).

(iii) The errors of the regression model satisfy

$$E(\varepsilon_t | X_{t-1}, \mathcal{F}_{-\infty}^{t-1}(X, Y)) = E(\varepsilon_t | X_{t-1}) = 0$$

and

$$\text{Var}(\varepsilon_t | X_{t-1}, \mathcal{F}_{-\infty}^{t-1}(X, Y)) = E(\varepsilon_t^2 | X_{t-1}) = 1,$$

where $\mathcal{F}_{-\infty}^{t-1}(X, Y)$ denotes the σ-algebra generated by the sequence $\{(X_{j-1}, Y_j), j = -\infty, \ldots, t-1\}$.

ASSUMPTION A4: The function $F(x, y)$ is continuous in (x, y), and twice continuously differentiable with respect to x and y. Let $L(x, y)$ denote generically the derivatives $\frac{\partial}{\partial x} F(x, y)$, $\frac{\partial}{\partial y} F(x, y)$, $\frac{\partial^2}{\partial x^2} F(x, y)$, $\frac{\partial^2}{\partial y^2} F(x, y)$ and $\frac{\partial^2}{\partial x \partial y} F(x, y)$. Then, $L(x, y)$ is continuous in (x, y) and satisfies $\sup_{x,y} |y^2 L(x, y)| < \infty$.

ASSUMPTION A5:

(i) For all $j = 1, ..., d : h_j / h_1 \to C_j$, with $0 < C_j < \infty$, and the bandwidth h_1 satisfies $(\log T)^{-1} T^\eta h_1^d \to \infty$ for $\eta = \frac{b-1-d-d/r-(1+b)/(s-1)}{b+3-d-(1+b)/(s-1)}$, where r is such that $b > \frac{1+(s-1)(1+d/r+d)}{s-2}$, and with b and s as defined in Assumption A1 and A3 respectively, $T h_1^{2d+\delta} \to \infty$ for some small $\delta > 0$, $T h_1^{2p+2} \to 0$ for odd p and $T h_1^{2p+4} \to 0$ for even p.

(ii) The kernel k is a symmetric probability density function on $[-1, 1]$, k is d times continuously differentiable, and $k^{(j)}(\pm 1) = 0$ for $j = 0, \ldots, d-1$.

Assumption A3 and the first condition in Assumption A5-(i) are taken from Hansen (2008), and they ensure suitable rates of convergence of the kernel estimators of $\mu(\cdot)$ and $\sigma(\cdot)$. Our next assumption states a linear expansion for the estimator $\widehat{\beta}$.

ASSUMPTION A6: The estimator $\widehat{\beta}$ satisfies:

$$\widehat{\beta} - \beta_0 = T^{-1} \sum_{t=1}^{T} r_t + o_P(T^{-1/2}),$$

for a strictly stationary process r_t such that $E(r_t \mid I_{t-1}) = 0$ a.s. and $E(\|r_t\|^2) < \infty$ (where $\|\cdot\|$ is the Euclidean norm).

A generic example of estimator $\widehat{\beta}$ that satisfies Assumption A6 is the semiparametric least squares estimator proposed in Yang (2006). A similar estimator can be developed in our more general context, but since the parameter β_0 is often of secondary importance, we omit details here. We show below that Assumption A6 can be simplified under some circumstances.

Henceforth, the following quantities appear in the asymptotic variance of $\widehat{\theta}_{ls}$:

$$
\begin{aligned}
g_{\theta t}(\theta) &= \frac{\partial}{\partial \theta} g(\theta, \sigma(X_{t-1}), X_{t-1}) \\
&= \left(\frac{\partial}{\partial \theta_1} g(\theta, \sigma(X_{t-1}), X_{t-1}), \dots, \frac{\partial}{\partial \theta_q} g(\theta, \sigma(X_{t-1}), X_{t-1}) \right)' \\
g_{ut}(\theta) &= \frac{\partial}{\partial u} g(\theta, u, X_{t-1})|_{u=\sigma(X_{t-1})}, \\
g_{xt}(\theta) &= \frac{\partial}{\partial x'} g(\theta, \sigma(X_{t-1}), x)|_{x=X_{t-1}}, \\
u_t &= \varepsilon_t - 0.5 g_{ut}(\theta_0)(\varepsilon_t^2 - 1), \\
S(\theta_0) &= E\left[w(X_{t-1}) \sigma^{-2}(X_{t-1}) g_{\theta t}(\theta_0) g_{\theta t}'(\theta_0) \right]
\end{aligned}
$$

and

$$v(\theta_0) = E\left[w(X_{t-1}) \sigma^{-2}(X_{t-1}) g_{\theta t}(\theta_0) g_{xt}(\theta_0) \frac{\partial}{\partial \beta'} \gamma(I_{t-1}, \beta_0) \right].$$

Theorem 3.1 *Assume A1-A6. Then, the following linear expansion holds:*

$$\widehat{\theta}_{ls} - \theta_0 = T^{-1} \sum_{t=1}^{T} w(X_{t-1}) s_{ls}(I_{t-1}, \varepsilon_t) + o_P(T^{-1/2}),$$

where

$$s_{ls}(I_{t-1}, \varepsilon_t) = S^{-1}(\theta_0) \left[\sigma^{-1}(X_{t-1}) g_{\theta t}(\theta_0) u_t - v(\theta_0) \frac{r_t}{w(X_{t-1})} \right].$$

The proof of Theorem 3.1 is given in the Appendix. From this theorem the asymptotic normality of $\sqrt{T}(\widehat{\theta}_{ls} - \theta_0)$ follows easily by means of a standard central limit theorem for dependent data (see e.g. Theorem 4.2 in Rio, 2000).

Corollary 3.1 *Assume A1-A6. Then,*

$$\sqrt{T}(\widehat{\theta}_{ls} - \theta_0) \to_d N(0, \sigma_{ls}^2),$$

where $\sigma_{ls}^2 = E[w^2(X_{t-1})s_{ls}(I_{t-1}, \varepsilon_t)s'_{ls}(I_{t-1}, \varepsilon_t)]$.

Notice that the influence function of $\widehat{\theta}_{ls}$, i.e. $w(X_{t-1})s_{ls}(I_{t-1}, \varepsilon_t)$, depends on the estimator $\widehat{\beta}$ used, that is, there is an asymptotic impact from estimating the factor X_{t-1} on the estimator $\widehat{\theta}_{ls}$. This shows that inferences that do not account for this term are in general misleading. There are two situations under which the impact of $\widehat{\beta}$ on $\widehat{\theta}_{ls}$ vanishes asymptotically. First, if $\widehat{\beta} - \beta_0 = o_P(T^{-1/2})$, the estimator of β_0 converges so fast, relative to $\widehat{\theta}_{ls}$, that it can be assumed to be known. This is for instance the case when $\widehat{\beta}$ is obtained from very high frequency data estimates. A second case is when $v(\theta_0) = 0$. In turn, this happens when all components of $g_{xt}(\theta_0)\frac{\partial}{\partial\beta'}\gamma(I_{t-1}, \beta_0)$ equal zero, which occurs in many applications. For example, when $g(\theta_0, \sigma(X_{t-1}), X_{t-1}) = g(\theta_0, \sigma(X_{t-1}), S_{t-1})$, where S_{t-1} denotes a subvector of $X_{t-1}(\beta_0)$ that is observed, because for every element of the vector $g_{xt}(\theta_0)$, either it equals zero (when the corresponding covariate is not observable), or the corresponding row of the matrix $\frac{\partial}{\partial\beta'}\gamma(I_{t-1}, \beta_0)$ equals zero (when that covariate is observed); so the resulting vector of the product equals zero, and hence $v(\theta_0) = 0$. A leading example is the linear model (1). In the second case, the assumption on $\widehat{\beta}$ can be weakened to $\widehat{\beta} - \beta_0 = O_P(T^{-1/2})$ and the influence function of $\widehat{\theta}_{ls}$ simplifies, as shown in the following corollary.

Corollary 3.2 *Assume A1-A5. If $\widehat{\beta} - \beta_0 = O_P(T^{-1/2})$ and all components of $g_{xt}(\theta_0)\frac{\partial}{\partial\beta'}\gamma(I_{t-1}, \beta_0)$ equal zero, then, it holds:*

$$\widehat{\theta}_{ls} - \theta_0 = T^{-1}\sum_{t=1}^{T} w(X_{t-1})s_{ls}(I_{t-1}, \varepsilon_t) + o_P(T^{-1/2}),$$

where $s_{ls}(I_{t-1}, \varepsilon_t) = S^{-1}(\theta_0)g_{\theta t}(\theta_0)\sigma^{-1}(X_{t-1})u_t$.

Example 2 (cont.) In the ICAPM model with a hedging component $g_{xt}(\theta_0)\partial\gamma(I_{t-1}, \beta_0)/\partial\beta' = \partial g_1(\theta_2, x)/\partial x\partial\gamma(I_{t-1}, \beta_0)/\partial\beta'$ will be generally non zero, and hence there will be an impact in the standard errors of $\widehat{\theta}_{ls}$ from estimating the factors. ▲

To illustrate our results, including estimation of standard errors, we consider an application to the popular linear model (1) with estimated factors; see Merton (1973, 1980).

4   The linear model with estimated factors

This section provides formulae for the estimators, influence functions, asymptotic variances σ_{ls}^2 and their estimators for the classical linear model with estimated factors. These formulae will be used in our simulations and empirical application below. We consider a factor generated by a linear index

model $X_{t-1}(\beta_0) = \beta_0' Z_t$, where Z_t is a vector of variables in I_{t-1}. The parameter β_0 is estimated by Ichimura's (1993) SLSE

$$\widehat{\beta} = \arg\min_{\beta \in B} \sum_{t=1}^{T} w_\beta(Z_{t-1}) \left(Y_t - \widehat{\mu}_\beta(\beta' Z_t) \right)^2,$$

where $w_\beta(Z_{t-1})$ is a weight function to remove the impact of low-density regions on the estimator $\widehat{\beta}$ and $\widehat{\mu}_\beta(\cdot)$ is a local polynomial estimator for the conditional mean of Y_t given $\beta' Z_t$. For identifiability purposes we normalize the first component of β to one, i.e. $\beta = (1, \beta_2')$. To simplify notation, we identify β_2 with β. It is well-known that under mild regularity conditions (Ichimura, 1993)

$$\widehat{\beta} - \beta_0 = \Lambda^{-1} T^{-1} \sum_{t=1}^{T} \sigma(X_{t-1}) \varepsilon_t \dot{\mu}_\beta(\beta_0' Z_t) w_\beta(Z_{t-1}) + o_P(T^{-1/2}),$$

where $\Lambda = E[w_\beta(Z_{t-1})\dot{\mu}_\beta(\beta_0' Z_t)\dot{\mu}_\beta'(\beta_0' Z_t)]$ and $\dot{\mu}_\beta(\beta_0' Z_t) = \partial E(Y_t | \beta' Z_t = \beta' z)|_{z=Z_t, \beta=\beta_0} / \partial \beta$. Then, Assumption A6 is satisfied under mild and known regularity conditions with $r_t = \Lambda^{-1}\sigma(X_{t-1})\varepsilon_t\dot{\mu}_\beta(\beta_0' Z_t)w_\beta(Z_{t-1})$.

Let us now focus on the linear model (1). With the estimator $\widehat{\beta}$, we compute fitted values $\widehat{X}_{t-1} = \widehat{\beta}' Z_t$, which are, in turn, used for estimating $\widehat{\sigma}$ and compute the SGLS estimator

$$\widehat{\theta}_{ls} = \left(\sum_{t=1}^{T} w(X_{t-1})\widehat{\sigma}^2(X_{t-1}) \right)^{-1} \left(\sum_{t=1}^{T} w(X_{t-1})Y_t \right). \tag{3}$$

Note that for this example, $g(\theta_0, \sigma(X_{t-1}), X_{t-1}) = g(\theta_0, \sigma(X_{t-1})) = \theta_0\sigma^2(X_{t-1})$ holds, then the relevant formulas for standard error computation and estimation are those of Corollary 3.2. A consistent estimator for the asymptotic variance of $\widehat{\theta}_{ls}$ is given by

$$\widehat{\sigma}_{ls}^2 = T^{-1} \sum_{t=1}^{T} w^2(\widehat{X}_{t-1})\widehat{s}_{ls}^2(I_{t-1}, \widehat{\varepsilon}_t), \tag{4}$$

where

$$\widehat{s}_{ls}(I_{t-1}, \widehat{\varepsilon}_t) = \frac{\widehat{\sigma}(\widehat{X}_{t-1})\widehat{u}_t}{\widehat{S}},$$

$$\widehat{S} = \frac{1}{T} \sum_{t=1}^{T} w(\widehat{X}_{t-1})\widehat{\sigma}^2(\widehat{X}_{t-1}),$$

$$\widehat{u}_t = \widehat{\varepsilon}_t - \widehat{\theta}_{ls}\widehat{\sigma}(\widehat{X}_{t-1})(\widehat{\varepsilon}_t^2 - 1)$$

and

$$\widehat{\varepsilon}_t = \frac{Y_t - \widehat{\theta}_{ls}\widehat{\sigma}^2(\widehat{X}_{t-1})}{\widehat{\sigma}(\widehat{X}_{t-1})}.$$

In the linear model estimation of the factors X_{t-1} does not have an asymptotic impact in the standard errors for $\widehat{\theta}_{ls}$ and in related inferences. Estimation of the conditional variance, however, has an asymptotic impact on the limiting distribution of $\widehat{\theta}_{ls}$. In extensions of this setting, such as in models with time-varying prices of risk or hedging components, the situation is different and, as shown in our Theorem 1, inferences are generally affected by the first step estimation of the factors. For an

illustration of this point, we provide in the Appendix expressions for estimates and standard errors for a time-varying linear model, $g(\theta_0, \sigma(x), x) = (\theta_{01} + \theta_{02}x)\sigma^2(x)$, with factors estimated as above.

The linear model has been traditionally estimated by the ordinary least squares estimator (OLSE)

$$\widehat{\theta}_{ols} = \left(\sum_{t=1}^{T} w(X_{t-1})\widehat{\sigma}^4(X_{t-1})\right)^{-1} \left(\sum_{t=1}^{T} w(X_{t-1})\widehat{\sigma}^2(X_{t-1})Y_t\right).$$

The typical asymptotic variance estimate obtained from OLSE theory will be inconsistent for the true asymptotic variance, which means that traditional inference that does not account for estimation of the conditional variance will be generally wrong (cf. Pagan and Ullah, 1988). More specifically, the standard t-test based on uncorrected OLSE inference uses $\widehat{s}_{ols}(I_{t-1}, \widehat{\varepsilon}_t) = \widehat{S}^{-1}\widehat{\sigma}^2(\widehat{X}_{t-1})\widehat{\varepsilon}_t$ in (4), with $\widehat{S} = T^{-1}\sum_{t=1}^{T} w(X_{t-1})\widehat{\sigma}^4(X_{t-1})$, whereas the correct one in our semiparametric setting should use $\widehat{s}_{ls}(I_{t-1}, \widehat{\varepsilon}_t) = \widehat{S}^{-1}\widehat{\sigma}^2(\widehat{X}_{t-1})\widehat{u}_t$. Note that when $\theta_0 = 0$, $\varepsilon_t = u_t$ and the difference between these standard errors estimates is asymptotically negligible. When $\theta_0 \neq 0$, standard OLSE inference that does not account for estimation of the conditional variance is misleading, and the difference of asymptotic standard errors can be large. For instance, for Gaussian errors it can be shown that the difference between the limits of the standard error estimates is $\sqrt{3}\theta_0$. These asymptotic results are confirmed in our simulations below.

5 Simulation Study

In this section we study the finite sample properties of the proposed estimator. We consider the standard linear model $\mu(X_{t-1}) = \theta_0\sigma^2(X_{t-1})$, where θ_0 is an unknown parameter. We generate data according to the model

$$Y_t = \theta_0(1 + 0.2X_{t-1}^2) + (1 + 0.2X_{t-1}^2)^{1/2}\varepsilon_t,$$

where the innovations ε_t are i.i.d. and standard normally distributed, and the parameter θ_0 varies over the set of values specified in the tables below.

In these simulations $X_{t-1} = \beta_0'Z_t$, where $Z_t = (Z_{1t}, Z_{2t}, Z_{3t})'$ is an i.i.d. 3-dimensional vector of independent $Uniform[0, 1]$ components and with $\beta_0' = (1, \sqrt{0.5}, \sqrt{0.5})$. The parameter β_0 is estimated by Ichimura's (1993) SLSE defined in (??). Once the estimation of β_0 has been done, the corresponding estimated index is used to obtain the SGLS estimator of θ_0 as defined in (3). The regression function, μ, and variance function, σ, are nonparametrically estimated by Nadaraya-Watson estimators. Throughout the simulations only data corresponding the [10%, 90%] of the range of the index variable were considered during the estimation process (this corresponds to choosing w as an indicator function in that range). For the choice of the smoothing parameter or bandwidth to estimate the functions μ and σ, we take the one obtained by regular cross-validation (c-v, in the table) and fixed bandwidths (0.20, 0.30, 0.40). Sample sizes 100 and 200 are considered. All results are based on 1000 simulated data sets.

Table 1 displays the observed mean squared error (MSE) of $\hat{\theta}_{ls}$, which decreases as the sample size increases. We have also computed the 95% confidence interval for θ_0 based on the asymptotic normal

distribution of the estimator. The empirical coverage and the average length of the confidence intervals are displayed in the table. The empirical coverage is close to the nominal one and the average length also behaves correctly. The choice of the smoothing parameter does not seem to have a big impact on the reported quantities. Table 2 shows the analogous results for the oracle, but practically unfeasible, estimator when it is assumed that the true value of the parameter β_0 is known. As expected, its MSE is slightly better than the one of the feasible and realistic estimator, achieving almost no difference for sample size 200, which is consistent with our asymptotic results. Similar comments apply for the empirical coverage and length of the confidence interval.

[**Table 1 (at the end of the manuscript) to be placed around here**]

[**Table 2 (at the end of the manuscript) to be placed around here**]

To illustrate the implications of our results for testing, suppose we aim to test for (a) $H_0 : \theta_0 = 0$ vs $H_1 : \theta_0 \neq 0$ and (b) $H_0 : \theta_0 = 2.5$ vs $H_1 : \theta_0 \neq 2.5$ in the context of the linear model. We compare the empirical size and power of three t-tests. First, the t-test derived from our estimator and standard errors $\widehat{\sigma}_{ls}$ (Test #1, in the table). Second, the analogous t-test associated to the oracle estimator that does not account for estimation of β_0 (Test #2). Finally, we also compare with the ordinary t-test under a linear model of the form $Y_t = \theta_0 \widehat{\sigma}^2(X_t) + \tilde{\varepsilon}_t$, which does not account for estimation of the conditional variance (Test #3).

The observed rejection proportions for test (a) are displayed in Table 3. The approximation of the level (row $\theta_0 = 0.00$) is correct for all test, and the power increases as the sample size increases and/or the true value of the parameter goes away from the value specified in the null hypothesis. In fact, the power of the feasible test #1 is very similar to the power of the oracle test #2, so the estimation of β_0 does not seem to impact the performance of the test. Test #3 also approximates the level correctly and produces good power. The smoothing parameter also seems not relevant.

On the other hand, Table 4 shows the results for test (b). Now, the row $\theta_0 = 2.50$ corresponds to the null hypothesis. Here we can see that the behavior of test #1 is correct, and very similar to the oracle test #2. However, test #3 fails to approximate the level, since the asymptotic variance of the estimate is not consistently estimated. The simulation results confirm our asymptotic results: test #3 is only valid when $\theta_0 = 0$. To show the impact of not taking into account the estimation of the conditional variance, Table 5 displays results for tests of the form $H_0 : \theta_0 = \theta$ for several values of θ and where the data were generated under the null hypothesis. The approximation of the level is incorrect unless $\theta = 0$, and the approximation gets worse as the true value of θ goes away from zero.

[**Table 3 (at the end of the manuscript) to be placed around here**]

[**Table 4 (at the end of the manuscript) to be placed around here**]

[**Table 5 (at the end of the manuscript) to be placed around here**]

6 Empirical Application

In this section we examine the monthly excess returns on the most comprehensive CRSP value-weighted index (including dividends)–the monthly continuously compounded return on the index minus the monthly return on the 30-day Treasury Bills–over the period January 1926-December 2008. In this application, however, we focus on the restricted period from January 1947 to December 2008, since it has been suggested in the literature that the Great Depression may lead to a structural change in the conditional volatility, invalidating some of the applied inferences. The excess return data were obtained from the CRSP, which includes the NYSE, the AMEX and the NASDAQ and can be considered the best available proxy for "the market". We also consider some predetermined variables in the information set which include: the spread in yields between Moody's Baa and Aaa rated bonds (Z_{1t}), the dividend yield on the S&P500 in excess of the 30-day Treasury bill rate (Z_{2t}), and the excess holding period on the 3-month Treasury bill (Z_{3t}). A number of studies have used information variables similar to these; see Keim and Stambaugh (1986), Campbell (1987), Fama and French (1988) and Harvey (1989, 2001) among others.

We aim to estimate the model:

$$\mu(X_{t-1}) = \theta_0 \sigma^2(X_{t-1}),$$

where $X_{t-1} = \beta_0' Z_t$. The parameter $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})'$ is estimated by Ichimura's (1993) SLSE. To estimate β_0, we normalize $\beta_{01} = 1$ for identification purposes, and use a Gaussian kernel with a bandwidth that is chosen simultaneously with the estimator by minimizing the SLS criteria (see Härdle, Hall and Ichimura, 1993). We approximate standard errors by bootstrap with 1000 Monte Carlo replications. The index estimates are $\widehat{\beta} = (1, 1.134, 5.098)$ with bootstrap standard errors (0,17.378,1092). In the index specification the 3-month Treasury bill (Z_{3t}) is highly not significant, but excluding this covariate has little impact on our subsequent inferences.

Then, we estimate θ_0 by our SGLS estimator and construct 95% confidence intervals based on our asymptotic results. The results are reported in Table 6. The regression function, μ, and variance function, σ, are nonparametrically estimated by Nadaraya-Watson estimators with a Gaussian kernel. Only data corresponding the $[10\%, 90\%]$ of the range of the index were considered during the estimation process. For the choice of the smoothing parameter or bandwidth to estimate the functions μ and σ, we take the one obtained by least squares cross-validation (c-v, in the table). To study the sensitivity of the estimator to the bandwidth, we report results for three choices of the bandwidth h (0.5×c-v, c-v and 1.5×c-v).

[Table 6 (at the end of the manuscript) to be placed around here]

We find a significant and positive market price of risk for the subperiod 1947-2008 for all values of the bandwidth at significance level of 5%. The sensitivity to the bandwidth parameter is low for the estimates. Our results stand in stark contrast with previous existing literature using monthly data and nonparametric or semiparametric methods, that find negative (or positive, but insignificant) coefficients; see e.g. Linton and Perron (2003) and references therein. The discrepancy between our findings and existing ones may be explained by the efficiency of our methods relative to existing ones.

For completeness, we plot the functions $\widehat{\mu}(\widehat{X}_{t-1})$ and $\widehat{\sigma}(\widehat{X}_{t-1})$, as a function of $\widehat{X}_{t-1}$, with 95% bootstrap confidence bands computed with 1000 bootstrap replications. In addition, we plot the pairs $(\widehat{\sigma}(\widehat{X}_{t-1}), \widehat{\mu}(\widehat{X}_{t-1}))$ for all t. Based on the reported results, the linear specification (1) seems to provide a good fit to this data set. We observe two different regimes in the behaviour of $\widehat{\mu}(\widehat{X}_{t-1})$ and $\widehat{\sigma}(\widehat{X}_{t-1})$ as a function of $\widehat{X}_{t-1}$; In the left tail distribution of $\widehat{X}_{t-1}$ return and risk seem flat, but there is a point after which both $\widehat{\mu}(\widehat{X}_{t-1})$ and $\widehat{\sigma}(\widehat{X}_{t-1})$ become simultaneously increasing as as a function of $\widehat{X}_{t-1}$.

[Figures 1-3 (at the end of the manuscript) to be placed around here]

We have done a number of robustness checks in our application. In addition to the bandwidth sensitivity mentioned above, we have considered estimates where Y_{t-1} is added to the set of covariates. In this case the estimate of θ_0 becomes 2.964, with a standard error of 0.865 and a 95% confidence interval [1.165, 4.763]. The estimate is positive and significantly larger than the estimate without including Y_{t-1}. The fit of the linear model is, however, much worse than in the first case (without including Y_{t-1}).

To conclude, this application shows how the combination of the robustness of our semiparametric specification with the added efficiency of our SGLS procedure leads to a positive and significant estimate for the coefficient θ_0 in linear specifications of the risk-return relation, in start contrast to most existing methods using nonparametric or semiparametric procedures.

7 Conclusions

An extensive empirical literature has documented sensitivity of risk-return relationships to parametric assumptions on the first two conditional moments of excess returns. Motivated by this fact, we have proposed semiparametric estimators for parametric risk-return restrictions that do not require parametric specifications for the conditional mean and variance. In addition, since different financial theories of asset pricing suggest different restrictions between risk and return, we have considered generic parametric risk-return restrictions. The main distinctive feature of our approach is its generality. Our approach is general because we allow for estimated factors, and generic risk-return restrictions that include, among others, parametric models for dynamic SR or hedging components.

We have provided sufficient conditions for the estimated factors not to have an impact in the asymptotic distribution of our semiparametric estimator. This is the case, for instance, for the classical linear model with estimated factors. Empirically important extensions of this setting, such as models with a parametric (time-varying) SR or models with hedging components, lead to situations where estimated factors invalidate standard inferences that do not account for this impact. We have supplied generic expressions for the asymptotic distribution of our semiparametric estimator in these more general cases as well. These expressions are more complicated than in the linear case. Alternative ways to carry out inferences in our general settings are certainly available, such as bootstrap methods, and the performance of these methods will be investigated in future research.

8 Appendix

8.1 Linear Model with Time-varying Price of Risk

The situation in the more general time-varying linear model $g(\theta_0, \sigma(x), x) = (\theta_{01} + \theta_{02}x)\sigma^2(x)$ is more complicated than in the linear model (where $\theta_{02} = 0$) , due to the impact of generated regressors on the asymptotic variance. Set $\bar{X}_{t-1} = (1, X_{t-1})'$. The parameter $\theta_0 = (\theta_{01}, \theta_{02})'$ is estimated by the SGLS

$$\widehat{\theta}_{ls} = \left(\sum_{t=1}^{T} w(X_{t-1})\widehat{\sigma}^2(X_{t-1})\bar{X}_{t-1}\bar{X}'_{t-1}\right)^{-1} \left(\sum_{t=1}^{T} w(X_{t-1})\bar{X}_{t-1}Y_t\right).$$

In this example the relevant formulas for standard error computation and estimation are those of Theorem 3.1, and they need to account for the impact of the first step estimator $\widehat{\beta}$. A consistent estimator for the asymptotic variance of $\widehat{\theta}_{ls}$ is given by

$$\widehat{\sigma}^2_{ls} = T^{-1}\sum_{t=1}^{T} w^2(\widehat{X}_{t-1})\widehat{s}_{ls}(I_{t-1}, \widehat{\varepsilon}_t)\widehat{s}'_{ls}(I_{t-1}, \widehat{\varepsilon}_t),$$

where

$$\widehat{s}_{ls}(I_{t-1}, \widehat{\varepsilon}_t) = \widehat{S}^{-1}\left[\widehat{\sigma}(\widehat{X}_{t-1})\bar{X}_{t-1}\widehat{u}_t - \widehat{v}(\widehat{\theta}_{ls})\widehat{r}_t\frac{\widehat{\sigma}^2(\widehat{X}_{t-1})}{w(\widehat{X}_{t-1})}\right],$$

$$\widehat{S} = T^{-1}\sum_{t=1}^{T} w^2(\widehat{X}_{t-1})\widehat{\sigma}^2(\widehat{X}_{t-1})\bar{X}_{t-1}\bar{X}'_{t-1},$$

$$\widehat{u}_t = \widehat{\varepsilon}_t - \widehat{\theta}'_{ls}\bar{X}_{t-1}\widehat{\sigma}(\widehat{X}_{t-1})(\widehat{\varepsilon}_t^2 - 1),$$

$$\widehat{v}(\widehat{\theta}_{ls}) = \widehat{\theta}_{2ls}T^{-1}\sum_{t=1}^{T} w(\widehat{X}_{t-1})\widehat{\sigma}^2(\widehat{X}_{t-1})\bar{X}_{t-1}Z'_t$$

and

$$\widehat{r}_t = \hat{\Lambda}^{-1}\widehat{\sigma}(\widehat{X}_{t-1})\widehat{\varepsilon}_t\dot{\mu}_{n,\beta}(\widehat{\beta}'Z_t)w_\beta(Z_{t-1})$$

$$\hat{\Lambda} = T^{-1}\sum_{t=1}^{T} \dot{\mu}_{n,\beta}(\widehat{\beta}'Z_t)\dot{\mu}'_{n,\beta}(\widehat{\beta}'Z_t)w_\beta(Z_{t-1})$$

15

and $\dot{\mu}_{n,\beta}(\widehat{\beta}'Z_t)$ is a consistent estimator for $\dot{\mu}_\beta(\beta_0'Z_t)$, e.g. the derivative of a local-polynomial estimator $\widehat{\mu}_\beta(\beta'Z_t)$ of $E(Y_t|\beta'Z_t = \beta'z)$.

8.2 Mathematical Proofs

Proof of Theorem 3.1. Define

$$R_T(\theta) = T^{-1}\sum_{t=1}^{T} w(\widehat{X}_{t-1})\widehat{\sigma}^{-2}(\widehat{X}_{t-1})\Big(Y_t - g(\theta, \widehat{\sigma}(\widehat{X}_{t-1}), \widehat{X}_{t-1})\Big)\frac{\partial}{\partial\theta}g(\theta, \widehat{\sigma}(\widehat{X}_{t-1}), \widehat{X}_{t-1}).$$

Then,

$$-R_T(\theta_0) = R_T(\widehat{\theta}_{ls}) - R_T(\theta_0) = \Big(\frac{\partial}{\partial\theta'}R_T(\theta_0)\Big)(\widehat{\theta}_{ls} - \theta_0) + o_P(\|\widehat{\theta}_{ls} - \theta_0\|)$$

$$= -S(\theta_0)(\widehat{\theta}_{ls} - \theta_0) + o_P(\|\widehat{\theta}_{ls} - \theta_0\|),$$

and hence

$$\widehat{\theta}_{ls} - \theta_0 = S^{-1}(\theta_0)R_T(\theta_0)(1 + o_P(1)).$$

Define

$$h(x, \sigma(x)) = w(x)\sigma^{-2}(x)\frac{\partial}{\partial\theta}g(\theta, \sigma(x), x).$$

Now, write

$$R_T(\theta_0)$$

$$= T^{-1}\sum_{t=1}^{T} h(X_{t-1}, \sigma(X_{t-1}))\Big(Y_t - g(\theta_0, \sigma(X_{t-1}), X_{t-1})\Big)$$

$$- T^{-1}\sum_{t=1}^{T} h(X_{t-1}, \sigma(X_{t-1}))\Big(g(\theta_0, \widehat{\sigma}(\widehat{X}_{t-1}), \widehat{X}_{t-1}) - g(\theta_0, \sigma(X_{t-1}), X_{t-1})\Big)$$

$$+ T^{-1}\sum_{t=1}^{T} \big[h(\widehat{X}_{t-1}, \widehat{\sigma}(\widehat{X}_{t-1})) - h(X_{t-1}, \sigma(X_{t-1}))\big]\Big(Y_t - g(\theta_0, \sigma(X_{t-1}), X_{t-1})\Big)$$

$$- T^{-1}\sum_{t=1}^{T} \big[h(\widehat{X}_{t-1}, \widehat{\sigma}(\widehat{X}_{t-1})) - h(X_{t-1}, \sigma(X_{t-1}))\big]$$

$$\times \Big(g(\theta_0, \widehat{\sigma}(\widehat{X}_{t-1}), \widehat{X}_{t-1}) - g(\theta_0, \sigma(X_{t-1}), X_{t-1})\Big)$$

$$= T_1 + T_2 + T_3 + T_4. \tag{5}$$

In what follows we will calculate the terms T_2, T_3 and T_4. It is easily seen that $T_4 = o_P(T^{-1/2})$. The term T_2 can be written as

$$-T^{-1}\sum_{t=1}^{T} w(X_{t-1})\sigma^{-2}(X_{t-1})\Big\{g_{ut}(\theta_0)(\widehat{\sigma}(X_{t-1}) - \sigma(X_{t-1})) \tag{6}$$

$$+ \frac{\partial}{\partial x'}g(\theta_0, \sigma(x), x)|_{x=X_{t-1}}(\widehat{X}_{t-1} - X_{t-1})\Big\}g_{\theta t}(\theta_0) + o_P(T^{-1/2}),$$

which follows from the fact that $\widehat{\beta} - \beta_0 = O_P(T^{-1/2})$.

The second term of (6) is equal to

$$- T^{-1} \sum_{t=1}^{T} w(X_{t-1}) \sigma^{-2}(X_{t-1}) g_{\theta t}(\theta_0) \frac{\partial}{\partial x'} g(\theta_0, \sigma(x), x)|_{x=X_{t-1}} \frac{\partial}{\partial \beta'} \gamma(I_{t-1}, \beta_0)(\widehat{\beta} - \beta_0)$$

$$+ o_P(T^{-1/2})$$

$$= -E\left[w(X_{t-1}) \sigma^{-2}(X_{t-1}) g_{\theta t}(\theta_0) \left\{ g_{ut}(\theta_0) \frac{\partial}{\partial x'} \sigma(x)|_{x=X_{t-1}} + g_{xt}(\theta_0) \right\} \frac{\partial}{\partial \beta'} \gamma(I_{t-1}, \beta_0) \right]$$

$$\times T^{-1} \sum_{t=1}^{T} r_t + o_P(T^{-1/2}). \tag{7}$$

In order to study the first term of (6), consider the class

$$\mathcal{F} = \left\{ x \to A_v(x) = w(x) \sigma^{-2}(x) \frac{\partial}{\partial u} g(\theta_0, u, x)|_{u=\sigma(x)} v(x) \frac{\partial}{\partial \theta} g(\theta_0, \sigma(x), x) : v \in C_M^{\alpha}(R_w) \right\},$$

where $C_M^{\alpha}(R_w)$ is the space of continuous functions v defined on the compact set R_w, for which

$$\|v\|_{\alpha} = \max_{k. \leq \underline{\alpha}} \sup_{x} |D^k v(x)| + \max_{k. = \underline{\alpha}} \sup_{x,x'} \frac{|D^k v(x) - D^k v(x')|}{\|x - x'\|^{\alpha - \underline{\alpha}}} \leq M < \infty,$$

where $\underline{\alpha}$ is the largest integer strictly smaller than α (which we choose later in the proof), $k = (k_1, \ldots, k_d)$,

$$D^k = \frac{\partial^{k.}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}},$$

and $k. = \sum k_i$. A sufficient condition for the class $\mathcal{F}$ to be Donsker is that

$$\int_0^{2M} \sqrt{\log N_{[]}(\delta, \mathcal{F}, \| \cdot \|_{2,\beta})} \, d\delta < \infty,$$

where for any function g,

$$\|g\|_{2,\beta}^2 = \int_0^1 \beta^{-1}(u) Q_g^2(u) du,$$

and where β^{-1} is the inverse cadlag of the decreasing function $u \to \beta_{\lfloor u \rfloor}$ ($\lfloor u \rfloor$ being the integer part of u, and β_t being the mixing coefficient) and Q_g is the inverse cadlag of the tail function $u \to P(\|g\| > u)$ (see Section 4.3 in Dedecker and Louhichi, 2002). Here, $N_{[]}(\delta, \mathcal{F}, \| \cdot \|_{2,\beta})$ is the δ-bracketing number of the class $\mathcal{F}$, i.e. it is the smallest number of δ-brackets needed to cover the space $\mathcal{F}$, where a δ-bracket is the set of all functions h such that $h_\ell \leq h \leq h_u$ and where (h_ℓ, h_u) satisfy $\|h_u - h_\ell\|_{2,\beta} \leq \delta$. In order to prove (8.2) note that it follows from Corollary 2.7.2 in Van der Vaart and Wellner (1996) that for any $\delta > 0$,

$$\log N_{[]}(\delta, C_M^{\alpha}(R_w), \| \cdot \|_2) \leq K \delta^{-d/\alpha},$$

where $\| \cdot \|_2$ is the L_2-norm. Let $v_1^L \leq v_1^U, \ldots, v_m^L \leq v_m^U$ be the $m = O(\exp(K \delta^{-d/\alpha}))$ brackets of

$C_M^\alpha(R_w)$. Then, for any $j = 1, \ldots, m$,

$$P\Big(\Big\|w(X_{t-1})\sigma^{-2}(X_{t-1})g_{ut}(\theta_0)\big[v_j^U(X_{t-1}) - v_j^L(X_{t-1})\big]g_{\theta t}(\theta_0)\Big\| > z\Big)$$
$$\leq P\Big(v_j^U(X_{t-1}) - v_j^L(X_{t-1}) > Kz\Big)$$
$$\leq \frac{1}{K^2 z^2}E\Big|v_j^U(X_{t-1}) - v_j^L(X_{t-1})\Big|^2 \leq \frac{\delta^2}{K^2 z^2}.$$

Hence,

$$\Big\|w(X_{t-1})\sigma^{-2}(X_{t-1})g_{ut}(\theta_0)\big[v_j^U(X_{t-1}) - v_j^L(X_{t-1})\big]g_{\theta t}(\theta_0)\Big\|_{2,\beta}^2$$
$$\leq \int_0^1 \beta^{-1}(u)\frac{\delta^2}{K^2 u}du \leq C\int_0^1 u^b\frac{\delta^2}{u}du = \frac{C\delta^2}{b},$$

where the latter inequality follows from assumption A1. It now follows that the class $\mathcal{F}$ is Donsker, provided $\alpha > d/2$. It can also be shown quite easily that $P(\widehat{\sigma} - \sigma \in C_M^\alpha(R_w)) \to 1$ under the assumptions of the theorem (if we restrict $\widehat{\sigma}$ and σ to R_w; see Lemma A.1 in Neumeyer and Van Keilegom, 2010). Finally, straightforward calculations, based on the uniform consistency of $\widehat{\sigma}$ lead to

$$\mathrm{Var}\Big(A_{\widehat{\sigma}-\sigma}(X_{t-1})\Big) \xrightarrow{P} 0$$

where the variance is taken with respect to X_{t-1}, conditionally on the function $\widehat{\sigma} - \sigma$. Hence, it follows from Corollary 2.3.12 in Van der Vaart and Wellner (1996) that the first term of (6) equals

$$-T^{-1}\sum_{t=1}^T A_{\widehat{\sigma}-\sigma}(X_{t-1}) = -E\Big[A_{\widehat{\sigma}-\sigma}(X_{t-1})\Big] + o_P(T^{-1/2})$$
$$= -E\Big[w(X_{t-1})\sigma^{-2}(X_{t-1})g_{ut}(\theta_0)(\widehat{\sigma}(X_{t-1}) - \sigma(X_{t-1}))g_{\theta t}(\theta_0)\Big] + o_P(T^{-1/2}),$$

where the expected value is taken with respect to X_{t-1}, conditionally on the function $\widehat{\sigma} - \sigma$. In order to simplify the notation, we restrict attention in what follows to the case $p = 0$ (i.e. local constant smoothing), but the proof for $p > 0$ follows along the same lines. Standard calculations show that

$$\widehat{\sigma}^2(x) - \sigma^2(x)$$
$$= f_X^{-1}(x)T^{-1}\sum_{t=1}^T K_h(x - \widehat{X}_{t-1})\Big\{(Y_t - \mu(x))^2 - \sigma^2(x)\Big\} + o_P(T^{-1/2})$$

uniformly in x. Hence,

$$\widehat{\sigma}(x) - \sigma(x)$$
$$= [2\sigma(x)f_X(x)]^{-1}T^{-1}\sum_{t=1}^T K_h(x - \widehat{X}_{t-1})\Big\{(Y_t - \mu(x))^2 - \sigma^2(x)\Big\} + o_P(T^{-1/2}).$$

This means that if we define

$$\xi(x) = -w(x)\sigma^{-2}(x)\frac{\partial}{\partial u}g(\theta_0, u, x)|_{u=\sigma(x)}\frac{\partial}{\partial \theta}g(\theta_0, \sigma(x), x),$$

18

we can write the first term of (6) as

$$E\left[\xi(X_{t-1})\left\{\widehat{\sigma}(X_{t-1}) - \sigma(X_{t-1})\right\}\right]$$

$$= T^{-1}\sum_{t=1}^{T} E\left[\xi(X)(2\sigma(X)f_X(X))^{-1}K_h(X - \widehat{X}_{t-1})\left\{(Y_t - \mu(X))^2 - \sigma^2(X)\right\}\right]$$

$$+ o_P(T^{-1/2}),$$

where the latter expected value is conditional on (I_{t-1}, Y_t) $(t = 1, \ldots, T)$ (and hence also on $\widehat{X}_{t-1}$), and where X has the same distribution as X_{t-1}. The latter expression equals

$$\frac{1}{2}T^{-1}\sum_{t=1}^{T}\int \xi(\widehat{X}_{t-1} + uh)\sigma^{-1}(\widehat{X}_{t-1} + uh)K(u)\left\{(Y_t - \mu(\widehat{X}_{t-1} + uh))^2\right.$$

$$\left. - \sigma^2(\widehat{X}_{t-1} + uh)\right\}du + o_P(T^{-1/2})$$

$$= \frac{1}{2}T^{-1}\sum_{t=1}^{T}\xi(\widehat{X}_{t-1})\sigma^{-1}(\widehat{X}_{t-1})\left\{(Y_t - \mu(\widehat{X}_{t-1}))^2 - \sigma^2(\widehat{X}_{t-1})\right\} + o_P(T^{-1/2})$$

$$= \frac{1}{2}T^{-1}\sum_{t=1}^{T}\xi(X_{t-1})\sigma^{-1}(X_{t-1})\left\{(Y_t - \mu(X_{t-1}))^2 - \sigma^2(X_{t-1})\right\}$$

$$- T^{-1}\sum_{t=1}^{T}\xi(X_{t-1})\sigma^{-1}(X_{t-1})\left\{(Y_t - \mu(X_{t-1}))\frac{\partial}{\partial x'}\mu(x)|_{x=X_{t-1}}\right.$$

$$\left. + \sigma(X_{t-1})\frac{\partial}{\partial x'}\sigma(x)|_{x=X_{t-1}}\right\}\frac{\partial}{\partial\beta'}\gamma(I_{t-1}, \beta_0)(\widehat{\beta} - \beta_0) + o_P(T^{-1/2})$$

$$= \frac{1}{2}T^{-1}\sum_{t=1}^{T}\xi(X_{t-1})\sigma^{-1}(X_{t-1})\left\{(Y_t - \mu(X_{t-1}))^2 - \sigma^2(X_{t-1})\right\}$$

$$- T^{-1}\sum_{t=1}^{T}\xi(X_{t-1})\frac{\partial}{\partial x'}\sigma(x)|_{x=X_{t-1}}\frac{\partial}{\partial\beta'}\gamma(I_{t-1}, \beta_0)(\widehat{\beta} - \beta_0) + o_P(T^{-1/2})$$

$$= -\frac{1}{2}T^{-1}\sum_{t=1}^{T}w(X_{t-1})\sigma^{-1}(X_{t-1})g_{ut}(\theta_0)g_{\theta t}(\theta_0)\left\{\varepsilon_t^2 - 1\right\}$$

$$+ E\left[w(X_{t-1})\sigma^{-2}(X_{t-1})g_{ut}(\theta_0)g_{\theta t}(\theta_0)\frac{\partial}{\partial x'}\sigma(x)|_{x=X_{t-1}}\frac{\partial}{\partial\beta'}\gamma(I_{t-1}, \beta_0)\right]T^{-1}\sum_{t=1}^{T}r_t$$

$$+ o_P(T^{-1/2}). \tag{8}$$

This together with (7) finishes the calculation of the term T_2 in (6). It remains to consider the term T_3. It can be easily shown that this term is $o_P(T^{-1/2})$ by using similar arguments as for T_2, and taking into account that $E(Y_t - g(\theta_0, \sigma(X_{t-1}), X_{t-1})|X_{t-1}) = 0$. Combining this with (8.2), (5), (7) and (8) leads to the result. □

19

References

Abel, A. B. (1988). Stock prices under time-varying dividend risk: An exact solution in an infinite-horizon general equilibrium model. *Journal of Monetary Economics*, 22, 375-393.

Abel, A. B. (1999). Risk premia and term premia in general equilibrium. *Journal of Monetary Economics*, 43, 3-33.

Backus, D. K. and Gregory, A. W. (1993). Theoretical relations between risk premiums and conditional variance. *Journal of Business and Economic Statistics*, 11, 177-185.

Bekaert, G. and Harvey, C. R. (1995). Time-Varying world market integration. *Journal of Finance*, 50, 403-444.

Bansal, R. and Yaron, A. (2004). Risks for the long run: a potential resolution of asset pricing puzzles. *The Journal of Finance*, 59, 1481-1509.

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

Brandt, M. W. and Kang, Q. (2004). On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics*, 72, 217-257.

Brandt, M. W. and Wang, L. (2010). Measuring the time-varying risk-return relation from the cross-section of equity returns. *Unpublished manuscript.*

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18, 373-399.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107, 205-251.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31, 281-318.

Cochrane, J. H. (2005). *Asset Pricing.* Princeton University Press, Princeton, NJ.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica*, 53, 363-384.

Dedecker, J. and Louhichi, L. (2002). Maximal inequalities and empirical central limit theorems. In: *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, editors). Birkhäuser, Boston.

De Santis, G. and Gerard, B. (1997). International asset pricing and portfolio diversification with time-varying risk. *The Journal of Finance*, 52, 1881-1912.

Engle, R., Lilien, D. and Robins, R. (1987). Estimation of time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, 55, 391-407.

Fama. E. and French, K. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96, 246-273.

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

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

Ferson, W. (1989). Changes in expected security returns, risk and level of interest rates. *The Journal of Finance*, 44, 1191-1217.

Ferson, W., Foerster, S. and Keim, D. (1993). General tests of latent variable models and mean-variance spanning. *The Journal of Finance*, 48, 131-156.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19, 3-29.

Gennotte, G. and Marsh, T. (1993). Variations in economic uncertainty and risk premiums on capital assets. *European Economic Review*, 37, 1021-1041.

Ghosh, A. and Linton, O. (2009). Consistent estimation of the risk-return trade off in the presence of measurement error. Working Paper 09-49, Departamento de Economía, Universidad Carlos III de Madrid.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76, 509-548.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48, 1779-1801.

Guo, H. and Whitelaw, R. F. (2006). Uncovering the risk-return relation in the stock market. *The Journal of Finance*, 61, 1433-1463.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24, 726-748.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, 21, 157-178.

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

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

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics*, 58, 71-120.

Keim, D. B. and Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17, 357-390.

Lettau, M. and Ludvigson, S. C. (2010). Measuring and modeling variation in the risk-return tradeoff. In: *Handbook of Financial Econometrics* (Y.

Ait-Sahalia and L. P. Hansen, editors). North-Holland, Amsterdam.

Linton, O. and Perron, B. (2003). The shape of the risk premium: Evidence from a semiparametric GARCH Model. *Journal of Business and Economic Statistics*, 21, 354-367.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk-return relation: A factor analysis approach.

Journal of Financial Economics, 83, 171-222.

Lundblad, C. (2007). The risk return tradeoff in the long run: 1836-2003. *Journal of Financial Economics*, 85, 123-150.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41, 867-887.

Merton, R. C. (1980). On estimating the expected return on the market. An explanatory investigation. *Journal of Financial Economics*, 8, 323-361.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, 101, 1067-1078.

Pagan, A. and Hong, Y. (1991). Nonparametric estimation and the risk premium. In: *Nonparametric Models in Econometrics and Statistics: Proceedings of the Fifth International in Economic Theory and Econometrics* (W. A. Barnett, J. A. Powell and G. E. Tauchen, editors). Cambridge University Press, Cambridge, New York and Melbourne.

Pagan, A. and A. Ullah (1988). The econometric analysis of models with risk premium. *Journal of Applied Econometrics* 3, 87-105.

Pástor, L., Sinha, M. and Swaminathan, B. (2008). Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *The Journal of Finance*, 63, 2859-2897.

Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. Springer.

Scruggs, J. and Glabadanidis, P. (2003). Risk premia and the dynamics covariance between stock and bond returns. *Journal of Financial and Quantitative Analysis*, 38, 295-316.

Tang, Y. and Whitelaw, R. F. (2011). Time-varying Sharpe ratios and market timing. *Quarterly Journal of Finance*, 1, 465-493.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer.

Veronesi, P. (2000). How does information quality affect stock returns?. *Journal of Finance*, 55, 807-837.

Whitelaw, R. F. (1994). Time variations and covariations in the expectation and volatility of stock market returns. *The Journal of Finance*, 49, 515-541.

Whitelaw, R. F. (2000). Stock market risk and return: An equilibrium approach. *The Review of Financial Studies*, 13, 521-547.

Yang, L. (2006). A semiparametric GARCH model for foreign exchange volatility. *Journal of Econometrics*, 130, 365-384.

Yogo, M. (2008). Asset prices under habit formation and reference-dependent preferences. *Journal of Business and Economic Statistics*, 26, 131-143.

Table 1: Observed mean square error (MSE) $\times 100$ of the estimator of θ_0, empirical coverage of the confidence intervals based on the asymptotic normality of the estimator and average length of the confidence intervals. The nominal confidence level is 0.95.

			MSE				CI coverage				CI average length			
T	θ_0	$h:$	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4
100	-0.50		1.97	2.14	1.99	1.92	0.953	0.942	0.946	0.949	0.528	0.533	0.529	0.527
100	-0.25		1.61	1.70	1.62	1.58	0.934	0.935	0.936	0.936	0.451	0.456	0.452	0.450
100	0.00		1.25	1.29	1.25	1.22	0.945	0.943	0.945	0.947	0.423	0.428	0.425	0.422
100	0.25		1.39	1.48	1.40	1.37	0.956	0.954	0.953	0.954	0.451	0.456	0.452	0.450
100	0.50		2.14	2.35	2.16	2.05	0.942	0.940	0.948	0.948	0.532	0.537	0.533	0.530
100	2.00		14.33	16.25	14.45	13.45	0.949	0.936	0.944	0.951	1.332	1.344	1.336	1.328
100	2.25		20.31	23.37	20.41	18.68	0.949	0.938	0.951	0.956	1.514	1.532	1.519	1.508
100	2.50		23.74	26.75	23.50	21.89	0.940	0.935	0.938	0.945	1.657	1.679	1.662	1.651
200	-0.50		0.94	1.00	0.95	0.92	0.945	0.939	0.947	0.946	0.368	0.371	0.369	0.368
200	-0.25		0.65	0.67	0.65	0.65	0.956	0.956	0.955	0.955	0.314	0.316	0.315	0.314
200	0.00		0.60	0.62	0.61	0.60	0.939	0.938	0.939	0.939	0.293	0.295	0.294	0.293
200	0.25		0.69	0.71	0.70	0.69	0.949	0.951	0.949	0.949	0.315	0.317	0.316	0.315
200	0.50		0.98	1.03	0.99	0.96	0.943	0.937	0.942	0.948	0.368	0.370	0.368	0.367
200	2.00		5.79	6.35	5.82	5.48	0.950	0.948	0.948	0.953	0.932	0.937	0.933	0.929
200	2.25		7.78	8.34	7.60	7.16	0.950	0.947	0.953	0.954	1.047	1.052	1.048	1.043
200	2.50		9.38	10.29	9.38	8.79	0.962	0.955	0.961	0.963	1.149	1.154	1.149	1.143

Table 2: Observed mean square error (MSE) $\times 100$ of the oracle estimator of θ_0 when the parameter β_0 is assumed to be known, empirical coverage of the confidence intervals based on the asymptotic normality of the estimator and average length of the confidence intervals. The nominal confidence level is 0.95.

			MSE				CI coverage				CI average length			
T	θ_0	$h:$	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4
100	-0.50		1.78	1.90	1.78	1.73	0.954	0.949	0.949	0.952	0.519	0.522	0.519	0.517
100	-0.25		1.51	1.57	1.51	1.47	0.945	0.946	0.948	0.949	0.448	0.452	0.449	0.447
100	0.00		1.19	1.23	1.20	1.18	0.950	0.949	0.948	0.950	0.420	0.423	0.420	0.419
100	0.25		1.49	1.56	1.49	1.46	0.935	0.933	0.937	0.938	0.449	0.454	0.450	0.448
100	0.50		1.97	2.14	2.02	1.96	0.948	0.939	0.947	0.952	0.525	0.529	0.525	0.524
100	2.00		11.10	12.01	11.17	10.71	0.939	0.942	0.936	0.934	1.273	1.279	1.270	1.266
100	2.25		13.74	15.47	14.35	13.75	0.961	0.951	0.952	0.959	1.426	1.437	1.425	1.419
100	2.50		18.07	19.86	18.55	17.77	0.933	0.926	0.920	0.920	1.553	1.562	1.552	1.546
200	-0.50		0.90	0.95	0.92	0.90	0.953	0.949	0.950	0.952	0.366	0.367	0.365	0.365
200	-0.25		0.64	0.66	0.64	0.64	0.955	0.955	0.956	0.956	0.312	0.313	0.312	0.311
200	0.00		0.59	0.60	0.59	0.59	0.948	0.948	0.949	0.949	0.292	0.294	0.293	0.292
200	0.25		0.72	0.74	0.72	0.71	0.942	0.945	0.944	0.945	0.313	0.315	0.314	0.313
200	0.50		0.94	0.98	0.94	0.93	0.943	0.939	0.941	0.941	0.363	0.364	0.363	0.363
200	2.00		4.64	4.92	4.78	4.70	0.950	0.948	0.948	0.951	0.892	0.895	0.891	0.890
200	2.25		6.32	6.61	6.51	6.45	0.940	0.936	0.936	0.937	0.993	0.994	0.991	0.989
200	2.50		7.63	8.00	7.90	7.81	0.935	0.930	0.930	0.930	1.079	1.080	1.078	1.076

Table 3: Empirical power for the test $H_0 : \theta_0 = 0$ versus $H_1 : \theta_0 \neq 0$. Test #1: realistic t-test, with estimated β_0; Test #2: oracle t-test, without estimating β_0; Test #3: ordinary t-test test, without taking into account the estimation of the conditional variance. The significance level is 0.05.

			Test #1				Test #2				Test #3			
T	θ_0	$h:$	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4
100	-0.50		0.993	0.993	0.993	0.993	0.997	0.997	0.997	0.997	0.999	0.997	0.999	0.999
100	-0.25		0.642	0.644	0.642	0.639	0.632	0.636	0.633	0.630	0.739	0.717	0.728	0.736
100	0.00		0.055	0.057	0.055	0.053	0.050	0.051	0.052	0.050	0.053	0.052	0.053	0.051
100	0.25		0.650	0.660	0.652	0.649	0.654	0.665	0.660	0.654	0.590	0.584	0.584	0.590
100	0.50		0.996	0.996	0.996	0.996	0.995	0.995	0.995	0.995	0.979	0.981	0.982	0.981
200	-0.50		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	-0.25		0.916	0.918	0.916	0.914	0.911	0.914	0.912	0.911	0.963	0.953	0.956	0.959
200	0.00		0.061	0.062	0.061	0.061	0.052	0.052	0.051	0.051	0.059	0.055	0.054	0.058
200	0.25		0.923	0.925	0.923	0.923	0.921	0.921	0.921	0.919	0.897	0.895	0.896	0.896
200	0.50		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Empirical power for the test $H_0 : \theta_0 = 2.5$ versus $H_1 : \theta_0 \neq 2.5$. Test #1: realistic t-test, with estimated β_0; Test #2: oracle t-test, without estimating β_0; Test #3: ordinary t-test test, without taking into account the estimation of the conditional variance. The significance level is 0.05.

			Test #1				Test #2				Test #3			
T	θ_0	h:	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4
100	2.00		0.274	0.255	0.273	0.289	0.366	0.345	0.373	0.385	0.853	0.849	0.863	0.861
100	2.25		0.092	0.091	0.087	0.087	0.157	0.136	0.163	0.166	0.646	0.657	0.655	0.655
100	2.50		0.060	0.065	0.062	0.055	0.067	0.074	0.080	0.080	0.460	0.482	0.479	0.474
100	2.75		0.093	0.122	0.095	0.079	0.050	0.057	0.050	0.042	0.326	0.366	0.336	0.324
100	3.00		0.249	0.281	0.245	0.221	0.102	0.132	0.107	0.096	0.332	0.362	0.344	0.337
200	2.00		0.502	0.470	0.492	0.526	0.635	0.607	0.638	0.649	0.970	0.970	0.971	0.970
200	2.25		0.137	0.129	0.133	0.138	0.253	0.236	0.259	0.267	0.787	0.800	0.798	0.793
200	2.50		0.038	0.045	0.039	0.037	0.065	0.070	0.070	0.070	0.512	0.530	0.525	0.521
200	2.75		0.162	0.176	0.156	0.137	0.055	0.072	0.063	0.060	0.357	0.394	0.374	0.363
200	3.00		0.372	0.395	0.365	0.346	0.135	0.148	0.136	0.125	0.422	0.446	0.431	0.423

Table 5: Empirical power for the test $H_0 : \theta_0 = \theta$ versus $H_1 : \theta_0 \neq \theta$ with data generated under the null hypothesis. Test #1: realistic t-test, with estimated β_0; Test #2: oracle t-test, without estimating β_0; Test #3: ordinary t-test test, without taking into account the estimation of the conditional variance. The significance level is 0.05.

			Test #1				Test #2				Test #3			
T	θ	h:	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4	c-v	0.2	0.3	0.4
100	-2.5		0.051	0.060	0.050	0.044	0.049	0.061	0.058	0.062	0.243	0.190	0.214	0.231
100	-2.0		0.041	0.050	0.047	0.040	0.050	0.061	0.058	0.052	0.245	0.207	0.225	0.242
100	-1.5		0.049	0.060	0.054	0.049	0.053	0.055	0.050	0.052	0.216	0.191	0.199	0.203
100	-1.0		0.050	0.058	0.049	0.046	0.039	0.050	0.044	0.045	0.145	0.133	0.131	0.137
100	-0.5		0.052	0.056	0.052	0.053	0.058	0.064	0.058	0.055	0.093	0.079	0.084	0.092
100	0.0		0.051	0.052	0.051	0.051	0.060	0.062	0.061	0.059	0.048	0.051	0.047	0.047
100	0.5		0.051	0.055	0.052	0.050	0.044	0.049	0.047	0.043	0.090	0.105	0.101	0.095
100	1.0		0.058	0.060	0.053	0.055	0.067	0.068	0.065	0.065	0.197	0.223	0.214	0.209
100	1.5		0.048	0.054	0.049	0.044	0.043	0.049	0.051	0.048	0.262	0.295	0.271	0.268
100	2.0		0.063	0.067	0.060	0.053	0.063	0.068	0.063	0.065	0.357	0.387	0.369	0.362
100	2.5		0.059	0.066	0.053	0.050	0.060	0.062	0.063	0.062	0.475	0.500	0.497	0.493
200	-2.5		0.056	0.055	0.049	0.045	0.074	0.074	0.073	0.074	0.199	0.163	0.181	0.195
200	-2.0		0.054	0.061	0.057	0.054	0.055	0.063	0.064	0.064	0.221	0.198	0.207	0.214
200	-1.5		0.041	0.042	0.042	0.040	0.046	0.046	0.046	0.046	0.186	0.166	0.178	0.182
200	-1.0		0.042	0.046	0.040	0.036	0.046	0.047	0.048	0.049	0.159	0.142	0.145	0.146
200	-0.5		0.042	0.045	0.048	0.043	0.057	0.056	0.057	0.059	0.072	0.065	0.064	0.065
200	0.0		0.042	0.041	0.042	0.042	0.042	0.043	0.042	0.041	0.049	0.047	0.053	0.051
200	0.5		0.042	0.047	0.041	0.038	0.048	0.049	0.049	0.049	0.089	0.100	0.096	0.089
200	1.0		0.053	0.056	0.050	0.049	0.058	0.064	0.060	0.057	0.194	0.214	0.208	0.202
200	1.5		0.043	0.054	0.048	0.040	0.044	0.051	0.044	0.045	0.262	0.282	0.275	0.271
200	2.0		0.047	0.052	0.048	0.044	0.050	0.064	0.060	0.057	0.402	0.425	0.407	0.406
200	2.5		0.041	0.047	0.042	0.036	0.059	0.060	0.064	0.064	0.531	0.571	0.550	0.543

Table 6: Parametric estimates 1947-2008.

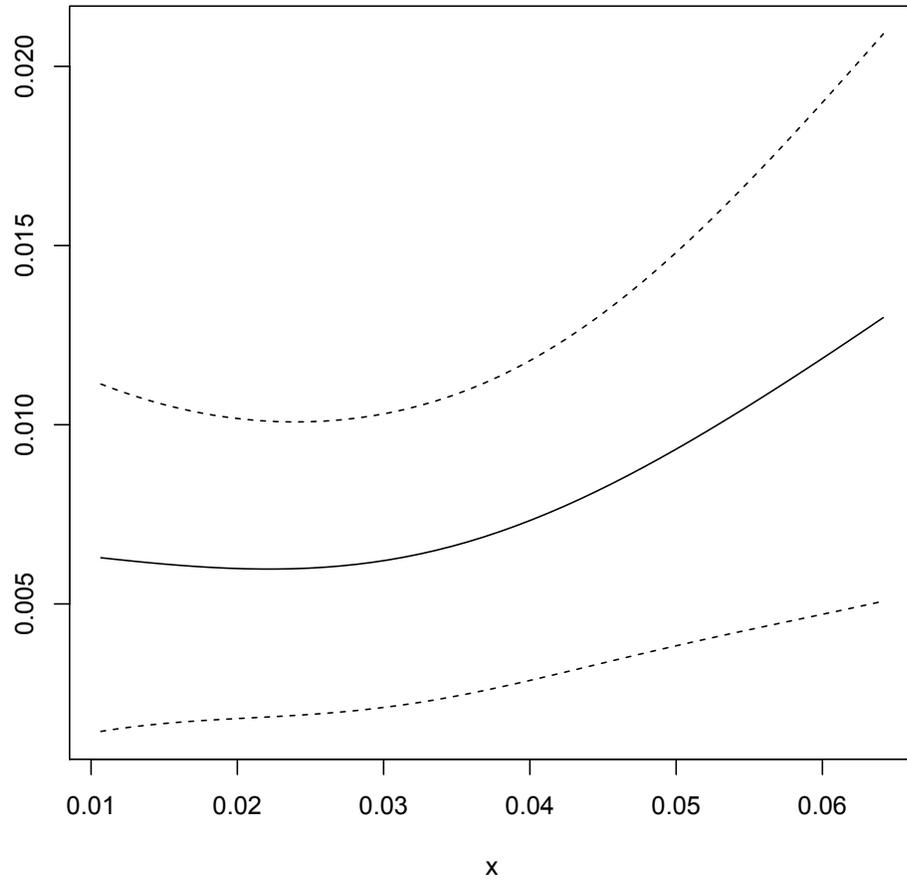| h	0.5×c-v	c-v	1.5×c-v
θ_0	2.664 (0.8725)	2.643 (0.865)	2.626 (0.860)
95% CI	[0.954, 4.370]	[0.946, 4.341]	[0.939, 4.313]

Note: Standard errors in parenthesis

Figure 1: Nonparametric estimator of conditional mean of excess returns: Nadaraya-Watson estimate of $E[Y_t|X_{t-1}]$ (solid line) with bandwidths chosen by Least-Squares cross-validation and 95% pointwise confidenceintervals based on 1000 bootstrapped replications (dotted lines).
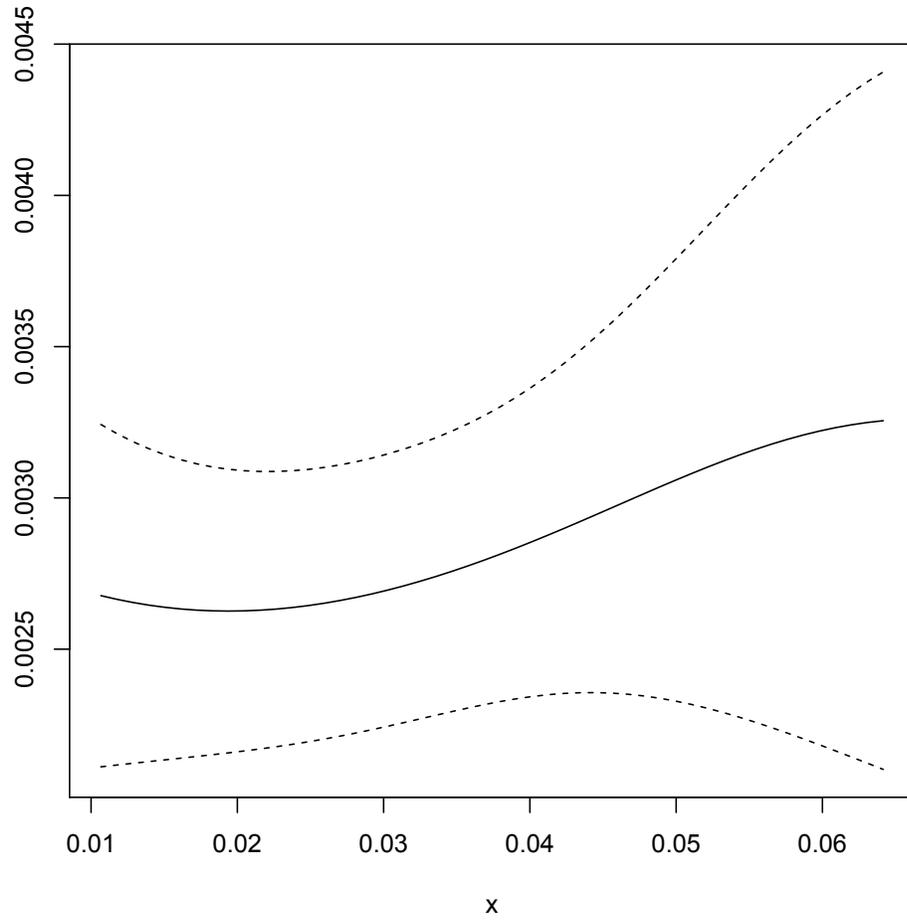
Figure 2: Nonparametric estimator of the conditional variance of excess returns. Nadaraya-Watson estimate of $\mathrm{Var}[Y_t|X_{t-1}]$ (solid line) with bandwidths chosen by Least-Squares cross-validation and 95% pointwise confidence intervals based on 1000 bootstrapped replications (dotted lines).
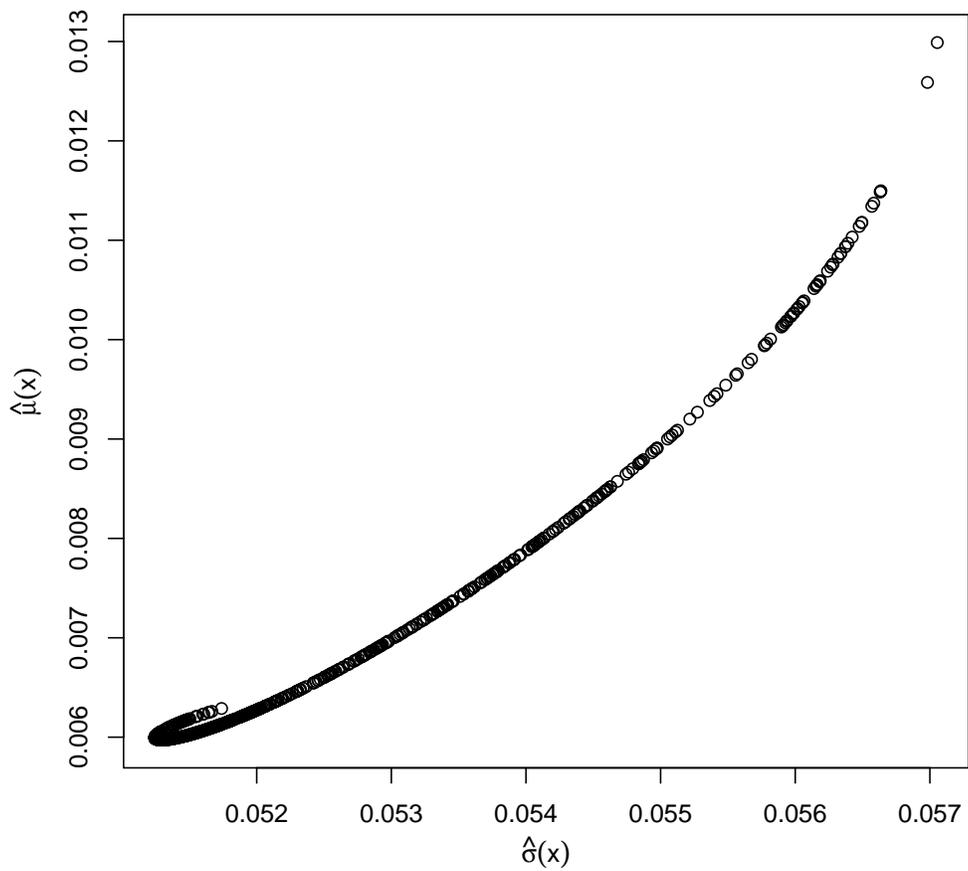
Figure 3: Nonparametric estimators of risk-return. Nadaraya-Watson estimates of pairs $(\widehat{\sigma}(\widehat{X}_{t-1}), \widehat{\mu}(\widehat{X}_{t-1}))$ for all t.