

A CONSISTENT NONPARAMETRIC TEST FOR CAUSALITY IN QUANTILE

KIHO JEONG

Kyungpook National University

WOLFGANG K. HÄRDLE

Humboldt-Universität zu Berlin

SONG SONG

Humboldt-Universität zu Berlin

and

University of California, Berkeley

This paper proposes a nonparametric test of Granger causality in quantile. Zheng (1998, *Econometric Theory* 14, 123–138) studied the idea to reduce the problem of testing a quantile restriction to a problem of testing a particular type of mean restriction in independent data. We extend Zheng's approach to the case of dependent data, particularly to the test of Granger causality in quantile. Combining the results of Zheng (1998) and Fan and Li (1999, *Journal of Nonparametric Statistics* 10, 245–271), we establish the asymptotic normal distribution of the test statistic under a β -mixing process. The test is consistent against all fixed alternatives and detects local alternatives approaching the null at proper rates. Simulations are carried out to illustrate the behavior of the test under the null and also the power of the test under plausible alternatives. An economic application considers the causal relations between the crude oil price, the USD/GBP exchange rate, and the gold price in the gold market.

1. INTRODUCTION

Whether movements in one economic variable cause reactions in another variable is an important issue in economic policy and also for financial investment decisions. A framework for investigating causality between economic indicators has been developed by Granger (1969). Testing for Granger causality between

The research was conducted while Jeong was visiting C.A.S.E.—Center for Applied Statistics and Economics—Humboldt-Universität zu Berlin in the summers of 2005 and 2007. Jeong is grateful for their hospitality during the visits. Jeong's work was supported by a Korean Research Foundation grant funded by the Korean government (MOEHRD) (KRF-2006-B00002), and Härdle and Song's work was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk." We thank the editor, two anonymous referees, and Holger Dette for concrete suggestions on improving the manuscript and restructuring the paper. Their valuable comments and suggestions are gratefully acknowledged. Address correspondence to Kiho Jeong, Kyungpook National University, Korea; e-mail: khjeong@knu.ac.kr.

economic time series has been since studied intensively in empirical macroeconomics and empirical finance. The majority of research results were obtained in the context of Granger causality in the conditional mean. The conditional mean, though, is a questionable element of analysis if the distributions of the variables involved are nonelliptic or fat tailed as is to be expected with, for example, financial returns. The focus of a causality analysis on the mean might result in unclear news. The conditional mean is only one element of an overall summary for the conditional distribution. A tail area causal relation may be quite different from a causality based on the center of the distribution. Lee and Yang (2007) explore money-income Granger causality in the conditional quantile and find that Granger causality is significant in tail quantiles, whereas it is not significant in the center of the distribution.

An illustrating motivation for the research presented here is from labor market analysis where one tries to find out how income depends on the age of the employee for different education levels, genders, and nationalities, and so on (discrimination effects); see, for example, Buchinsky (1995). In particular, the effect of education on income is summarized by the basic claim of Day and Newburger (2007): At most ages, more education equates with higher earnings, and the payoff is most notable at the highest educational level, which is actually from the point of view of mean regression. However, whether this difference is significant or not is still questionable, especially for different ends of the (conditional) income distribution. Härdle, Ritov, and Song (2009) show that for the 0.20 quantile confidence bands for income given “university,” “apprenticeship,” and “low education” status do not differ significantly from one another although they become progressively lower, which indicates that high education does not equate to higher earnings significantly for the lower tails of income, whereas increasing age seems to be the main driving force. For the conditional median, the bands for “university” and “low education” differ significantly. For the 0.80 quantiles, all conditional quantiles differ, which indicates that higher education is associated with higher earnings. However, these findings do not necessarily indicate causalities. To answer the question “Does education Granger cause income in various conditional quantiles?” the concept of Granger causality in means cannot be used to estimate or test for these effects. Hence the need for the concept of Granger causality in quantiles and the need to develop tests for these effects emerge.

Another motivation comes from controlling and monitoring downside market risk and investigating large comovements between financial markets. These are important for risk management and portfolio/investment diversification (Hong, Liu, and Wang, 2009). Various other risk management tasks are described in Bollerslev (2001) and Campbell and Cochrane (1999) indicating the importance of Granger causality in quantile. Yet another motivation comes from the well-known robustness properties of the conditional quantile: like the parallel boxplot—calculated across an explanatory variable—the set of conditional quantiles characterizes the entire distribution in more detail.

Based on the kernel method, we propose a nonparametric test for Granger causality in quantile. Testing conditional quantile restrictions by nonparametric estimation techniques in dependent data situations has not been considered in the literature before. This paper intends to fill this literature gap. In an unpublished working paper that has been independently carried out from ours, Lee and Yang (2007) also propose a test for Granger causality in the conditional quantile. Their test, however, relies on linear quantile regression and thus is subject to possible functional misspecification of quantile regression. Recently, Hong et al. (2009) investigated Granger causality in value at risk (VaR) with a corresponding (kernel-based) test. Their method, however, offers two possible improvements. The first is that it needs a parametric specification of VaR, again subject to misspecification errors. The second is that their test does not directly check causality but rather a necessary condition for causality.

The problem of testing conditional mean restrictions using nonparametric estimation techniques has been actively studied for dependent data. Among the related work, the testing procedures of Fan and Li (1999) and Li (1999) use the general hypothesis of the form $E(\varepsilon|z) = 0$, where ε and z are the regression error term and the vector of regressors, respectively. They consider the distance measure of $J = E[\varepsilon E(\varepsilon|z) f(z)]$ to construct kernel-based procedures. For the advantages of using this distance measure in kernel-based testing procedures, see Li and Wang (1998) and Hsiao and Li (2001). A feasible test statistic based on J has a second-order degenerate U -statistic as the leading term under the null hypothesis. Generalizing the result of Hall (1984) for independent data, Fan and Li (1999) establish the asymptotic normal distribution for a general second-order degenerate U -statistic with dependent data.

All the results stated previously on testing mean restrictions are however irrelevant when testing quantile restrictions. Zheng (1998) proposed an idea to transform quantile restrictions to mean restrictions in independent data. Following his idea, one can use the existing technical results on testing mean restrictions in testing quantile restrictions. In this paper, by combining Zheng's idea and the results of Fan and Li (1999) and Li (1999), we derive a test statistic for Granger causality in quantile and establish the asymptotic normal distribution of the proposed test statistic under a β -mixing process. Our testing procedure can be extended to several hypothesis testing problems with conditional quantile in dependent data; for example, testing a parametric regression functional form, testing the insignificance of a subset of regressors, and testing semiparametric versus nonparametric regression models.

The paper is organized as follows. Section 2 presents the test statistic. Section 3 establishes the asymptotic normal distribution under the null hypothesis of no causality in quantile. Section 4 displays a fairly extensive simulation study to illustrate the behavior of the test under the null, in addition to the power of the test under plausible alternatives. Section 5 considers the causal relations between the crude oil and gold prices as an economic application. Section 6 concludes the paper. Technical proofs are given in the Appendix.

2. NONPARAMETRIC TEST FOR GRANGER CAUSALITY IN QUANTILE

To simplify the exposition, we assume a bivariate case, or that only $\{y_t, w_t\}$ are observable. Granger causality in mean (Granger, 1988) is defined as follows.

1. w_t does not cause y_t in mean with respect to $\{y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}\}$ if

$$E(y_t | y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}) = E(y_t | y_{t-1}, \dots, y_{t-p}) \quad \text{and}$$

2. w_t is a prima facie cause in mean of y_t with respect to $\{y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}\}$ if

$$E(y_t | y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}) \neq E(y_t | y_{t-1}, \dots, y_{t-p}).$$

Motivated by the definition of Granger causality in mean, we define Granger causality in quantile as follows.

1. w_t does not cause y_t in the θ -quantile with respect to $\{y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}\}$ if

$$Q_\theta(y_t | y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}) = Q_\theta(y_t | y_{t-1}, \dots, y_{t-p}). \quad (1)$$

2. w_t is a prima facie cause in the θ -quantile of y_t with respect to $\{y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}\}$ if

$$Q_\theta(y_t | y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q}) \neq Q_\theta(y_t | y_{t-1}, \dots, y_{t-p}), \quad (2)$$

where $Q_\theta(y_t|\cdot)$ is the θ th ($0 < \theta < 1$) conditional quantile of y_t given \cdot , which depends on t .

Denote $x_t \equiv (y_{t-1}, \dots, y_{t-p})$, $z_t \equiv (y_{t-1}, \dots, y_{t-p}, w_{t-1}, \dots, w_{t-q})$, and the conditional distribution function y_t given $z_t(x_t)$ by $F_{y_t|z_t}(y_t|z_t)(F_{y_t|x_t}(y_t|x_t))$, which is abbreviated as $F_{y|z}(y|z)$ ($F_{y|x}(y|x)$) later, and $v_t = (x_t, z_t)$. In this paper, $F_{y|z}(y|z)$ is assumed to be absolutely continuous in y for almost all $v = (x, z)$. Denote $Q_\theta(z_t) \equiv Q_\theta(y_t|z_t)$ and $Q_\theta(x_t) \equiv Q_\theta(y_t|x_t)$. Then we have, with probability 1,

$$F_{y|z}\{Q_\theta(z_t)|z_t\} = \theta, \quad v = (x, z) \quad \text{and}$$

from the definitions (1) and (2), the hypotheses to be tested are

$$H_0 : P\{F_{y|z}(Q_\theta(x_t)|z_t) = \theta\} = 1 \quad \text{a.s.} \quad (3)$$

$$H_1 : P\{F_{y|z}(Q_\theta(x_t)|z_t) = \theta\} < 1 \quad \text{a.s.} \quad (4)$$

Zheng (1998) proposed an idea to reduce the problem of testing a quantile restriction to a problem of testing a particular type of mean restriction. The null hypothesis (3) is true if and only if $E[1\{y_t \leq Q_\theta(x_t)|z_t\}] = \theta$ or $1\{y_t \leq Q_\theta(x_t)\} = \theta + \varepsilon_t$ where $E(\varepsilon_t|z_t) = 0$ and $1(\cdot)$ is the indicator function. For a list of related literature we refer to Li and Wang (1998) and Zheng (1998). Although various distance measures can be used to consistently test the hypothesis (3), we consider the following distance measure:

$$J \equiv E \left[\left\{ F_{y|z}(Q_\theta(x_t)|z_t) - \theta \right\}^2 f_z(z_t) \right], \tag{5}$$

with $f_{z_t}(z_t)$ being the marginal density function of z_t , which is sometimes abbreviated as $f_z(z_t)$. Note that $J \geq 0$ and the equality holds if, and only if, H_0 is true, with strict inequality holding under H_1 . Thus J can be used as a proper candidate for consistent testing of H_0 (Li, 1999, p. 104). Because $E(\varepsilon_t|z_t) = F_{y|z}\{Q_\theta(x_t)|z_t\} - \theta$ we have

$$J = E \{ \varepsilon_t E(\varepsilon_t|z_t) f_z(z_t) \}. \tag{6}$$

The test is based on a sample analogue of $E\{\varepsilon E(\varepsilon|z) f_z(z)\}$. Including the density function $f_z(z)$ avoids the problem of trimming on the boundary of the density function; see Powell, Stock, and Stoker (1989) for an analogue approach. The density-weighted conditional expectation $E(\varepsilon|z) f_z(z)$ can be estimated by kernel methods

$$\hat{E}(\varepsilon_t|z_t) \hat{f}_z(z_t) = \frac{1}{(T-1)h^m} \sum_{s \neq t}^T K_{ts} \varepsilon_s, \tag{7}$$

where $m = p + q$ is the dimension of z , $K_{ts} = K\{(z_t - z_s)/h\}$ is the kernel function, and h is a bandwidth. Then we have a sample analogue of J as

$$\begin{aligned} J_T &\equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \varepsilon_t \varepsilon_s \\ &= \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} [1\{y_t \leq Q_\theta(x_t)\} - \theta] [1\{y_s \leq Q_\theta(x_s)\} - \theta]. \end{aligned} \tag{8}$$

The θ th conditional quantile of y_t given x_t , $Q_\theta(x_t)$, can also be estimated by the nonparametric kernel method

$$\hat{Q}_\theta(x_t) = \hat{F}_{y|x}^{-1}(\theta|x_t), \tag{9}$$

where

$$\hat{F}_{y|x}(y_t|x_t) = \frac{\sum_{s \neq t} L_{ts} 1(y_s \leq y_t)}{\sum_{s \neq t} L_{ts}} \tag{10}$$

is the Nadaraya–Watson kernel estimator of $F_{y|x}(y_t|x_t)$ with the kernel function of $L_{ts} = L(x_t - x_s)/a$ and the bandwidth parameter of a . The unknown error ε can be estimated as

$$\hat{\varepsilon}_t \equiv I \left\{ y_t \leq \hat{Q}_\theta(x_t) \right\} - \theta. \tag{11}$$

Replacing ε by $\hat{\varepsilon}$, we have a feasible kernel-based test statistic of J ,

$$\begin{aligned} \hat{J}_T &\equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \hat{\varepsilon}_t \hat{\varepsilon}_s \\ &= \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \left[1 \left\{ y_t \leq \hat{Q}_\theta(x_t) \right\} - \theta \right] \left[1 \left\{ y_s \leq \hat{Q}_\theta(x_s) \right\} - \theta \right]. \end{aligned} \tag{12}$$

3. THE LIMITING DISTRIBUTIONS OF THE TEST STATISTIC

Two existing works are useful in deriving the limiting distribution of the test statistic; one is Theorem 2.3 of Franke and Mwita (2003) on the uniform convergence rate of a nonparametric quantile estimator; another is Lemma 2.1 of Li (1999) on the asymptotic distribution of a second-order degenerate U -statistic, which is derived from Theorem 2.1 of Fan and Li (1999). We restate these results in lemmas subsequently for ease of reference. We collect the assumptions needed for Theorem 3.1.

(A1)

- (a) $\{y_t, w_t\}_{t=1}^T$ is strictly stationary and absolutely regular with mixing coefficients $\beta(\tau) = \mathcal{O}(\rho^\tau)$ for some $0 < \rho < 1$.
- (b) For some integer $v \geq 2$, $f_y, f_z,$ and f_x all are bounded and belong to \mathfrak{A}_v^∞ (see (D2) later in this section).
- (c) Use $\mu_s^t(z)$ ($\mu_s^t(\varepsilon)$) to denote the σ algebra generated by (z_s, \dots, z_t) ($(\varepsilon_s, \dots, \varepsilon_t)$) for $s \leq t$. With probability 1, $E \left[\varepsilon_t | \mu_{-\infty}^t(z), \mu_{-\infty}^{t-1}(\varepsilon) \right] = 0$, that is, the error ε_t is a martingale difference process. The terms $E \left[\varepsilon_t^{4+\eta} \right] < \infty$ and $E \left[\left| \varepsilon_{t_1}^{i_1} \varepsilon_{t_2}^{i_2} \dots \varepsilon_{t_l}^{i_l} \right|^{1+\xi} \right] < \infty$ for some arbitrarily small $\eta > 0$ and $\xi > 0$, where $2 \leq l \leq 4$ is an integer, $0 \leq i_j \leq 4$, and $\sum_{j=1}^l i_j \leq 8$. The terms $\sigma_\varepsilon^2(z) = E(\varepsilon_t^2 | z_t = z)$ and $\mu_{\varepsilon 4}(z) = E \left[\varepsilon_t^4 | z_t = z \right]$ all satisfy some Lipschitz conditions: $|a(u+v) - a(u)| \leq D(u) \|v\|$ with $E \left[|D(z)|^{2+\eta'} \right] < \infty$ for some small $\eta' > 0$, where $a(\cdot) = \sigma_\varepsilon^2(\cdot), \mu_{\varepsilon 4}(\cdot)$.
- (d) Let $f_{\tau_1, \dots, \tau_l}(\cdot)$ be the joint probability density function of $(z_{\tau_1}, \dots, z_{\tau_l})$ ($1 \leq l \leq 3$). Then $f_{\tau_1, \dots, \tau_l}(\cdot)$ is bounded and satisfies a Lipschitz condition: $|f_{\tau_1, \dots, \tau_l}(z_1 + u_1, z_2 + u_2, \dots, z_l + u_l) - f_{\tau_1, \dots, \tau_l}(z_1, z_2, \dots, z_l)| \leq D_{\tau_1, \dots, \tau_l}(z_1, \dots, z_l) \|u\|$, where $u = (u_1, \dots, u_l), z = (z_1, \dots, z_l)$, and $D_{\tau_1, \dots, \tau_l}(\cdot)$ is integrable and satisfies the condition that $\int \int \int D_{\tau_1, \dots, \tau_l}(z_1, \dots, z_l) \|z\|^{2\xi} dz_1, \dots, dz_l < M < \infty$ and $\int \int \int D_{\tau_1, \dots, \tau_l}(z_1, \dots, z_l) f_{\tau_1, \dots, \tau_l}(z_1, \dots, z_l) dz_1, \dots, dz_l < M < \infty$ for some $\xi > 1$.

- (e) For any y and x satisfying $0 < F_{y|x}(y|x) < 1$ and $f_x(x) > 0$, $F_{y|x}$ and $f_x(x)$ are continuous and bounded in x and y ; for fixed y , the conditional distribution function $F_{y|x}$ and the conditional density function $f_{y|x}$ belong to \mathfrak{A}_3^∞ ; $f_{y|x}(Q_\theta(x)|x) > 0$ for all x ; $f_{y|x}$ is uniformly bounded in x and y by, say, c_f .
- (f) For some compact set G , there are $\varepsilon > 0$ and $\gamma > 0$ such that $f_x \geq \gamma$ for all x in the ε -neighborhood $\{x \mid \|x - u\| < \varepsilon, u \in G\}$ of G . For the compact set G and some compact neighborhood Θ_0 of 0 , the set $\Theta = \{v = Q_\theta(x) + \mu \mid x \in G, \mu \in \Theta_0\}$ is compact, and for some constant $c_0 > 0$, $f_{y|x}(y|x) \geq c_0$ for all $x \in G, v \in \Theta$.
- (g) There is an increasing sequence s_T of positive integers such that for some finite A ,

$$\frac{T}{s_T} \beta^{2s_T/(3T)}(s_T) \leq A, \quad 1 \leq s_T \leq \frac{T}{2} \quad \text{for all } T \geq 1.$$

(A2)

- (a) We use product kernels for both $L(\cdot)$ and $K(\cdot)$. Let l and k be their corresponding univariate kernel which is bounded and symmetric. Then $l(\cdot)$ is nonnegative, $l(\cdot) \in \Upsilon_v$, $k(\cdot)$ is nonnegative, and $k(\cdot) \in \Upsilon_2$.
- (b) $h = \mathcal{O}(T^{-\alpha'})$ for some $0 < \alpha' < (7/8)m$.
- (c) $a = \mathcal{O}(1)$ and $\tilde{S}_T = T a^p (s_T \log T)^{-1} \rightarrow \infty$ for some $s_T \rightarrow \infty$.
- (d) A positive number δ exists such that for $r = 2 + \delta$ and a generic number M_0

$$\int \int \left| \frac{1}{h^m} K \left(\frac{z_1 - z_2}{h} \right) \right|^r dF_z(z_1) dF_z(z_2) \leq M_0 < \infty \quad \text{and}$$

$$E \left| \frac{1}{h^m} K \left(\frac{z_1 - z_2}{h} \right) \right|^r \leq M_0 < \infty.$$

- (e) For some $\delta' (0 < \delta' < \delta)$, $\beta(T) = \mathcal{O}(T^{-(2+\delta')/\delta'})$.

The following definitions are due to Robinson (1988).

DEFINITION (D1). $\Upsilon_\lambda, \lambda \geq 1$ is the class of even functions $k : R \rightarrow R$ satisfying $\int_R u^i k(u) du = \delta_{i0} (i = 0, 1, \dots, \lambda - 1)$,

$$k(u) = \mathcal{O} \left((1 + |u|^{\lambda+1+\varepsilon})^{-1} \right), \quad \text{for some } \varepsilon > 0,$$

where δ_{ij} is the Kronecker's delta.

DEFINITION (D2). $\mathfrak{A}_\mu^\alpha, \alpha > 0, \mu > 0$ is the class of functions $g : R^m \rightarrow R$ satisfying that g is $(d - 1)$ -times partially differentiable for $d - 1 \leq \mu \leq d$; for some $\rho > 0, \sup_{y \in \phi_{z\rho}} |g(y) - g(z) - G_g(y, z)| / |y - z|^\mu \leq D_g(z)$ for all z , where $\phi_{z\rho} = \{y \mid |y - z| < \rho\}$; $G_g = 0$ when $d = 1$; G_g is a $(d - 1)$ th degree

homogeneous polynomial in $y - z$ with coefficients being the partial derivatives of g at z of orders 1 through $d - 1$ when $d > 1$; and $g(z)$, its partial derivatives of order $d - 1$ and less, and $D_g(z)$ have finite α th moments.

The functions in \mathfrak{A}_μ^α are thus expanded in a Taylor series with a local Lipschitz condition on the remainder, (α, μ) depending simultaneously on smoothness and moment properties. Bounded functions in $\text{Lip}(\mu)$ (the Lipschitz class of degree μ) for $0 < \mu \leq 1$ are in \mathfrak{A}_μ^α ; for $\mu > 1$, \mathfrak{A}_μ^α contains the bounded and $(d - 1)$ -times boundedly differentiable functions whose $(d - 1)$ th partial derivatives are in $\text{Lip}(\mu - d + 1)$. In applying \mathfrak{A}_μ^α to f and F , we take $\alpha = \infty$.

Conditions (A1)(a)–(d) and (A2)(a) and (b) are adopted from conditions (D1) and (D2) of Li (1999), which are used to derive the asymptotic normal distribution of a second-order degenerate U -statistic. Assumption (A1)(a) requires $\{y_t, w_t\}_{t=1}^T$ to be a stationary absolutely regular process with geometric decay rate. Assumptions (A1)(b)–(d) are mainly some smoothness and moment conditions; these conditions are quite weak in the sense that they are similar to those used in Fan and Li (1996) for the independent data case. However, for autoregressive conditionally skedastic (ARCH) or generalized autoregressive conditionally heteroskedastic (GARCH) type error processes as considered in Engle (1982) and Bollerslev (1986), the error term ε_t may not have finite fourth moments in some situations. For example, let $\varepsilon_t | \varepsilon_{t-1} \sim N(0, \alpha_0 + \alpha_1 \varepsilon_{t-1}^2)$. Engle (1982) showed that ε_t does not have a finite fourth moment if $\alpha_1 > 1/\sqrt{3}$. Thus, Assumption (A1)(c) will be violated in such a case.

Assumption (A2)(a) requires $L(\cdot)$ to be a ν th- ($\nu \geq 2$) order kernel. This condition together with (A1)(b) ensures that the bias in the kernel estimation (of the null model) is bounded. The requirement that k is a nonnegative second-order kernel function in (A2)(b) is a quite weak and standard assumption.

Conditions (A1)(e)–(g) and (A2)(c) are technical conditions (A1), (A2), (B1), (B2), (C1), and (C2) of Theorem 2.3 of Franke and Mwita (2003), which are required to get the uniform convergence rate of the nonparametric kernel estimator of the conditional distribution function and corresponding conditional quantile with mixing data. Because the simple ARCH models (Engle, 1982; Masry and Tjøstheim, 1995, 1997), their extensions (Diebolt and Guegan, 1993), and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions, Assumption (A1)(g) is usually satisfied. There also exist simple methods to determine the mixing rates for various classes of random processes, for example, Gaussian, Markov, autoregressive moving average, ARCH, and GARCH. Hence the assumption of a known mixing rate is reasonable and has been adopted in many studies, for example, Györfi, Härdle, Sarda, and Vieu (1989), Irlle (1997), Meir (2000), Modha and Masry (1998), Roussas (1988), and Yu (1993). Auestad and Tjøstheim (1990) provided excellent discussions on the role of mixing for model identification in nonlinear time series analysis. But since the restriction of Assumption (A1)(c) as discussed before, ARCH or GARCH type processes may not satisfy all assumptions here. Finally conditions (A2)(d)

and (e) are adopted from conditions of Lemma 3.2 of Yoshihara (1976), which are required to get the asymptotic equivalence of the nondegenerate U -statistic and its projection under the β -mixing process. They are technical assumptions and are quite standard.

LEMMA 3.1 (Franke and Mwita, 2003). *Suppose conditions (A1)(e)–(g) and (A2)(c) hold. The bandwidth sequence is such that $a = \mathcal{O}(1)$ and $\tilde{S}_T = Ta^p(s_T \log T)^{-1} \rightarrow \infty$ for some $s_T \rightarrow \infty$. Let $S_T = a^2 + \tilde{S}_T^{-1/2}$. Then for the non-parametric kernel estimator of the conditional quantile of $\hat{Q}_\theta(x_t)$, equation (9), we have*

$$\sup_{\|x\| \in G} \left| \hat{Q}_\theta(x) - Q_\theta(x) \right| = \mathcal{O}(S_T) + \mathcal{O}\left(\frac{1}{Ta^p}\right) \quad a.s. \tag{13}$$

LEMMA 3.2 (Li, 1999). *Let $L_t = (\varepsilon_t, z_t)^T$ be a stochastic process that satisfies conditions (A1)(a)–(d). $\varepsilon_t \in \mathbb{R}$, $z_t \in \mathbb{R}^m$, and $K(\cdot)$ be the kernel function with h being the smoothing parameter that satisfies conditions (A2)(a) and (b). Define*

$$\sigma_\varepsilon^2(z) = E[\varepsilon_t^2 | z_t = z] \quad \text{and} \tag{14}$$

$$J_T \equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \varepsilon_t \varepsilon_s. \tag{15}$$

Then

$$Th^{m/2} J_T \rightarrow N(0, \sigma_0^2) \quad \text{in distribution,} \tag{16}$$

where $\sigma_0^2 = 2E\{\sigma_\varepsilon^4(z_t) f_z(z_t)\} \left\{ \int K^2(u) du \right\}$ and $f_z(\cdot)$ is the marginal density function of z_t .

We consider testing for local departures from the null that converge to the null at the rate $T^{-1/2}h^{-m/4}$. More precisely we consider the sequence of local alternatives

$$H_{1T} : F_{y|z}\{Q_\theta(x_t) + d_T l(z_t) | z_t\} = \theta, \tag{17}$$

where $d_T = T^{-1/2}h^{-m/4}$ and the function $l(\cdot)$ and its first-order derivatives are bounded.

THEOREM 3.1. *Assume the conditions (A1) and (A2). Then*

(i) *Under the null hypothesis (3), $Th^{m/2} \hat{J}_T \xrightarrow{L} N(0, \sigma_0^2)$ in distribution, where*

$$\sigma_0^2 = 2E\left\{ \sigma_\varepsilon^4(z_t) f_z(z_t) \right\} \left\{ \int K^2(u) du \right\} \quad \text{and}$$

$$\sigma_\varepsilon^2(z_t) = E(\varepsilon_t^2 | z_t) = \theta(1 - \theta).$$

(ii) *Under the null hypothesis (3), $\hat{\sigma}_0^2 \equiv 2\theta^2(1 - \theta)^2 / (T(T - 1)h^m) \sum_{s \neq t} K_{ts}^2$ is a consistent estimator of $\sigma_0^2 = 2E\{\sigma_\varepsilon^4(z_t) f_z(z_t)\} \int K^2(u) du$. Thus*

$$Th^{m/2} \hat{J}_T / \hat{\sigma}_0 = \sqrt{\frac{T}{T-1}} \frac{\sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \left[I \{y_t \leq \hat{Q}_\theta(x_t)\} - \theta \right] \left[I \{y_s \leq \hat{Q}_\theta(x_s)\} - \theta \right]}{\sqrt{2\theta(1-\theta)} \sqrt{\sum_{s \neq t} K_{ts}^2}}$$

(iii) Under the alternative hypothesis (4),

$$\hat{J}_T \rightarrow E\{[F_{y|z}(Q_\theta(x_t)|z_t) - \theta]^2 f_z(z_t)\} > 0 \quad \text{in probability.}$$

(iv) Under the local alternatives (A.2) in the Appendix, $Th^{m/2} \hat{J}_T \rightarrow N(\mu, \sigma_1^2)$ in distribution, where

$$\mu = E \left[f_{y|z}^2 \{Q_\theta(z_t)|z_t\} l^2(z_t) f_z(z_t) \right],$$

$$\sigma_1^2 = 2E \left\{ \sigma_v^4(z_t) f_z(z_t) \right\} \int K^2(u) du, \quad \text{and}$$

$$\sigma_v^2(z_t) = E(v_t^2|z_t) \quad \text{with } v_t \equiv I\{y_t \leq Q_\theta(x_t)\} - F(Q_\theta(x_t)|z_t).$$

Theorem 3.1 generalizes the results of Zheng (1998) for independent data to the weakly dependent data case. A detailed proof of Theorem 3.1 is given in the Appendix. The main difficulty in deriving the asymptotic distribution of the statistic defined in equation (12) is that a nonparametric quantile estimator is included in the indicator function that is not differentiable with respect to the quantile argument and thus prevents taking a Taylor expansion around the true conditional quantile $Q_\theta(x_t)$. To circumvent the problem, Zheng (1998) made use of the work of Sherman (1994) on uniform convergence of U -statistics indexed by parameters. In this paper, we bound the test statistic by the statistics in which the nonparametric quantile estimator in the indicator function is replaced with sums of the true conditional quantile and upper and lower bounds consistent with the uniform convergence rate of the nonparametric quantile estimator, $1(y_t \leq Q_\theta(x_t) - C_T)$ and $1(y_t \leq Q_\theta(x_t) + C_T)$.

An important further step is to show that the differences of the ideal test statistic J_T given in equation (8) and the statistics having the indicator functions obtained from the first step stated previously are asymptotically negligible. We may directly show that the second moments of the differences are asymptotically negligible by using the result of Yoshihara (1976) on the bound of moments of U -statistics for absolutely regular processes. However, it is tedious to get bounds on the second moments with dependent data. In the proof we use instead the fact that the differences are second-order degenerate U -statistics. Thus by using the result on the asymptotic normal distribution of the second-order degenerate U -statistic of Fan and Li (1999), we can derive the asymptotic variance that is based on the independent and identically distributed (i.i.d.) sequence having the same marginal distributions as weakly dependent variables in the test statistic. With this little

trick we only need to show that the asymptotic variance is $\mathcal{O}(1)$ in an i.i.d. situation. For details refer to the Appendix.

4. SIMULATION

We generate bivariate data $\{y_t, w_t\}_{t=1}^T$ according to the following model:

$$y_t = \frac{1}{2}y_{t-1} + cw_{t-1}^2 + \varepsilon_{1t},$$

$$w_t = 1 + \frac{1}{2}w_{t-1} + \varepsilon_{2t},$$

where ε_{1t} and ε_{2t} are independent standard normal random variables. Here $c = 0$ corresponds to the hypothetical model; that is, w_t does not cause y_t in the θ quantile with respect to $\{y_{t-1}, w_{t-1}\}$. All the coefficients are set such that the corresponding time series is stationary and β -mixing with corresponding densities bounded to satisfy the assumptions discussed before. We use different values of $c \in [0, 1]$ to investigate the power of the test, such that the higher c is, the stronger the causality of w_t on y_t is. Without loss of generality, we choose $\theta = 0.1, 0.5, 0.9$ and $T = 500, 1,000, 5,000$ here with the bandwidth h and a as in (7) and (10) as for a typical Nadaraya–Watson type estimator. We consider the nominal 0.05 significance level and repeat the test 500 times to generate the power.

Table 1 displays the power performance of the test for different combinations of T , c , and θ . First, obviously the power is very sensitive to the choice of T ; that is, the larger T is, for the same c and θ , the larger the power is. From a technical point of view, this makes sense, because the more data we have, the more evidence we can draw from to detect the “causality” effect. Our asymptotic result, Theorem 3.1, needs the plug-in estimation of the asymptotic covariance matrix that is used to normalize the test statistic. Note that such an estimator is model-dependent and under the alternative is consistent with a different value than the one under the null. As a result, the power deteriorates for small T . On the other hand, whether the causality effect exists or not is the nature of the series, which is independent of the sample size used in this technical test. Enhancing the power performance for small-sample data using the simulation-based method deserves further research. Second, as discussed before, the higher c is, the stronger the causality of w_t on y_t is, which is confirmed by the larger and larger power values. Third, for different quantiles θ , we find that the powers with respect to $\theta = 0.5$ are usually larger than the powers with respect to $\theta = 0.1$ and 0.9 .

5. APPLICATION TO COMMODITY PRICES

In financial and commodity markets, it has been argued that the covariation of the tails may be different from that of the rest of the distribution. The gold market is one of the most important markets in the world, where trading takes place 24 hours a day around the globe and transactions involving billions of dollars of

TABLE 1. Power performance for different combinations of T , c , and θ

c	Power (θ 0.1)	c	Power (θ 0.5)	c	Power (θ 0.9)
$T = 500$					
0.00	0.024	0.00	0.108	0.00	0.010
0.03	0.030	0.03	0.288	0.03	0.020
0.06	0.058	0.06	0.796	0.06	0.108
0.09	0.190	0.09	0.991	0.09	0.585
0.12	0.414	0.12	1.000	0.12	0.950
0.15	0.696	0.15	1.000	0.15	0.994
0.18	0.888	0.18	1.000	0.18	1.000
0.21	0.962	0.21	1.000	0.21	1.000
0.24	0.988	0.24	1.000	0.24	1.000
0.27	1.000	0.27	1.000	0.27	1.000
0.30	1.000	0.30	1.000	0.30	1.000
$T = 1,000$					
0.00	0.014	0.00	0.130	0.00	0.018
0.01	0.022	0.01	0.144	0.01	0.024
0.02	0.038	0.02	0.296	0.02	0.024
0.03	0.026	0.03	0.564	0.03	0.040
0.04	0.060	0.04	0.788	0.04	0.108
0.05	0.110	0.05	0.946	0.05	0.284
0.06	0.196	0.06	0.990	0.06	0.506
0.07	0.356	0.07	1.000	0.07	0.838
0.08	0.530	0.08	1.000	0.08	0.950
0.09	0.676	0.09	1.000	0.09	0.994
0.10	0.816	0.10	1.000	0.10	0.996
0.11	0.906	0.11	1.000	0.11	1.000
0.12	0.958	0.12	1.000	0.12	1.000
0.13	0.972	0.13	1.000	0.13	1.000
0.14	0.994	0.14	1.000	0.14	1.000
0.15	0.998	0.15	1.000	0.15	1.000
0.16	1.000	0.16	1.000	0.16	1.000
$T = 5,000$					
0.00	0.020	0.00	0.116	0.00	0.026
0.01	0.028	0.01	0.328	0.01	0.046
0.02	0.124	0.02	0.904	0.02	0.142
0.03	0.490	0.03	1.000	0.03	0.728
0.04	0.924	0.04	1.000	0.04	0.988
0.05	1.000	0.05	1.000	0.05	1.000
0.06	1.000	0.06	1.000	0.06	1.000
0.07	1.000	0.07	1.000	0.07	1.000
0.08	1.000	0.08	1.000	0.08	1.000
0.09	1.000	0.09	1.000	0.09	1.000
0.10	1.000	0.10	1.000	0.10	1.000

TABLE 2. Unit root tests

Variable	Test type	Time trend term	Test statistics	CR value 5%	Unit root	Unit root after differencing
LN Oil	DF	no	0.86955	-1.94160	yes	no
	ADF	no	0.72255	-1.94160	yes	no
	PP	no	0.73107	-1.94160	yes	no
	KPSS	no	2.16221	0.14600	yes	no
	DF	include	-0.81819	-2.86386	yes	no
	ADF	include	-1.03287	-2.86386	yes	no
	PP	include	-0.94355	-2.86386	yes	no
	KPSS	include	2.16221	0.14600	yes	no
GBP	DF	no	-0.12461	-1.94160	yes	no
	ADF	no	-0.16186	-1.94160	yes	no
	PP	no	-0.12506	-1.94160	yes	no
	KPSS	no	5.26720	0.14600	yes	no
	DF	include	-1.53295	-2.86386	yes	no
	ADF	include	-1.51000	-2.86386	yes	no
	PP	include	-1.53853	-2.86386	yes	no
	KPSS	include	5.26720	0.14600	yes	no
LN Gold	DF	no	0.45931	-1.94160	yes	no
	ADF	no	1.03139	-1.94160	yes	no
	PP	no	0.69975	-1.94160	yes	no
	KPSS	no	3.50910	0.14600	yes	no
	DF	include	-1.98422	-2.86386	yes	no
	ADF	include	-1.36627	-2.86386	yes	no
	PP	include	-1.66336	-2.86386	yes	no
	KPSS	include	3.50910	0.14600	yes	no

Note: "LN Oil", "GBP", and "LN Gold" refer to the logarithmic Brent crude oil price, USD/GBP exchange rate, and logarithmic NYMEX spot gold price, respectively. The "Test types" DF, ADF, PP, and KPSS refer to unit root tests of, respectively, Dickey-Fuller (Fuller, 1976), augmented Dickey-Fuller (Fuller, 1976), Phillips-Perron (Phillips & Perron, 1988), and (Kwaikowski et al., 1992).

gold are carried out each day. Understanding the mechanism of gold price changes is important for many outstanding issues in international economics and finance. Market participants are increasingly concerned with their exposure to large gold price fluctuations with special interest in which factors drive the changes. In this section, we apply the quantile causality test to investigate relations between the Brent crude oil, USD/GBP exchange rate and NYMEX spot gold prices (in USD per barrel and per ounce, respectively). The data, as seen in Figure 1, obtained from Datastream, are daily observations from 20 February 1997 to 17 July 2009 ($T = 3,237$). We use the USD/GBP instead of USD/EUR because the euro was only introduced as a new currency from 1 January 1999. As indicated by Table 2, we assume differenced logarithmic data are stationary and β -mixing with corresponding densities bounded. Because a long memory effect is not expected, we choose $p = q = 1$ and $m = 2$.

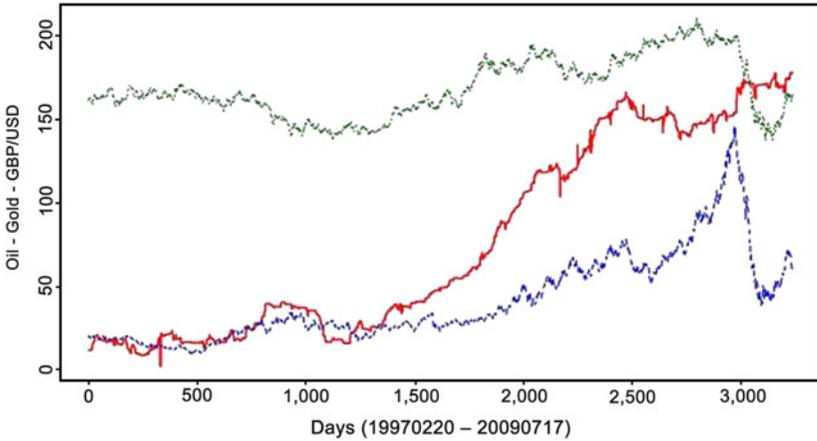


FIGURE 1. Plot of the gold prices, oil price, and exchange rate time series from 20 February 1997 to 17 July 2009.

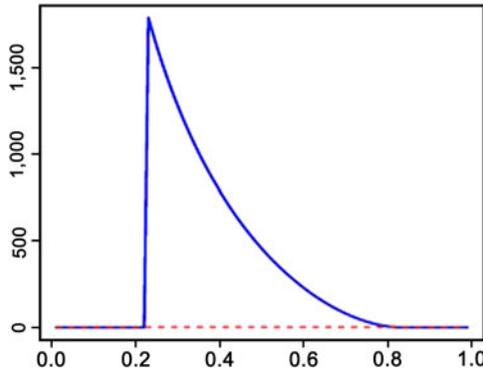


FIGURE 2. Test statistics with respect to different quantiles for the oil-gold prices causality test.

Figures 2 and 3 present results of testing whether oil prices Granger cause gold prices and whether the USD/GBP exchange rate Granger causes gold prices at the various quantiles, respectively, where logarithmic returns instead of the raw observations are used. The solid line and dotted line represent the standardized test statistics with respect to different quantiles (x -axis) and the critical value 1.96, respectively. In Figures 2 and 3, because the test statistic exceeds the critical value when $0.22 \leq \theta \leq 0.80$, we conclude that the oil price and exchange rate changes do not cause the gold price change in $\theta < 0.22$ or $\theta > 0.80$, whereas it is a prima facie cause in the $0.22 \leq \theta \leq 0.80$ quantile, respectively. For example, the oil price and USD/GBP exchange rate increases suggest that investors are wary of the U.S. dollar’s weakness and inflation. Because gold is typically bought as an

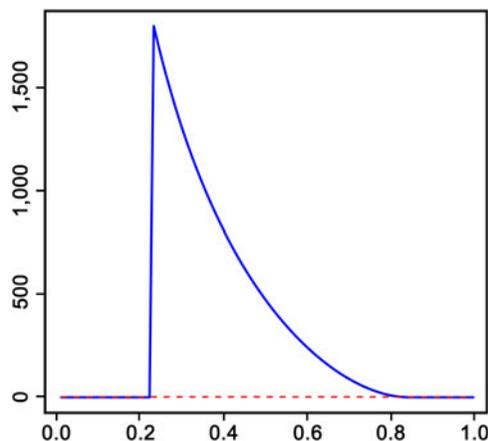


FIGURE 3. Test statistics with respect to different quantiles for the exchange rate-gold prices causality test.

alternative to the dollar among safe-haven assets, investors seeking safety from inflation risk and currency devaluation will cause the gold price to rise. However, the extreme low and high changes of the gold market may be caused by speculation. This is consistent with most of the empirical findings in the literature that the codependency may be stronger in the center than in the tails. By combining results from Figures 2 and 3, we find that the oil price and exchange rate changes have a significant predictive power for nonextreme gold price changes, which is, however, not significant for extreme changes. This finding could help to make it possible to use the gold price and GBP to hedge oil price changes in a more precise way with more careful investigation of their relations, which deserves further research.

6. CONCLUSION

By extending the Zheng (1998) idea to dependent data, we propose a consistent test for Granger causality in conditional quantile. The appealing feature of our proposed test is that it can investigate Granger causality in various conditional quantiles. The benefit of this is illustrated in the commodity market application where the causal relationships among the oil price, USD/GBP exchange rate, and gold price were shown to be different between a tail area and in the center of the distribution. We also illustrate that oil price and USD/GBP changes have significant predictive power on nonextreme gold price changes.

The test can be extended in a number of ways to test conditional quantile restrictions with dependent data: First, it can be extended to test functional misspecification, or the insignificance of a subset of regressors in quantile regression function, and second, it can also be used to test some semiparametric versus nonparametric models in quantile regression models.

REFERENCES

- Auestad, B. & D. Tjøstheim (1990) Identification of nonlinear time series: First order characterisation and order determination. *Biometrika* 77, 669–687.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. (2001) Financial econometrics: Past developments and future challenges. *Journal of Econometrics* 100, 41–51.
- Buchinsky, M. (1995) Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics* 65, 65–154.
- Campbell, J. & J. Cochrane (1999) By force of habit: A consumption-based explanation of aggregate stock market behaviour. *Journal of Political Economy* 107, 205–251.
- Day, J.C. & E.C. Newburger (2002) The big payoff: Educational attainment and synthetic estimates of work-life earnings. Special studies. Current population reports. Statistical report p23–210, U.S. Department of Commerce, U.S. Census Bureau.
- Diebolt, J. & D. Guégan (1993) Tail behavior of the stationary density of general nonlinear autoregressive processes of order 1. *Journal of Applied Probability* 30, 315–329.
- Engle, R. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables, parametric and semi-parametric functional forms. *Econometrica* 64, 865–890.
- Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U -statistics of absolutely regular processes with applications to model specification tests. *Journal of Nonparametric Statistics* 10, 245–271.
- Franke, J. & P. Mwita (2003) Nonparametric Estimates for Conditional Quantiles of Time Series. *Wirtschaftsmathematik* 87, University of Kaiserslautern.
- Fuller, W. (1976) *Introduction to Statistical Time Series*. Wiley.
- Granger, C. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Granger, C. (1988) Some recent developments in a concept of causality. *Journal of Econometrics* 39, 199–211.
- Györfi, L., W. Härdle, P. Sarda, & P. Vieu (1989) *Nonparametric Curve Estimation from Time Series*. Springer-Verlag.
- Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.
- Härdle, W., Y. Ritov, & S. Song (2009) Bootstrap Partial Linear Quantile Regression and Confidence Bands. SFB649 Discussion paper 2010-002, Humboldt Universität zu Berlin.
- Härdle, W. & T. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84, 986–995.
- Hong, Y., Y. Liu, & S. Wang (2009) Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics* 150, 271–287.
- Hsiao, C. & Q. Li (2001) A consistent test for conditional heteroskedasticity in time-series regression models. *Econometric Theory* 17, 188–221.
- Irle, A. (1997) On the consistency in nonparametric estimation under mixing assumptions. *Multivariate Analysis* 60, 123–147.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159–178.
- Lee, T. & W. Yang (2007) Money-Income Granger-Causality in Quantiles. Manuscript, University of California, Riverside.
- Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
- Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics* 87, 145–165.

Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory* 11, 258–289.

Masry, E. & D. Tjøstheim (1997) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–252.

Meir, R. (2000) Nonparametric time series prediction through adaptive model selection. *Machine Learning* 39, 5–34.

Modha, D. & E. Masry (1998) Memory-universal prediction of stationary random processes. *IEEE Transactions on Information Theory* 44, 117–133.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Powell, J., J. Stock, & T. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

Robinson, P.M. (1988) Root- n -consistent semiparametric regression. *Econometrica* 56, 931–954.

Roussas, G.G. (1988) Nonparametric estimation in mixing sequences of random variables. *Journal of Statistical Planning and Inference* 18, 135–149.

Sherman, R. (1994) Maximal inequalities for degenerate U -processes with applications to optimization estimators. *Annals of Statistics* 22, 439–459.

Yoshihara, K. (1976) Limiting behavior of u -statistics for stationary absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 35, 237–252.

Yu, B. (1993) Density estimation in the l_1 norm for dependent data with applications. *Annals of Statistics* 21, 711–735.

Zheng, J. (1998) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* 14, 123–138.

APPENDIX

Proof of Theorem 3.1(i). In the proof, we use several approximations to \hat{J}_T . We define them now and recall a few already defined statistics for convenience of reference.

$$\hat{J}_T \equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \hat{\varepsilon}_t \hat{\varepsilon}_s, \tag{A.1}$$

$$J_T \equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \varepsilon_t \varepsilon_s, \tag{A.2}$$

$$J_{TU} \equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \varepsilon_t U \varepsilon_s U, \tag{A.3}$$

$$J_{TL} \equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \varepsilon_t L \varepsilon_s L, \tag{A.4}$$

where $\hat{\varepsilon}_t = I \{y_t \leq \hat{Q}_\theta(x_t)\} - \theta$,
 $\varepsilon_t = I \{y_t \leq Q_\theta(x_t)\} - \theta$,
 $\varepsilon_{tU} = I \{y_t + C_T \leq Q_\theta(x_t)\} - \theta$,
 $\varepsilon_{tL} = I \{y_t - C_T \leq Q_\theta(x_t)\} - \theta$,

and C_T is an upper bound consistent with the uniform convergence rate of the nonparametric estimator of conditional quantile given in equation (13). The proof of Theorem 3.1(i) consists of three steps.

Step 1. Asymptotic normality.

$$Th^{m/2}J_T \rightarrow N(0, \sigma_0^2), \tag{A.5}$$

where $\sigma_0^2 = 2E\{\theta^2(1-\theta)^2f(z_t)\} \left\{ \int K^2(u)du \right\}$ under the null.

Step 2. Conditional asymptotic equivalence. Suppose that both $Th^{m/2}(J_T - J_{TU}) = \mathcal{O}_p(1)$ and $Th^{m/2}(J_T - J_{TL}) = \mathcal{O}_p(1)$.

$$\text{Then } Th^{m/2}(\hat{J}_T - J_T) = \mathcal{O}_p(1). \tag{A.6}$$

Step 3. Asymptotic equivalence.

$$Th^{m/2}(J_T - J_{TU}) = \mathcal{O}_p(1) \quad \text{and} \quad Th^{m/2}(J_T - J_{TL}) = \mathcal{O}_p(1). \tag{A.7}$$

The combination of steps 1–3 yields Theorem 3.1(i).

Proof of Step 1. Because J_T is a degenerate U -statistic of order 2, the result follows from Lemma 3.2. ■

Proof of Step 2. The proof of step 2 is motivated by the technique of Härdle and Stoker (1989) that was used in treating trimming an indicator function asymptotically. Suppose that the following two statements hold:

$$Th^{m/2}(J_T - J_{TU}) = \mathcal{O}_p(1) \quad \text{and} \tag{A.8}$$

$$Th^{m/2}(J_T - J_{TL}) = \mathcal{O}_p(1). \tag{A.9}$$

Use C_T to denote an upper bound consistent with the uniform convergence rate of the nonparametric estimator of conditional quantile given in equation (13). Suppose that

$$\sup|\hat{Q}_\theta(x) - Q_\theta(x)| \leq C_T. \tag{A.10}$$

If inequality (A.10) holds, then the following statements also hold:

$$\{Q_\theta(x) > y_t + C_T\} \subset \{\hat{Q}_\theta(x) > y_t\} \subset \{Q_\theta(x) > y_t - C_T\}, \tag{A.11}$$

$$1(Q_\theta(x) > y_t + C_T) \leq 1(\hat{Q}_\theta(x) > y_t) \leq 1(Q_\theta(x) > y_t - C_T), \tag{A.12}$$

$$J_{TU} \leq \hat{J}_T \leq J_{TL}, \tag{A.13}$$

$$|Th^{m/2}(J_T - \hat{J}_T)| \leq \max\{|Th^{m/2}(J_T - J_{TU})|, |Th^{m/2}(J_T - J_{TL})|\}. \tag{A.14}$$

Using (A.10) and (A.14), we have the following inequality:

$$\begin{aligned} & P\left\{|Th^{m/2}(J_T - \hat{J}_T)| > \delta \mid \sup|\hat{Q}_\theta(x) - Q_\theta(x)| \leq C_T\right\} \\ & \leq P\left\{\max\{|Th^{m/2}(J_T - J_{TU})|, |Th^{m/2}(J_T - J_{TL})|\} > \delta \mid \sup|\hat{Q}_\theta(x) - Q_\theta(x)| \leq C_T\right\} \\ & \quad \text{for all } \delta > 0. \end{aligned} \tag{A.15}$$

Invoking Lemma 3.1 and condition (A2)(c), we have

$$P\left\{\sup|\hat{Q}_\theta(x) - Q_\theta(x)| \leq C_T\right\} \rightarrow 1 \quad \text{as } T \rightarrow \infty. \tag{A.16}$$

By (A.8) and (A.9), as $T \rightarrow \infty$, we have

$$P\left\{\max\{|Th^{m/2}(J_T - J_{TU})|, |Th^{m/2}(J_T - J_{TL})|\} > \delta\right\} \rightarrow 0 \quad \text{for all } \delta > 0. \tag{A.17}$$

Therefore, as $T \rightarrow \infty$,

$$\text{the right-hand side of the inequality (A.15)} \times P\left\{\sup|\hat{Q}_\theta(x) - Q_\theta(x)| \leq C_T\right\} \rightarrow 0;$$

$$\text{the left-hand side of the inequality (A.15)} \times P\left\{\sup|\hat{Q}_\theta(x) - Q_\theta(x)| \leq C_T\right\}$$

$$= P\left\{|Th^{m/2}(J_T - \hat{J}_T)| > \delta\right\} \rightarrow 0.$$

In summary, we have that if both $Th^{m/2}(J_T - J_{TU}) = \mathcal{O}_p(1)$ and $Th^{m/2}(J_T - J_{TL}) = \mathcal{O}_p(1)$, then $Th^{m/2}(\hat{J}_T - J_T) = \mathcal{O}_p(1)$. ■

Proof of Step 3. In the remaining proof, we focus on showing that $Th^{m/2}(J_T - J_{TU}) = \mathcal{O}_p(1)$, with the proof of $Th^{m/2}(J_T - J_{TL}) = \mathcal{O}_p(1)$ being treated similarly. The proof of step 3 is close in line with the proof in Zheng (1998). Denote

$$H_T(s, t, g) \equiv K_{ts}\{1(y_t \leq g(x_t)) - \theta\}\{1(y_s \leq g(x_s)) - \theta\} \quad \text{and} \tag{A.18}$$

$$J[g] \equiv \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T H_T(s, t, g). \tag{A.19}$$

Then we have $J_T \equiv J[Q_\theta]$ and $J_{TU} \equiv J[Q_\theta - C_T]$. We decompose $H_T(s, t, g)$ into three parts:

$$\begin{aligned} H_T(s, t, g) &= K_{ts}\{1(y_t \leq g(x_t)) - F(g(x_t)|z_t)\}\{1(y_s \leq g(x_s)) - F(g(x_s)|z_s)\} \\ &\quad + 2 \times K_{ts}\{1(y_t \leq g(x_t)) - F(g(x_t)|z_t)\}\{F(g(x_s)|z_s) - \theta\} \\ &\quad + K_{ts}\{F(g(x_t)|z_t) - \theta\}\{F(g(x_s)|z_s) - \theta\} \\ &= H_{1T}(s, t, g) + 2H_{2T}(s, t, g) + H_{3T}(s, t, g). \end{aligned} \tag{A.20}$$

Then let $J_j[g] = 1/(T(T-1)h^m) \sum_{t=1}^T \sum_{s \neq t}^T H_{jT}(s, t, g)$, $j = 1, 2, 3$. We will treat $J_j[Q_\theta] - J_j[Q_\theta - C_T]$ for $j = 1, 2, 3$ separately.

(1) $Th^{m/2}[J_1(Q_\theta) - J_1(Q_\theta - C_T)] = \mathcal{O}_p(1)$. By simple manipulation, we have

$$\begin{aligned} &J_1(Q_\theta) - J_1(Q_\theta - C_T) \\ &= \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T [H_{1T}(s, t, Q_\theta) - H_{1T}(s, t, Q_\theta - C_T)] \\ &= \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \left\{ [1(y_t \leq Q_\theta(x_t)) - F(Q_\theta(x_t)|z_t)] \right. \\ &\quad \left. \times [1(y_s \leq Q_\theta(x_s)) - F(Q_\theta(x_s)|z_s)] \right\} \end{aligned}$$

$$\begin{aligned}
 & -[1(y_t \leq (Q_\theta(x_t) - C_T)) - F((Q_\theta(x_t) - C_T)|z_t)] \\
 & \times [1(y_s \leq (Q_\theta(x_s) - C_T)) - F((Q_\theta(x_s) - C_T)|z_s)] \}.
 \end{aligned}
 \tag{A.21}$$

To avoid tedious work to get bounds on the second moment of $J_1(Q_\theta) - J_1(Q_\theta - C_T)$ with dependent data, we note that the right-hand side of (A.21) is a degenerate U -statistic of order 2. Thus we can apply Lemma 3.2 and have

$$Th^{m/2} [J_1(Q_\theta) - J_1(Q_\theta - C_T)] \rightarrow N(0, \sigma_2^2) \quad \text{in distribution,}
 \tag{A.22}$$

where the definition of the asymptotic variance σ_2^2 is based on the i.i.d. sequence having the same marginal distributions as weakly dependent variables in (A.21). That is,

$$\sigma_2^2 = 2h^{-m} \tilde{E} [H_{1T}(s, t, Q_\theta) - H_{1T}(s, t, Q_\theta - C_T)]^2,$$

where the notation \tilde{E} is an expectation evaluated at an i.i.d. sequence having the same marginal distribution as the mixing sequences in (A.21) (Fan and Li, 1999, p. 248). Now, to show that $Th^{m/2} [J_1(Q_\theta) - J_1(Q_\theta - C_T)] = \mathcal{O}_p(1)$, we only need to show that the asymptotic variance $\sigma_2^2(z)$ is $\mathcal{O}(1)$ with i.i.d. data. Use Λ_T to denote an upper bound consistent with the integral over K_{ts} being of the order $\mathcal{O}(h^m)$. We have

$$\begin{aligned}
 & \tilde{E} [H_{1T}(s, t, Q_\theta) - H_{1T}(s, t, Q_\theta - C_T)]^2 \\
 & \leq \Lambda_T \tilde{E} \{ [1_t(Q_\theta) - F_t(Q_\theta)] [1_s(Q_\theta) - F_s(Q_\theta)] \\
 & \quad - [1_t(Q_\theta - C_T) - F_t(Q_\theta - C_T)] [1_s(Q_\theta - C_T) - F_s(Q_\theta - C_T)] \}^2 \\
 & \leq \Lambda_T \tilde{E} \{ F_t(Q_\theta) [1 - F_t(Q_\theta)] F_s(Q_\theta) [1 - F_s(Q_\theta)] \} \\
 & \quad + \tilde{E} \{ F_t(Q_\theta - C_T) [1 - F_t(Q_\theta - C_T)] F_s(Q_\theta - C_T) [1 - F_s(Q_\theta - C_T)] \} \\
 & \quad - 2E \{ [F_t(\min(Q_\theta, Q_\theta - C_T)) - F_t(Q_\theta)] F_t(Q_\theta - C_T) \} \\
 & \quad \quad \times [F_s(\min(Q_\theta, Q_\theta - C_T)) - F_s(Q_\theta)] F_s(Q_\theta - C_T) \} \\
 & = \Lambda_T \tilde{E} \{ [F_t(Q_\theta) - F_t(Q_\theta) F_t(Q_\theta)] [F_s(Q_\theta) - F_s(Q_\theta) F_s(Q_\theta)] \} \\
 & \quad - \Lambda_T \tilde{E} \{ [F_t(\min(Q_\theta, Q_\theta - C_T)) - F_t(Q_\theta) F_t(Q_\theta - C_T)] \\
 & \quad \quad \times [F_s(\min(Q_\theta, Q_\theta - C_T)) - F_s(Q_\theta) F_s(Q_\theta - C_T)] \} \\
 & \quad + \Lambda_T \tilde{E} \{ [F_t(Q_\theta - C_T) - F_t(Q_\theta - C_T) F_t(Q_\theta - C_T)] \\
 & \quad \quad \times [F_s(Q_\theta - C_T) - F_s(Q_\theta - C_T) F_s(Q_\theta - C_T)] \} \\
 & \quad - \Lambda_T \tilde{E} \{ [F_t(\min(Q_\theta, Q_\theta - C_T)) - F_t(Q_\theta) F_t(Q_\theta - C_T)] \\
 & \quad \quad \times [F_s(\min(Q_\theta, Q_\theta - C_T)) - F_s(Q_\theta) F_s(Q_\theta - C_T)] \} \\
 & \leq \Lambda_T C_T.
 \end{aligned}
 \tag{A.23}$$

Thus we have that $\sigma_2^2 = \mathcal{O}(C_T) = \mathcal{O}(1)$, and so

$$Th^{m/2} [J_1(Q_\theta) - J_1(Q_\theta - C_T)] = \mathcal{O}_p(1).
 \tag{A.24}$$

(2) $Th^{m/2} [J_2(Q_\theta) - J_2(Q_\theta - C_T)] = O_p(1)$. Note that $H_{2T}(s, t, Q_\theta) = 0$ because of $F_{y|z}(Q_\theta(x_s)|z_s) - \theta = 0$. Then we have

$$\begin{aligned}
 J_2(Q_\theta) - J_2(Q_\theta - C_T) &= -J_2(Q_\theta - C_T) \\
 &= -\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \\
 &\quad \times \{1(y_t \leq Q_\theta(x_t) - C_T) - F_{y|z}(Q_\theta(x_t) - C_T|z_t)\} \\
 &\quad \times \{F_{y|z}(Q_\theta(x_s) - C_T|z_s) - \theta\}.
 \end{aligned} \tag{A.25}$$

By taking a Taylor expansion of $F_{y|z}(Q_\theta(x_s) - C_T|z_s)$ around $Q_\theta(x_s)$, it equals

$$\begin{aligned}
 &-\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \\
 &\quad \times \{1(y_t \leq Q_\theta(x_t) - C_T) - F_{y|z}(Q_\theta(x_t) - C_T|z_t)\} \\
 &\quad \times (-C_T) f_{y|z}(\bar{Q}_\theta(x_s)|z_s),
 \end{aligned} \tag{A.26}$$

where \bar{Q}_θ is between Q_θ and $Q_\theta - C_T$. Thus we have

$$\begin{aligned}
 &(J_2(Q_\theta) - J_2(Q_\theta - C_T))^2 \\
 &\leq \left[\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \right. \\
 &\quad \left. \times \{1(y_t \leq Q_\theta(x_t) - C_T) - F_{y|z}(Q_\theta(x_t) - C_T|z_t)\} \right]^2 \Lambda^2 C_T^2 \\
 &= \Lambda^2 C_T^2 \left[\frac{1}{T} \sum_{t=1}^T \{1(y_t \leq Q_\theta(x_t) - C_T) - F_{y|z}(Q_\theta(x_t) - C_T)\} \hat{f}_z(z_t) \right]^2 \\
 &\equiv \Lambda^2 C_T^2 \left\{ \frac{1}{T} \sum_{t=1}^T u_t \hat{f}_z(z_t) \right\}^2 \\
 &= \Lambda^2 C_T^2 T^{-2} \sum_{t=1}^T u_t^2 \hat{f}_z^2(z_t) + \Lambda^2 C_T^2 T^{-2} \sum_{t=1}^T \sum_{s \neq t}^T u_t u_s \hat{f}_z(z_t) \hat{f}_z(z_s) \\
 &\equiv J_{21} + J_{22},
 \end{aligned} \tag{A.27}$$

where the inequality holds because of Assumption (A.1)(e).

$$\begin{aligned}
 E|J_{21}| &= \Lambda^2 C_T^2 T^{-1} \left[T^{-1} \sum_{t=1}^T E \left\{ u_t^2 \hat{f}_z^2(z_t) \right\} \right] \\
 &= O\left(C_T^2 T^{-2} h^{-m}\right),
 \end{aligned} \tag{A.28}$$

where the second equality is derived by using Lemma C.3(iii) of Li (1999).

$$\begin{aligned}
 J_{22} &= \Lambda^2 C_T^2 \left[T^{-2} \sum_{t=1}^T \sum_{s \neq t}^T u_t u_s f_z(z_t) f_z(z_s) \right. \\
 &\quad + 2T^{-2} \sum_{t=1}^T \sum_{s \neq t}^T u_t u_s f_z(z_t) \left\{ \hat{f}_z(z_s) - f_z(z_s) \right\} \\
 &\quad \left. + T^{-2} \sum_{t=1}^T \sum_{s \neq t}^T u_t u_s \left\{ \hat{f}_z(z_t) - f_z(z_t) \right\} \left\{ \hat{f}_z(z_s) - f_z(z_s) \right\} \right] \\
 &\equiv \Lambda C_T^2 (J_{221} + J_{222} + J_{223}). \tag{A.29}
 \end{aligned}$$

Following the line of the proof of Lemma A.2(i) of Li (1999) we have that

$$\begin{aligned}
 J_{221} &= \mathcal{O}_p(T^{-2}), \quad J_{222} = \mathcal{O}_p(T^{-1}), \quad \text{and} \quad J_{223} = \mathcal{O}_p(T^{-1}); \quad \text{thus} \\
 J_{22} &= \mathcal{O}_p(C_T^2 T^{-1}). \tag{A.30}
 \end{aligned}$$

Thus, combining (A.28) and (A.30), we have

$$\begin{aligned}
 Th^{m/2} [J_2(Q_\theta) - J_2(Q_\theta - C_T)] &= \mathcal{O}_p(C_T) + \mathcal{O}_p(C_T T^{1/2} h^{m/2}) \\
 &= \mathcal{O}_p(1). \tag{A.31}
 \end{aligned}$$

(3) $Th^{m/2} [J_3(Q_\theta) - J_3(Q_\theta - C_T)] = \mathcal{O}_p(1)$. Noting that $H_{3T}(s, t, Q_\theta) = 0$ because of $F(Q_\theta(x_j)|z_j) - \theta = 0$ for $j = t, s$, we have

$$\begin{aligned}
 &J_3(Q_\theta) - J_3(Q_\theta - C_T) \\
 &= -\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \\
 &\quad \times \{F(Q_\theta(x_t) - C_T|z_t) - \theta\} \{F(Q_\theta(x_s) - C_T|z_s) - \theta\} \\
 &= \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) C_T^2 f_{y|z}(\bar{Q}_\theta(x_t)|z_t) f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \\
 &= C_T^2 \frac{1}{T} \sum_{t=1}^T f_{y|z}(\bar{Q}_\theta(x_t)|z_t) f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \hat{f}_z(z_t). \tag{A.32}
 \end{aligned}$$

Thus, we have

$$\begin{aligned}
 &E|J_3(Q_\theta) - J_3(Q_\theta - C_T)| \\
 &\leq \Lambda C_T^2 \frac{1}{T} \sum_{t=1}^T E|\hat{f}_z(z_t)| \\
 &\leq \Lambda C_T^2 \frac{1}{T} \sum_{t=1}^T E|f_z(z_t)| + \Lambda C_T^2 \frac{1}{T} \sum_{t=1}^T E|\hat{f}_z(z_t) - f_z(z_t)| \\
 &= \mathcal{O}(C_T^2). \tag{A.33}
 \end{aligned}$$

Finally, we have

$$\begin{aligned} Th^{m/2} [J_3(Q_\theta) - J_3(Q_\theta - C_T)] &= \mathcal{O}_p \left(Th^{m/2} C_T^2 \right) \\ &= \mathcal{O}_p(1). \end{aligned} \tag{A.34}$$

By combining (A.24), (A.31), and (A.34), we have the result of step 3. ■

Proof of Theorem 3.1(ii). Because

$$\begin{aligned} \sigma_0^2 &= 2\theta^2(1-\theta)^2 E\{f_z(z_t)\} \int K^2(u) du \quad \text{and} \\ \hat{\sigma}_0^2 &\equiv 2\theta^2(1-\theta)^2 \frac{1}{T(T-1)h^m} \sum_{s \neq t} K_{ts}^2, \end{aligned}$$

it is enough to show that

$$\begin{aligned} \sigma_T^2 &\equiv \frac{1}{T(T-1)h^m} \sum_{s \neq t} K_{ts}^2 \\ &= E\{f_z(z_t)\} \int K^2(u) du + \mathcal{O}_p(1). \end{aligned} \tag{A.35}$$

Note that σ_T^2 is a nondegenerate U -statistic of order 2 with kernel

$$H_T(z_t, z_s) = \frac{1}{h^m} K^2 \left(\frac{z_t - z_s}{h} \right). \tag{A.36}$$

Because Assumptions (A2)(d) and (e) satisfy the conditions of Lemma 3.2 of Yoshihara (1976) on the asymptotic equivalence of the U -statistic and its projection under β -mixing, we have for $\gamma = 2(\delta - \delta')/\delta'(2 + \delta) > 0$

$$\begin{aligned} \sigma_T^2 &\equiv \frac{1}{T(T-1)} \sum_{s \neq t} H_T(z_t, z_s) \\ &= \int \int H_T(z_1, z_2) dF_z(z_1) dF_z(z_2) \\ &\quad + 2T^{-1} \sum_{t=1}^T \left[\int H_T(z_t, z_2) dF_z(z_2) - \int \int H_T(z_1, z_2) dF_z(z_1) dF_z(z_2) \right] \\ &\quad + \mathcal{O}_p(T^{-1-\gamma}) \\ &= \int \int H_T(z_1, z_2) dF_z(z_1) dF_z(z_2) + \mathcal{O}_p(1) \\ &= \int \int \frac{1}{h^m} K^2 \left(\frac{z_1 - z_2}{h} \right) dF_z(z_1) dF_z(z_2) + \mathcal{O}_p(1) \\ &= \int K^2(u) du \int f_z^2(z) dz + \mathcal{O}_p(1). \end{aligned} \tag{A.37}$$

The result of Theorem 3.1(ii) follows from (A.37). ■

Proof of Theorem 3.1(iii). The proof of Theorem 3.1(iii) consists of two steps.

Step 1. Show that $\hat{J}_T = J_T + \mathcal{O}_p(1)$ under the alternative hypothesis (4).

Step 2. Show that $J_T = J + \mathcal{O}_p(1)$ under the alternative hypothesis (4),

where $J = E\{[F_{y|z}(Q_\theta(x_t)|z_t) - \theta]^2 f_z(z_t)\}$. The combination of steps 1 and 2 yields Theorem 3.1(iii).

Proof of Step 1. We note that the results of step 2 and $Th^{m/2} [J_1(Q_\theta) - J_1(Q_\theta - C_T)] = \mathcal{O}_p(1)$ of step 3 in the proof of Theorem 3.1(i) still hold under the alternative hypothesis (4). Thus we focus on showing that $J_2(Q_\theta) - J_2(Q_\theta - C_T) = \mathcal{O}_p(1)$ and $J_3(Q_\theta) - J_3(Q_\theta - C_T) = \mathcal{O}_p(1)$.

We begin with showing that $J_2(Q_\theta) - J_2(Q_\theta - C_T) = \mathcal{O}_p(1)$. By the same procedures as in (A.27), we can show that $J_2(Q_\theta - C_T) = \mathcal{O}_p(T^{-1}h^{-m/2})$. Thus it remains to show that $J_2(Q_\theta) = \mathcal{O}_p(1)$. By taking a Taylor expansion of $F_{y|z}(Q_\theta(x_s)|z_s)$ around $Q_\theta(x_s)$, we have

$$\begin{aligned} J_2(Q_\theta) &= -\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \\ &\quad \times \{1(y_t \leq Q_\theta(x_t)) - F_{y|z}(Q_\theta(x_t)|z_t)\} \times f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \\ &= \frac{1}{T} \sum_{t=1}^T \{1(y_t \leq Q_\theta(x_t)) - F_{y|z}(Q_\theta(x_t))\} f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \hat{f}_z(z_t) \\ &\equiv \frac{1}{T} \sum_{t=1}^T u_t f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \hat{f}_z(z_t). \end{aligned} \tag{A.38}$$

By similar arguments as in (A.26) and (A.31), we have

$$J_2(Q_\theta) = \mathcal{O}(T^{-1}h^{-m}). \tag{A.39}$$

Next, we show that $Th^{m/2} [J_3(Q_\theta) - J_3(Q_\theta - C_T)] = \mathcal{O}_p(1)$ under the alternative hypothesis (4). Because $F(Q_\theta(x_j)|z_j) - \theta \neq 0$ for $j = t, s$ under the alternative hypothesis, we have

$$\begin{aligned} &J_3(Q_\theta) - J_3(Q_\theta - C_T) \\ &= \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \{F(Q_\theta(x_t)|z_t) - \theta\} \{F(Q_\theta(x_s)|z_s) - \theta\} \\ &\quad - \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \\ &\quad \times \{F(Q_\theta(x_t) - C_T|z_t) - \theta\} \{F(Q_\theta(x_s) - C_T|z_s) - \theta\} \\ &= \frac{1}{T} \sum_{t=1}^T \{F(Q_\theta(x_t)|z_t) - \theta\} \{F(Q_\theta(x_s)|z_s) - \theta\} \hat{f}_z(z_t) \\ &\quad - \frac{1}{T} \sum_{t=1}^T \{F(Q_\theta(x_t) - C_T|z_t) - \theta\} \{F(Q_\theta(x_s) - C_T|z_s) - \theta\} \hat{f}_z(z_t). \end{aligned} \tag{A.40}$$

By taking a Taylor expansion of $F_{y|z}(Q_\theta(x_j) - C_T|z_j)$ around $Q_\theta(z_j)$ for $j = t, s$, we have

$$\begin{aligned}
 J_3(Q_\theta) - J_3(Q_\theta - C_T) &= \frac{1}{T} \sum_{t=1}^T \{F(Q_\theta(x_t)|z_t) - \theta\} C_T f_{y|z}(\bar{Q}_\theta(x_t)|z_t) \hat{f}_z(z_t) \\
 &\quad + \frac{1}{T} \sum_{t=1}^T C_T f_{y|z}(\bar{Q}_\theta(x_t)|z_t) \{F(Q_\theta(x_s)|z_s) - \theta\} \hat{f}_z(z_t) \\
 &\quad - \frac{1}{T} \sum_{t=1}^T C_T^2 f_{y|z}(\bar{Q}_\theta(x_t)|z_t) f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \hat{f}_z(z_t). \tag{A.41}
 \end{aligned}$$

We further take a Taylor expansion of $F_{y|z}(Q_\theta(x_j)|z_j)$ around $Q_\theta(z_j)$ for $j = t, s$ and have

$$\begin{aligned}
 J_3(Q_\theta) - J_3(Q_\theta - C_T) &= \frac{1}{T} \sum_{t=1}^T f_{y|z}(\bar{Q}_\theta(x_t, z_t)|z_t) C_T f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \hat{f}_z(z_t) \\
 &\quad + \frac{1}{T} \sum_{t=1}^T C_T f_{y|z}(\bar{Q}_\theta(x_t)|z_t) f_{y|z}(\bar{Q}_\theta(x_s, z_s)|z_s) \hat{f}_z(z_t) \\
 &\quad - \frac{1}{T} \sum_{t=1}^T C_T^2 f_{y|z}(\bar{Q}_\theta(x_t)|z_t) f_{y|z}(\bar{Q}_\theta(x_s)|z_s) \hat{f}_z(z_t), \tag{A.42}
 \end{aligned}$$

where $\bar{Q}_\theta(x_s, z_s)$ is between $Q_\theta(x_s)$ and $Q_\theta(z_s)$. Then by using the same procedures as in (A.30), we have

$$J_3(Q_\theta) - J_3(Q_\theta - C_T) = \mathcal{O}(C_T). \tag{A.43}$$

Now we have the result of step 1 for the proof of Theorem 3.1(iii). ■

Proof of Step 2. Using (7) and the uniform convergence rate of the kernel regression estimator under a β -mixing process, we have

$$\begin{aligned}
 J_T &= \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \varepsilon_t \varepsilon_s \\
 &= \frac{1}{T} \sum_{t=1}^T \hat{E}(\varepsilon_t|z_t) \hat{f}_z(z_t) \varepsilon_t \\
 &= \frac{1}{T} \sum_{t=1}^T E(\varepsilon_t|z_t) f_z(z_t) \varepsilon_t + \frac{1}{T} \sum_{t=1}^T \left\{ \hat{E}(\varepsilon_t|z_t) \hat{f}_z(z_t) - E(\varepsilon_t|z_t) f_z(z_t) \right\} \varepsilon_t \\
 &= \frac{1}{T} \sum_{t=1}^T E(\varepsilon_t|z_t) f_z(z_t) \varepsilon_t + \mathcal{O}_p(1) \\
 &= E \left[E(\varepsilon_t|z_t) f_z(z_t) \varepsilon_t \right] + \mathcal{O}_p(1) \\
 &= J + \mathcal{O}_p(1). \tag{A.44}
 \end{aligned}$$

■

Proof of Theorem 3.1(iv). The proof of Theorem 3.1(iv) is close in line with the proof in Zheng (1998). The proof of Theorem 3.1(iv) consists of two steps.

Step 1. Show that $\hat{J}_T = J_T + \mathcal{O}_p(T^{-1}h^{-m/2})$ under the alternative hypothesis (A.2).

Step 2. Show that $Th^{m/2}J_T \rightarrow N(\mu, \sigma_1^2)$ under the alternative hypothesis (A.2),

$$\text{where } \mu = E\left[f_{y|z}^2\{Q_\theta(z_t)|z_t\}l^2(z_t)f_z(z_t)\right], \quad \sigma_1^2 = 2E\left\{\sigma_v^4(z_t)f_z(z_t)\right\} \\ \int K^2(u)du, \text{ and } \sigma_v^2(z_t) = E(v_t^2|z_t) \text{ with } v_t \equiv I\{y_t \leq Q_\theta(x_t)\} - F(Q_\theta(x_t)|z_t).$$

Proof of Step 1. The results of step 1 in the proof of Theorem 3.1(iii) show that, under the general alternative hypothesis (4), the elements consisting of $\hat{J}_T - J_T$ are all $\mathcal{O}_p(T^{-1}h^{-m/2})$ except for $J_2(Q_\theta(x))$, the order of which is $\mathcal{O}(T^{-1}h^{-m})$ as in (A.39). Thus we need to show that $J_2(Q_\theta(x)) = \mathcal{O}_p(T^{-1}h^{-m/2})$ under the local alternative hypothesis (A.2). Taking a Taylor expansion of $F_{y|z}\{Q_\theta(z_t) + d_T l(z_t)|z_t\}$ around $d_T = 0$, we have

$$F_{y|z}\{Q_\theta(z_t) + d_T l(z_t)|z_t\} = \theta + d_T f_{y|z}\{Q_\theta(z_t)|z_t\}l(z_t) + \mathcal{O}_p(d_T^2). \tag{A.45}$$

By similar procedures as in (A.38) and (A.39), we have

$$J_2(Q_\theta(x)) = -\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s \neq t}^T \frac{1}{h^m} K\left(\frac{z_t - z_s}{h}\right) \{1(y_t \leq Q_\theta(x_t)) - F_{y|z}(Q_\theta(x_t)|z_t)\} \\ \times d_T f_{y|z}\{Q_\theta(z_t)|z_t\}l(z_t) + \mathcal{O}_p(d_T^2) \\ = -d_T \frac{1}{T} \sum_{t=1}^T \{1(y_t \leq Q_\theta(x_t)) - F_{y|z}(Q_\theta(x_t)|z_t)\} \\ \times f_{y|z}\{Q_\theta(z_t)|z_t\}l(z_t)\hat{f}_z(z_t) + \mathcal{O}_p(d_T^2) \\ \equiv -d_T \frac{1}{T} \sum_{t=1}^T u_t f_{y|z}\{Q_\theta(z_t)|z_t\}l(z_t)\hat{f}_z(z_t) + \mathcal{O}_p(d_T^2) \\ = \mathcal{O}_p(d_T^2). \tag{A.46}$$

■

Proof of Step 2. Taking a Taylor expansion of $F_{y|z}\{Q_\theta(z_t) + d_T l(z_t)|z_t\}$ around $d_T = 0$, we have

$$J_T(Q_\theta(x)) = \frac{1}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \{1(y_t \leq Q_\theta(x_t))F(Q_\theta(x_t)|z_t)\} \\ \times \{1(y_s \leq Q_\theta(x_s)) - F(Q_\theta(x_s)|z_s)\} \\ - \frac{2d_T}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} \{1(y_t \leq Q_\theta(x_t)) - F(Q_\theta(x_t)|z_t)\} \\ \times f_{y|z}\{Q_\theta(z_s)|z_s\}l(z_s)$$

$$\begin{aligned}
 & + \frac{d_T^2}{T(T-1)h^m} \sum_{t=1}^T \sum_{s \neq t}^T K_{ts} f_{y|z} \{Q_\theta(z_t)|z_t\} l(z_t) f_{y|z} \{Q_\theta(z_s)|z_s\} l(z_s) \\
 & + \mathcal{O}_p \left(d_T^2 \right) \\
 & = T_{1T} - 2d_T T_{2T} + d_T^2 T_{3T} + \mathcal{O}_p \left(d_T^2 \right). \tag{A.47}
 \end{aligned}$$

Noting that T_{1T} is a degenerate U -statistic of order 2, by Lemma 3.2, we have

$$Th^{m/2} T_{1T} \rightarrow N \left(0, \sigma_1^2 \right) \quad \text{in distribution,} \tag{A.48}$$

Similarly to the proof for (A.31), we can show that $T_{2T} = \mathcal{O} \left\{ (Th^m)^{-1} \right\}$, and so $d_T T_{2T} = \mathcal{O} \left\{ (Th^{m/2})^{-1} \right\}$. And by the same procedures as in (A.44), we have

$$T_{3T} \rightarrow E \left[f_{y|z}^2 \{Q_\theta(z_t)|z_t\} l^2(z_t) f_z(z_t) \right] \quad \text{in probability.} \tag{A.49}$$

Thus,

$$Th^{m/2} J_T \rightarrow N \left(\mu, \sigma_1^2 \right), \tag{A.50}$$

where $\mu = E \left[f_{y|z}^2 \{Q_\theta(z_t)|z_t\} l^2(z_t) f_z(z_t) \right]$. ■



Bootstrap confidence bands and partial linear quantile regression[☆]

Song Song^{a,b,*}, Ya'acov Ritov^c, Wolfgang K. Härdle^a

^a Humboldt-Universität zu Berlin, Germany

^b The University of Texas at Austin, United States

^c The Hebrew University of Jerusalem, Israel

ARTICLE INFO

Article history:

Received 13 June 2011

Available online 31 January 2012

JEL classification:

C14

C21

C31

J01

J31

J71

AMS subject classifications:

62F40

62G08

62G86

Keywords:

Bootstrap

Quantile regression

Confidence bands

Nonparametric fitting

Kernel smoothing

Partial linear model

ABSTRACT

In this paper bootstrap confidence bands are constructed for nonparametric quantile estimates of regression functions, where resampling is done from a suitably estimated empirical distribution function (edf) for residuals. It is known that the approximation error for the confidence band by the asymptotic Gumbel distribution is logarithmically slow. It is proved that the bootstrap approximation provides an improvement. The case of multidimensional and discrete regressor variables is dealt with using a partial linear model. An economic application considers the labor market differential effect with respect to different education levels.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Quantile regression, as first introduced by Koenker and Bassett [25], is “gradually developing into a comprehensive strategy for completing the regression prediction” as claimed by Koenker and Hallock [26]. Quantile smoothing is an effective method to estimate quantile curves in a flexible nonparametric way. Since this technique makes no structural assumptions on the underlying curve, it is very important to have a device for understanding when observed features are significant and deciding between functional forms. For example, a question often asked in this context is whether or not an observed peak or valley is actually a feature of the underlying regression function or is only an artifact of the observational noise. For such issues, confidence bands (i.e., uniform over location) give an idea about the global variability of the estimate.

The nonparametric quantile estimate could be obtained either using a check function such as a robustified local linear smoother [10,35,36], or through estimating the conditional distribution function using the double-kernel local linear

[☆] The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin is gratefully acknowledged. Ya'acov Ritov's research is supported by an ISF grant and a Humboldt Award. We thank Thorsten Vogel and Alexandra Spitz-Oener for sharing their data, comments and suggestions.

* Correspondence to: The University of Texas at Austin, 78751 Austin, United States.

E-mail address: ssoonngg123@gmail.com (S. Song).

technique [11,35,36]. Besides these, [17] proposed a weighted version of the Nadaraya–Watson estimator, which was further studied by Cai [5]. In the previous work the theoretical focus has mainly been on obtaining consistency and asymptotic normality of the quantile smoother, and thereby providing the necessary ingredients to construct its pointwise confidence intervals. This, however, is not sufficient to get an idea about the global variability of the estimate; neither can it be used to correctly answer questions about the curve's shape, which contains the lack of fit test as an immediate application. This motivates us to construct the confidence bands.

To this end, [22] used strong approximations of the empirical process and extreme value theory. However, the very poor convergence rate of extremes of a sequence of n independent normal random variables is well documented and was first noticed and investigated by Fisher and Tippett [12], and discussed in greater detail by Hall [16]. In the latter paper it was shown that the rate of the convergence to its limit (the suprema of a stationary Gaussian process) can be no faster than $(\log n)^{-1}$. For example, the supremum of a nonparametric quantile estimate can converge to its limit no faster than $(\log n)^{-1}$. These results may make extreme value approximation of the distributions of suprema somewhat doubtful, for example in the context of the uniform confidence band construction for a nonparametric quantile estimate.

This paper proposes and analyzes a bootstrap-based method of obtaining the confidence bands for nonparametric quantile estimates. The method is simple to implement, does not rely on the evaluation of quantities which appear in asymptotic distributions, and takes the bias properly into account (at least asymptotically). Additionally, we show that the bootstrap distribution can approximate the true one (w.r.t. the $\|\cdot\|_\infty$ norm, details in Theorem 2.1) up to $n^{-2/5}$, which represents a significant improvement relative to $(\log n)^{-1}$, which is based on the asymptotic Gumbel distribution, as studied by Härdle and Song [22]. Previous research by Hahn [15] showed consistency of a bootstrap approximation to the cumulative distribution function (cdf) without assuming independence of the error and regressor terms. Ref. [23] showed bootstrap methods for median regression models based on a smoothed least-absolute-deviations (SLAD) estimate.

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a sequence of independent identically distributed bivariate random variables with joint pdf $f(x, y)$, joint cdf $F(x, y)$, conditional pdf $f(y|x), f(x|y)$, conditional cdf $F(y|x), F(x|y)$ for Y given X and X given Y respectively, and marginal pdf $f_X(x)$ for X , $f_Y(y)$ for Y . With some abuse of notation we use the letters f and F to denote different pdfs and cdfs respectively. The exact distribution will be clear from the context. At the first stage we assume that $x \in J^* = (a, b)$ for some $0 < a < b < 1$. Let $l(x)$ denote the p -quantile curve, i.e. $l(x) = F_{Y|x}^{-1}(p)$.

In economics, discrete or categorical regressors are very common. An example is from labor market analysis where one tries to find out how revenues depend on the age of the employee w.r.t. different education levels, labor union statuses, genders and nationalities, i.e. in econometric analysis one targets the differential effects. For example, [4] examined the US wage structure by quantile regression techniques. This motivates the extension to multivariate covariables by partial linear modelling (PLM). This is convenient especially when we have categorical elements of the X vector. Partial linear models, which were first considered by Green and Yandell [14,8,34,32], are gradually developing into a class of commonly used and studied semiparametric regression models, which can retain the flexibility of nonparametric models and ease the interpretation of linear regression models while avoiding the “curse of dimensionality”. Recently [29] used penalized quantile regression for variable selection of partially linear models with measurement errors.

In this paper, we propose an extension of the quantile regression model to $x = (u, v)^\top \in \mathbb{R}^d$ with $u \in \mathbb{R}^{d-1}$ and $v \in J^* \subset \mathbb{R}$. The quantile regression curve we consider is $\tilde{l}(x) = F_{Y|x}^{-1}(p) = u^\top \beta + l(v)$. The multivariate confidence band can then be constructed, based on the univariate uniform confidence band, plus the estimated linear part which we will prove is more accurately (\sqrt{n} consistency) estimated. This makes various tasks in economics, e.g. labor market differential effect investigation, multivariate model specification tests and the investigation of the distribution of income and wealth across regions or countries or the distribution across households possible. Additionally, since the natural link between quantile and expectile regression was developed by Newey and Powell [30], we can further extend our result into expectile regression for various tasks, e.g. demography risk research or expectile-based Value at Risk (EVAR) as in [28]. For high-dimensional modelling, [2] recently investigated high-dimensional sparse models with L_1 penalty. Additionally, our result might also be further extended to intersection bounds (one side confidence bands), which is similar to the work of Chernozhukov et al. [6].

The rest of this article is organized as follows. To keep the main idea transparent, in Section 2, as an introduction to the more complicated situation, the bootstrap approximation rate for the (univariate) confidence band is presented through a coupling argument. An extension to multivariate covariance X with partial linear modelling is shown in Section 3 with the actual type of confidence bands and their properties. In Section 4, we compare via a Monte Carlo study the bootstrap uniform confidence band with the one based on the asymptotic theory and investigate the behavior of partial linear estimates with the corresponding confidence band. In Section 5, an application considers the labor market differential effect. The discussion is restricted to the semiparametric extension. We do not discuss the general nonparametric regression. We conjecture that this extension is possible under appropriate conditions. Section 6 contains concluding remarks. All proofs are sketched in the Appendix.

2. Bootstrap confidence bands in the univariate case

Suppose $Y_i = l(X_i) + \varepsilon_i$, $i = 1, \dots, n$, where ε_i has the (conditional) distribution function $F(\cdot|X_i)$. For simplicity, but without any loss of generality, we assume that $F(0|X_i) = p$. $F(\xi|x)$ is smooth as a function of x and ξ for any x , and for any ξ in the neighborhood of 0. We assume:

- (A1) X_1, \dots, X_n are an i.i.d. sample, and $\inf_x f_X(x) = \lambda_0 > 0$. The quantile function satisfies $\sup_x |l^{(j)}(x)| \leq \lambda_j < \infty, j = 1, 2$.
 (A2) The distribution of Y given X has a density and $\inf_{x,t} f(t|x) \geq \lambda_3 > 0$, continuous at all $x \in J^*$, and at t only in a neighborhood of 0. More exactly, we have the following Taylor expansion at $x' = x, t = 0$, for some $A(\cdot)$ and $f_0(\cdot)$:

$$F(t|x') = F(0|x) + \left. \frac{\partial F(t|x')}{\partial x'} \right|_{x'=x,t=0} t + \left. \frac{\partial F(t|x')}{\partial t} \right|_{x'=x,t=0} (x' - x) + R(t, x'; x)$$

$$\stackrel{\text{def}}{=} p + f_0(x)t + A(x)(x' - x) + R(t, x'; x), \tag{1}$$

where

$$\sup_{t,x,x'} \frac{|R(t, x'; x)|}{t^2 + |x' - x|^2} < \infty.$$

Let K be a symmetric density function with compact support and $d_K = \int u^2 K(u)du < \infty$. Let $l_h(\cdot) = l_{n,h}(\cdot)$ be the nonparametric p -quantile estimate of Y_1, \dots, Y_n with weight function $K\{(X_i - \cdot)/h\}$ for some global bandwidth $h = h_n (K_h(u) = h^{-1}K(u/h))$, that is, a solution of

$$\frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{Y_i < l_h(x)\}}{\sum_{i=1}^n K_h(x - X_i)} < p \leq \frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{Y_i \leq l_h(x)\}}{\sum_{i=1}^n K_h(x - X_i)}. \tag{2}$$

Generally, the bandwidth may also depend on x . A local (adaptive) bandwidth selection though deserves future research.

Note that by assumption (A1), $l_h(x)$ is the quantile of a discrete distribution, which is equivalent to a sample of size $\mathcal{O}_p(nh)$ from a distribution with p -quantile whose bias is $\mathcal{O}(h^2)$ relative to the true value. Let δ_n be the local rate of convergence of the function l_h , essentially $\delta_n = h^2 + (nh)^{-1/2} = \mathcal{O}(n^{-2/5})$ with optimal bandwidth choice $h = h_n = \mathcal{O}(n^{-1/5})$ as in [36]. We employ also an auxiliary estimate $l_g \stackrel{\text{def}}{=} l_{n,g}$, essentially one similar to $l_{n,h}$ but with a slightly larger bandwidth $g = g_n = h_n n^\zeta$ (a heuristic explanation of why it is essential to oversmooth g is given later), where ζ is some small number. The asymptotically optimal choice of ζ as shown later is 4/45.

- (A3) The estimate l_g satisfies

$$\sup_{x \in J^*} |l_g''(x) - l''(x)| = \mathcal{O}_p(1),$$

$$\sup_{x \in J^*} |l_g'(x) - l'(x)| = \mathcal{O}_p(\delta_n/h). \tag{3}$$

Assumption (A3) is only stated to overwrite the issue here. It actually follows from the assumptions on (g, h) . A sequence $\{a_n\}$ is slowly varying if $n^{-\alpha} a_n \rightarrow 0$ for any $\alpha > 0$. With some abuse of notation we will use S_n to denote any slowly varying function which may change from place to place, e.g. $S_n^2 = S_n$ is a valid expression (since if S_n is a slowly varying function, then S_n^2 is slowly varying as well). λ_i and C_i are generic constants throughout this paper and the subscripts have no specific meaning. Note that there is no S_n term in (3) exactly because the bandwidth g_n used to calculate l_g is slightly larger than that used for l_h . We want to smooth it such that l_g , as an estimate of the quantile function, has a slightly worse rate of convergence, but its derivatives converge faster.

We also consider a family of estimates $\hat{F}(\cdot|X_i), i = 1, \dots, n$, estimating respectively $F(\cdot|X_i)$ and satisfying $\hat{F}(0|X_i) = p$. For example we can take the distribution with a point mass $[\sum_{j=1}^n K\{\alpha_n(X_j - X_i)\}]^{-1} K\{(X_j - X_i)/h\}$ on $Y_j - l_h(X_i), j = 1, \dots, n$, i.e.

$$\hat{F}(\cdot|X_i) = \frac{\sum_{j=1}^n K_h(X_j - X_i) \mathbf{1}\{Y_j - l_h(X_i) \leq \cdot\}}{\sum_{j=1}^n K_h(X_j - X_i)}. \tag{4}$$

We additionally assume:

- (A4) $f_X(x)$ is twice continuously differentiable and $f(t|x)$ is continuous in x , Hölder-continuous in t and uniformly bounded in x and t by, say, λ_4 .

For the precision of $\hat{F}(\cdot|X_i)$'s approximation around 0, we employ the following lemma from Franke and Mwita [13]:

Lemma 2.1 ([13, Lemma A.3-5]). *If assumptions (A1, A2, A4) hold, then for $|t| < S_n \delta_n, \delta_n \rightarrow 0, i = 1, \dots, n, X_i \in J^*$,*

$$\sup_{|t| < S_n \delta_n, i=1, \dots, n, X_i \in J^*} |\hat{F}(t|X_i) - F(t|X_i)| = \mathcal{O}_p\{S_n \delta_n\}. \tag{5}$$

Let $F^{-1}(\cdot|\cdot)$ and $\hat{F}^{-1}(\cdot|\cdot)$ be the inverse function of the conditional cdf and its estimate. We consider the following bootstrap procedure. Let U_1, \dots, U_n be i.i.d. uniform $[0, 1]$ variables. Let

$$Y_i^* = l_g(X_i) + \hat{F}^{-1}(U_i|X_i), \quad i = 1, \dots, n \tag{6}$$

be the bootstrap sample. We couple this sample to an unobserved hypothetical sample from the true conditional distribution

$$Y_i^\# = l(X_i) + F^{-1}(U_i|X_i), \quad i = 1, \dots, n. \tag{7}$$

Note that the vectors (Y_1, \dots, Y_n) and $(Y_1^\#, \dots, Y_n^\#)$ are equally distributed given X_1, \dots, X_n . We are really interested in the exact values of Y_i^* and $Y_i^\#$ only when they are near the appropriate quantile, that is, only if $|U_i - p| < S_n \delta_n$. But then, by Eq. (1), Lemma 2.1 and the inverse function theorem, we have

$$\max_{i:|F^{-1}(U_i|X_i)-F^{-1}(p)|<S_n\delta_n} |F^{-1}(U_i|X_i) - \hat{F}^{-1}(U_i|X_i)| = \max_{i:|Y_i^\#-l(X_i)|<S_n\delta_n} |Y_i^\# - l(X_i) - Y_i^* + l_g(X_i)| = \mathcal{O}_p\{S_n\delta_n\}. \tag{8}$$

Let now $q_{hi}(Y_1, \dots, Y_n)$ be the solution of the local quantile as given by (2) at X_i , with bandwidth h , i.e. $q_{hi}(Y_1, \dots, Y_n) \stackrel{\text{def}}{=} l_h(X_i)$ for data set $\{(X_i, Y_i)\}_{i=1}^n$. Note that by (3), if $|X_i - X_j| = \mathcal{O}(h)$, then

$$\max_{|X_i-X_j|<ch} |l_g(X_i) - l_g(X_j) - l(X_i) + l(X_j)| = \mathcal{O}_p(\delta_n). \tag{9}$$

Let l_h^* and $l_h^\#$ be the local bootstrap quantile and its coupled sample analogue. Then

$$\begin{aligned} l_h^*(X_i) - l_g(X_i) &= q_{hi}[\{Y_j^* - l_g(X_i)\}_{j=1}^n] \\ &= q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\}_{j=1}^n], \end{aligned} \tag{10}$$

while

$$l_h^\#(X_i) - l(X_i) = q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]. \tag{11}$$

From (8)–(11) we conclude that

$$\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) + l(X_i)| = \mathcal{O}_p(\delta_n). \tag{12}$$

Based on (12), we obtain the following theorem (the proof is given in the Appendix):

Theorem 2.1. *If assumptions (A1–A4) hold, then*

$$\sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) + l(x)| = \mathcal{O}_p(\delta_n) = \mathcal{O}_p(n^{-2/5}).$$

Remark. Theorem 2.1 indicates that the r.v. $l_h^*(x) - l_g(x)$ approximates the one of $l_h^\#(x)$ up to $n^{-2/5}$ (w.r.t. the $\|\cdot\|_\infty$ norm). Thus a number of replications of $l_h^*(x)$ can be used as the basis for simultaneous error bars.

Although Theorem 2.1 is stated with a fixed bandwidth, in practice, to take care of the heteroscedasticity effect, we construct confidence bands with the width depending on the densities, which is motivated by the counterpart based on the asymptotic theory as in [22]. Thus we have the following corollary.

Corollary 2.1. *Let $d_{*\alpha}$ be defined by $P^*(|l_h^*(x) - l_g(x)| > d_{*\alpha}^*) = \alpha$, where P^* is the bootstrap distribution conditioned on the sample. If (A1)–(A4) hold, then the confidence interval $l_h(x) \pm d_{*\alpha}^*$ has an asymptotic uniform coverage of $1 - \alpha$, in the sense that $P(\sup_{x \in J^*} |l_h(x) - l(x)| > d_{*\alpha}^*) \rightarrow \alpha$.*

In practice we would use the approximate $(1 - \alpha) \times 100\%$ confidence band over \mathbb{R} given by

$$l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_{*\alpha}^*, \tag{13}$$

where $d_{*\alpha}^*$ is based on the bootstrap sample (defined later) and $\hat{f}\{l_h(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$ with use of $f(y|x) = f(x, y)/f_X(x)$.

Below is the summary of the basic steps for the bootstrap procedure.

- (1) Given (X_i, Y_i) , $i = 1, \dots, n$, compute the local quantile smoother $l_h(x)$ of Y_1, \dots, Y_n with bandwidth h and obtain residuals $\hat{\varepsilon}_i = Y_i - l_h(X_i)$, $i = 1, \dots, n$.

(2) Compute the conditional edf:

$$\hat{F}(t|x) = \frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{\hat{\varepsilon}_i \leq t\}}{\sum_{i=1}^n K_h(x - X_i)}.$$

(3) For each $i = 1, \dots, n$, generate random variables $\varepsilon_{i,b}^* \sim \hat{F}(t|X_i)$, $b = 1, \dots, B$ and construct the bootstrap sample $Y_{i,b}^*$, $i = 1, \dots, n$, $b = 1, \dots, B$ as follows:

$$Y_{i,b}^* = l_g(X_i) + \varepsilon_{i,b}^*.$$

(4) For each bootstrap sample $\{(X_i, Y_{i,b}^*)\}_{i=1}^n$, compute l_h^* and the random variable

$$d_b \stackrel{\text{def}}{=} \sup_{x \in J^*} \left[\hat{f}\{l_h^*(x)|x\} \sqrt{\hat{f}_X(x)} |l_h^*(x) - l_g(x)| \right] \tag{14}$$

where $\hat{f}\{l(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$.

(5) Calculate the $(1 - \alpha)$ quantile d_α^* of d_1, \dots, d_B .

(6) Construct the bootstrap uniform confidence band centered around $l_h(x)$, i.e. $l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_\alpha^*$.

While bootstrap methods are well-known tools for assessing variability, more care must be taken to properly account for the type of bias encountered in nonparametric curve estimation. The choice of bandwidth is crucial here. In our experience the bootstrap works well with a rather crude choice of g ; one may, however, specify g more precisely. Since the main role of the pilot bandwidth is to provide a correct adjustment for the bias, we use the goal of bias estimation as a criterion. Recall that the bias in the estimation of $l(x)$ by $l_h^\#(x)$ is given by

$$b_h(x) = E l_h^\#(x) - l(x).$$

The bootstrap bias of the estimate constructed from the resampled data is

$$\hat{b}_{h,g}(x) = E l_h^*(x) - l_g(x). \tag{15}$$

Note that in (15) the expected value is computed under the bootstrap estimation. The following theorem gives an asymptotic representation of the mean squared error for the problem of estimating $b_h(x)$ by $\hat{b}_{h,g}(x)$. It is then straightforward to find g to minimize this representation. Such a choice of g will make the quantiles of the original and coupled bootstrap distributions close to each other. In addition to the technical assumptions before, we also need:

(A5) l and f are four times continuously differentiable.

(A6) K is twice continuously differentiable.

Theorem 2.2. Under assumptions (A1–A6), for any $x \in J^*$

$$E \left[\left\{ \hat{b}_{h,g}(x) - b_h(x) \right\}^2 \mid X_1, \dots, X_n \right] \sim h^4 (C_1 g^4 + C_2 n^{-1} g^{-5}) \tag{16}$$

in the sense that the ratio between the RHS and the LHS tends in probability to 1 for some constants C_1, C_2 .

An immediate consequence of Theorem 2.2 is that the rate of convergence of g should be $n^{-1/9}$, see also [20]. This makes precise the previous intuition which indicated that g should slightly oversmooth. Under our assumptions, reasonable choices of h will be of the order $n^{-1/5}$ as in [36]. Hence, (16) shows once again that g should tend to zero more slowly than h . Note that Theorem 2.2 is not stated uniformly over h . The reason is that we are only trying to give some indication of how the pilot bandwidth g should be selected.

We summarize how to select the bandwidth h for the local quantile smoother and g for the oversmoothed estimate as below.

- 1 Select h as in [36] which is also quoted below.
 - Use ready-made and sophisticated methods to select h_{mean} , the optimal bandwidth choice for regression mean estimation; we use the technique of Ruppert et al. [33].
 - Use $h = h_{\text{mean}} \{p(l-p)/\phi(\Phi^{-1}(p))^2\}^{1/5}$ to obtain all other h 's (w.r.t. different p 's) from h_{mean} . ϕ and Ψ are the PDF and CDF of standard normal distributions respectively.
- 2 According to Theorem 2.2, select g as $g = n^{4/45} h$.

3. Bootstrap confidence bands in PLMs

The case of multivariate regressors may be handled via a semiparametric specification of the quantile regression curve. More specifically we assume that with $x = (u, v)^\top \in \mathbb{R}^d$, $v \in \mathbb{R}$:

$$\tilde{l}(x) = u^\top \beta + l(v).$$

In this section we show how to proceed in this multivariate setting and how – based on [Theorem 2.1](#) – a multivariate confidence band may be constructed. We first describe the numerical procedure for obtaining estimates of β and l , where l denotes – as in the earlier sections – the one-dimensional conditional quantile curve. We then move on to the theoretical properties. First note that the PLM quantile estimation problem can be seen as estimating (β, l) in

$$\begin{aligned} y &= u^\top \beta + l(v) + \varepsilon \\ &= \tilde{l}(x) + \varepsilon \end{aligned} \tag{17}$$

where the p -quantile of ε conditional on both u and v is 0.

In order to estimate β , let a_n denote an increasing sequence of positive integers and set $b_n = a_n^{-1}$. For each $n = 1, 2, \dots$, partition the unit interval $[0, 1]$ for v in a_n intervals I_{ni} , $i = 1, \dots, a_n$, of equal length b_n and let m_{ni} denote the midpoint of I_{ni} . In each of these small intervals I_{ni} , $i = 1, \dots, a_n$, $l(v)$ can be considered as being approximately constant, and hence (17) can be considered as a linear model. This observation motivates the following two stage estimation procedure.

- (1) A linear quantile regression inside each partition is used to estimate $\hat{\beta}_i$, $i = 1, \dots, a_n$. Their weighted mean yields $\hat{\beta}$. More exactly, consider the parametric quantile regression of y on u , $\mathbf{1}(v \in [0, b_n))$, $\mathbf{1}(v \in [b_n, 2b_n))$, \dots , $\mathbf{1}(v \in [1 - b_n, 1])$. That is, let

$$\psi(t) \stackrel{\text{def}}{=} (p - 1)t\mathbf{1}(t < 0) + pt\mathbf{1}(t \geq 0).$$

Then let

$$\hat{\beta} = \arg \min_{\beta} \min_{l_1, \dots, l_{a_n}} \sum_{i=1}^{a_n} \psi \left\{ Y_i - \beta^\top U_i - \sum_{j=1}^{a_n} l_j \mathbf{1}(V_i \in I_{ni}) \right\}.$$

- (2) Calculate the smooth quantile estimate as in (2) from $(V_i, Y_i - U_i^\top \hat{\beta})_{i=1}^n$, and name it as $\tilde{l}_h(v)$.

The following theorem states the asymptotic distribution of $\hat{\beta}$.

Theorem 3.1. *If assumption (A1) holds, for the above two stage estimation procedure, there exist positive definite matrices D , C , such that*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N\{0, p(1 - p)D^{-1}CD^{-1}\} \text{ as } n \rightarrow \infty,$$

where $C = \text{plim}_{n \rightarrow \infty} C_n$ and $D = \text{plim}_{n \rightarrow \infty} D_n$ with $C_n = \frac{1}{n} \sum_{i=1}^n U_i^\top U_i$ and $D_n = \frac{1}{n} \sum_{j=1}^n f\{l(V_j)|V_j\}U_j^\top U_j$ respectively.

Note that $l(v)$, $\tilde{l}_h(v)$ (quantile smoother based on $(v, y - u^\top \beta)$) and $\tilde{l}_h(v)$ can be treated as zeros (w.r.t. θ , $\theta \in I$ where I is a possibly infinite, or possibly degenerate, interval in \mathbb{R}) of the functions

$$\tilde{H}(\theta, v) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(v, \tilde{y})\psi(\tilde{y} - \theta)d\tilde{y}, \tag{18}$$

$$\tilde{H}_n(\theta, v) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n K_h(v - V_i)\psi(\tilde{Y}_i - \theta), \tag{19}$$

$$\tilde{\tilde{H}}_n(\theta, v) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n K_h(v - V_i)\psi(\tilde{\tilde{Y}}_i - \theta), \tag{20}$$

where

$$\tilde{Y}_i \stackrel{\text{def}}{=} Y_i - U_i^\top \beta,$$

$$\tilde{\tilde{Y}}_i \stackrel{\text{def}}{=} Y_i - U_i^\top \hat{\beta} = Y_i - U_i^\top \beta + U_i^\top (\beta - \hat{\beta}) \stackrel{\text{def}}{=} \tilde{Y}_i + Z_i.$$

From [Theorem 3.1](#) we know that $\hat{\beta} - \beta = \mathcal{O}_p(1/\sqrt{n})$ and $\|Z_i\|_\infty = \mathcal{O}_p(1/\sqrt{n})$. Under the following assumption, which is satisfied by exponential and generalized hyperbolic distributions, also used in [\[18\]](#):

- (A7) The conditional densities $f(\cdot|\tilde{y})$, $\tilde{y} \in \mathbb{R}$, are uniformly local Lipschitz continuous of order $\tilde{\alpha}$ (ulL- $\tilde{\alpha}$) on J , uniformly in $\tilde{y} \in \mathbb{R}$, with $0 < \tilde{\alpha} \leq 1$, and $(nh)/\log n \rightarrow \infty$,

for some constant C_3 not depending on n , Lemma 2.1 in [22] shows a.s. as $n \rightarrow \infty$:

$$\sup_{\theta \in I} \sup_{v \in J^*} |\tilde{H}_n(\theta, v) - \tilde{H}(\theta, v)| \leq C_3 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\}.$$

Observing that $\sqrt{h/\log n} = \mathcal{O}(1)$, we then have

$$\begin{aligned} \sup_{\theta \in I} \sup_{v \in J^*} |\tilde{\tilde{H}}_n(\theta, v) - \tilde{H}(\theta, v)| &\leq \sup_{\theta \in I} \sup_{v \in J^*} |\tilde{H}_n(\theta, v) - \tilde{H}(\theta, v)| + \underbrace{\sup_{\theta \in I} \sup_{v \in J^*} |\tilde{H}_n(\theta, v) - \tilde{\tilde{H}}_n(\theta, v)|}_{\leq \mathcal{O}_p(1/\sqrt{n}) \sup_{v \in J} |n^{-1} \sum K_h|} \\ &\leq C_4 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \end{aligned} \tag{21}$$

for a constant C_4 which can be different from C_3 . To show the uniform consistency of the quantile smoother, we shall reduce the problem of strong convergence of $\tilde{l}_h(v) - l(v)$, uniformly in v , to an application of the strong convergence of $\tilde{H}_n(\theta, v)$ to $\tilde{H}(\theta, v)$, uniformly in v and θ . For our result on $\tilde{l}_h(\cdot)$, we shall also require

$$(A8) \inf_{v \in J^*} \left| \int \psi\{y - l(v) + \varepsilon\} dF(y|v) \right| \geq \tilde{q}|\varepsilon|, \quad \text{for } |\varepsilon| \leq \delta_1,$$

where δ_1 and \tilde{q} are some positive constants, see also [19]. This assumption is satisfied if a constant \tilde{q} exists giving $f\{l(v)|v\} > \tilde{q}/p, x \in J$. Ref. [22] showed:

Lemma 3.1. Under assumptions (A7) and (A8), we have a.s. as $n \rightarrow \infty$

$$\sup_{v \in J^*} |\tilde{l}_h(v) - l(v)| \leq C_5 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{22}$$

with another constant C_5 not depending on n . If we consider the bandwidth $h = \mathcal{O}(n^{-1/5})$ and then skip the slow varying function $\log n$, then $(nh/\log n)^{-1/2} = \mathcal{O}(n^{-2/5}) < \mathcal{O}(n^{-1/5}) \leq h^{\tilde{\alpha}}$, (22) can be further simplified to

$$\sup_{v \in J^*} |\tilde{l}_h(v) - l(v)| \leq C_5 \{h^{\tilde{\alpha}}\}.$$

Since the proof is essentially the same as Theorem 2.1 of the above mentioned reference, it is omitted here.

The convergence rate for the parametric part $\mathcal{O}_p(n^{-1/2})$ (Theorem 3.1) is smaller than the bootstrap approximation error for the nonparametric part $\mathcal{O}_p(n^{-2/5})$ as shown in Theorem 2.1. This makes the construction of uniform confidence bands for multivariate $x \in \mathbb{R}^d$ with a partial linear model possible.

Proposition 3.1. Under the assumptions (A1)–(A8), an approximate $(1 - \alpha) \times 100\%$ confidence band over $\mathbb{R}^{d-1} \times [0, 1]$ is

$$u^\top \hat{\beta} + \tilde{l}_h(v) \pm \left[\hat{f}\{\tilde{l}_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_\alpha^*,$$

where $\hat{f}\{\tilde{l}_h(x)|x\}, \hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}, f_X(x)$.

Note that here we actually only require that the convergence rate of the parametric part, which is typically $\mathcal{O}_p(n^{-1/2})$, is smaller than the bootstrap approximation error for the nonparametric part $\mathcal{O}_p(n^{-2/5})$. This makes construction for the uniform confidence bands of more general semiparametric models possible instead of just the partial linear model shown here and similar results could be obtained easily.

4. A Monte Carlo study

This section is divided into two parts. First we concentrate on a univariate regressor variable x , check the validity of the bootstrap procedure together with settings in the specific example, and compare it with asymptotic uniform bands. Secondly we incorporate the partial linear model to handle the multivariate case of $x \in \mathbb{R}^d$.

Below is the summary of the simulation procedure.

- (1) Simulate $(X_i, Y_i), i = 1, \dots, n$ according to their joint pdf $f(x, y)$.

In order to compare with earlier results in the literature, we choose the joint pdf of bivariate data $\{(X_i, Y_i)\}_{i=1}^n, n = 1000$ as

$$f(x, y) = f_{y|x}(y - \sin x) \mathbf{1}(x \in [0, 1]), \tag{23}$$

where $f_{y|x}(x)$ is the pdf of $N(0, x)$ with an increasing heteroscedastic structure. Thus the theoretical quantile is $l(x) = \sin(x) + \sqrt{x}\Phi^{-1}(p)$. Based on this normality property, all the assumptions can be seen to be satisfied.

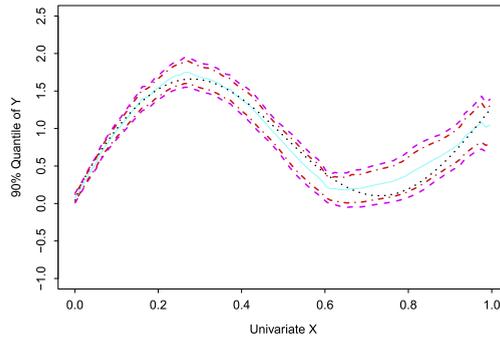


Fig. 1. The real 0.9 quantile curve (black dotted line), 0.9 quantile estimate (cyan solid line) with corresponding 95% uniform confidence band from asymptotic theory (magenta dashed lines) and confidence band from bootstrapping (red dashed–dot lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2) Compute the local quantile smoother $l_h(x)$ of Y_1, \dots, Y_n with bandwidth h and obtain residuals $\hat{\varepsilon}_i = Y_i - l_h(X_i)$, $i = 1, \dots, n$.

If we choose $p = 0.9$, then $\Phi^{-1}(p) = 1.2816$, $l(x) = \sin(x) + 1.2816\sqrt{x}$. Set $h = 0.05$.

(3) Compute the conditional edf:

$$\hat{F}(t|x) = \frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{\hat{\varepsilon}_i \leq t\}}{\sum_{i=1}^n K_h(x - X_i)}.$$

The choice of kernel functions plays a minor role here. Section 3.4.3 and Table 3.3 of Härdle et al. [21] discuss the efficiencies of different kernels. The Epanechnikov kernel would be the optimal one; however, the differences among various kernels are small. Thus, we just use the Gaussian kernel to assure numerical stability. This is also convenient because the optimal bandwidth suggested by Yu and Jones [36] is also calculated based on the Gaussian kernel.

(4) For each $i = 1, \dots, n$, generate random variables $\varepsilon_{i,b}^* \sim \hat{F}(t|x)$, $b = 1, \dots, B$ and construct the bootstrap sample $Y_{i,b}^*$, $i = 1, \dots, n$, $b = 1, \dots, B$ as follows:

$$Y_{i,b}^* = l_g(X_i) + \varepsilon_{i,b}^*,$$

with $g = 0.2$.

(5) For each bootstrap sample $\{(X_i, Y_{i,b}^*)\}_{i=1}^n$, compute l_h^* and the random variable

$$d_b \stackrel{\text{def}}{=} \sup_{x \in J^*} \left[\hat{f}\{l_h^*(x)|x\} \sqrt{\hat{f}_X(x)} |l_h^*(x) - l_g(x)| \right], \tag{24}$$

where $\hat{f}\{l(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$ with use of $f(y|x) = f(x, y)/f_X(x)$.

(6) Calculate the $(1 - \alpha)$ quantile d_α^* of d_1, \dots, d_B .

(7) Construct the bootstrap uniform confidence band centered around $l_h(x)$, i.e. $l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_\alpha^*$.

Fig. 1 shows the theoretical 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from the asymptotic theory and the confidence band from the bootstrap. The real 0.9 quantile curve is marked as the black dotted line. We then compute the classic local quantile estimate $l_h(x)$ (cyan solid) with its corresponding 95% uniform confidence band (magenta dashed) based on asymptotic theory according to Härdle and Song [22]. The 95% confidence band from the bootstrap is displayed as red dashed–dot lines. At first sight, the quantile smoother, together with two corresponding bands, all capture the heteroscedastic structure quite well, and the width of the bootstrap confidence band is similar to the one based on asymptotic theory in [22]. Fig. 2 presents the bootstrap confidence bands constructed using different oversmoothing bandwidths w.r.t. the same (but different from the one used for Fig. 1) randomly generated data set, namely, 1/2, 1 and 2 times (from left to right) of the oversmoothing bandwidth $g = n^{4/45}h$ used before. As we can see, when we deviate from $g = n^{4/45}h$, the bootstrap confidence bands get wider.

We now extend x to the multivariate case and use a different quantile function to verify our method. Choose $x = (u, v)^T \in \mathbb{R}^d$, $v \in \mathbb{R}$, and generate the data $\{(U_i, V_i, Y_i)\}_{i=1}^n$, $n = 1000$ with

$$y = 2u + v^2 + \varepsilon - 1.2816, \tag{25}$$

where u and v are uniformly distributed random variables in $[0, 2]$ and $[0, 1]$ respectively. ε has a standard normal distribution. The theoretical 0.9-quantile curve is $\tilde{l}(x) = 2u + v^2$. Since the choice of a_n is uncertain here, we test different

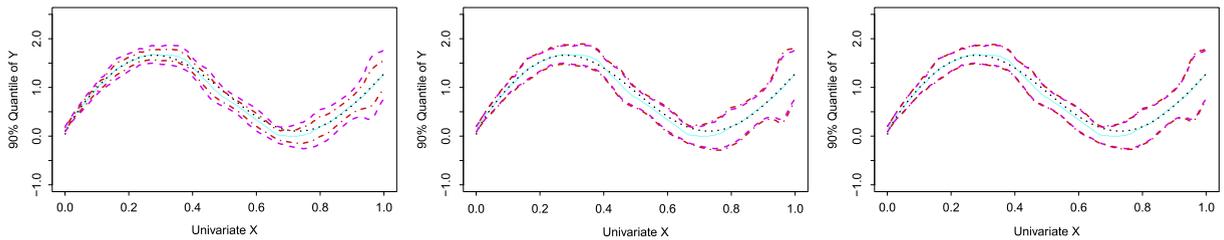


Fig. 2. The real 0.9 quantile curve (black dotted line), 0.9 quantile estimate (cyan solid line) with corresponding 95% uniform confidence band from asymptotic theory (magenta dashed lines) and confidence band from bootstrapping (red dashed-dot lines). The left, middle and right plots correspond to the oversmoothing bandwidth set as $n^{4/45}h/2$, $n^{4/45}h$ and $2n^{4/45}h$ respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

SSE of $\hat{\beta}$ with respect to a_n for different numbers of observations.

a_n	$n = 1000$	$n = 8000$	$n = 261148$
$n^{1/3}/8$			3.6×10^{-3}
$n^{1/3}/4$	5.4×10^{-1}	4.0×10^{-2}	3.3×10^{-3}
$n^{1/3}/2$	6.1×10^{-1}	3.5×10^{-2}	3.2×10^{-3}
$n^{1/3}$	6.2×10^{-1}	3.6×10^{-2}	3.1×10^{-3}
$n^{1/3} \cdot 2$	8.0×10^{-1}	3.9×10^{-2}	2.9×10^{-3}
$n^{1/3} \cdot 4$	4.9×10^{-1}	3.6×10^{-2}	2.8×10^{-3}
$n^{1/3} \cdot 8$			3.4×10^{-3}

choices of a_n for different n by simulation. To this end, we modify the theoretical model as follows:

$$y = 2u + v^2 + \varepsilon - \Phi^{-1}(p)$$

such that the real β is always equal to 2 no matter if p is 0.01 or 0.99. The result is displayed in Fig. 3 for $n = 1000$, $n = 8000$, $n = 261148$ (number of observations for the data set used in the following application part including both uncensored and censored observations). Different lines correspond to different a_n , i.e. $n^{1/3}/8$, $n^{1/3}/4$, $n^{1/3}/2$, $n^{1/3}$, $n^{1/3} \cdot 2$, $n^{1/3} \cdot 4$ and $n^{1/3} \cdot 8$. At first, it seems that the choice of a_n does not matter too much. To further investigate this, we calculate the SSE ($\sum_1^{99} \{\hat{\beta}(i/100) - \beta\}$) where $\hat{\beta}(i/100)$ denotes the estimate corresponding to the $i/100$ quantile. The results are displayed in Table 1. Obviously a_n has much less effect than n on SSE. Considering the computational cost, which increases with a_n , and the estimation performance, empirically we suggest $a_n = n^{1/3}$. Certainly this issue is far from settled and needs further investigation.

Thus for the specific model (25), we have $a_n = 10$, $\hat{\beta} = 1.997$, $h = 0.2$ and $g = 0.7$. In Fig. 4 the theoretical 0.9 quantile curve with respect to v , and the 0.9 quantile estimate with corresponding uniform confidence band are displayed. The real 0.9 quantile curve is marked as the black dotted line. We then compute the quantile smoother $l_h(x)$ (magenta solid). The 95% bootstrap uniform confidence band is displayed as red dashed lines and covers the true quantile curve quite well.

5. A labor market application

Our intuition of the effect of education on income is summarized by Day and Newburger’s basic claim [7]: “At most ages, more education equates with higher earnings, and the payoff is most notable at the highest educational levels”, which is actually from the point of view of mean regression. However, whether this difference is significant or not is still questionable, especially for different ends of the (conditional) income distribution. To this end, a careful investigation of quantile regression is necessary. Since different education levels may reflect different productivity, which is unobservable and may also result from different ages, abilities etc., to study the labor market differential effect with respect to different education levels, a semiparametric partial linear quantile model is preferred, which can retain the flexibility of the nonparametric models for the age and other unobservable factors and ease the interpretation of the education factor.

We use the administrative data from the German National Pension Office (Deutsche Rentenversicherung Bund) for the following group: West German part, males, born between 1939 and 1942 who began receiving a pension in 2004 or 2005 (when they were 62–66 years old) with at least 30 yearly uncensored observations. Since different people entered into the pension system and stopped receiving job earnings at different ages, we only consider those earnings recorded by the pension system when they were between 25 and 59 years old. For example, we consider person A’s yearly earnings when he was 25–59 (entering into the pension system at 25), person B’s when he was 27–59 (entering into the pension system at 27), and person C’s when he was 30–59 (entering into the pension system at 30). In total, $n = 128429$ observations are available. We have the following three education categories: “low education”, “apprenticeship” and “university” for the variable u (we assign them the numerical values 1, 2 and 3 respectively); the variable v is the age of the employee. “Low education” means without post-secondary education in Germany. “Apprenticeship” means part of Germany’s dual education

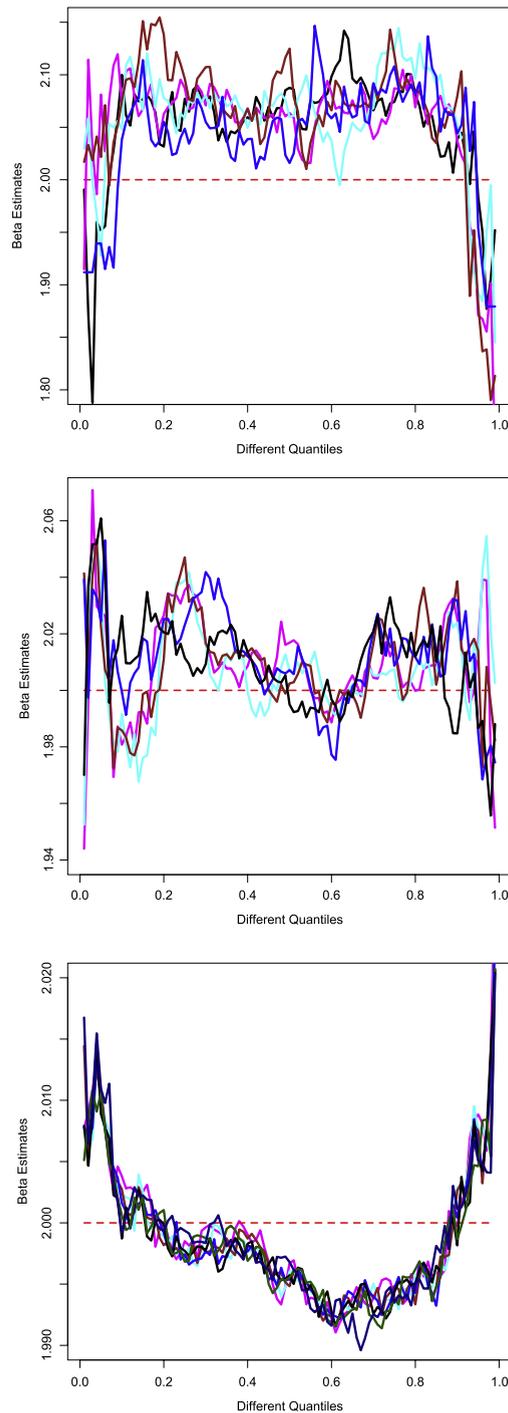


Fig. 3. $\hat{\beta}$ with respect to different quantiles for different numbers of observations, i.e. $n = 1000$ (top), $n = 8000$ (middle), $n = 261\,148$ (bottom). Different lines in the same plot correspond to different q_n , i.e. $n^{1/3}/8$, $n^{1/3}/4$, $n^{1/3}/2$, $n^{1/3}$, $n^{1/3} \cdot 2$, $n^{1/3} \cdot 4$ and $n^{1/3} \cdot 8$.

system. Depending on the profession, a person may work for three to four days a week in the company and then spend one or two days at a vocational school (Berufsschule). “University” in Germany also includes technical colleges (applied universities). Since the level and structure of wages differ substantially between East and West Germany, we concentrate on West Germany only here (which we usually refer to simply as Germany). Our data have several advantages over the most often used German Socio-Economics Panel (GSOEP) data for analyzing wages in Germany. Firstly, they are available for a much longer period, as opposed to from 1984 only for the GSOEP data. Secondly, and more importantly, they have a much larger sample size. Thirdly, wages are likely to be measured much more precisely. Fourthly, we observe a complete earnings

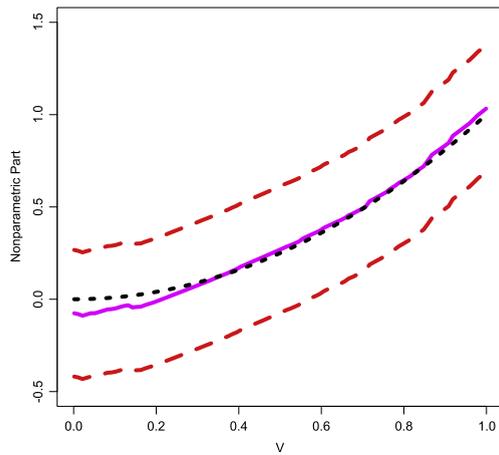


Fig. 4. Nonparametric part smoothing, real 0.9 quantile curve (black dotted line) with respect to v , 0.9 quantile smoother (magenta solid line) with corresponding 95% bootstrap uniform confidence band (red dashed lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

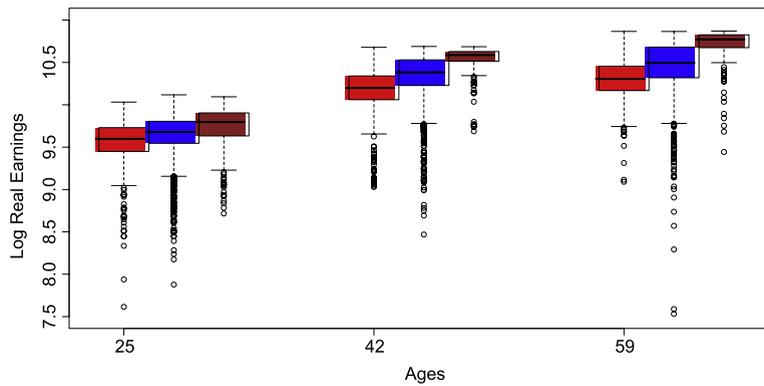


Fig. 5. Boxplots for “low education” (red), “apprenticeship” (blue) and “university” (brown) groups corresponding to different ages. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

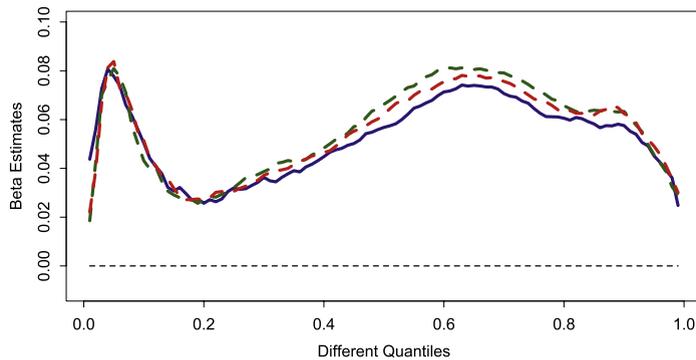


Fig. 6. $\hat{\beta}$ corresponding to different quantiles with 6, 13, 25 partitions.

history from the individual's first job until his retirement, therefore this is a true panel, not a pseudo-panel. There are also several drawbacks. For example, some very wealthy individuals are not registered in the German pension system, e.g. if their monthly income is more than some threshold (which may vary for different years due to the inflation effect), the individual has the right not to be included in the public pension system, and thus is not recorded. Besides this, it is also right-censored at the highest level of earnings that is subject to social security contributions, so the censored observations in the data are only for those who actually decided to stay within the public system. Because of the combination of truncation and censoring, this paper focuses on the uncensored data only, and we should not draw inferences from the very high quantile, i.e. we only consider the 0.80 quantiles here. Recently, similar data were also used to investigate the German wage structure as in [9].

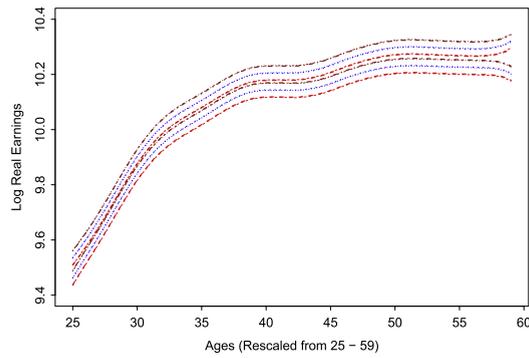


Fig. 7. 95% bootstrap (thick) and asymptotic (thin) uniform confidence bands for 0.20-quantile smoothers w.r.t. 3 different education levels. The “low education”, “apprenticeship” and “university” levels are marked as red dashed, blue dotted and brown dashed-dot lines respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

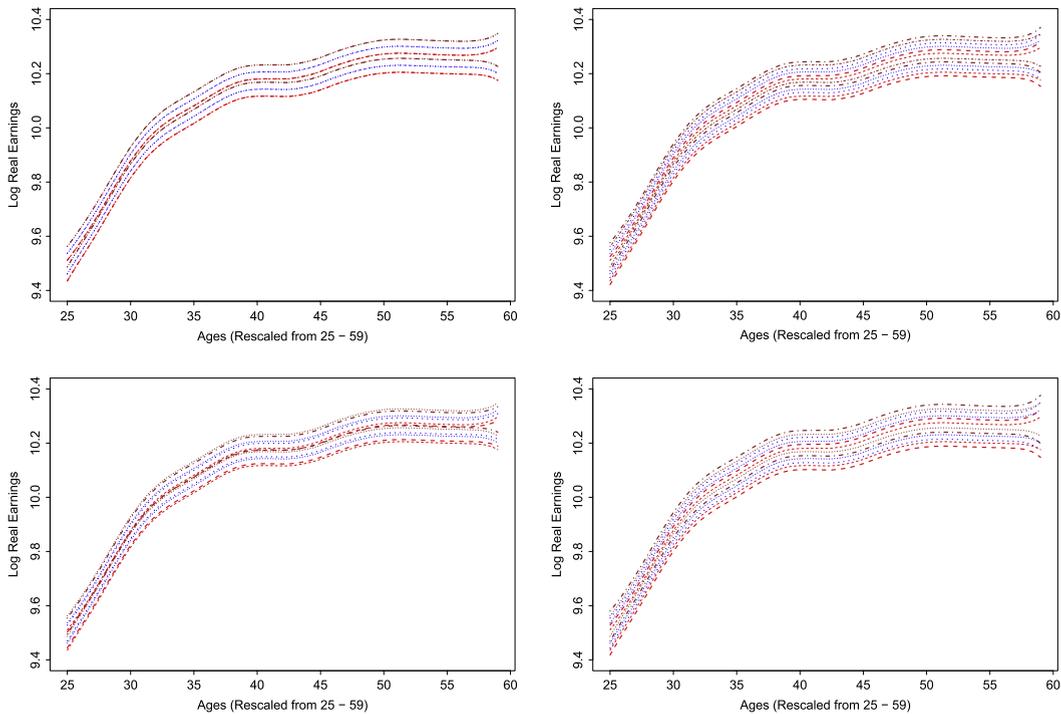


Fig. 8. 95% bootstrap (thick) and asymptotic (thin) uniform confidence bands for 0.20-quantile smoothers w.r.t. 3 different education levels with the oversmoothing bandwidth set as $g/2$, $g/4$, $2g$ and $4g$ (from left to right, up to down) respectively. The “low education”, “apprenticeship” and “university” levels are marked as red dashed, blue dotted and brown dashed-dot lines respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Following from Becker’s [1] human capital model, a log transformation is performed first on the hourly real wages (unit: EUR, at year 2000 prices). Fig. 5 displays the boxplots for the “low education”, “apprenticeship” and “university” groups corresponding to different ages. In the data all ages (25–59) are reported as integers and are categorized in one-year groups. We rescaled them to the interval $[0, 1]$ by dividing by 40, with corresponding bandwidths h of 0.041, 0.039, 0.041 for the 0.20, 0.50, 0.80 nonparametric quantile smoothers respectively. Correspondingly, as discussed before, we choose $g = n^{4/45}h$, thus 0.12, 0.11, 0.12 for the corresponding oversmothers respectively. To detect whether a differential effect for different education levels exists, we compare the corresponding uniform confidence bands, i.e. differences indicate that the differential effect may exist for different education levels in the German labor market for that specific labor group.

Following an application of the partial linear model in Section 3, Fig. 6 displays $\hat{\beta}$ with respect to different quantiles for 6, 13, and 25 partitions, respectively. At first, the $\hat{\beta}$ curve is quite surprising, since it is not, as in mean regression, a positive constant, but rather varies a lot, e.g. $\hat{\beta}(0.20) = 0.026$, $\hat{\beta}(0.50) = 0.057$ and $\hat{\beta}(0.80) = 0.061$. Furthermore, it is robust to different numbers of partitions. It seems that the differences between the “low education” and “university” groups are different for different tails of the wage distribution. To judge whether these differences are significant, we use the uniform

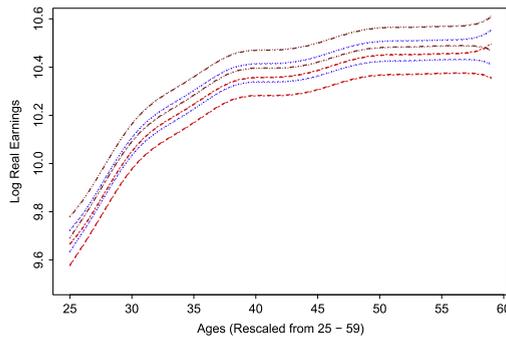


Fig. 9. 95% bootstrap (thick) and asymptotic (thin) uniform confidence bands for 0.50-quantile smoothers w.r.t. 3 different education levels. The “low education”, “apprenticeship” and “university” levels are marked as red dashed, blue dotted and brown dashed–dot lines respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

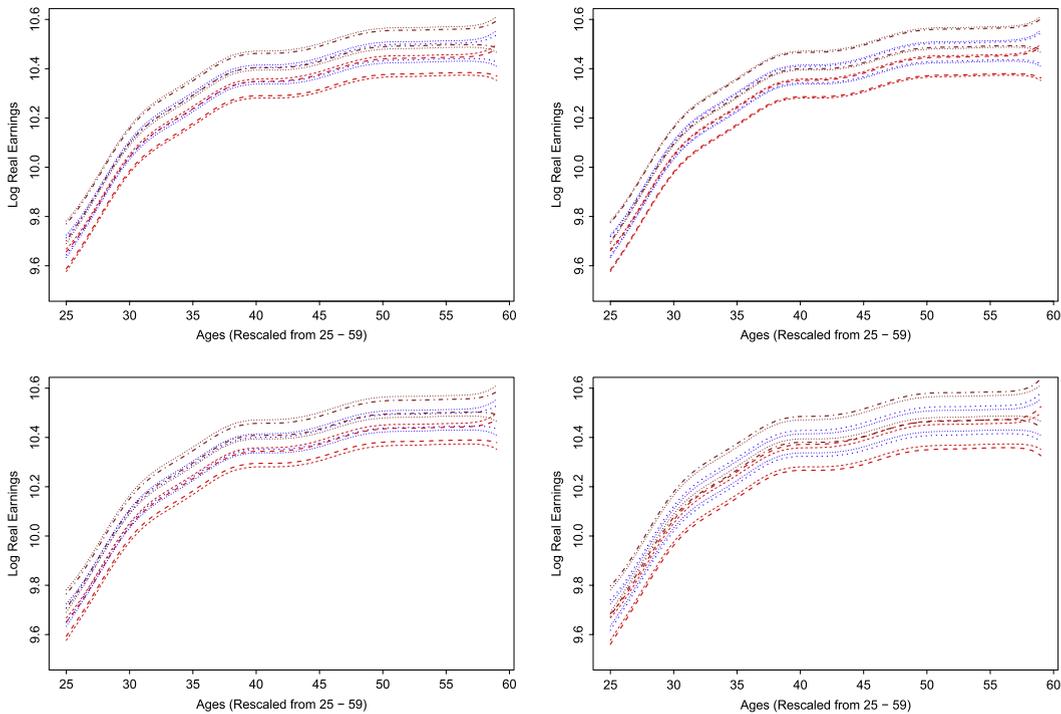


Fig. 10. 95% bootstrap (thick) and asymptotic (thin) uniform confidence bands for 0.50-quantile smoothers w.r.t. 3 different education levels with the oversmoothing bandwidth set as $g/2$, $g/4$, $2g$ and $4g$ (from left to right, up to down) respectively. The “low education”, “apprenticeship” and “university” levels are marked as red dashed, blue dotted and brown dashed–dot lines respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

confidence band techniques discussed in Section 2 which are displayed in Figs. 7–11 corresponding to the 0.20, 0.50 and 0.80 quantiles respectively.

The 95% uniform confidence bands from bootstrapping for the “low education” group are marked as red dashed lines, while the ones for “apprenticeship” and “university” are displayed as blue dotted and brown dashed–dot lines, respectively. The corresponding asymptotic bands studied in [22] are also added for reference (thin lines with the same style and color), which overlap with the bootstrap bands for large samples as here. For the 0.20 quantile in Fig. 7, the bands for “university”, “apprenticeship” and “low education” do not differ significantly from one another although they become progressively lower, which indicates that high education does not equate to higher earnings significantly for the lower tails of wages, while increasing age seems to be the main driving force. For the 0.50 quantile in Fig. 9, the bands for “university” and “low education” differ significantly from one another although not from that for “apprenticeship”. However, for the 0.80 quantiles in Fig. 11, all the bands differ significantly (except on the right boundary because of the nonparametric method’s boundary effect) resulting from the relatively large $\hat{\beta}(0.80) = 0.061$, which indicates that high education is significantly associated with higher earnings for the upper tails of wages.

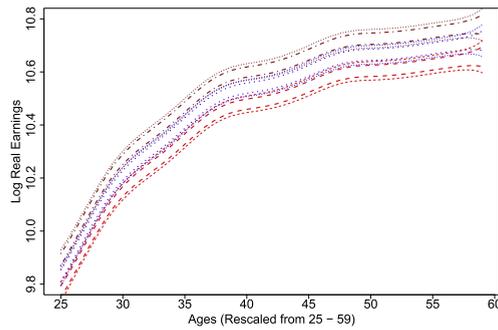


Fig. 11. 95% bootstrap (thick) and asymptotic (thin) uniform confidence bands for 0.80-quantile smoothers w.r.t. 3 different education levels. The “low education”, “apprenticeship” and “university” levels are marked as red dashed, blue dotted and brown dashed–dot lines respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

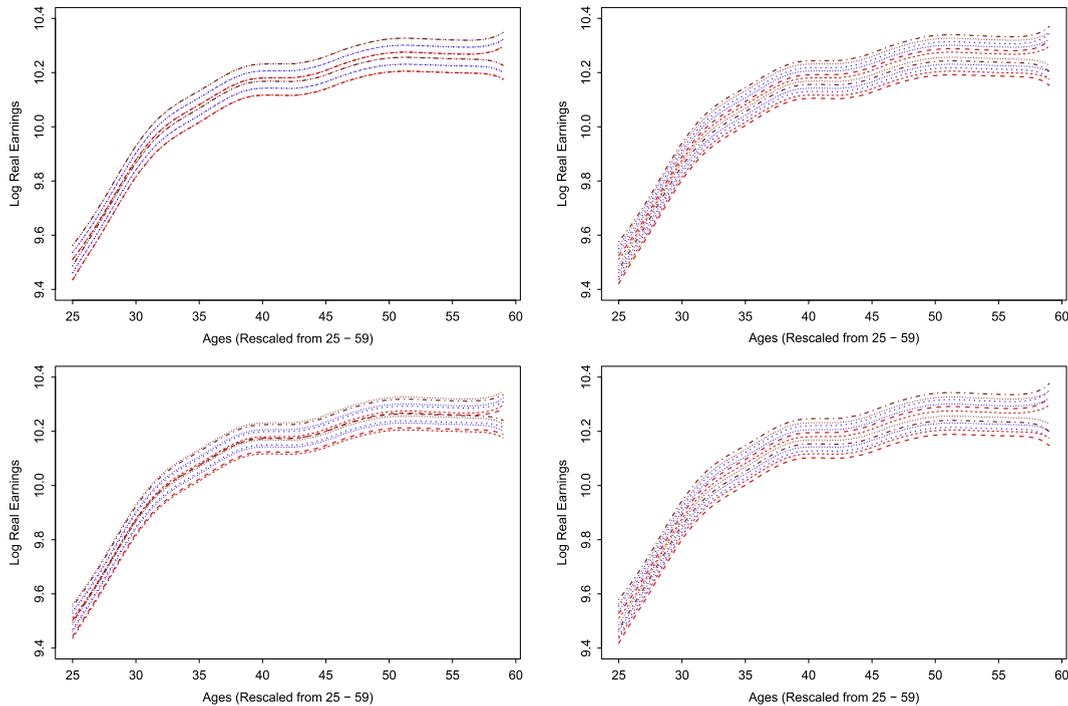


Fig. 12. 95% bootstrap (thick) and asymptotic (thin) uniform confidence bands for 0.80-quantile smoothers w.r.t. 3 different education levels with the oversmoothing bandwidth set as $g/2$, $g/4$, $2g$ and $4g$ (from left to right, up to down) respectively. The corresponding line styles and colors are the same as in Fig. 7. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Coupled with Figs. 7, 9 and 11, Figs. 8, 10 and 12 present the corresponding bootstrap confidence bands constructed using different oversmoothing bandwidths, namely, half, quarter, twice and quadruple (from left to right, up to down) of the oversmoothing bandwidth $g = n^{4/45}h$ used before. The corresponding asymptotic bands are also added for reference (thin lines with the same style and color). As we can see, in practice, for the typically large labor economic data set, the bootstrap confidence bands are quite robust to the choice of the oversmoothing bandwidth.

If we investigate the explanations for the differences in different tails of the income distribution, maybe the most prominent reason is the rapid development of technology, which has been extensively studied. The point is that technology does not simply increase the demand for upper-end labor relative to that of lower-end labor, but instead asymmetrically affects the bottom and the top of the wage distribution, resulting in its strong asymmetry.

6. Conclusions

In this paper we construct confidence bands for nonparametric quantile estimates of regression functions. The method is based on bootstrapping, where resampling is done from a suitably estimated empirical distribution function (edf) for residuals. It is proven that the bootstrap approximation provides an improvement over the confidence bands constructed

via the asymptotic Gumbel distribution. We also propose a partial linear model to handle the case of multidimensional and discrete regressor variables. An economic application considering the labor market differential effect with respect to various education levels is studied. The conclusions from the point of view of quantile regression are consistent with those of the (grouped) mean regression, but in a more careful way in the sense that we provide formal statistical tools to judge these uniformly. The partial linear quantile regression techniques, together with confidence bands, developed in this paper display very interesting findings compared with classic (mean) methods and will bring in more contributions to the differential analysis of the labor market.

Appendix

Proof of Theorem 2.1. We start by proving Eq. (8). Write first $\hat{F}^{-1}(U_i|X_i) = F^{-1}(U_i|X_i) + \Delta_i$. Fix any i such that $|F^{-1}(U_i|X_i) - F^{-1}(p)| \leq S_n \delta_n$, which, by Eq. (1), implies that $|U_i - p| < S_n \delta_n$. Lemma 2.1 gives

$$\max_i |\hat{F}(S_n^2 \delta_n | X_i) - F(S_n^2 \delta_n | X_i)| = \mathcal{O}_p(S_n \delta_n). \tag{26}$$

Together with $F(\pm S_n^2 \delta_n | X_i) = p \pm \mathcal{O}_p(S_n^2 \delta_n)$, again by Eq. (1), we have $\hat{F}(\pm S_n^2 \delta_n | X_i) = p \pm \mathcal{O}_p(S_n^2 \delta_n)$ and thus

$$\begin{aligned} \hat{F}(-S_n^2 \delta_n | X_i) &= p - \mathcal{O}_p(S_n^2 \delta_n) \leq p - S_n \delta_n < U_i < p + S_n \delta_n \\ &< p + \mathcal{O}_p(S_n^2 \delta_n) = \hat{F}(S_n^2 \delta_n | X_i). \end{aligned}$$

Since $\hat{F}(\cdot | X_i)$ is monotone non-decreasing, $|\hat{F}^{-1}(U_i | X_i)| \leq S_n^2 \delta_n$, which means, by $S_n^2 = S_n$,

$$|\hat{F}^{-1}(U_i | X_i)| \leq S_n \delta_n. \tag{27}$$

Apply now Lemma 2.1 again to Eq. (27), and obtain

$$\begin{aligned} S_n \delta_n &\geq |\hat{F}\{\hat{F}^{-1}(U_i | X_i) | X_i\} - F\{\hat{F}^{-1}(U_i | X_i) | X_i\}| \\ &= |U_i - F\{F^{-1}(U_i | X_i) + \Delta_i | X_i\}| \\ &= |F\{F^{-1}(U_i | X_i) | X_i\} - F\{F^{-1}(U_i | X_i) + \Delta_i | X_i\}| \\ &\geq f_0(X_i) |\Delta_i|. \end{aligned} \tag{28}$$

Hence $|\Delta_i| < S_n \delta_n$, and we summarize it as

$$\max_{i: |F^{-1}(U_i | X_i) - F^{-1}(p)| < S_n \delta_n} |F^{-1}(U_i | X_i) - \hat{F}^{-1}(U_i | X_i)| = \mathcal{O}_p\{S_n \delta_n\}.$$

To show Eq. (12), define

$$\begin{aligned} Z_{1j} &\stackrel{\text{def}}{=} Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i), \\ Z_{2j} &\stackrel{\text{def}}{=} Y_j^\# - l(X_j) + l(X_j) - l(X_i). \end{aligned}$$

Thus $q_{hi}[\{(Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i))\}_{j=1}^n]$ and $q_{hi}[\{(Y_j^\# - l(X_j) + l(X_j) - l(X_i))\}_{j=1}^n]$ can be seen as $l_h(X_i)$ for data sets $\{(X_i, Z_{1i})\}_{i=1}^n$ and $\{(X_i, Z_{2i})\}_{i=1}^n$ respectively. Similarly to Härdle and Song [22], they can be treated as zeros (w.r.t. $\theta, \theta \in I$ where I is a possibly infinite, or possibly degenerate, interval in \mathbb{R}) of the functions

$$\tilde{G}_n(\theta, X_i) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n K_h(X_i - X_j) \psi(Z_{1j} - \theta), \tag{29}$$

$$\tilde{\tilde{G}}_n(\theta, X_i) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n K_h(X_i - X_j) \psi(Z_{2j} - \theta). \tag{30}$$

From (8) and (9), we have

$$\begin{aligned} \max_i &|[(Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i))\}_{j=1}^n] - [(Y_j^\# - l(X_j) + l(X_j) - l(X_i))\}_{j=1}^n]| \\ &= \mathcal{O}_p\{S_n \delta_n\} + \mathcal{O}_p(\delta_n) = \mathcal{O}_p(\delta_n). \end{aligned} \tag{31}$$

Thus

$$\sup_{\theta \in I} \max_i |\tilde{G}_n(\theta, X_i) - \tilde{\tilde{G}}_n(\theta, X_i)| \leq \mathcal{O}_p(\delta_n) \max \left| n^{-1} \sum K_h \right| = \mathcal{O}_p(\delta_n).$$

To show the difference of the two quantile smoothers, we shall reduce the strong convergence of $q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\}_{j=1}^n] - q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]$, for any i , to an application of the strong convergence of $\tilde{G}(\theta, X_i)$ to $\tilde{G}_n(\theta, X_i)$, uniformly in θ , for any i . Under assumptions (A7) and (A8), in a similar spirit to Härdle and Song [22], we get

$$\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)| = \mathcal{O}_p(\delta_n).$$

To show the supremum of the bootstrap approximation error, without loss of generality, based on assumption (A1), we reorder the original observations $\{X_i, Y_i\}_{i=1}^n$, such that $X_1 \leq X_2 \leq \dots \leq X_n$. First decompose:

$$\begin{aligned} \sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)| &= \max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) + l(X_i)| \\ &\quad + \max_i \sup_{x \in [X_i, X_{i+1}]} |l_h^*(x) - l_g(x) - l_h^\#(x) + l(x)|. \end{aligned} \tag{32}$$

From assumption (A1) we know $l'(\cdot) \leq \lambda_1$ and $\max_i (X_{i+1} - X_i) = \mathcal{O}_p(S_n/n)$. By the mean value theorem, we conclude that the second term of (32) is of a lower order than the first term. Together with Eq. (12) we have

$$\sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)| = \mathcal{O}\{\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)|\} = \mathcal{O}_p(\delta_n),$$

which means that the supremum of the approximation error over all x is of the same order of the maximum over the discrete observed X_i . \square

Proof of Theorem 2.2. The proof of (16) uses methods related to those in the proof of Theorem 3 of Härdle and Marron [20], so only the main steps are explicitly given. The first step is a bias-variance decomposition,

$$E \left[\left\{ \hat{b}_{h,g}(x) - b_h(x) \right\}^2 \mid X_1, \dots, X_n \right] = \mathcal{V}_n + \mathcal{B}_n^2, \tag{33}$$

where

$$\begin{aligned} \mathcal{V}_n &= \text{Var}[\hat{b}_{h,g}(x) \mid X_1, \dots, X_n], \\ \mathcal{B}_n &= E[\hat{b}_{h,g}(x) - b_h(x) \mid X_1, \dots, X_n]. \end{aligned}$$

Following the uniform Bahadur representation techniques for quantile regression as in Theorem 3.2 of Kong et al. [27], we have the following linear approximation for the quantile smoother as a local polynomial smoother corresponding to a specific loss function:

$$l_h^\#(x) - l(x) = L_n + \mathcal{O}_p(L_n),$$

where

$$L_n = \frac{n^{-1} \sum K_h(x - X_i) \psi \{Y_i - l(x)\}}{f \{l(x) \mid x\} f_X(x)}$$

for

$$\begin{aligned} \psi(u) &= p \mathbf{1}\{u \in (0, \infty)\} - (1 - p) \mathbf{1}\{u \in (-\infty, 0)\} \\ &= p - \mathbf{1}\{u \in (-\infty, 0)\}, \end{aligned}$$

$$\begin{aligned} l(x - t) - l(x) &= l'(x)(-t) + l''(x)t^2 + \mathcal{O}(t^2), \\ \{l(x - t) - l(x)\}' &= l''(x)(-t) + l'''(x)t^2 + \mathcal{O}(t^2), \\ f(x - t) &= f(x) + f'(x)(-t) + f''(x)(t^2) + \mathcal{O}(t^2), \\ f'(x - t) &= f'(x) + f''(x)(-t) + f'''(x)t^2 + \mathcal{O}(t^2), \end{aligned}$$

$$\int K_h(t) t dt = 0,$$

$$\int K_h(t) t^2 dt = h^2 d_K,$$

$$\int K_h(t) \mathcal{O}(t^2) dt = \mathcal{O}(h^2).$$

Then we have

$$\mathcal{B}_n = \mathcal{B}_{n1} + \mathcal{O}(\mathcal{B}_{n1}),$$

where

$$\mathcal{B}_{n1} = \frac{\int K_g(x-t)\mathcal{U}_h(t)dt - \mathcal{U}_h(x)}{f_X(x)f\{l(x)|x\}}$$

for

$$\begin{aligned} \mathcal{U}_h(x) &= \int K_h(x-s)\psi\{l(s)-l(x)\}f(s)ds \\ &= \int K_h(t)\psi\{l(x-t)-l(x)\}f(x-t)dt. \end{aligned}$$

By differentiation, a Taylor expansion and properties of the kernel K (see assumption (A2)),

$$\mathcal{U}'_h(x) = \int K_h(t)[\psi'\{l(x-t)-l(x)\}'f(x-t) + \psi\{l(x-t)-l(x)\}f'(x-t)]dt.$$

Here ψ' is the derivative of ψ except the 0 point, which actually does not matter since there is integration afterwards. Collecting terms, we get

$$\begin{aligned} \mathcal{U}'_h(x) &= \int K_h(t)\{\psi'l''(x)f'_X(x)t^2 + \psi'l'''f_X(x)t^2 + af'''(x)t^2 + \mathcal{O}(t^2)\}dt \\ &= \int K_h(t)\{C_0t^2 + o(t^2)\}dt = h^2d_K \cdot C_0 + \mathcal{O}(h^2), \end{aligned}$$

where a is a constant with $|a| < 1$ and $C_0 = \psi'l''(x)f'_X(x) + \psi'l'''f_X(x) + af'''(x)$.

Hence, by another substitution and Taylor expansion, for the first term in the numerator of \mathcal{B}_{n1} , we have

$$\mathcal{B}_{n2} = g^2h^2(d_K)^2 \cdot C_0 + \mathcal{O}(g^2h^2).$$

Thus, along almost all sample sequences,

$$\mathcal{B}_n^2 = C_1g^4h^4 + \mathcal{O}(g^4h^4) \tag{34}$$

for $C_1 = (d_K)^4C_0^2/[f'_X(x)f^2\{l(x)|x\}]$.

For the variance term, calculation in a similar spirit shows that

$$\mathcal{V}_n = \mathcal{V}_{n1} + \mathcal{O}(\mathcal{V}_{n1}),$$

where

$$\mathcal{V}_{n1} = \frac{\int K_g^2(x-t)\mathcal{W}_h(t)dt - \left\{\int K_g(x-t)\mathcal{U}_h(t)dt\right\}^2}{f_X(x)f\{l(x)|x\}}$$

for

$$\begin{aligned} \mathcal{W}_h(x) &= \int K_h^2(x-s)\psi\{l(s)-l(x)\}^2f(s)ds \\ &= \int K_h^2(t)\psi\{l(x-t)-l(x)\}^2f(x-t)dt. \end{aligned}$$

Hence, by Taylor expansion, collecting items and similar calculation, we have

$$\mathcal{V}_n = n^{-1}h^4g^{-5}C_2 + \mathcal{O}(n^{-1}h^4g^{-5}) \tag{35}$$

for a constant C_2 . This, together with (33) and (34), completes the proof of Theorem 2.2. \square

Proof of Theorem 3.1. In the case where the function l is known, the estimate $\hat{\beta}_l$ is

$$\hat{\beta}_l = \operatorname{argmin}_{\beta} \sum_{i=1}^n \psi\{Y_i - l(V_i) - U_i^\top \beta\}.$$

Since l is unknown, in each of these small intervals I_{ni} , $l(V_i)$ could be regarded as a constant $\alpha = l(m_{ni})$ for some i whose corresponding interval I_{ni} covers V_i . From assumption (A1), we know that $|l(V_i) - \alpha_i| \leq \lambda_1 b_n < \infty$. If we define our first step estimate $\hat{\beta}_i$ inside each small interval as

$$(\hat{\alpha}_i, \hat{\beta}_i) = \operatorname{argmin}_{\alpha, \beta} \sum \psi(Y_i - \alpha - U_i^\top \beta),$$

$|\{Y_i - l(V_i) - U_i^\top \beta\} - (Y_i - \alpha - U_i^\top \beta)| \leq \lambda_1 b_n < \infty$ indicates that we could treat $\hat{\beta}_i$ as $\hat{\beta}_i$ inside each partition. If we use d_i to denote the number of observations inside partition I_{ni} (based on the i.i.d. assumption as in assumption (A1), on average $d_i = n/a_n$). For each of the $\hat{\beta}_i$'s inside interval I_{ni} , various parametric quantile regression works, e.g. the convex function rule in [31,24], yield

$$\sqrt{d_i}(\hat{\beta}_i - \beta) \xrightarrow{\mathcal{L}} N\{0, p(1-p)D_i'^{-1}(p)C_i'D_i'^{-1}(p)\} \quad (36)$$

with the matrices $C_i' = d_i^{-1} \sum_{i=1}^{d_i} U_i^\top U_i$ and $D_i'(p) = d_i^{-1} \sum_{i=1}^{d_i} f\{l(V_i)|V_i\}U_i^\top U_i$.

To get $\hat{\beta}$, our second step is to take the weighted mean of $\hat{\beta}_1, \dots, \hat{\beta}_{a_n}$ as

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^{a_n} d_i(\hat{\beta}_i - \beta)^2 = \sum_{i=1}^{a_n} d_i \hat{\beta}_i / n.$$

Note that under this construction, $\hat{\beta}_1, \dots, \hat{\beta}_{a_n}$ are independent but not identical. Thus we intend to use the Lindeberg condition for the central limit theorem. To this end, we use s_n^2 to denote $\text{Var}(\sum_{i=1}^{a_n} d_i \hat{\beta}_i / n)$, and we need to further check whether the following ‘‘Lindeberg condition’’ holds:

$$\lim_{a_n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^{a_n} \int_{(|d_i \hat{\beta}_i / n - \beta| > \varepsilon s_n)} (\hat{\beta}_i - \beta)^2 dF = 0, \quad \text{for all } \varepsilon > 0. \quad (37)$$

Since

$$\begin{aligned} \text{Var} \left\{ \sum_{i=1}^{a_n} d_i(\hat{\beta}_i - \beta) / n \right\} &= \sum_i^{a_n} p(1-p) \left\{ \left[n/d_i \sum_{j=1}^{d_i} f\{l(V_j)|v\}U_j^\top U_j \right]^{-1} \right. \\ &\quad \left. \times \sum_{i=1}^{d_i} U_i^\top U_i \left[n/d_i \sum_{j=1}^{d_i} f\{l(V_j)|v\}U_j^\top U_j \right]^{-1} \right\} \\ &\approx p(1-p) \left[\sum_{j=1}^n f\{l(V_j)|v\}U_j^\top U_j \right]^{-1} \sum_{i=1}^n U_i^\top U_i \left[\sum_{j=1}^n f\{l(V_j)|v\}U_j^\top U_j \right]^{-1} \\ &\stackrel{\text{def}}{=} \frac{1}{n} p(1-p) D_n^{-1} C_n D_n^{-1}, \end{aligned}$$

where $D_n = \frac{1}{n} \sum_{j=1}^n f\{l(V_j)|V_j\}U_j^\top U_j$ and $C_n = \frac{1}{n} \sum_{i=1}^n U_i^\top U_i$, together with the normality of $\hat{\beta}_i$ as in (36) and properties of the tail of the normal distribution, e.g. Exe. 14.3–14.4 of Borak et al. [3], (37) follows.

Thus as $n, a_n \rightarrow \infty$ (although at a lower rate than n), together with $C = \text{plim}_{n \rightarrow \infty} C_n$, $D = \text{plim}_{n \rightarrow \infty} D_n$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N\{0, p(1-p)D^{-1}CD^{-1}\}. \quad \square \quad (38)$$

References

- [1] G. Becker, Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, third ed., The University of Chicago Press, 1994.
- [2] A. Belloni, V. Chernozhukov, L_1 -penalized quantile regression in high-dimensional sparse models, *Annals of Statistics* 39 (1) (2011) 82–130.
- [3] S. Borak, W. Härdle, B. Lopez, *Statistics of Financial Markets Exercises and Solutions*, Springer-Verlag, Heidelberg, 2010.
- [4] M. Buchinsky, Quantile regression, Box–Cox transformation model, and the US wage structure, 1963–1987, *Journal of Econometrics* 65 (1995) 109–154.
- [5] Z.W. Cai, Regression quantiles for time series, *Econometric Theory* 18 (2002) 169–192.
- [6] V. Chernozhukov, S. Lee, A.M. Rosen, Intersection bounds: estimation and inference, CEMMAP Working Paper, 19/09, 2009.
- [7] J.C. Day, E.C. Newburger, The big payoff: educational attainment and synthetic estimates of work-life earnings. Special studies, current population reports, Statistical Report, US Department of Commerce Economics and Statistics Administration, US CENSUS BUREAU, 2002, pp. 23–210.
- [8] L. Denby, Smooth regression functions. Statistical Report 26, AT&T Bell Laboratories, 1986.
- [9] C. Dustmann, J. Ludsteck, U. Schönberg, Revisiting the German wage structure, *The Quarterly Journal of Economics* 124 (2) (2009) 843–881.
- [10] J. Fan, T.C. Hu, Y.K. Troung, Robust nonparametric function estimation, *Scandinavian Journal of Statistics* 21 (1994) 433–446.
- [11] J. Fan, Q. Yao, H. Tong, Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, *Biometrika* 83 (1996) 189–206.
- [12] R.A. Fisher, L.H.C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society* 24 (1928) 180–190.
- [13] J. Franke, P. Mwita, Nonparametric estimates for conditional quantiles of time series, Report in *Wirtschaftsmathematik* 87, University of Kaiserslautern, 2003.
- [14] P.J. Green, B.S. Yandell, Semi-parametric generalized linear models, in: *Proceedings 2nd International GLIM Conference*, in: *Lecture Notes in Statistics* 32, vol. 32, Springer, New York, 1985, pp. 44–55.
- [15] J. Hahn, Bootstrapping quantile regression estimators, *Econometric Theory* 11 (1) (1995) 105–121.
- [16] P. Hall, On convergence rates of suprema, *Probability Theory and Related Fields* 89 (1991) 447–455.
- [17] P. Hall, R. Wolff, Q. Yao, Methods for estimating a conditional distribution function, *Journal of the American Statistical Association* 94 (1999) 154–163.
- [18] W. Härdle, P. Janssen, R. Serfling, Strong uniform consistency rates for estimators of conditional functionals, *Annals of Statistics* 16 (1988) 1428–1429.

- [19] W. Härdle, S. Luckhaus, Uniform consistency of a class of regression function estimators, *Annals of Statistics* 12 (1984) 612–623.
- [20] W. Härdle, J. Marron, Bootstrap simultaneous error bars for nonparametric regression, *Annals of Statistics* 19 (1991) 778–796.
- [21] W. Härdle, M. Müller, S. Sperlich, A. Werwatz, *Nonparametric and Semiparametric Models*, Springer Verlag, Heidelberg, 2004.
- [22] W. Härdle, S. Song, Confidence bands in quantile regression, *Econometric Theory* 26 (2010) 1–22. with corrections forthcoming or obtained from the authors.
- [23] J.L. Horowitz, Bootstrap methods for median regression models, *Econometrica* 66 (6) (1998) 1327–1351.
- [24] K. Knight, Comparing conditional quantile estimators: first and second order considerations, Unpublished Manuscript, 2001.
- [25] R. Koenker, G.W. Bassett, Regression quantiles, *Econometrica* 46 (1978) 33–50.
- [26] R. Koenker, K.F. Hallock, Quantile regression, *The Journal of Economic Perspectives* 15 (4) (2001) 143–156.
- [27] E. Kong, O. Linton, Y. Xia, Uniform Bahadur representation for local polynomial estimates of M -regression and its application to the additive model, *Econometric Theory* 26 (2010) 1529–1564.
- [28] C. Kuan, J. Yeh, Y. Hsu, Assessing value at risk with care, the conditional autoregressive expectile models, *Journal of Econometrics* 150 (2009) 261–270.
- [29] H. Liang, R. Li, Variable selection for partially linear models with measurement errors, *Journal of the American Statistical Association* 104 (485) (2009) 234–248.
- [30] W. Newey, J. Powell, Asymmetric least squares estimation and testing, *Econometrica* 55 (1987) 816–847.
- [31] D. Pollard, Asymptotics for least absolute deviation estimators, *Econometric Theory* 7 (1991) 186–199.
- [32] P.M. Robinson, Semiparametric econometrics: a survey, *Journal of Applied Econometrics* 3 (1988) 35–51.
- [33] D. Ruppert, S.J. Sheather, M.P. Wand, An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association* 90 (432) (1995) 1257–1270.
- [34] P.E. Speckman, Regression analysis for partially linear models, *Journal of the Royal Statistical Society, Series B* 50 (1988) 413–436.
- [35] K. Yu, M.C. Jones, A comparison of local constant and local linear regression quantile estimation, *Computational Statistics and Data Analysis* 25 (1997) 159–166.
- [36] K. Yu, M.C. Jones, Local linear quantile regression, *Journal of the American Statistical Association* 93 (1998) 228–237.



Difference based ridge and Liu type estimators in semiparametric regression models[☆]

Esra Akdeniz Duran^{a,1}, Wolfgang Karl Härdle^b, Maria Osipenko^{c,*}

^a Department of Statistics, Gazi University, Turkey

^b Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, Germany

^c CASE, School of Business and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Germany

ARTICLE INFO

Article history:

Received 4 March 2011

Available online 7 September 2011

AMS subject classifications:

62G08

62J07

Keywords:

Difference based estimator

Differencing estimator

Differencing matrix

Liu estimator

Liu type estimator

Multicollinearity

Ridge regression estimator

Semiparametric model

ABSTRACT

We consider a difference based ridge regression estimator and a Liu type estimator of the regression parameters in the partial linear semiparametric regression model, $y = X\beta + f + \varepsilon$. Both estimators are analyzed and compared in the sense of mean-squared error. We consider the case of independent errors with equal variance and give conditions under which the proposed estimators are superior to the unbiased difference based estimation technique. We extend the results to account for heteroscedasticity and autocovariance in the error terms. Finally, we illustrate the performance of these estimators with an application to the determinants of electricity consumption in Germany.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Semiparametric partial linear models have received considerable attention in statistics and econometrics. They have a wide range of applications, from biomedical studies to economics. In these models, some explanatory variables have a linear effect on the response while others are entering nonparametrically. Consider the semiparametric regression model:

$$y_i = x_i^\top \beta + f(t_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

where y_i 's are observations at t_i , $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq 1$ and $x_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$ are known p -dimensional vectors with $p \leq n$. In many applications, t_i 's are values of an extra univariate "time" variable at which responses y_i are observed. In the case $t_i \in \mathbb{R}^k$, $t_i = (t_{i1}, \dots, t_{ik})^\top$, the triples $(y_1, x_1, t_1), \dots, (y_n, x_n, t_n)$ should be ordered using one of the algorithms mentioned in [30], Appendix A, or in [8, Section 2.2].

[☆] This research was supported by Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk'.

* Corresponding author.

E-mail addresses: esraakdeniz@gmail.com (E. Akdeniz Duran), maria.osipenko@wiwi.hu-berlin.de (M. Osipenko).

¹ Dr. Esra Akdeniz Duran was a research associate at Humboldt-Universität zu Berlin, Germany during this research.

In Eq. (1), $\beta = (\beta_1, \dots, \beta_p)^\top$ is an unknown p -dimensional parameter vector, $f(\cdot)$ is an unknown smooth function and ε 's are independent and identically distributed random errors with $E(\varepsilon|x, t) = 0$ and $\text{Var}(\varepsilon|x, t) = \sigma^2$. We shall call $f(t)$ the smooth part of the model and assume that it represents a smooth unparameterized functional relationship.

The goal is to estimate the unknown parameter vector β and the nonparametric function $f(t)$ from the data $\{y_i, x_i, t_i\}_{i=1}^n$. In vector/matrix notation, (1) is written as

$$y = X\beta + f + \varepsilon \tag{2}$$

where $y = (y_1, \dots, y_n)^\top, X = (x_1, \dots, x_n), f = \{f(t_1), \dots, f(t_n)\}^\top, \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$.

Semiparametric models are by design more flexible than standard linear regression models since they combine both parametric and nonparametric components. There exist various goodness-of-fit tests to identify the nonparametric part in this kind of models; see [8] and the references therein. Estimation techniques for semiparametric partially linear models are based on different nonparametric regression procedures. The most important approaches to estimate β and f are given in [12,4,7,6,5,14,24,15,33].

In practice, researchers often encounter the problem of multicollinearity. In case of multicollinearity, we know that the $(p \times p)$ matrix $X^\top X$ has one or more small eigenvalues; the estimates of the regression coefficients can therefore have large variances: the least squares estimator performs poorly in this case. Hoerl and Kennard [17] proposed the ridge regression estimator and it has become the most common method to overcome this particular weakness of the least squares estimator. For the purpose of this paper, we will employ the biased estimator that was proposed by Liu [20] to combat the multicollinearity. The Liu estimator combines the Stein [26] estimator with the ridge regression estimator; see also [1,13].

The condition number is a measure of multicollinearity. If $X^\top X$ is ill-conditioned with a large condition number, the ridge regression estimator or Liu estimator can be used to estimate β , [21]. We consider difference based ridge and Liu type estimators in comparison to the unbiased difference based approach. We give theoretical conditions that determine superiority among the estimation techniques in the mean squared error matrix sense.

We use data on monthly electricity consumption and its determinants (income, electricity and gas prices, temperature) for Germany. The purpose is to understand electricity consumption as a linear function of income and price and a nonlinear function of temperature: semiparametric approach is therefore necessary here. The data reveal a high condition number of 20.5; we therefore expect a more precise estimation with Ridge or Liu type estimators. We show how our theoretically derived conditions can be implemented for a given data set and be used to determine the appropriate biased estimation technique.

The paper is organized as follows. In Section 2, the model and the differencing estimator is defined. We introduce difference based ridge and Liu type estimators in Section 3. In Section 4, the differencing estimator proposed by Yatchew [30] and the difference based Liu type estimator are compared in terms of the mean squared error. In Section 5, both biased regression methodologies in semiparametric regression models are compared in terms of the mean squared error. Section 6 relaxes the assumption of i.i.d. errors and replicates the results of the previous sections in the presence of heteroscedasticity and autocorrelation. Section 7 gives a real data example to show the performance of the proposed estimators.

2. The model and differencing estimator

In this section, we introduce a difference based technique for the estimation of the linear coefficient vector in a semiparametric regression. This technique has been used to remove the nonparametric component in the partially linear model by various authors (e.g. [30,32,19,3]).

Consider the semiparametric regression model (2). Let $d = (d_0, d_1, \dots, d_m)^\top$ be an $m + 1$ vector where m is the order of differencing and d_0, d_1, \dots, d_m are differencing weights that minimize

$$\sum_{k=1}^m \left(\sum_{j=1}^{m-k} d_j d_{k+j} \right)^2,$$

such that

$$\sum_{j=0}^m d_j = 0 \quad \text{and} \quad \sum_{j=0}^m d_j^2 = 1 \tag{3}$$

are satisfied.

Let us define the $(n - m) \times n$ differencing matrix D to have first and last rows $(d^\top, 0_{n-m-1}^\top), (0_{n-m-1}^\top, d^\top)$ respectively, with i -th row $(0_i, d^\top, 0_{n-m-i-1}^\top), i = 1, \dots, (n - m - 1)$, where 0_r indicates an r -vector of all zero elements

$$D = \begin{pmatrix} d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & \cdots & 0 \\ 0 & d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & & & \\ 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m & 0 \\ 0 & 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m \end{pmatrix}.$$

Applying the differencing matrix to (2) permits direct estimation of the parametric effect. Eubank et al. [6] showed that the parameter vector in (2) can be estimated with parametric efficiency. If f is an unknown function with bounded first derivative, then Df is essentially 0, so that applying the differencing matrix we have

$$\begin{aligned} Dy &= DX\beta + Df + D\varepsilon \approx DX\beta + D\varepsilon \\ \tilde{y} &\approx \tilde{X}\beta + \tilde{\varepsilon} \end{aligned} \tag{4}$$

where $\tilde{y} = Dy$, $\tilde{X} = DX$ and $\tilde{\varepsilon} = D\varepsilon$. Constraints (3) ensure that the nonparametric effect is removed and $\text{Var}(\tilde{\varepsilon}) = \text{Var}(\varepsilon) = \sigma^2$. With (4), a simple differencing estimator of the parameter β in the semiparametric regression model results:

$$\begin{aligned} \hat{\beta}_{(0)} &= \{(DX)^\top(DX)\}^{-1}(DX)^\top Dy \\ &= (\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{y}. \end{aligned} \tag{5}$$

Thus, differencing allows one to perform inferences on β as if there were no nonparametric component f in model (2), [9]. We will also use the modified estimator of σ^2 proposed by Eubank et al. [7]

$$\hat{\sigma}^2 = \frac{\tilde{y}^\top(I - P^\perp)\tilde{y}}{\text{tr}\{D^\top(I - P^\perp)D\}} \tag{6}$$

with $P^\perp = \tilde{X}(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top$, I ($p \times p$) identity matrix and $\text{tr}(\cdot)$ denoting the trace function for a square matrix.

3. Difference based ridge and Liu type estimator

As an alternative to $\hat{\beta}_{(0)}$ in (5), [27] propose:

$$\hat{\beta}_{(1)}(k) = (\tilde{X}^\top\tilde{X} + kI)^{-1}\tilde{X}^\top\tilde{y}, \quad k \geq 0;$$

here k is the ridge-biasing parameter selected by the researcher. We call $\hat{\beta}_{(1)}(k)$ a difference based ridge regression estimator of the semiparametric regression model.

From the least squares perspective, the coefficients β are chosen to minimize

$$(\tilde{y} - \tilde{X}\beta)^\top(\tilde{y} - \tilde{X}\beta). \tag{7}$$

Adding to the least squares objective (7) a penalizing function of the squared norm $\|\eta\hat{\beta}_{(0)} - \beta\|^2$ for the vector of regression coefficients, yields a conditional objective:

$$L = (\tilde{y} - \tilde{X}\beta)^\top(\tilde{y} - \tilde{X}\beta) + (\eta\hat{\beta}_{(0)} - \beta)^\top(\eta\hat{\beta}_{(0)} - \beta). \tag{8}$$

Minimizing (8) with respect to β , we obtain the estimator $\hat{\beta}_{(2)}(\eta)$ an alternative to $\hat{\beta}_{(0)}$ in (5):

$$\hat{\beta}_{(2)}(\eta) = (\tilde{X}^\top\tilde{X} + I)^{-1}(\tilde{X}^\top\tilde{y} + \eta\hat{\beta}_{(0)}), \tag{9}$$

where η , $0 \leq \eta \leq 1$, is a biasing parameter and when $\eta = 1$, $\hat{\beta}_{(2)}(\eta) = \hat{\beta}_{(0)}$. The formal resemblance between (9) and the Liu estimator motivated [1,18,29] to call it the difference based Liu type estimator of the semiparametric regression model.

4. Mean squared error matrix (MSEM) comparison of $\hat{\beta}_{(0)}$ with $\hat{\beta}_{(2)}(\eta)$

In this section, the objective is to examine the difference of the mean square error matrices of $\hat{\beta}_{(0)}$ and $\hat{\beta}_{(2)}(\eta)$. We note that for any estimator $\tilde{\beta}$ of β , its mean squared error matrix (MSEM) is defined as $\text{MSEM}(\tilde{\beta}) = \text{Cov}(\tilde{\beta}) + \text{Bias}(\tilde{\beta})\text{Bias}(\tilde{\beta})^\top$, where $\text{Cov}(\tilde{\beta})$ denotes the variance–covariance matrix and $\text{Bias}(\tilde{\beta}) = E(\tilde{\beta}) - \beta$ is the bias vector. The expected value of $\hat{\beta}_{(2)}(\eta)$ can be written as

$$E\{\hat{\beta}_{(2)}(\eta)\} = \beta - (1 - \eta)(\tilde{X}^\top\tilde{X} + I)^{-1}\beta.$$

The bias of the $\hat{\beta}_{(2)}(\eta)$ is given as

$$\text{Bias}\{\hat{\beta}_{(2)}(\eta)\} = -(1 - \eta)(\tilde{X}^\top\tilde{X} + I)^{-1}\beta. \tag{10}$$

Denoting $F_\eta = (\tilde{X}^\top\tilde{X} + I)^{-1}(\tilde{X}^\top\tilde{X} + \eta I)$ and observing F_η and $(\tilde{X}^\top\tilde{X})^{-1}$ are commutative, we may write $\hat{\beta}_{(2)}(\eta)$ as

$$\begin{aligned} \hat{\beta}_{(2)}(\eta) &= F_\eta\hat{\beta}_{(0)} = F_\eta(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{y} \\ &= (\tilde{X}^\top\tilde{X})^{-1}F_\eta\tilde{X}^\top\tilde{y}. \end{aligned}$$

Setting $S = (D^T \tilde{X})^T (D^T \tilde{X})$ and $U = (\tilde{X}^T \tilde{X})^{-1}$ we may write $\text{Cov}\{\hat{\beta}_{(2)}(\eta)\}$ as

$$\text{Cov}\{\hat{\beta}_{(2)}(\eta)\} = \sigma^2 F_\eta USUF_\eta^T, \tag{11}$$

$$\text{Cov}\{\hat{\beta}_{(0)}\} = \sigma^2 USU. \tag{12}$$

Using (11) and (12), the difference $\Delta_1 = \text{Cov}\{\hat{\beta}_{(0)}\} - \text{Cov}\{\hat{\beta}_{(2)}(\eta)\}$ can be expressed as

$$\begin{aligned} \Delta_1 &= \sigma^2 (USU - F_\eta USUF_\eta^T) \\ &= \sigma^2 F_\eta \{F_\eta^{-1} USU (F_\eta^T)^{-1} - USU\} F_\eta^T \\ &= \sigma^2 (1 - \eta^2) (U^{-1} + I)^{-1} \left\{ \frac{1}{1 + \eta} (US + SU) + USU \right\} (U^{-1} + I)^{-1}. \end{aligned} \tag{13}$$

Let $\tau = \frac{1}{1+\eta} > 0$, $M = USU$, $N = US + SU$. Since $M = L^T L$ and $\text{rank}(L) = p < n - m$, then M is a $(p \times p)$ positive definite matrix, where $L = D^T \tilde{X} (\tilde{X}^T \tilde{X})^{-1}$ and $N = US + SU$ is a symmetric matrix. Thus, we may write (13) as

$$\begin{aligned} \Delta_1 &= \sigma^2 (1 - \eta^2) H (M + \tau N) H \\ &= \sigma^2 (1 - \eta^2) H (Q^T)^{-1} (Q^T M Q + \tau Q^T N Q) Q^{-1} H \\ &= \sigma^2 (1 - \eta^2) H (Q^T)^{-1} (I + \tau E) Q^{-1} H, \end{aligned}$$

where $I + \tau E = \text{diag}(1 + \tau e_{11}, \dots, 1 + \tau e_{pp})$ and $H = (U^{-1} + I)^{-1}$. Since M is a positive definite and N is a symmetric matrix, a nonsingular matrix Q exists such that $Q^T M Q = I$ and $Q^T N Q = E$; here E is a diagonal matrix and its diagonal elements are the roots of the polynomial equation $|M^{-1}N - eI| = 0$ (see [11, pp. 408] and [16, pp. 563]) and since $N = US + SU \neq 0$, there is at least one diagonal element of E that is nonzero. Let $e_{ii} < 0$ for at least one i ; then positive definiteness of $I + \tau E$ is guaranteed by

$$0 < \tau < \min_{e_{ii} < 0} \left| \frac{1}{e_{ii}} \right|. \tag{14}$$

Hence $1 + \tau e_{ii} > 0$ for all $i = 1, \dots, p$ and therefore $I + \tau E$ is a positive definite matrix. Consequently, Δ_1 becomes a positive definite matrix, as well. It is now evident that the estimator $\hat{\beta}_{(2)}(\eta)$ has a smaller variance compared with the estimator $\hat{\beta}_{(0)}$ if and only if (14) is satisfied.

Next, we give necessary and sufficient conditions for the difference based Liu type estimator $\hat{\beta}_{(2)}(\eta)$ to be superior to $\hat{\beta}_{(0)}$ in the mean squared error matrix (MSEM) sense.

The proof of the next theorem requires the following lemma.

Lemma 4.1 (Farebrother [10]). *Let A be a positive definite $(p \times p)$ matrix, b a $(p \times 1)$ nonzero vector and δ a positive scalar. Then $\delta A - bb^T$ is non-negative if and only if $b^T A^{-1} b \leq \delta$.*

Let us compare the performance of $\hat{\beta}_{(2)}(\eta)$ with the differencing estimator $\hat{\beta}_{(0)}$ with respect to the MSEM criterion. In order to do that, define $\Delta_2 = \text{MSEM}\{\hat{\beta}_{(0)}\} - \text{MSEM}\{\hat{\beta}_{(2)}(\eta)\}$. Observe that

$$\text{MSEM}\{\hat{\beta}_{(0)}\} = \text{Cov}\{\hat{\beta}_{(0)}\} = \sigma^2 USU \tag{15}$$

and

$$\text{MSEM}\{\hat{\beta}_{(2)}(\eta)\} = \sigma^2 F_\eta USUF_\eta^T + (1 - \eta)^2 (U^{-1} + I)^{-1} \beta \beta^T (U^{-1} + I)^{-1}. \tag{16}$$

Then from (15) and (16) one derives

$$\begin{aligned} \Delta_2 &= \sigma^2 F_\eta \{F_\eta^{-1} USU (F_\eta^T)^{-1} - USU\} F_\eta^T - (1 - \eta)^2 (U^{-1} + I)^{-1} \beta \beta^T (U^{-1} + I)^{-1}, \\ &= H \left\{ \sigma^2 (1 - \eta^2) (M + \tau N) - (1 - \eta)^2 \beta \beta^T \right\} H, \\ &= (1 - \eta)^2 H \left\{ \sigma^2 \frac{1 + \eta}{1 - \eta} (M + \tau N) - \beta \beta^T \right\} H. \end{aligned}$$

Applying Lemma 4.1 and assuming condition (14) to be satisfied, we see Δ_2 is positive definite if and only if

$$\beta^T (M + \tau N)^{-1} \beta \leq \sigma^2 \frac{1 + \eta}{1 - \eta}, \quad 0 < \eta < 1.$$

Now we may state the following theorem.

Theorem 4.1. *Consider the two estimators $\hat{\beta}_{(2)}(\eta)$ and $\hat{\beta}_{(0)}$ of β . Let $W = \frac{1+\eta}{1-\eta} (M + \tau N)$ be a positive definite matrix. Then the biased estimator $\hat{\beta}_{(2)}(\eta)$ is MSEM superior to $\hat{\beta}_{(0)}$ if and only if*

$$\beta^T W^{-1} \beta \leq \sigma^2.$$

5. MSEM comparison of $\hat{\beta}_{(1)}(k)$ and $\hat{\beta}_{(2)}(\eta)$

Let us now compare the MSEM performance of

$$\begin{aligned} \hat{\beta}_{(1)}(k) &= (\tilde{X}^T \tilde{X} + kI)^{-1} \tilde{X}^T \tilde{y} \\ &= S_k \tilde{X}^T D y \\ &= A_1 y \end{aligned} \tag{17}$$

with

$$\begin{aligned} \hat{\beta}_{(2)}(\eta) &= (\tilde{X}^T \tilde{X} + I)^{-1} (\tilde{X}^T y + \eta \hat{\beta}_{(0)}) \\ &= (\tilde{X}^T \tilde{X})^{-1} (\tilde{X}^T \tilde{X} + I)^{-1} (\tilde{X}^T \tilde{X} + \eta I) \tilde{X}^T \tilde{y} \\ &= U F_\eta \tilde{X}^T D y \\ &= A_2 y. \end{aligned} \tag{18}$$

The MSEM of the difference based ridge regression estimator $\hat{\beta}_{(1)}(k)$ is given by

$$\begin{aligned} \text{MSEM}\{\hat{\beta}_{(1)}(k)\} &= \text{Cov}\{\hat{\beta}_{(1)}(k)\} + \text{Bias}\{\hat{\beta}_{(1)}(k)\} \text{Bias}\{\hat{\beta}_{(1)}(k)\}^T \\ &= S_k (\sigma^2 S + k^2 \beta \beta^T) S_k^T \\ &= \sigma^2 (A_1 A_1^T) + d_1 d_1^T, \end{aligned}$$

where $S_k = (\tilde{X}^T \tilde{X} + kI)^{-1}$ and $d_1 = \text{Bias}\{\hat{\beta}_{(1)}(k)\} = -k S_k \beta$; see [27]. The MSEM in (16) may be written as

$$\text{MSEM}\{\hat{\beta}_{(2)}(\eta)\} = \sigma^2 (A_2 A_2^T) + d_2 d_2^T,$$

with $d_2 = \text{Bias}\{\hat{\beta}_{(2)}(\eta)\} = -(1 - \eta)(U^{-1} + I)^{-1} \beta$.

Define

$$\Delta_3 = \text{MSEM}\{\hat{\beta}_{(1)}(k)\} - \text{MSEM}\{\hat{\beta}_{(2)}(\eta)\} = \sigma^2 (A_1 A_1^T - A_2 A_2^T) + (d_1 d_1^T - d_2 d_2^T). \tag{19}$$

For the following proofs we employ the following lemma.

Lemma 5.1 (Trenkler and Toutenburg [28]). Let $\tilde{\beta}_{(j)} = A_j y, j = 1, 2$ be the two linear estimators of β . Suppose the difference $\text{Cov}(\tilde{\beta}_{(1)}) - \text{Cov}(\tilde{\beta}_{(2)})$ of the covariance matrices of the estimators $\tilde{\beta}_{(1)}$ and $\tilde{\beta}_{(2)}$ is positive definite. Then $\text{MSEM}(\tilde{\beta}_{(1)}) - \text{MSEM}(\tilde{\beta}_{(2)})$ is positive definite if and only if $d_2^T \{\text{Cov}(\tilde{\beta}_{(1)}) - \text{Cov}(\tilde{\beta}_{(2)}) + d_1 d_1^T\}^{-1} d_2 < 1$.

Theorem 5.1. The sampling variance of $\hat{\beta}_{(2)}(\eta)$ is smaller than that of $\hat{\beta}_{(1)}(k)$, if and only if $\lambda_{\min}(G_2^{-1} G_1) > 1$, where λ_{\min} is the minimum eigenvalue of $G_2^{-1} G_1$ and $G_j = \sigma^2 A_j A_j^T, j = 1, 2$.

Proof. Consider the difference

$$\begin{aligned} \Delta^* &= \text{Cov}\{\hat{\beta}_{(1)}(k)\} - \text{Cov}\{\hat{\beta}_{(2)}(\eta)\} \\ &= \sigma^2 (A_1 A_1^T - A_2 A_2^T), \\ &= G_1 - G_2 \end{aligned}$$

with $G_1 = (D^T \tilde{X} W_k U)^T = V^T V, W_k = I + kU$ and $G_2 = (\tilde{X} F_\eta^T U)^T (\tilde{X} F_\eta^T U)$. Since $\text{rank}(V) = p < n - m$, G_1 is a $(p \times p)$ positive definite matrix and G_2 is a symmetric matrix. Hence, a nonsingular matrix O exists such that $O^T G_1 O = I$ and $O^T G_2 O = \Lambda$, with Λ diagonal matrix with diagonal elements roots λ of the polynomial equation $|G_1 - \lambda G_2| = 0$ (see [16, p. 563] or [25, p. 160]). Thus, we may write $\Delta^* = (O^T)^{-1} (O^T G_1 O - O^T G_2 O) O^{-1} = (O^T)^{-1} (\Lambda - I) O^{-1}$ or $O^T \Delta^* O = \Lambda - I$. If $G_1 - G_2$ is positive definite, then $O^T G_1 O - O^T G_2 O = \Psi - I$ is positive definite. Hence $\lambda_i - 1 > 0, i = 1, 2, \dots, p$, so we get $\lambda_{\min}(G_2^{-1} G_1) > 1$.

Now let $\lambda_{\min}(G_2^{-1} G_1) > 1$ hold. Furthermore, with G_2 positive definite and G_1 symmetric, we have $\lambda_{\min} < \frac{v^T G_1 v}{v^T G_2 v} < \lambda_{\max}$ for all nonzero $(p \times 1)$ vectors v , so $G_1 - G_2$ is positive definite; see [23, p. 74]. It is obvious that $\text{Cov}\{\hat{\beta}_{(2)}(\eta)\} - \text{Cov}\{\hat{\beta}_{(1)}(k)\}$ is positive definite for $0 \leq \eta \leq 1, k \geq 0$ if and only if $\lambda_{\min}(G_2^{-1} G_1) > 1$. \square

Theorem 5.2. Consider $\hat{\beta}_{(1)}(k) = A_1 y$ and $\hat{\beta}_{(2)}(\eta) = A_2 y$ of β . Suppose that the difference $\text{Cov}\{\hat{\beta}_{(1)}(k)\} - \text{Cov}\{\hat{\beta}_{(2)}(\eta)\}$ is positive definite. Then

$$\Delta_3 = \text{MSEM}\{\hat{\beta}_{(1)}(k)\} - \text{MSEM}\{\hat{\beta}_{(2)}(\eta)\}$$

is positive definite if and only if

$$d_2^T \{\sigma^2 (A_1 A_1^T - A_2 A_2^T) + d_1 d_1^T\}^{-1} d_2 < 1$$

with $A_1 = S_k \tilde{X}^T D, A_2 = U F_\eta \tilde{X}^T D$.

Proof. The difference between the MSEMs of $\widehat{\beta}_{(2)}(\eta)$ and $\widehat{\beta}_{(1)}(k)$ is given by

$$\begin{aligned} \Delta_3 &= \text{MSEM}\{\widehat{\beta}_{(1)}(k)\} - \text{MSEM}\{\widehat{\beta}_{(2)}(\eta)\} \\ &= \sigma^2(A_1A_1^\top - A_2A_2^\top) + (d_1d_1^\top - d_2d_2^\top) \\ &= \text{Cov}\{\widehat{\beta}_{(1)}(k)\} - \text{Cov}\{\widehat{\beta}_{(2)}(\eta)\} + (d_1d_1^\top - d_2d_2^\top). \end{aligned}$$

Applying Lemma 5.1 yields the desired result. \square

It should be noted that all results reported above are based on the assumption that k and η are non-stochastic. The theoretical results indicate that the $\widehat{\beta}_{(2)}(\eta)$ is not always better than the $\widehat{\beta}_{(1)}(k)$, and vice versa. For practical purposes, we have to replace these unknown parameters by some suitable estimators.

6. The heteroscedasticity and correlated error case

Up to this point, independent errors with equal variance were assumed. The error term might also exhibit autocorrelation. To account for these effects, we extend the results in this section and consider the more general case of heteroscedasticity and autocovariance in the error terms.

Consider now observations $\{y_t, x_t, t_t\}_{t=1}^T$ and the semiparametric partial linear model $y_t = x_t^\top \beta + f(t_t) + \varepsilon_t$, $t = 1, \dots, T$. Let $E(\varepsilon\varepsilon^\top | x, t) = \Omega$ not necessarily diagonal. To keep the structure of the errors for later inference, we define an $(n \times n)$ permutation matrix P as in [32]. Consider a permutation:

$$\begin{pmatrix} 1 & t_{(1)} \\ \dots & \dots \\ i & t_{(i)} \\ \dots & \dots \\ n & t_{(n)} \end{pmatrix}$$

where $i = 1, \dots, n$ is the index of the ordered nonparametric variable and $t_{(i)} = 1, \dots, T$ corresponding time index of the observations. Then P is defined for $i, j = 1, \dots, n$:

$$P_{ij} = \begin{cases} 1, & j = t_{(i)} \\ 0, & \text{otherwise.} \end{cases}$$

We can now rewrite the model after reordering and differencing:

$$DPy = DPX\beta + DPf(x) + DP\varepsilon, \quad E(\varepsilon\varepsilon^\top | x, t) = \Omega. \tag{20}$$

Then, with $\widetilde{X} = DPX$ and $\widetilde{y} = DPy$ from (20), $\widehat{\beta}_{(0)}$ is given:

$$\widehat{\beta}_{(0)} = (\widetilde{X}^\top \widetilde{X})^{-1} \widetilde{X}^\top \widetilde{y} \tag{21}$$

with

$$\begin{aligned} \text{Cov}(\widehat{\beta}_{(0)}) &= (\widetilde{X}^\top \widetilde{X})^{-1} \widetilde{X}^\top DP\Omega D^\top P^\top \widetilde{X} (\widetilde{X}^\top \widetilde{X})^{-1} \\ &= U\widetilde{X}^\top DP\Omega D^\top P^\top \widetilde{X}U. \end{aligned} \tag{22}$$

We will use a heteroscedasticity and autocovariance consistent estimator described in [22] for the interior matrix of (22), which is in our case:

$$DP\widehat{\Omega}D^\top P^\top = \{\widehat{DP\varepsilon}(\widehat{DP\varepsilon})^\top\} \odot \left\{ \sum_{\ell=0}^{\mathcal{L}} \left(1 - \frac{\ell}{\mathcal{L} + 1}\right) H^\ell \right\} \tag{23}$$

with $\widehat{DP\varepsilon} = \widetilde{y} - \widetilde{X}\widehat{\beta}_{(0)}$, \odot denoting the elementwise matrix product, \mathcal{L} the maximum lag of nonzero autocorrelation in the errors and H^0 the identity matrix. Let L_ℓ be a matrix with ones on the ℓ th diagonal; then H^ℓ , $\ell = 1, \dots, \mathcal{L}$ are such that:

$$H_{ij}^\ell = \begin{cases} 0, & \text{if } \{DP(L_\ell + L_\ell^\top)D^\top P^\top\}_{ij} = 0, \\ 1, & \text{otherwise and } i, j = 1, \dots, p. \end{cases}$$

Plugging (23) in (22), we obtain a consistent estimator for $\text{Cov}(\widehat{\beta}_{(0)})$; see [31] for details.

Denoting $\widetilde{S} = \widetilde{X}^\top DP\Omega D^\top P^\top \widetilde{X}$, we can write down $\text{Cov}\{\widehat{\beta}_{(1)}(k)\}$ and $\text{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ in model (20).

$$\text{Cov}\{\widehat{\beta}_{(1)}(k)\} = S_k \widetilde{S} S_k \tag{24}$$

$$\text{Cov}\{\widehat{\beta}_{(2)}(\eta)\} = F_\eta U \widetilde{S} U F_\eta. \tag{25}$$

Using (22) and (25), the difference $\Delta_1 = \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ can be expressed as

$$\begin{aligned} \Delta_1 &= (U\widetilde{S}U - F_\eta U\widetilde{S}U F_\eta^\top) \\ &= F_\eta \{F_\eta^{-1} U\widetilde{S}U (F_\eta^\top)^{-1} - U\widetilde{S}U\} F_\eta^\top \\ &= (1 - \eta^2)(U^{-1} + I)^{-1} \left\{ \frac{1}{1 + \eta} (U\widetilde{S} + \widetilde{S}U) + U\widetilde{S}U \right\} (U^{-1} + I)^{-1}, \end{aligned} \tag{26}$$

with $\tau = \frac{1}{1+\eta} > 0$, $\widetilde{M} = U\widetilde{S}U$, $\widetilde{N} = U\widetilde{S} + \widetilde{S}U$. Since \widetilde{M} is a $(p \times p)$ positive definite matrix and \widetilde{N} is a symmetric matrix, a nonsingular matrix T exists such that $T^\top \widetilde{M} T = I$ and $T^\top \widetilde{N} T = \widetilde{E}$; here \widetilde{E} is a diagonal matrix and its diagonal elements are the roots of the polynomial equation $|\widetilde{M}^{-1} \widetilde{N} - \widetilde{e}I| = 0$ (see [11, pp. 408] and [16, pp. 563]) and we may write (26) as

$$\begin{aligned} \Delta_1 &= (1 - \eta^2) H (\widetilde{M} + \tau \widetilde{N}) H \\ &= (1 - \eta^2) H (T^\top)^{-1} (T^\top \widetilde{M} T + \tau T^\top \widetilde{N} T) T^{-1} H \\ &= (1 - \eta^2) H (T^\top)^{-1} (I + \tau \widetilde{E}) T^{-1} H, \end{aligned}$$

where $I + \tau \widetilde{E} = \text{diag}(1 + \tau \widetilde{e}_{11}, \dots, 1 + \tau \widetilde{e}_{pp})$ and $H = (U^{-1} + I)^{-1}$. Since $\widetilde{N} = U\widetilde{S} + \widetilde{S}U \neq 0$, there is at least one diagonal element of \widetilde{E} that is nonzero.

Let $\widetilde{e}_{ii} < 0$ for at least one i ; then positive definiteness of $I + \tau \widetilde{E}$ is guaranteed by

$$0 < \tau < \min_{\widetilde{e}_{ii} < 0} \left| \frac{1}{\widetilde{e}_{ii}} \right|. \tag{27}$$

Hence $1 + \tau \widetilde{e}_{ii} > 0$ for all $i = 1, \dots, p$ and therefore $I + \tau \widetilde{E}$ is a positive definite matrix. Consequently, Δ_1 becomes a positive definite matrix, as well. It is now evident that the estimator $\widehat{\beta}_{(2)}(\eta)$ has a smaller variance compared with the estimator $\widehat{\beta}_{(0)}$ if and only if (27) is satisfied.

With

$$\begin{aligned} \Delta'_1 &= \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(1)}(k)\} \\ &= k^2 S_k \left\{ \frac{1}{k} (U\widetilde{S} + \widetilde{S}U) + U\widetilde{S}U \right\} S_k \\ &= k^2 S_k \left(\frac{1}{k} \widetilde{N} + \widetilde{M} \right) S_k \end{aligned}$$

and analogous argumentation as above obtained for $\widehat{\beta}_{(1)}(k)$:

$$0 < \frac{1}{k} < \min_{\widetilde{e}_{ii} < 0} \left| \frac{1}{\widetilde{e}_{ii}} \right|. \tag{28}$$

The next theorem extends the results of Theorem 3.1 in [27] and Theorem 4.1 of Section 4 to the more general case of (20).

Theorem 6.1. Consider the estimators $\widehat{\beta}_{(i)}(x)$, $i = \{1, 2\}$; $x = \{k, \eta\}$ and $\widehat{\beta}_{(0)}$ of β . Let $W_1 = \widetilde{M} + \tau \widetilde{N}$, $W_2 = \frac{1+\eta}{1-\eta} (\widetilde{M} + \tau \widetilde{N})$ be positive definite (alternative: assume that (27) and (28) hold). Then the biased estimator $\widehat{\beta}_{(i)}(x)$ is MSEM superior to $\widehat{\beta}_{(0)}$ if and only if

$$\beta^\top W_i^{-1} \beta \leq 1.$$

Proof. Consider the differences

$$\begin{aligned} \Delta_2 &= \text{MSEM}(\widehat{\beta}_{(0)}) - \text{MSEM}\{\widehat{\beta}_{(2)}(\eta)\} \\ &= \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(2)}(\eta)\} - \text{Bias}\{\widehat{\beta}_{(2)}(\eta)\} \text{Bias}\{\widehat{\beta}_{(2)}(\eta)\}^\top \\ &= F_\eta \{F_\eta^{-1} U\widetilde{S}U (F_\eta^\top)^{-1} - U\widetilde{S}U\} F_\eta^\top - (1 - \eta)^2 (U^{-1} + I)^{-1} \beta \beta^\top (U^{-1} + I)^{-1} \\ &= (1 - \eta)^2 H \left\{ \frac{1 + \eta}{1 - \eta} (\widetilde{M} + \tau \widetilde{N}) - \beta \beta^\top \right\} H \\ &= (1 - \eta)^2 H (W_2 - \beta \beta^\top) H. \end{aligned}$$

$$\begin{aligned} \Delta'_2 &= \text{MSEM}(\widehat{\beta}_{(0)}) - \text{MSEM}\{\widehat{\beta}_{(1)}(k)\} \\ &= \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(1)}(k)\} - \text{Bias}\{\widehat{\beta}_{(1)}(k)\} \text{Bias}\{\widehat{\beta}_{(1)}(k)\}^\top \\ &= S_k\{k(\widetilde{S}U + U\widetilde{S}) + k^2U\widetilde{S}U - k^2\beta\beta^\top\}S_k \\ &= k^2S_k\left(\frac{1}{k}\widetilde{N} + \widetilde{M} - \beta\beta^\top\right)S_k \\ &= k^2S_k(W_1 - \beta\beta^\top)S_k. \end{aligned}$$

With Lemma 4.1, the assertion follows. \square

Theorem 6.1 gives conditions under which the biased estimator $\widehat{\beta}_{(i)}(x)$, $i = \{1, 2\}$; $x = \{k, \eta\}$ is superior to $\widehat{\beta}_{(0)}$ in the presence of heteroscedasticity and autocorrelation in the data.

Note that for comparison of the biased estimators Theorem 5.1 can be extended straight forwardly to the general case by exchanging G_1 and G_2 by $\widetilde{G}_1 = \widetilde{A}_1\widetilde{\Omega}\widetilde{A}_1^\top$ and $\widetilde{G}_2 = \widetilde{A}_2\widetilde{\Omega}\widetilde{A}_2^\top$ correspondingly, with $\widetilde{A}_1 = S_k\widetilde{X}^\top DP$, $\widetilde{A}_2 = UF_\eta\widetilde{X}^\top DP$. Hence, the sampling variance of $\widehat{\beta}_{(2)}(\eta)$ is always smaller than that of $\widehat{\beta}_{(1)}(k)$, if and only if $\lambda_{\min}(\widetilde{G}_2^{-1}\widetilde{G}_1) > 1$, where λ_{\min} is the minimum eigenvalue of $\widetilde{G}_2^{-1}\widetilde{G}_1$.

Now, we give a generalized version of Theorem 5.2.

Theorem 6.2. Consider $\widehat{\beta}_{(1)} = \widetilde{A}_1y$ and $\widehat{\beta}_{(2)} = \widetilde{A}_2y$ of β . Suppose that the difference $\text{Cov}\{\widehat{\beta}_{(1)}\} - \text{Cov}\{\widehat{\beta}_{(2)}\}$ is positive definite. Then

$$\Delta_3 = \text{MSEM}(\widehat{\beta}_{(1)}) - \text{MSEM}(\widehat{\beta}_{(2)})$$

is positive definite if and only if

$$d_2^\top(\widetilde{A}_1\widetilde{\Omega}\widetilde{A}_1^\top - \widetilde{A}_2\widetilde{\Omega}\widetilde{A}_2^\top + d_1d_1^\top)^{-1}d_2 < 1.$$

Proof. The difference between the MSEMs of $\widehat{\beta}_{(2)}(\eta)$ and $\widehat{\beta}_{(1)}(k)$ is given by

$$\begin{aligned} \Delta_3 &= \text{MSEM}(\widehat{\beta}_{(1)}) - \text{MSEM}(\widehat{\beta}_{(2)}) \\ &= \widetilde{A}_1\widetilde{\Omega}\widetilde{A}_1^\top - \widetilde{A}_2\widetilde{\Omega}\widetilde{A}_2^\top + d_1d_1^\top - d_2d_2^\top \\ &= \text{Cov}(\widehat{\beta}_{(1)}) - \text{Cov}(\widehat{\beta}_{(2)}) + d_1d_1^\top - d_2d_2^\top. \end{aligned}$$

Applying Lemma 5.1 yields the desired result. \square

We note that in order to use the criteria above, one has to estimate the parameters. The estimation of Ω is thereby the most challenging. However, as long as estimator (23) is available, all considered criteria can be evaluated on the real data and can be used for practical purposes.

7. Determinants of electricity demand

The empirical study example is motivated by the importance of explaining variation in electricity consumption. Since electricity is a non-storable good, electricity providers are interested in understanding and hedging demand fluctuations.

Electricity consumption is known to be influenced negatively by the price of electricity and positively by the income of the consumers. As electricity is frequently used for heating and cooling, the effect of the air temperature must also be present. Both heating by low temperatures and cooling by high temperatures result in higher electricity consumption and motivate the use of a nonparametric specification for the temperature effect. Thus we consider the semiparametric regression model defined in (1)

$$y = f(t) + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_{13}x_{13} + \varepsilon, \tag{29}$$

where y is the log monthly electricity consumption per person (aggregated electricity consumption was divided by population interpolated linearly from quarterly data), t is cumulated average temperature index for the corresponding month taken as average of 20 German cities computed from the data of German weather service (Deutscher Wetterdienst), x_1 is the log GDP per person interpolated linearly from quarterly data, detrended and deseasonalized and x_2 is the log rate of electricity price to the gas price, detrended. The data for 199601-201009 comes from EUROSTAT. Reference prices for electricity were computed as an average of electricity tariffs for consumer groups IND-Ie and HH-Dc, for gas—IND-I3-2 and HH-D3 with reference period 2005S1. Time series of prices were obtained by scaling with electricity price or correspondingly gas price indices. x_3, x_4, \dots, x_{13} are dummy variables for the monthly effects.

The model in (29) includes both parametric effects and a nonparametric effect. The only nonparametric effect is implied by the temperature variable. From Fig. 1, we can see that the effect of t on y is likely to be nonlinear, while the effects of other variables are roughly linear. The dummy variables enter into the linear part in the specification of the semiparametric regression as well.

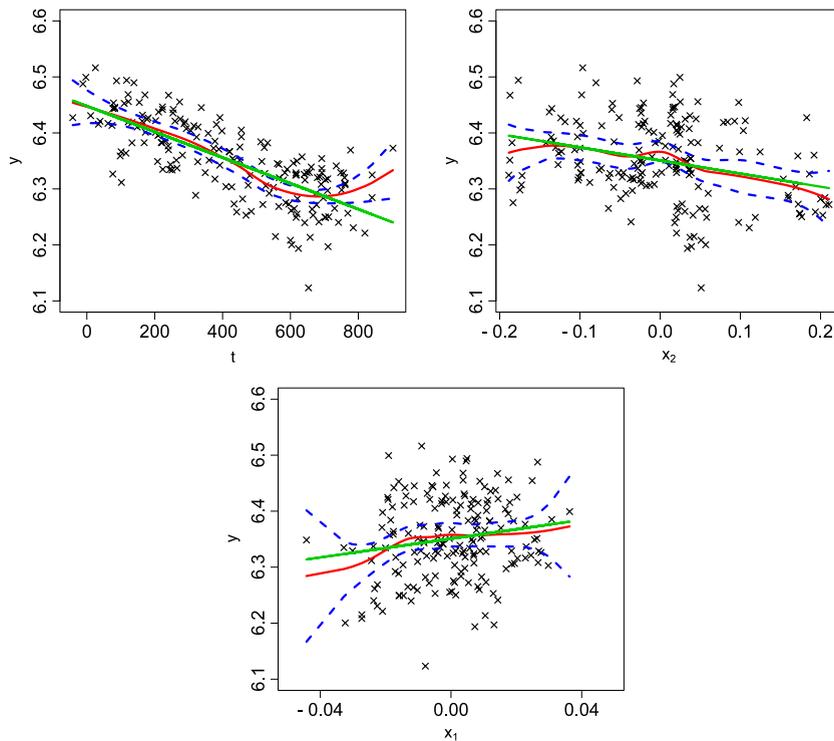


Fig. 1. Plots of individual exp. variables vs. dependent variable, linear fit (green), local polynomial fit (red), 95% confidence bands (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We note that the condition number of $X^T X$ of these explanatory variables is 20.5, which justifies the use of $\hat{\beta}_{(1)}(k)$ and $\hat{\beta}_{(2)}(\eta)$; see [2].

Throughout the paper, we use fifth-order differencing ($m = 5$). Results for other orders of differencing were similar.

The admissible regions for the biasing parameters η and k for MSEM superiority were $\eta \geq 0.923$ and $k \leq 0.0085$. These bounds were determined using the estimated parameters and the inequalities from Theorem 4.1 and Theorem 3.1 in [27], respectively. Under more general assumptions on Ω and resulting heteroscedasticity and autocovariance consistent Newey–West covariance estimator, defined in (23), the admissible region for η (Theorem 6.1 and restriction (27)) was shrunk to $\eta \geq 0.927$. For $\hat{\beta}_{(1)}(k)$, no admissible values of k were found, since admissible $k \geq 1.57$ of (28) do not satisfy the condition of Theorem 6.1 (see Table 2).

Alternatively, we used a scalar mean squared error (SMSE), defined as the trace of the corresponding MSEM, to compare the estimators. The bounds for k and η can then be calculated only numerically using a grid on $[0, 1]$ for the biasing parameters and determining the regions where SMSEs of the proposed estimators are lower. SMSE superiority of $\hat{\beta}_{(1)}(k)$ and $\hat{\beta}_{(2)}(\eta)$ over $\hat{\beta}_{(0)}$ under general Ω is given for $k \leq 0.0267$ and $\eta \geq 0.384$ compared to $k \leq 0.0123$ and $\eta \geq 0.708$ by standard assumptions; see Fig. 2 which depicts SMSE of the estimators and the corresponding η and k under standard and general assumptions. Thus the SMSE superiority intervals for η and k become even larger in the case of the general form of Ω .

Our computations here are performed with R 2.10.1 and the codes are available on www.quantlet.org.

Results of different estimation procedures can be found in Table 1. We note that regardless of the estimator type, the effect of income is positive and the effect of relative price is negative as expected from an economic perspective, as in [4]. However, the R^2 obtained by difference based methods is higher and SMSE lower for Liu type and ridge difference based estimator. The values of biasing parameters for which conditions of Theorems 5.1 and 5.2 are satisfied are given in Table 3. The superiority of $\hat{\beta}_{(2)}(\eta)$ over $\hat{\beta}_{(1)}(k)$ is assured for the zone of values marked by plus.

Returning to our semiparametric specification, we may now remove the estimated parametric effect from the dependent variable and analyze the nonparametric effect. We use a local linear estimator of f to model the nonparametric effect of temperature. The resulting plots are presented in Fig. 3 where we also include the linear effect. We notice that all differencing procedures result in similar estimators of f , regardless of notable differences in the coefficients of the linear part. The estimator of f is consistent with findings e.g. of [4] for US electricity data.

In both specifications, f is different from the linear effect and therefore including temperature as a linear effect is misleading.

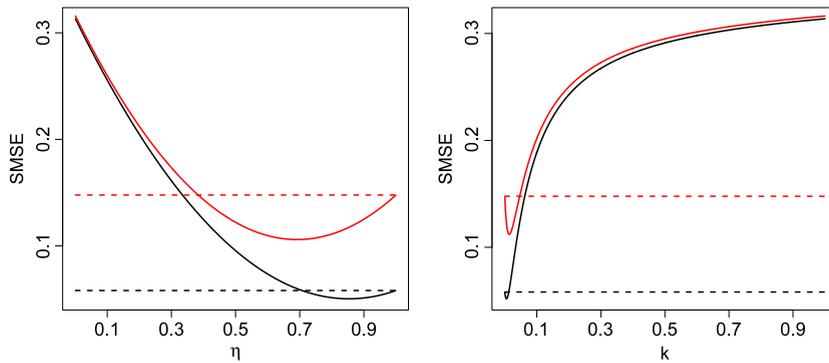


Fig. 2. SMSE of $\hat{\beta}_{(2)}(\eta)$ in dependence of η (left) and $\hat{\beta}_{(1)}(k)$ in dependence of k (right) against that of $\hat{\beta}_{(0)}$ (dashed) under standard assumptions (black) and under generalized assumptions (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results of OLS, difference based and Liu type difference based estimations.

	$\hat{\beta}_{OLS}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{(1)}(10^{-3})$	$\hat{\beta}_{(2)}(0.95)$
x_1	0.634	0.578*	0.550*	0.562*
x_2	-0.152***	-0.160***	-0.158***	-0.161***
x_3	0.030***	0.030*	0.030*	0.030*
x_4	-0.043***	-0.040**	-0.040**	-0.040**
x_5	0.011	0.031	0.031	0.031
x_6	-0.051**	-0.014	-0.013	-0.014
x_7	-0.054*	-0.014	-0.013	-0.014
x_8	-0.079**	-0.065	-0.064	-0.065
x_9	-0.036	-0.037	-0.036	-0.037
x_{10}	-0.052	-0.044	-0.043	-0.044
x_{11}	-0.049	-0.013	-0.012	-0.013
x_{12}	-0.000	0.040	0.040	0.040
x_{13}	-0.001	0.016	0.016	0.016
t	$-13 \cdot 10^{-5}$ ***	-	-	-
R^2	0.729	0.749	0.749	0.749

* Indicates significance on 10%.
 ** Indicates significance on 5%.
 *** Indicates significance on 1%.

Table 2
Standard errors of the estimators in comparison to Newey–West standard errors for the effects of x_1 (income) and x_2 (relative price).

$\hat{\Omega}$	$\hat{\beta}_{(0)}$ $\hat{\sigma}^2 I$	$\hat{\Omega}_{NW}$	$\hat{\beta}_{(1)}(10^{-3})$ $\hat{\sigma}^2 I$	$\hat{\Omega}_{NW}$	$\hat{\beta}_{(2)}(0.95)$ $\hat{\sigma}^2 I$	$\hat{\Omega}_{NW}$
x_1	0.215	0.347	0.209	0.337	0.205	0.215
x_2	0.034	0.047	0.034	0.047	0.034	0.034
SMSE	0.058	0.148	0.056	0.141	0.054	0.058

8. Conclusion

We proposed a difference based Liu type estimator and a difference based ridge regression estimator for the partial linear semiparametric regression model.

The results show that in case of multicollinearity, the proposed estimator, $\hat{\beta}_{(2)}(\eta)$ is superior to the difference based estimator $\hat{\beta}_{(0)}$. We gave bounds on the value of η which ensure the superiority of the proposed estimator. The two biased estimators $\hat{\beta}_{(2)}(\eta)$ and $\hat{\beta}_{(1)}(k)$ for different values of η and k can be compared in terms of MSEM with the theoretical results above.

Finally, an application to electricity consumption has been provided to show properties of the proposed estimator based on the mean square error criterion. We could estimate the linear effects of the linear determinants as well as the nonparametric effect f of a cumulated average temperature index.

Thus, the theoretical results obtained allow us to tackle the problem of multicollinearity in real applications of semiparametric models. Moreover, we are able to get estimators of the linear effects with lower standard errors by tuning parameters k and η accordingly.

Table 3

Admissible biasing parameters η and k marked by plus if they satisfy conditions of Theorems 5.1 and 5.2, i.e. $\widehat{\beta}_{(2)}(\eta)$ is superior to $\widehat{\beta}_{(1)}(k)$.

$\eta \cdot 10^2$	$k \cdot 10^4$												
	1	2	3	4	5	6	7	8	9	10	11	12	13
9.23–9.23	–	–	–	–	–	–	–	–	–	–	–	–	–
9.24–9.24	+	–	–	–	–	–	–	–	–	–	–	–	–
9.25–9.25	+	+	–	–	–	–	–	–	–	–	–	–	–
9.26–9.26	+	+	+	–	–	–	–	–	–	–	–	–	–
9.27–9.27	+	+	+	+	–	–	–	–	–	–	–	–	–
9.28–9.28	+	+	+	+	+	–	–	–	–	–	–	–	–
9.29–9.30	+	+	+	+	+	+	–	–	–	–	–	–	–
9.31–9.31	+	+	+	+	+	+	+	–	–	–	–	–	–
9.32–9.32	+	+	+	+	+	+	+	+	–	–	–	–	–
9.34–9.35	+	+	+	+	+	+	+	+	+	–	–	–	–
9.36–9.37	+	+	+	+	+	+	+	+	+	+	–	–	–
9.38–9.39	+	+	+	+	+	+	+	+	+	+	+	–	–
9.40–9.43	+	+	+	+	+	+	+	+	+	+	+	+	–
9.44–9.56	+	+	+	+	+	+	+	+	+	+	+	+	+
9.57–9.61	+	+	+	+	+	+	+	+	+	+	+	+	–
9.62–9.65	+	+	+	+	+	+	+	+	+	+	+	+	–
9.66–9.69	+	+	+	+	+	+	+	+	+	+	+	–	–
9.70–9.72	+	+	+	+	+	+	+	+	+	+	–	–	–
9.73–9.76	+	+	+	+	+	+	+	+	+	–	–	–	–
9.77–9.79	+	+	+	+	+	+	+	–	–	–	–	–	–
9.80–9.82	+	+	+	+	+	+	–	–	–	–	–	–	–
9.83–9.85	+	+	+	+	+	–	–	–	–	–	–	–	–
9.86–9.88	+	+	+	+	–	–	–	–	–	–	–	–	–
9.89–9.91	+	+	+	–	–	–	–	–	–	–	–	–	–
9.92–9.94	+	+	–	–	–	–	–	–	–	–	–	–	–
9.95–9.97	+	–	–	–	–	–	–	–	–	–	–	–	–
9.98–9.99	–	–	–	–	–	–	–	–	–	–	–	–	–

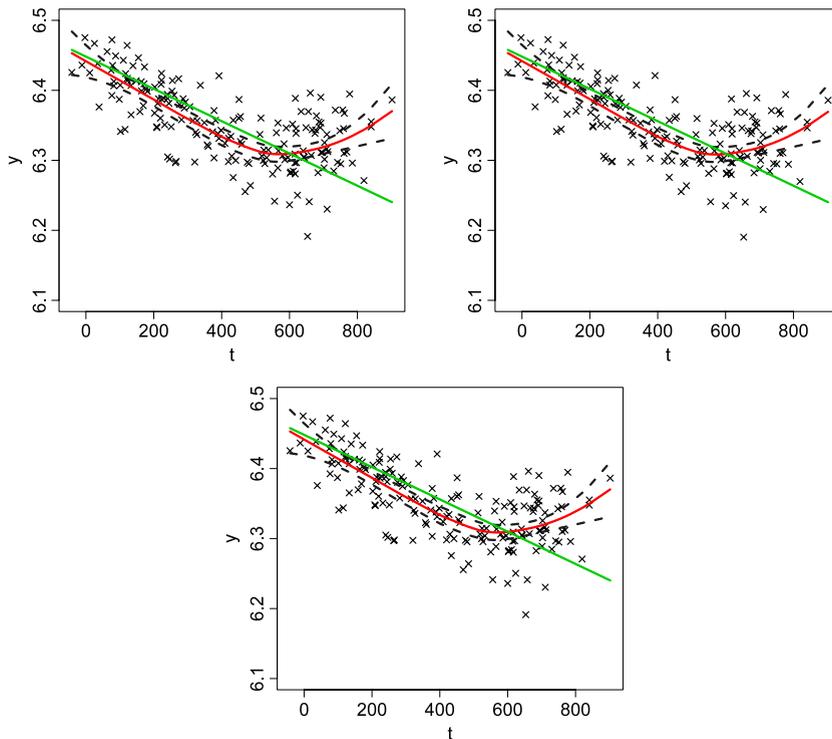


Fig. 3. Estimated f nonlinear effect of t on y via differenced based (left), Liu-type differenced based (right) and difference-based ridge (center) approaches.

References

[1] F. Akdeniz, S. Kaçiranlar, On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE, *Communications in Statistics Theory and Methods* 24 (7) (1995) 1789–1797.
 [2] D. Belsley, E. Kuh, R. Welsch, *Regression Diagnostics*, Wiley, New York, 1980.

- [3] L. Brown, M. Levine, Variance estimation in nonparametric regression via the difference sequence method, *Annals of Statistics* 35 (2007) 2219–2232.
- [4] R.F. Engle, C. Granger, J. Rice, A. Weiss, Semiparametric estimates of the relation between weather and electricity sales, *Journal of American Statistical Association* 81 (1986) 310–320.
- [5] R. Eubank, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York, 1999.
- [6] R. Eubank, E. Kambour, J. Kim, K. Klipple, C. Reese, M. Schimek, Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society Series B* 50 (3) (1988) 413–436.
- [7] R. Eubank, E. Kambour, J. Kim, K. Klipple, C. Reese, M. Schimek, Estimation in partially linear models, *Computational Statistics and Data Analysis* 29 (1998) 27–34.
- [8] J. Fan, L. Huang, Goodness-of-fit tests for parametric regression models, *Journal of American Statistical Association* 96 (2001) 640–652.
- [9] J. Fan, Y. Wu, Semiparametric estimation of covariance matrices for longitudinal data, *Journal of American Statistical Association* 103 (2008) 1520–1533.
- [10] R. Farebrother, Further results on the mean square error of ridge regression, *Journal of the Royal Statistical Society Series B* 38 (1976) 248–250.
- [11] F. Graybill, *Matrices with Applications in Statistics*, Duxbury Classic, 1983.
- [12] P. Green, C. Jennison, A. Seheult, Analysis of field experiments by least squares smoothing, *Journal of the Royal Statistical Society Series B* 47 (1985) 299–315.
- [13] M. Gruber, *Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators*, Marcell Dekker, Inc., New York, 1985.
- [14] W. Härdle, H. Liang, J. Gao, *Partially Linear Models*, Physika Verlag, Heidelberg, 2000.
- [15] W. Härdle, M. Müller, S. Sperlich, A. Werwatz, *Nonparametric and Semiparametric Models*, Springer Verlag, Heidelberg, 2004.
- [16] D. Harville, *Matrix Algebra from a Statistician's Perspective*, Springer Verlag, New York, 1997.
- [17] A. Hoerl, R. Kennard, Ridge regression: biased estimation for orthogonal problems, *Technometrics* 12 (1970) 55–67.
- [18] M. Hubert, P. Wijekoon, Improvement of the Liu estimation in linear regression model, *Statistical Papers* 47 (3) (2006) 471–479.
- [19] K. Klipple, R. Eubank, Difference-based variance estimators for partially linear models, *Festschrift in honor of Distinguished Professor Mir Masoom Ali on the occasion of his retirement*, 2007, pp. 313–323.
- [20] K. Liu, A new class of biased estimate in linear regression, *Communications in Statistics Theory and Methods* 22 (1993) 393–402.
- [21] K. Liu, Using Liu type estimator to combat multicollinearity, *Communications in Statistics Theory and Methods* 32 (5) (2003) 1009–1020.
- [22] W. Newey, K. West, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55 (3) (1987) 703–708.
- [23] C. Rao, *Linear Statistical Inference and its Applications*, Wiley, New York, 1973.
- [24] D. Ruppert, M. Wand, R. Carroll, R. Gill, *Semiparametric Regression*, Cambridge University Press, 2003.
- [25] J. Schott, *Matrix Analysis for Statistics*, second ed., Wiley Inc., New Jersey, 2005.
- [26] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: *Proc. Third Berkeley Symp. Math. Statist. Prob.*, vol. 1, 1956, pp. 197–206.
- [27] G. Tabakan, F. Akdeniz, Difference-based ridge estimator of parameters in partial linear model, *Statistical Papers* 51 (2010) 357–368.
- [28] G. Trenkler, H. Toutenburg, Mean square matrix comparisons between two biased estimators—an overview of recent results, *Statistical Papers* 31 (1990) 165–179.
- [29] H. Yang, J. Xu, An alternative stochastic restricted Liu estimator in linear regression, *Statistical Papers* 50 (2009) 639–647.
- [30] A. Yatchew, An elementary estimator of the partial linear model, *Economics Letters* 57 (1997) 135–143.
- [31] A. Yatchew, *Differencing methods in nonparametric regression: simple techniques for the applied econometrician*, 1999. <http://www.economics.utoronto.ca/yatchew/>.
- [32] A. Yatchew, *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, 2003.
- [33] J. You, G. Chen, Y. Zhou, Statistical inference of partially linear regression models with heteroscedastic errors, *Journal of Multivariate Analysis* 98 (8) (2007) 1539–1557.



Modelling and forecasting liquidity supply using semiparametric factor dynamics[☆]

Wolfgang Karl Härdle, Nikolaus Hautsch, Andrija Mihoci^{*}

Humboldt Universität zu Berlin and C.A.S.E. – Center for Applied Statistics and Economics, Spandauer Str. 1, D-10178 Berlin, Germany
Center for Financial Studies (CFS), Frankfurt, Germany

ARTICLE INFO

Article history:

Received 21 April 2010
Received in revised form 21 March 2012
Accepted 2 April 2012
Available online 10 April 2012

JEL classification:

C14
C32
C53
G11

Keywords:

Limit order book
Liquidity risk
Semiparametric model
Factor structure
Prediction

ABSTRACT

We model the dynamics of ask and bid curves in a limit order book market using a dynamic semiparametric factor model. The shape of the curves is captured by a factor structure which is estimated nonparametrically. Corresponding factor loadings are modelled jointly with best bid and best ask quotes using a vector error correction specification. Applying the framework to four stocks traded at the Australian Stock Exchange (ASX) in 2002, we show that the suggested model captures the spatial and temporal dependencies of the limit order book. We find spillover effects between both sides of the market and provide evidence for short-term quote predictability. Relating the shape of the curves to variables reflecting the current state of the market, we show that the recent liquidity demand has the strongest impact. In an extensive forecasting analysis we show that the model is successful in forecasting the liquidity supply over various time horizons during a trading day. Moreover, it is shown that the model's forecasting power can be used to improve optimal order execution strategies.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Due to technological progress in the organization of trading systems and exchanges, electronic limit order book trading has become the dominant trading form for equities. Open limit order books provide important information on the current liquidity supply as reflected by the offered price-quantity relationships on both sides of the market. These supply and demand schedules provide valuable information on traders' price expectations in the spirit of the seminal paper by [Glosten \(1994\)](#), reflect the current implied costs of trading as well as demand and supply elasticities. However, while the dynamic behavior of liquidity demand, as reflected by trading intensities and trade sizes, has been already studied in various papers (see, e.g., [Hautsch and Huang \(2012\)](#) and [Brownlees et al. \(2009\)](#)), the stochastic properties of liquidity supply is still widely unknown. An obvious reason is that liquidity supply is reflected by high-dimensional bid and ask schedules which are not straightforwardly modelled in a dynamic setting. Consequently, it is a widely open question whether and to which extent liquidity supply might be predictable.

[☆] This work was supported by the Deutsche Forschungsgemeinschaft via Collaborative Research Center 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin, Germany.

^{*} Corresponding author. Tel.: +49 30 2093 5623.

E-mail address: mihoci@cms.hu-berlin.de (A. Mihoci).

The paper's major idea is to capture the shape of high-dimensional ask and bid curves by a lower-dimensional factor structure which is estimated non-parametrically. We propose a dynamic semiparametric factor model where the shape of order schedules is captured by a non-parametric factor structure while the curves' dynamic behavior is driven by time-varying factor loadings. The latter are modelled parametrically employing a vector error correction model (VECM). We show that the model captures the dynamics of high-dimensional order curves very well and is sufficiently parsimonious to produce valuable out-of-sample predictions. Moreover, the schedule of market depth posted around best quotes reveals strong serial dependence and thus is predictable. This structure is induced by the inventory character of order volume which is strongly persistent over time.

By providing empirical evidence on the dynamics and predictability of order book schedules, this paper fills a gap in empirical literature and complements recent (mostly theoretical) work on order splitting and dynamic order submission strategies. For instance, the question of how to reduce the costs of trading by optimally splitting a large order over time (e.g., over the course of a trading day) is of high relevance in financial practice. [Obizhaeva and Wang \(2005\)](#) and [Engle and Ferstenberg \(2007\)](#) analyze optimal splitting strategies whose implementations ultimately require predictions of future liquidity demand and supply. [Bertsimas and Lo \(1998\)](#) and [Almgren and Chriss \(2000\)](#) derive optimal execution strategies by minimizing expected costs of executing, an order in the context of static price impact functions. Optimal execution in a limit order book market is analyzed by [Alfonsi et al. \(2010\)](#). They allow for general shapes of order book curves and derive explicit optimal execution strategies in discrete time. By providing insights into the actual form of order book curves and their dynamic behavior, our results can be used as valuable inputs in theoretical frameworks.

While to the best of our knowledge our study is the first which models the shapes and dynamics of a complete (high-dimensional) order book, there is a substantial body of empirical literature on the dynamics of limit order books and the analysis of traders' order submission strategies, such as, e.g., [Biais et al. \(1995\)](#), [Griffiths et al. \(2000\)](#), [Ahn et al. \(2001\)](#), [Rinaldo \(2004\)](#), [Hollifield et al. \(2004\)](#), [Bloomfield et al. \(2005\)](#), [Degryse et al. \(2005\)](#), [Hall and Hautsch \(2006, 2007\)](#), [Large \(2007\)](#), [Hasbrouck and Saar \(2009\)](#) or [Cao et al. \(2009\)](#).

An important aspect in this literature is to analyze the question of how to optimally balance risks and gains of a trader's decision whether to post a market order or a limit order. As recently illustrated by [Chacko et al. \(2008\)](#), a limit order can be ultimately seen as an American option and transaction costs are rents that a monopolistic market maker extracts from impatient investors who trade via aggressive limit orders or market orders. Consequently, the analysis of liquidity risks (see, e.g., [Johnson, \(2008\)](#), [Liu \(2009\)](#), [Garvey and Wu \(2009\)](#), [Goyenko et al. \(2009\)](#)) and transaction costs (see, e.g. [Chacko et al. \(2008\)](#), [Hasbrouck \(2009\)](#)) are in the central focus of recent literature.

Given the objective to capture not only the volume around the best quotes but also pending quantities more deeply in the book, the underlying problem becomes inherently high-dimensional. A typical graphical snapshot of ask and bid curves for four stocks traded at the Australian Securities Exchange (ASX) in 2002, is given by [Fig. 1](#). The curse of dimensionality applies immediately as soon as time variations of the order curve shapes have to be taken into account. As shown by [Fig. 1](#) and as illustrated in more detail in the sequel of the paper, order volume is not necessarily only concentrated around the best quotes but can be substantially dispersed over a wider range of price levels. This is a typical scenario for moderately liquid markets as that of the ASX. In such a context, the dynamic modelling of all volume levels individually becomes complicate and intractable.

We suggest reducing the high dimensionality of the order book by means of a factor decomposition using the so-called Dynamic Semiparametric Factor Model (DSFM) proposed by [Fengler et al. \(2007\)](#), [Brüggemann et al. \(2008\)](#), [Park et al. \(2009\)](#) and [Cao et al. \(2009\)](#). Accordingly, we model the shape of the book in terms of underlying latent factors which are defined on a grid space around the best ask or bid quotes and can depend on additional explanatory variables capturing, e.g., the state of the market. In order to avoid specific functional forms for the shape of the curves, the factors as well as the corresponding loadings are estimated nonparametrically using B-splines. Then, in a second step, we model the multivariate dynamics of the factor loadings together with the best bid and the best ask price using a VEC specification.

Using this framework we aim answering the following research questions: (i) How many factors are required to model order book curves reasonably well? (ii) What does the shape of the factors look like? (iii) What do the dynamics of the estimated factor loadings look like? (iv) Does there exist evidence for a strong cross-dependence between both sides of the order book? (v) Can quotes be predictable in the short run? (vi) Does the shape of the order book curves depend on past price movements, past trading volume as well past volatility? (vii) How successful is the model in predicting future liquidity supply and can it be used to improve order execution strategies?

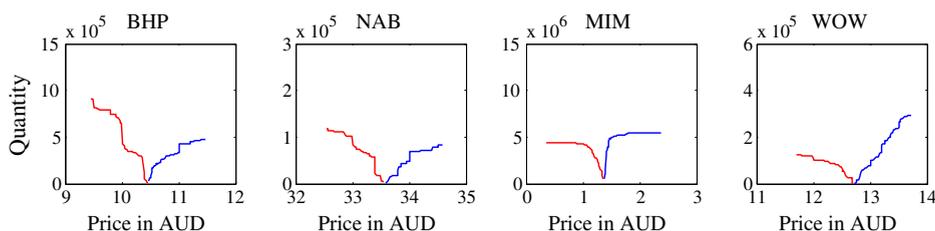


Fig. 1. Limit order books for selected stocks traded at the ASX on July 8, 2002 at 10:15. Red: bid curve, blue: ask curve.

Using limit order book data from four stocks traded at the ASX covering two months in 2002, we show that approximately 95% of the order book variations observed on 5-min intervals can be explained by two underlying time-varying factors. While the first factor captures the overall slope of the curves, the second one is associated with its curvature. Knowing the shape of the order book can help us to predict quotes in the short run. Further empirical results show relatively weak spill-over effects between the bid and the ask side of the market. It turns out that recent liquidity demand represented by the cumulative buy/sell trading observed over the past 5 min has an effect on the shape of the curve but does not induce a higher explanatory power. Similar evidence is shown for the impact of past returns and corresponding (realized) volatility. Moreover, we find that factor loadings follow highly persistent though stationary dynamics.

To evaluate the model's forecasting power, we perform an extensive out-of-sample forecasting analysis which is in line with a typical scenario in financial practice. In particular, at every 5-min interval during a trading day, the model is re-estimated and used to produce forecasts for the pending volume on each price level for all future 5-min intervals during the remainder of the trading day. We show that our approach is able to outperform a naive prediction, where the current order book is used as a predictor for the remaining day. These results can be used to improve intra-day order execution strategies by reducing implied transaction costs.

The remainder of the paper is structured as follows: After the data description in Section 2, the Dynamic Semiparametric Factor Model (DSFM) is introduced in Section 3. Empirical results regarding the modelling and forecasting of liquidity supply are provided in Sections 4 and 5, respectively. Section 6 concludes.

2. Data

2.1. Trading at the ASX and descriptive statistics

The Australian Stock Exchange (ASX) is a continuous double auction electronic market, where the continuous auction trading period is preceded and followed by a call auction. Normal trading takes place continuously on all stocks between 10:09 a.m. and 4:00 p.m. from Monday to Friday. During continuous trading, any buy (sell) order entered that has a price that is greater than (less than) or equal to existing queued buy (sell) orders, will be executed immediately. If an order cannot be executed completely, the remaining volume enters the queues as a limit order. Limit orders are queued in the buy and sell queues according to a strict price-time priority order. Orders can be entered, deleted and modified without restriction.

For order prices below 10 cents, the minimum tick size is 0.1 cents, for order prices above 10 cents and below 50 cents it is 0.5 cents, whereas for orders priced 50 cents and above it is 1 cent. Note that there might be orders which are entered with an undisclosed or hidden volume if the total value of the order exceeds AUD 200,000. Since this applies only to a small fraction of the posted volumes, we can safely neglect the occurrence of hidden volume in our empirical study. For more details on the data, see Hall and Hautsch (2007) using the same data base as well as the official description of the trading rules of the Stock Exchange Automated Trading System (SEATS) on the ASX on www.asxonline.com.

We select four companies traded at the ASX covering the period from July 8 to August 16, 2002 (30 trading days), namely Broken Hill Proprietary Limited (BHP), National Australia Bank Limited (NAB), MIM and Woolworths (WOW). The number of market and limit orders for the selected stocks is given in Table 1.

We observe more buy orders than sell orders implying that the bid side of the limit order book was changing more frequently than the ask side. BHP and NAB are significantly more actively traded than MIM and WOW shares. Aggregated over all stocks, 20.08% (23.98%) of all bid (ask) limit orders have been changed (after posting), whereas 13.70% (14.89%) have been cancelled. Furthermore, for both traded as well as posted quantities we find that on average sell volumes are higher than buy volumes (not reported here). Hence, confirming the result above, liquidity variations on the bid side are higher than that of the ask side. This finding might be explained by the fact that during the analyzed period the market generally went down creating more sell activities than buy activities.

The original dataset contains all limit order book records as well as the corresponding order curves represented by the underlying price-volume combinations. The latter is the particular object of interest for the remainder of the analysis.

Table 1

Total number of market and limit orders for selected stocks traded at the ASX from July 8 to August 16, 2002.

Orders	BHP	NAB	MIM	WOW
<i>Market orders</i>				
(i) Buy	28,030	16,304	4115	7260
(ii) Sell	16,755	15,142	2789	6464
<i>Limit orders</i>				
(i) Buy (bid side)	50,012	28,850	9551	13,234
– Changed	8009	7561	1637	3203
– Cancelled	5202	4725	2044	1951
(ii) Sell (ask side)	32,053	25,953	6474	11,318
– Changed	6891	6261	1862	3164
– Cancelled	4692	3863	1178	1554

2.2. Notation and data preprocessing

The underlying limit order book data contains identification attributes regarding $r=1,\dots,R$ different orders as well as quantities demanded and offered for different price levels $j=1,\dots,J$, at any time point $t=1,\dots,T$. Particularly, at any t , we observe $J=101$ price levels on a fixed minimum tick size grid originating from the best bid and ask quote.

Since the order book dynamics are found to be very persistent, we choose a sampling frequency of 5 min without losing too much information on the liquidity supply. To remove effects due to market opening and closure, the first 15 min and last 5 min are discarded. Hence, at each trading day, starting at 10:15 and ending at 15:55, we select per stock 69 price–quantity vectors, in total $T=2070$ vectors over the whole sample period. Denote $\tilde{Y}_{t,j}^b$ and $\tilde{Y}_{t,j}^a$ as the pending bid and ask volumes at bid and ask limit prices $\tilde{S}_{t,j}^b$ and $\tilde{S}_{t,j}^a$, respectively at time point t .

We define the best bid price at time t as the highest buy price $\tilde{S}_{t,101}^b$, and similarly, the best ask price at t as the lowest sell price $\tilde{S}_{t,1}^a$. The corresponding quantities at best bid and ask prices are then $\tilde{Y}_{t,101}^b$ and $\tilde{Y}_{t,1}^a$, respectively, yielding the mid-quote price to be defined as $\tilde{S}_t^* = (\tilde{S}_{t,101}^b + \tilde{S}_{t,1}^a)/2$. The absolute price deviations from the best bid and ask price at level j and time t are given by $\tilde{S}_{t,j}^b = \tilde{S}_{t,j}^b - \tilde{S}_{t,101}^b$ and $\tilde{S}_{t,j}^a = \tilde{S}_{t,j}^a - \tilde{S}_{t,1}^a$, respectively and constitute a fixed price grid. To measure spreads between individual price levels in *relative* terms, i.e., in relation to the prevailing best bid and ask price, we define so-called 'relative price levels' as $S_{t,j}^b = \tilde{S}_{t,j}^b / \tilde{S}_{t,101}^b$ and $S_{t,j}^a = \tilde{S}_{t,j}^a / \tilde{S}_{t,1}^a$, respectively.

In order to investigate to which extent order book information might reveal information to predict high-frequency returns, we regress 1 min and 5 min mid-quote returns, respectively, on lagged order imbalances

$$\tilde{Y}_{t-1,j}^b / (\tilde{Y}_{t-1,j}^b + \tilde{Y}_{t-1,j}^a)$$

and

$$\tilde{Y}_{t-1,j}^a / (\tilde{Y}_{t-1,j}^b + \tilde{Y}_{t-1,j}^a),$$

respectively, for $j=1,\dots,101$. Fig. 2 shows the implied R^2 values in dependence of the number of included imbalance levels. It turns out that order book imbalances indeed reveal short-term predictability. Interestingly, even levels far apart from the market have still distinct prediction power pushing the R^2 to values of approximately 10%. These findings show that the order book itself reveals predictive content for future price movements which could be exploited in trading strategies.

In order to account for intra-day seasonality effects, we adjust the order volumes correspondingly. To avoid to seasonally adjust all individual volume series separately, we assume that the seasonality impact on quoted volumes at all levels is identical and is well captured by the seasonalities in market depth on the best bid and ask levels $\tilde{Y}_{t,101}^b$ and $\tilde{Y}_{t,1}^a$, respectively. Assuming a multiplicative impact of the seasonality factor, the seasonally adjusted quantities are computed for both sides of the market at price level j , and time t as

$$Y_{t,j}^b = \frac{\tilde{Y}_{t,j}^b}{s_t^b} \tag{1}$$

$$Y_{t,j}^a = \frac{\tilde{Y}_{t,j}^a}{s_t^a}, \tag{2}$$

with s_t^b and s_t^a representing the seasonality components at time t for the bid and the ask side, respectively.

The non-stochastic seasonal trend factors s_t^b and s_t^a are specified parametrically using a flexible Fourier series approximation as proposed by Gallant (1981) and are given by

$$s_t^b = \delta^b \cdot \bar{t} + \sum_{m=1}^{M^b} \{ \delta_{c,m}^b \cos(\bar{t} \cdot 2\pi m) + \delta_{s,m}^b \sin(\bar{t} \cdot 2\pi m) \} \tag{3}$$

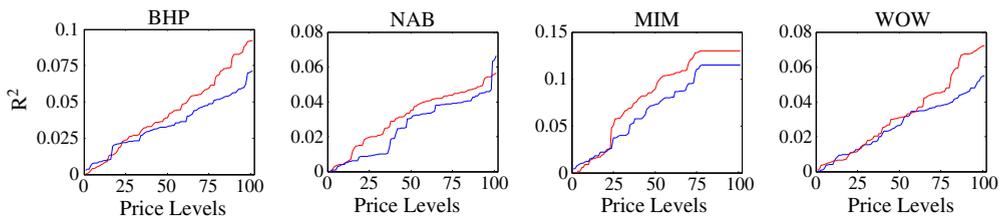


Fig. 2. Coefficients of determination (R^2) implied by linear regression of 1 min (red) and 5 min (blue) mid-quote returns on lagged order imbalances for selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days). The horizontal axis depicts the number of included imbalance levels.

$$s_{\bar{t}}^a = \delta^a \cdot \bar{t} + \sum_{m=1}^{M^a} \left\{ \delta_{c,m}^a \cos(\bar{t} \cdot 2\pi m) + \delta_{s,m}^a \sin(\bar{t} \cdot 2\pi m) \right\}. \tag{4}$$

Here δ^b , δ^a , $\delta_{c,m}^b$, $\delta_{c,m}^a$ and $\delta_{s,m}^b$ and $\delta_{s,m}^a$ are coefficients to be estimated, and \bar{t} denotes a normalized time trend mapping the time of the day on a (0,1] intervals. The polynomial orders M^b and M^a are selected according to the Bayes information criterion (BIC). For all stocks we select $M^b = M^a = 1$, except for the bid side for BHP ($M^b = 2$). The resulting intra-day seasonality patterns for both sides of all order book markets are plotted in Fig. 3.

For all stocks, we observe that the liquidity supply increases before market closure. We attribute this finding to traders' pressure and willingness to close positions overnight. Posting aggressive limit orders on the best levels (or even within the spread) maximizes the execution probability and avoids crossing the spread. Moreover, weak evidence for a 'lunch time dip' is presented which, however, is only observed for the more liquid stocks (NAB and BHP). In contrast, for the less liquid stocks, the amount of posted volume nearly monotonically increases over the course of the day.

3. The dynamic semiparametric factor model

Recall that the object of interest is the high-dimensional object of seasonally adjusted level-dependent order volume inventories $(Y_{t,j}^b, Y_{t,j}^a) \in \mathbb{R}^{202}$, observed on a 5-min frequency. Proposing a suitable statistical model requires finding an appropriate way of reducing the high dimension without losing too much information on the spatial and dynamic structure of the process. Moreover, applicability of the model requires computational tractability as well as numerical stability.

A common way to reduce the dimensionality of multivariate processes is to apply a factor decomposition. The underlying idea is that the high-dimensional process is ideally driven by only a few common factors which contain most underlying information. Factor models are often applied in the asset pricing literature to extract underlying common risk factors. In this spirit, a successful parametric factor model has been proposed, for instance, by Nelson and Siegel (1987) to model yield curves. In this framework, the shape of the curve is parametrically captured by Laguerre polynomials.

Limit order book curves inherently reflect traders' price expectations and the supply and demand in the market (see, e.g. Glosten (1994) for a theoretical framework). As there is no obvious parametric form for ask and bid curves and we want to avoid imposing assumptions on functional form, we prefer to capture the curve's spatial structure in a nonparametric way. A natural and powerful class of models for these kind of problems is the class of Dynamic Semiparametric Factor Models (DSFMs) proposed by Fengler et al. (2007), Brüggemann et al. (2008), Park et al. (2009) and Cao et al. (2009). The DSFM model successfully combines the advantages of a nonparametric approach for cross-sectionally ('spatially') fitting a curve and that of a parametric time series model for modelling persistent multivariate dynamics.

Assume that the observable J -dimensional random vector, $Y_{t,j}$, can be modelled based on the following orthogonal L -factor model,

$$Y_{t,j} = m_{0,j} + Z_{t,1}m_{1,j} + \dots + Z_{t,L}m_{L,j} + \varepsilon_{t,j}, \tag{5}$$

where $m(\cdot) = (m_0, m_1, \dots, m_L)^\top$ denotes the time-invariant factors, a tuple of functions with the property $m_l : \mathbb{R}^d \rightarrow \mathbb{R}$, $l = 0, \dots, L$, $Z_t = (1_T, Z_{t,1}, \dots, Z_{t,L})^\top$ denotes the time series of factor loadings, and $\varepsilon_{t,j}$ represents a white noise error term. The time index is denoted by $t = 1, \dots, T$, whereas the cross-sectional index is $j = 1, \dots, J$. Note that this type of factor model is rather restrictive, because it does not take explanatory variables into account.

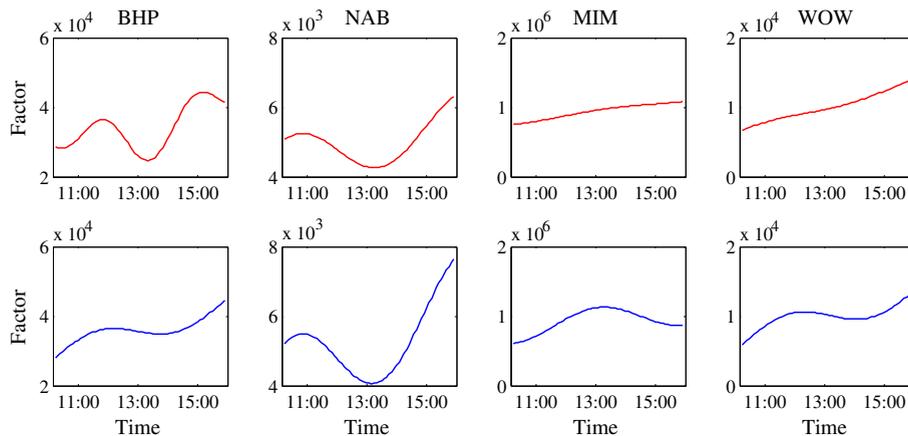


Fig. 3. Estimated intra-day seasonality factors for quantities offered at best bid prices (red) and for quantities supplied at best ask prices (blue) across selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days).

The DSFM is a generalization of the factor model given in Eq. (5) and allows the factors m_i to depend upon explanatory variables, $X_{t,j}$. Its analytical form is given by

$$Y_{t,j} = \sum_{l=0}^L Z_{t,l} m_l(X_{t,j}) + \varepsilon_{t,j} = Z_t^\top m(X_{t,j}) + \varepsilon_{t,j}, \tag{6}$$

where the processes $X_{t,j}$, $\varepsilon_{t,j}$ and Z_t are assumed to be independent. Moreover, the number of underlying factors L should not exceed the dimension of the object, J . The main idea of the DSFM is that L is significantly smaller than J resulting in a severe dimension reduction of the process.

As suggested by Park et al. (2009), the estimation of the factors m_i is performed using a series estimator. For $K \geq 1$, appropriate functions $\psi_k : [0, 1]^d \rightarrow \mathbb{R}$, $k = 1, \dots, K$, which are normalized such that $\int \psi_k^2(x) dx = 1$ holds, are selected. Park et al. (2009) select tensor B-spline basis functions for ψ_k , whereas Fengler et al. (2007) use a kernel smoothing approach. In the present study, we follow the former strategy and employ tensor B-spline basis functions.

After selecting the functions ψ_k , the factors $m(\cdot) = (m_0, m_1, \dots, m_L)^\top$ are approximated by $A\psi$, where $A = (a_{l,k}) \in \mathbb{R}^{(L+1) \times K}$ is a coefficient matrix, and $\psi(\cdot) = (\psi_1, \dots, \psi_K)^\top$ denotes a vector of selected functions. Here, K denotes the number of knots used for the tensor B-spline functions and is interpretable as a bandwidth parameter. Thus, the first part in the right-hand side of (6), which incorporates all factors and factor loadings, can be rewritten as

$$Z_t^\top m(X_{t,j}) = \sum_{l=0}^L Z_{t,l} m_l(X_{t,j}) = \sum_{l=0}^L Z_{t,l} \sum_{k=1}^K a_{l,k} \psi_k(X_{t,j}) = Z_t^\top A \psi(X_{t,j}). \tag{7}$$

In modelling liquidity supply we use either the ‘relative price levels’ on the bid side $S_{t,j}^b$ or those on the ask side $S_{t,j}^a$ as the most important explanatory variable $X_{t,j}$. When focusing on the LOB shape predictability, we add key (weakly exogenous) trading variables, namely the past 5-min aggregated trading volume on both sides of the market, the past 5-min log mid-quote return as well as the past 5-min volatility, see Section 3.

The coefficient matrix A and time series of factor loadings Z_t can be estimated using least squares. Hence, the estimated matrix \hat{A} and factor loadings $\hat{Z}_t = (1_T, \hat{Z}_{t,1}, \dots, \hat{Z}_{t,L})$ are defined as minimizers of the sum of squared residuals, $S(A, Z_t)$

$$(\hat{Z}_t, \hat{A}) = \arg \min_{Z_t, A} S(A, Z_t) \tag{8}$$

$$= \arg \min_{Z_t, A} \sum_{t=1}^T \sum_{j=1}^J \{Y_{t,j} - Z_t^\top A \psi(X_{t,j})\}^2. \tag{9}$$

To find a solution of the minimization problem stated in Eq. (9), a Newton–Raphson algorithm is used. As shown by Park et al. (2009), this algorithm is shown to converge to a solution at a geometric rate under some weak conditions on the initial choice $\{\text{vec}(A)^{(0)}, Z_t^{(0)}\}$. Moreover, Park et al. (2009) prove that the difference between the estimated loadings \hat{Z}_t and the true loadings Z_t are asymptotically negligible. Consequently, it is justified to use in a second step multivariate time series specifications in order to model the dynamics of the factor loadings. Note that due to the estimation complexity, the coefficients of the seasonal trend factors in Eqs. (1) and (2) are not estimated jointly with the unknown parameters (matrix A) and the factor loadings.

The selection of the number of time-invariant factors (L) and the number of knots K is performed by evaluating the proportion of explained variance (EV) given by

$$EV(L) = 1 - RV(L) = 1 - \frac{\sum_{t=1}^T \sum_{j=1}^J \left\{ Y_{t,j} - \sum_{l=0}^L \hat{Z}_{t,l} \hat{m}_l(X_{t,j}) \right\}^2}{\sum_{t=1}^T \sum_{j=1}^J \{Y_{t,j} - \bar{Y}\}^2}. \tag{10}$$

Moreover, the knots used in the tensor B-spline functions should be specified in advance. We choose linearly spaced knots, with a starting point determined by the minimal value of the explanatory variable (corrected by -5%), and the end point corresponding to the maximal value (corrected by 5%). Sensitivity analysis shows that the results are quite stable regarding the choice of grid points.

Because of the use of tensor B-spline functions for the demand and supply curves, which are monotonous in the price levels, our estimated first factor \hat{m}_1 and the estimated quantities $\hat{Y}_{t,j}$ are adjusted for extreme price levels. Correspondingly, for the bid side we keep constant the first (lowest) ten level values, and analogously, for the ask side we fix the last (highest) ten level values.

The model's goodness-of-fit is evaluated using the root mean squared error (RMSE) criterion,

$$RMSE = \sqrt{\frac{1}{TJ} \sum_{t=1}^T \sum_{j=1}^J \left\{ Y_{t,j} - \sum_{l=0}^L \hat{Z}_{t,j} \hat{m}_l(X_{t,j}) \right\}^2}. \quad (11)$$

4. Modelling limit order book dynamics

To model order book dynamics we follow a two step procedure for each stock individually. Employing the DSFM approach in the first step, we model the shape of order book curves in dependence of relative price levels. In the following step, the dynamics of the estimated factor loadings is analysed jointly with the best bid quotes, best ask quotes and the bid-ask spread in a parametric multivariate time series context. This procedure allows us to study the cross-dependency between both sides of the market, the interactions between the limit order book and the quotes, as well as the impact of the bid-ask spread on liquidity supply. Moreover, we investigate whether the order book shape itself is predictable by additional covariates, particularly, the past trading volume, past (realized) volatility as well as past log returns.

4.1. Limit order book modelling using the DSFM

We distinguish between two implementation methods of the DSFM:

- (i) Separated approach: Separate analysis of both sides of the limit order book, i.e., the bid side $Y_{t,j}^b \in \mathbb{R}^{101}$, and the ask side, $Y_{t,j}^a \in \mathbb{R}^{101}$.
- (ii) Combined approach: Simultaneous modelling of both sides of the limit order book with the bid side reversed, i.e. $(-Y_{t,j}^b, Y_{t,j}^a) \in \mathbb{R}^{202}$.

To model the limit order book in dependence of the relative price levels using the DSFM, i.e., the relative price deviations from the best bid price and best ask price, $S_{t,j}^b$ and $S_{t,j}^a$, respectively, we impose $K=20$ knots for the B-spline functions in case of the separated approach and $K=40$ knots in case of the combined approach. Using more knots does not result in significant improvements of the explained variance or in the corresponding RMSE, as defined in Eqs. (10) and (11).

Empirical results, available from the authors upon request, show that up to approximately 95% of the explained variation in order curves can be explained using $L=2$ factors, whereas the marginal contribution of a potentially third factor is only very small. Consequently, a two-factor DSFM specification is sufficient to capture the curve dynamics and is used in the sequel of the analysis. Furthermore, comparing the performance of the two alternative DSFM specifications, it turns out that in almost all cases the DSFM-separated approach outperforms the DSFM-combined approach in terms of a higher proportion of explained variance and lower values of the root mean squared error. We observe that at almost every price level the DSFM-separated approach outperforms the DSFM-combined approach. Therefore, the remainder of the analysis will rely on the DSFM-separated approach with two factors.

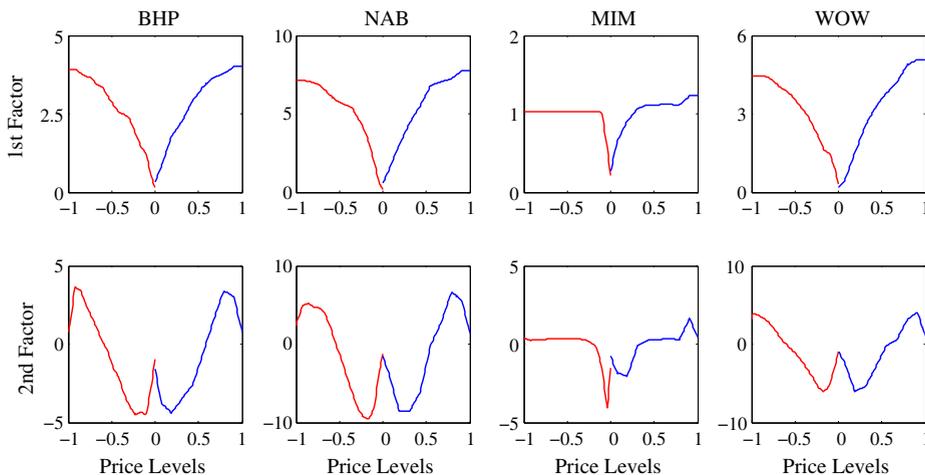


Fig. 4. Estimated first and second factor of the limit order book depending on relative price levels using the DSFM-separated approach with two factors for selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days). Red: bid curve, blue: ask curve.

Fig. 4 depicts the nonparametric estimates of the first and second factor \hat{m}_1 and \hat{m}_2 in dependence of the relative price grids. The first factor obviously captures the overall slope of the curve which is associated with the average trading costs for all volume levels on the corresponding sides of the market. In contrast, the second factor captures order curve fluctuations around the overall slope and thus can be interpreted as a ‘curvature’ factor in the spirit of Nelson and Siegel (1987). The shape of this factor reveals that the curve’s curvature is particularly distinct for levels close to the best quotes and for levels very deep in the book where the curve seems to spread out. The shapes of the estimated factors are remarkably similar for all stocks except for MIM. For the latter stock, the shapes of both factors are quite similar and significantly deviate from those reported for the other stocks. This finding is explained by the peculiarities of MIM for which the relative tick size is larger than for the other stocks. This implies that liquidity is concentrated on relatively few price levels around the best ask and bid quotes whereas the book flattens out for higher levels. This pattern is clearly revealed by the corresponding factors shown in Fig. 4.

However, a priori it is unclear whether modelling order book curves based on all 101 price levels is most appropriate in a prediction context. Besides the well-known trade-off between in-sample fit and out-of-sample prediction performance, we also face the difficulty that the predictive information revealed by order book volume might depend on the distance to the best quotes. For instance, if price levels far away from the market may contain information that help predicting books in the future, this information should be taken into account. However, if they contain virtually only noise (e.g., because of stale orders) it would be more optimal to ignore this information in order to extract a more precise factor structure on lower price levels only. Since optimizing this choice in an (out-of-sample) prediction context is tedious and computationally cumbersome, we restrict ourselves to the quite common proceeding of performing model selection based on in-sample information. Accordingly, we evaluate the model implied explained variance when not the full grid of 101 levels but just 25, 50 and 75 levels are employed. It turns out that the explained variance remains widely unchanged with the model fit increasing with the number of incorporated levels. This is particularly important in the context of order books of less liquid stocks. Therefore, we proceed by extracting the factor structure employing the entire book.

Time series plots of the corresponding factor loadings \hat{Z}_t^b and \hat{Z}_t^a are shown in Fig. 5. We observe that the loadings strongly vary over time reflecting time variations in the shape of the book. The series reveal clustering structures indicating a relatively high persistence in the processes. This result is not very surprising given the fact that order book inventories do not change too severely during short time horizons. Observing order book volumes on even higher frequencies than 5 min further increases this persistence, ultimately driving the processes toward unit root processes. Naturally, this behavior is particularly distinct for less frequently traded stocks and less severe for highly active stocks (cf. Hautsch and Huang (2012) for corresponding results for more liquid assets).

The high persistence is confirmed by autocorrelation functions of \hat{Z}_t^b and \hat{Z}_t^a (not shown in the paper) and corresponding unit root and stationarity tests. According to the Schmidt–Phillips test (see Schmidt and Phillips (1992), H_0 : unit root) for all processes the null hypothesis of a unit root can be rejected at the 5% significance level (test statistics for all estimated factor loadings are in the range $[-201.53, -53.88]$, whereas the critical value equals -25.20). Conversely, testing the null hypothesis of stationarity using the KPSS test (see Kwiatkowski et al. (1992), H_0 : weak stationarity) implies no rejections for the majority of the processes. Nevertheless, in five cases we have to reject stationarity. Finally, to test for possible cointegration between the factor loadings, we perform Johansen (1991) trace test (not shown in the paper) but do not find significant evidence for common stochastic trends underlying the order book.

A graphical illustration for the goodness-of-fit of the model, depicting the estimated vs. the actually observed limit order book curve, would suggest that the model fits the observed curves very well (no illustrations provided here). This is particularly true

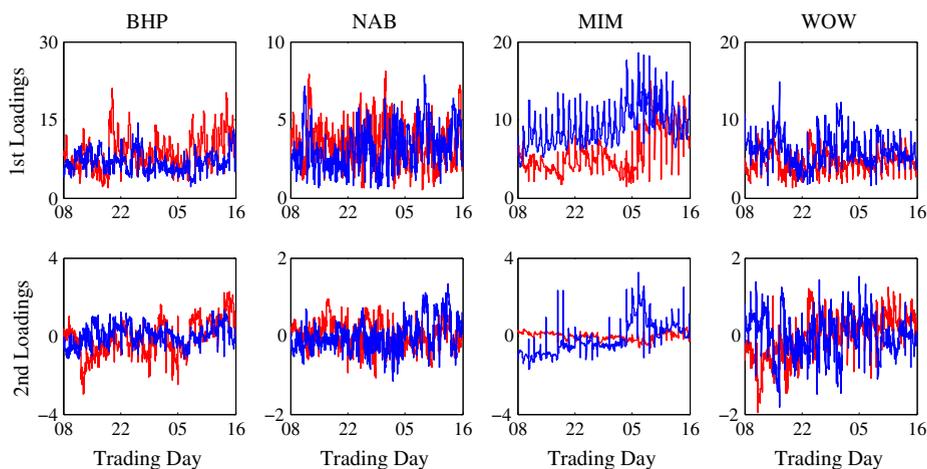


Fig. 5. Estimated first and second factor loadings of the limit order book depending on relative price levels using the DSFM-separated approach with two factors for selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days). Red: bid curve, blue: ask curve.

for price levels close to the best ask and bid quotes, at any chosen trading day and stock. Slight deviations are observed for price levels deeply in the book. However, the latter case is less relevant for most applications in practice.

4.2. Modelling limit order book dynamics

Our approach, stipulated under the philosophy *smooth in space and parametric in time*, allows us to investigate the limit order book dynamics in a multivariate time series modelling context, as well as to relate the order book dynamics to the time evolution of additional covariates. Formally, for each stock we focus on the dynamics of the four estimated stationary factor loadings. Including the best bid and the best ask price returns, we consider a (six dimensional) vector of endogenous variables

$$z_t = \left(\hat{z}_{1,t}^b, \hat{z}_{2,t}^b, \hat{z}_{1,t}^a, \hat{z}_{2,t}^a, \Delta \log \tilde{S}_{t,101}^b, \Delta \log \tilde{S}_{t,1}^a \right)^\top,$$

where $\hat{z}_{1,t}^b$, $\hat{z}_{2,t}^b$, $\hat{z}_{1,t}^a$ and $\hat{z}_{2,t}^a$ denote the estimated first (1) and second (2) factor loadings for the bid (*b*) and ask side (*a*), respectively. We denote by $\Delta \log \tilde{S}_{t,101}^b$ the best bid price return, and similarly, by $\Delta \log \tilde{S}_{t,1}^a$ the best ask price return.

Following Engle and Patton (2004) and Hautsch and Huang (2012), the bid-ask spread $(\log \tilde{S}_{t-1,101}^b - \log \tilde{S}_{t-1,1}^a)$ serves as a natural cointegration relationship between the two integrated ask and bid series. As all other endogenous variables are shown to be stationary, we obtain a vector error correction (VEC) specification of order q with the spread as the only cointegration relationship, i.e.,

$$z_t = c + \Gamma_1 z_{t-1} + \dots + \Gamma_q z_{t-q} + \gamma (\log \tilde{S}_{t-1,101}^b - \log \tilde{S}_{t-1,1}^a) + \varepsilon_t. \quad (12)$$

Here c denotes a vector with constants, vector $\gamma = (\gamma_1, \dots, \gamma_6)^\top$ collects parameters associated with the lagged bid-ask spread and ε_t represents a white noise error term. The matrices $\Gamma_1, \Gamma_2, \dots, \Gamma_q$ are parameter matrices associated with lagged endogenous variables. Technically, we determine the order q according to the BIC.

Estimation results show that in all cases, a maximum lag order of $q=4$ is sufficient. In particular, the following model orders are selected: BHP and WOW ($q=3$), NAB ($q=2$), MIM ($q=4$). For sake of brevity we refrain from showing all parameter estimates here, but just report the estimates of matrix Γ_1 and vector γ for BHP, NAB, MIM and WOW, respectively, which contain the most relevant information for an economic interpretation (5% significance is denoted by an asterisk (*)):

$$\begin{bmatrix} 0.95^* & 0.63^* & -0.05 & -0.26^* & 3.03 & 18.08 \\ 0.02^* & 0.79 & 0.00 & 0.04 & 10.68 & -16.12 \\ 0.04^* & 0.00 & 0.75^* & 0.02 & -59.60 & 67.60 \\ -0.00 & 0.04 & 0.02^* & 0.77^* & -13.99 & 13.55 \\ 0.00^* & 0.00 & -0.00^* & 0.00^* & -0.59 & 0.29 \\ 0.00^* & 0.00 & -0.00^* & 0.00^* & -0.26 & -0.04 \end{bmatrix}, \begin{bmatrix} -95.70 \\ -34.13 \\ 86.83 \\ -13.21 \\ -0.42 \\ 0.02 \end{bmatrix},$$

$$\begin{bmatrix} 0.71^* & 0.16 & -0.04 & -0.21 & 123.78^* & -124.07^* \\ 0.04^* & 0.78^* & -0.00 & 0.07^* & -22.39^* & 21.91 \\ 0.04 & 0.13 & 0.73^* & 0.18 & -88.56^* & 86.91^* \\ -0.03^* & -0.03 & 0.03^* & 0.71^* & 26.03^* & -25.46^* \\ 0.00^* & 0.00^* & -0.00^* & -0.00 & 0.21 & -0.34 \\ 0.00 & 0.00 & -0.00^* & 0.00 & 0.29 & -0.41 \end{bmatrix}, \begin{bmatrix} -174.41^* \\ 9.26 \\ 47.60 \\ -20.59 \\ -0.85 \\ -0.04 \end{bmatrix},$$

$$\begin{bmatrix} 0.90^* & 1.29^* & -0.00 & 0.55^* & -46.92 & 50.79 \\ 0.00 & 0.93^* & -0.01^* & -0.01 & 1.12 & -1.49 \\ -0.02 & 1.23^* & 0.99^* & 0.48^* & 31.56 & -44.50^* \\ 0.00 & 0.04 & 0.03^* & 0.84^* & 6.73 & -5.89 \\ 0.00 & 0.00 & -0.00 & -0.00 & 0.40 & -0.58 \\ 0.00 & 0.00 & -0.00 & -0.00 & 0.90 & -1.09 \end{bmatrix}, \begin{bmatrix} 62.01^* \\ 0.25 \\ -44.50^* \\ -21.66^* \\ -0.28 \\ -0.18 \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} 0.74^* & -0.02 & 0.12^* & 0.38^* & 28.87 & -37.11 \\ 0.04^* & 0.82^* & -0.02^* & -0.04 & 2.53 & -3.58 \\ 0.04 & 0.03 & 0.87^* & 0.19^* & -70.61^* & 72.84^* \\ -0.03^* & 0.02 & 0.02^* & 0.83^* & 12.81 & -13.70 \\ -0.00^* & -0.00^* & 0.00 & 0.00^* & 0.02 & -0.15 \\ -0.00^* & -0.00^* & 0.00 & 0.00^* & 0.21 & -0.34 \end{bmatrix}, \begin{bmatrix} -27.14 \\ -6.33 \\ 59.98 \\ -4.04 \\ -0.51 \\ 0.05 \end{bmatrix}.$$

The estimation results can be summarised as follows:

Firstly, we observe strong own-process dynamics, but only relatively weak (mostly insignificant) cross-dependencies between the endogenous variables. The latter are most pronounced for less frequently traded stocks (MIM and WOW). Overall, the quite weak inter-dependencies between the processes on the ask and bid side indicate that time variations in the liquidity schedule on the one side is almost unaffected by that on the other side.

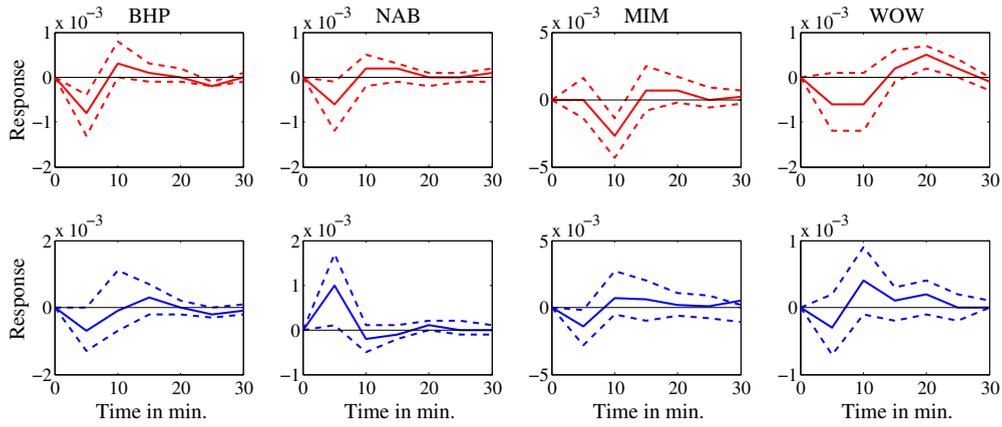


Fig. 6. Orthogonalized impulse-response analysis: responses of the best bid quote return to a one standard deviation shock in the estimated first bid factor loadings (upper panel) and response of the best ask quote return to a one standard deviation shock in the estimated first ask factor loadings (lower panel). We employ the DSFM-separated approach with two factors and a VEC specification for selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days). The response variable always enters the VEC specification in the first position. 95% confidence intervals are shown with dashed lines.

Secondly, the major finding is that quote changes are short-run predictable given the shape of the book. More precisely, changes in the factor loading have a short term impact on the quote changes, up to, say, 5–10 min. The impact is significant for the frequently traded stocks, and less severe for less liquid stocks. In particular, a shock on the bid side resulting in upward rotation of the bid curve (inducing a higher sell pressure) leads to an instantaneous decrease in the best bid quote followed by a significant increase of the price within the next few minutes, see, e.g. Fig. 6. This is driven by a growing buy pressure reflected by an increase of bid depth at and behind the market. Fig. 6 depicts the impulse responses of ask and bid quotes driven by a shock in the order book slope. While these effects are quite distinct on the bid side, they are, however, less pronounced on the ask side. A shock on the ask side, however, has a more neutral effect on the price, see, e.g. Fig. 6. However, note that the predictability of quotes only holds over comparably short horizons. Therefore, for daily order execution strategies, as discussed in Section 5, these effects are only of limited use.

Thirdly, we find slight evidence for asymmetric reactions of slope factor loadings on changes of the bid-ask spread. In particular, we observe that rising spreads tend to reduce the order aggressiveness on the bid side while the converse is true on the ask side. Hence, we conclude that as the bid and ask curves move apart, the price is (on average) decreasing. Similarly, as the bid-ask spread shrinks, the price is expected to increase. This re-confirms our finding in Chapter 2, that liquidity variations on the bid side are higher than that of the ask side with more sell activities than buy activities.

4.3. Drivers of the order book shape

In this section, we analyse whether the shape of order book curves is predictable based on key (weakly exogenous) trading variables. We select three variables for which we expect to observe the strongest impact on the book's shape, namely the past 5-min aggregated trading volume on both sides of the market representing the recent liquidity demand, the past 5-min log mid-quote return as well as the past 5-min volatility.

The buy and sell trading volumes at time t are given by the sum of traded quantities from all market orders r, \tilde{Q}_t^b and \tilde{Q}_t^s , over 5 min interval, namely, $\tilde{Q}_t^b = \sum_{r=1}^{R_t^b} \tilde{Q}_r^b$ and $\tilde{Q}_t^s = \sum_{r=1}^{R_t^s} \tilde{Q}_r^s$, where R_t^b and R_t^s denote the number of buy and sell orders over the interval $(t-1, t]$, respectively. Correspondingly, log returns r_t and volatility V_t are computed as

$$r_t = \log \frac{\tilde{S}_t^*}{\tilde{S}_{t-1}^*} \tag{13}$$

$$V_t = r_t^2, \tag{14}$$

where \tilde{S}_t^* and \tilde{S}_{t-1}^* denote the mid-quotes observed at t and $t-1$, respectively. Note that the trading volumes as well as the volatility are seasonally adjusted following the procedure explained above. Moreover, the used nonparametric procedure requires the variables to be standardized between -1 and 1 . This standardization is performed based on the minimum and maximum observations of the corresponding variables. Finally, as commonly known, nonparametric regression becomes computationally cumbersome for a high number of regressors. To keep our approach computationally tractable and to avoid problems due to the curse of dimensionality, we include the regressors only individually (together with the relative price distances). This ultimately yields a three-dimensional problem.

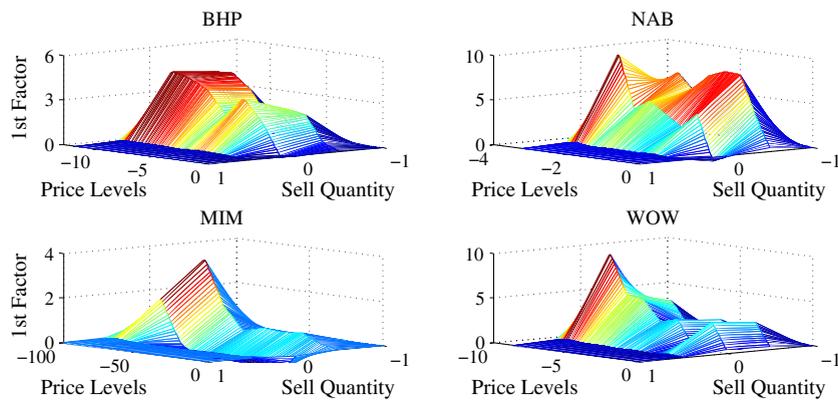


Fig. 7. Estimated first factors of the bid side with respect to relative price levels and the past log traded sell volume using the DSFM-separated approach with two factors for selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days).

Figs. 7 and 8 show the estimated first factors for the bid and the ask side in dependence of the past 5-min sell and buy trading volumes, respectively. As expected, we observe that the past liquidity demand influences the order book curve. A high trading volume implies that a non-trivial part of the pending volume in the book is removed. Thus, most of the observed variation of the factor's shape is induced by the fact that either quoted price levels close to the best quotes have been completely absorbed and the remaining volume is correspondingly 'shifted down' in relation to the new best quote or, alternatively, only a part of the pending volume on the best quotes is removed changing the distribution of the pending volumes across the (relative) price levels.

As expected, the curve flattens in the area of high volumes. Strikingly, we also observe a decaying pattern if the volume sizes decline. Actually, in all pictures, the maximum slope (and thus the highest level of liquidity supply) is observed for magnitudes of the standardized volume between -1 and 0 , i.e., comparably small (though not zero) trading volumes. This pattern might be technically explained by the standardization procedure based on extreme values or by the usual boundary problems of non-parametric regression. On the other hand, note that due the curse-of-dimensionality problem we cannot simultaneously control for other variables. For instance, very small market-side-specific trading volumes can indicate the occurrence of market imbalances or, alternatively, might be associated with wide spreads. Both scenarios could force investors to post rather limit orders than market orders which might explain the decaying shape of the figures after having observed small trading volumes.

To evaluate whether the inclusion of past trading volume further increases the model's goodness-of-fit, we calculated the corresponding RMSEs. Comparing the results (range from 4.37 to 10.42) with that for the basis model (range from 0.18 to 3.49) shows that the included regressors yield higher estimation errors. Hence, obviously the inclusion of additional regressors ultimately generates more noise overcompensating a possibly higher explanatory power. Similar results are also found for the past log returns and past volatility serving as regressors. The inclusion of log returns yields smaller estimation errors than the inclusion of volatility. However, the overall performance is lower than in the cases above. Because of this reason, we refrain from showing corresponding graphs of the estimated factors.

A possible reason for the declining model performance in case of included regressors might be the lower dimensionality of the regressors in comparison with that of the limit order book. Note that the included regressors do not reveal any variation across

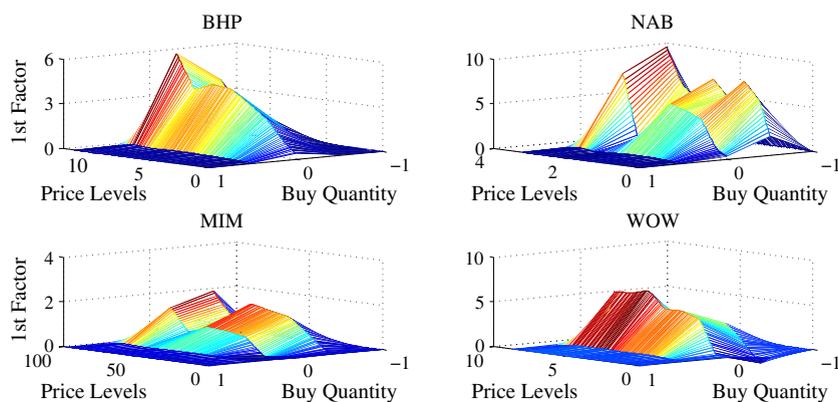


Fig. 8. Estimated first factors of the ask side with respect to relative price levels and the past log traded buy volume using the DSFM-separated approach with two factors for selected stocks traded at the ASX from July 8 to August 16, 2002 (30 trading days).

the levels of the book. Consequently, the explanatory variables cannot improve the model's spatial fit but just its dynamic fit. Obviously, the latter is not sufficient to obtain an overall reduction of estimation errors.

5. Forecasting liquidity supply

5.1. Setup

The aim of this section is to analyse the model's forecasting performance in a realistic setting mimicking the situation in financial applications. We consider an investor observing the limit order book at 5-minute snapshots together with the history over the past 10 trading days. It is assumed that during a trading day an investor updates limit order book every 5 min and requires producing forecasts for all (5 min) intervals of the remainder of the day. Such information might be useful in order to optimally balance order execution during the course of a day. Since we do not exceed beyond the end of the trading day (in order to avoid overnight effects), the forecasting horizon h subsequently declines if we approach market closure. Hence, starting at 10:30, we produce multi-step forecasts for all remaining $h = 66$ intervals during the day. Correspondingly, at 15:50, we are left with a horizon of $h = 1$. Since quotes themselves – according to our results above – are only predictable over short horizons which are virtually irrelevant for the present analysis, we do not explicitly incorporate this information here.

Consequently, the model is re-estimated every 5 min exploiting past information over a fixed window of 10 trading days (including the recent observation). Due to the length of the estimation period, we do not produce forecasts for the first two weeks of our sample but focus on the period between July 22 and August 16, 2002, thereby covering the period of 20 trading days. In accordance with our in-sample results reported in the previous section, we choose the DSFM-Separated approach based on two factors without additional regressors as underlying specification.

A natural benchmark to evaluate our model is the naive forecast. In this context, we assume that the investor has no appropriate prediction model but just uses the current liquidity supply as a forecast for the remainder of the day. More formally, we suppose that our investor can use the following two approaches in order to forecast liquidity supply $\hat{Y}_{t'+hj}$ at a given time point t' from July 22 at 10:25 until August 16, 2002, at 15:50, $t' = 693, \dots, 2069 = T - 1$, over a forecasting horizon $1 \leq h \leq 66$, and over the absolute price level j :

- (i) DSFM approach: Firstly, the factors and factor loadings are estimated using the DSFM-Separated approach with two factors, $K = 20$ knots used for the B-spline basis functions, and with past 690 observed (de-seasonalized) limit order book curves. More precisely, at time point t' , relative price levels $S_{t'-691:t',j}^b$ and $S_{t'-691:t',j}^a$ and de-seasonalized observed bid and ask sides $Y_{t'-691:t',j}^b$ and $Y_{t'-691:t',j}^a$ enter the estimation procedures. This yields estimates for the bid (ask) side, 66 times per day for each stock, in total 1320 times over 20 days. Secondly, since we do not account for (short-term) quote return predictability but only forecast the liquidity supply, we employ a simple 4-dimensional VAR(p) model for the four time-varying factor loadings. When fitted to the entire time series (30 trading days) and according to the BIC, a maximum lag order $p = 4$ is sufficient. In particular, the following VAR(p) models are selected: BHP and MIM – VAR(4), NAB – VAR(2), WOW – VAR(3). Using this specifications, we forecast the factor loadings over the forecasting period $\hat{Z}_{t'+h}$. Then, the predicted factor loadings together with the estimated time-invariant factors \hat{m}_t are used to predict the order book.
- (ii) Naive approach: Among all historical 690 limit order book curves, only the last one at time t' , $(Y_{t',j}^b, Y_{t',j}^a)$, is selected as the h -step ahead forecast.

The predictions are evaluated using the root mean squared prediction error (RMSPE), i.e., a version of the in-sample RMSE (11) where the sum over the sampling periods t and the sample size T are replaced by the forecasting horizons h and H , respectively. Since future quotes and relative price grids are not predicted by the model, we assume that quotes themselves follow random walk processes and the spread remains constant. Future quotes are therefore predicted using the current one. Consequently, the predicted future relative price grid remains constant.

5.2. Forecasting results

Fig. 9 shows the RMSPEs for each required forecasting horizon h during a trading day implied by the DSFM as well as the naive model. The following results can be summarized: First, overall the DSFM forecasts outperform the naive ones. Nevertheless, the naive forecast is a serious competitor which is hard to beat. This result is not surprising given the high persistence in liquidity supply. Secondly, the model's forecasting performance is obviously higher on the bid side than on the ask side. This result might be explained by the fact that during the sample period we observe a downward market inducing higher activities on the bid side than on the ask side. This is confirmed by the descriptive statistics shown above. Thirdly, the DSFM outperforms the naive model particularly over horizons up to 1 to 2 hours. For longer horizons, the picture is less clear.

Analyzing average RMSPEs (averaged over all forecasting horizons and both sides of the market) as reported by Table 2 indicate that the overall prediction performance of the DSFM approach is significantly higher than that of the benchmark.

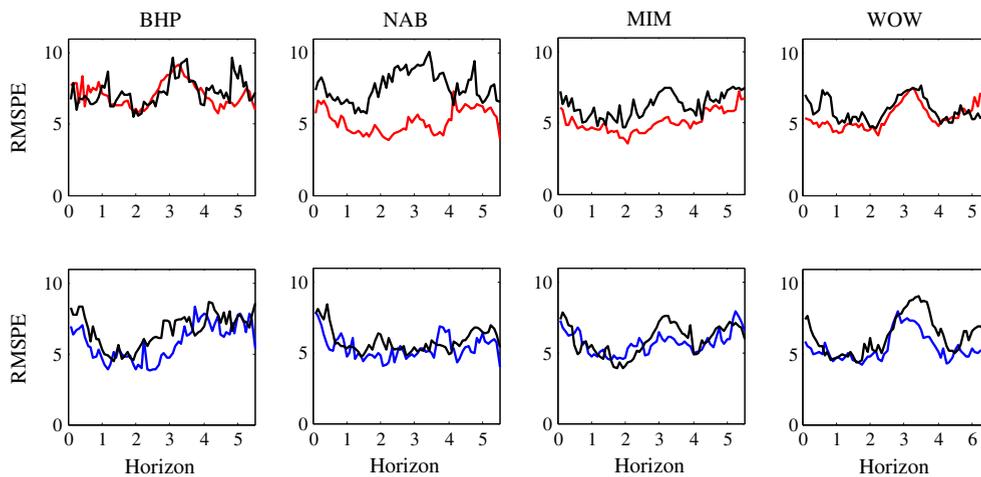


Fig. 9. Root mean squared prediction errors (RMSPEs) implied by the DSFM-separated approach with two factors for the bid side (red) as well as the ask side (black) and by the naive approach (black) for all intra-day forecasting horizons (in hours) for selected stocks traded at the ASX. Prediction period: July 22 to August 16, 2002 (20 trading days).

5.3. Financial and economic applications

The results in the previous section show that the DSFM is able to successfully predict liquidity supply over various forecasting horizons during a day. In this subsection, we apply these results to two practical examples. The first one is devoted to an order execution strategy, whereas the second one deals with forecasts of demand and supply elasticities.

Example 1. (Trading Strategy)

Suppose an institutional investor decides to buy (sell) a certain number of shares v over the course of a trading day, starting from 10:30 until 15:40. The size of the traded quantity for BHP, NAB and WOW is chosen as to be 5 or 10 times the average pending volume at the best bid (ask) level. In case of MIM, where liquidity supply is much more concentrated at the first level and the book is very thin for higher levels (see the empirical results in the previous sections), we choose the traded volume as being 2 and 5 times the average first level depth. This yields to the following quantities in the respective two cases of high (a) and very high (b) liquidity demand:

- (a) BHP – 175,000 shares; NAB – 25,000 shares; WOW – 50,000 shares; MIM – 1,860,000 shares
- (b) BHP – 350,000 shares; NAB – 50,000 shares; WOW – 100,000 shares; MIM – 4,650,000 shares.

To reduce the computational burden, we assume that trading is only performed on a 5 min grid throughout the day corresponding to 63 possible trading time points. Moreover, suppose that the investor makes her trading decision at 10:30 but does not monitor the market anymore during the day. Consequently, her forecasting horizon covers $h = 63$ periods at each trading day. Then, she has to decide between two execution strategies:

- (i) Splitting the buy (sell) order of size v in a 5 minute frequency proportionally over the trading day resulting into 63 trades of size $v/63$ each.
- (ii) Placing orders not proportionally but at those m (5 minute interval) time points throughout the day where the DSFM-based predicted implied trading costs c of the volume v are smallest (among all 63 possible periods). Then, the volume v is split over the m time points according to the relative proportions of expected trading costs. Hence, at interval i , $w_i \cdot v$ shares are traded, with $w_i = c_i / \sum_{j=1}^m c_j$ for $i = 1, \dots, m$.

Table 2

Average root mean squared prediction errors (RMPSEs) of both limit order book sides implied by the DSFM-separated approach with two factors and the naive model for selected stocks traded at the ASX in the period from July 22 to August 6, 2002 (20 forecasting days).

Approach	BID				ASK			
	BHP	NAB	MIM	WOW	BHP	NAB	MIM	WOW
Naive	7.11	7.59	6.03	6.08	6.50	5.96	5.96	6.19
DSFM	7.18	5.10	4.84	5.33	5.56	5.46	5.63	5.45

Strategy (i) can be seen as a special case of strategy (ii) if m is chosen as $m=63$ and the volume v is just equally split. Conversely, for $m=1$, we obtain the extreme case, where the entire quantity is traded once requiring to severely 'walk up' the book. The DSFM predictions of trading costs are computed based on the predicted order book shape at each point and the effective costs to buy or to sell the quantity v while using the ask and bid quotes prevailing at 10:25 (in accordance with the assumption of a random walk). Note that we do not optimize over the quantity underlying the predicted trading costs but just fix it at v corresponding to the maximally possible trade size per time point. Consequently, our strategy selects those trading points where the execution of the entire quantity v is expected to be cheapest and thus covers also the hypothetical (limiting) case of putting all weight w_i on a single point implying a 'one-shot' execution. Of course, an even more sophisticated (and optimized) strategy would require the prediction of trading costs for relative proportions of v which are themselves simultaneously optimized. However, this would substantially increase the numerical and computational burden and is beyond the scope of the current study.

To implement these strategies, we consider 20 forecasting days covering the period from July 22 to August 16, 2002. Fig. 10 shows the average percentage reduction in trading costs of strategy (ii) in excess of the equal-splitting ('naive') strategy (i) for various choices of $m \in [1,63]$. In most cases we observe that a strategic placing of orders according to DSFM predictions yield excess gains of approximately 10 basis points on average. Overall, the selling strategies are more beneficial than the buying strategies confirming the findings on prediction errors above. This is most striking for BHP where we observe a significant difference between sell-based and buy-based profits if the number of trading points are low. Apart from this observation we find a generally non-monotonic behavior of the curves implying losses if m is small, increasing (and positive) gains for a higher number of trading points and a convergence to zero for m reaching the upper limit of 63. This pattern indicates that trading the daily position using only a few large market orders is inferior to an equal-splitting strategy as the underlying transactions have to walk up the book too severely and cause huge price impacts. For higher values of m , the strategic placement according to DSFM predictions become profitable where in the limit of $m=63$, relative benefits are only due to a strategic (non-equal) weighting scheme. However, for MIM we observe a significantly different pattern implying the highest gains for m being small and nearly monotonically declining profits if m is increasing. This pattern is obviously induced by the peculiarities of the MIM order book which is extremely deep on the first level and makes 'one-shot' executions of large volumes quite beneficial. Overall, the patterns are very similar for the two classes of daily quantities, where as expected the relative gains become smaller with higher traded daily volume.

Overall, our findings indicate that the model is successful in predicting times where the market is sufficiently deep in order to execute a large orders. The fact that the model performs reasonably well is promising for more elaborate practical applications of the DSFM. Moreover, note that the reported results are valid under the assumption that there are no transaction fees. Actually, in practice, a proportional splitting strategy induces higher transaction costs as a complete execution via a market order. This component is not taken into account here and would even increase the performance of the DSFM-based execution strategy. Finally, predictions of trading costs could be further improved by exploiting possible predictive information of the limit order book for future returns. Our descriptive statistics reported above show that order book imbalances have indeed (slight) prediction power. We will leave these issues, however, for future research.

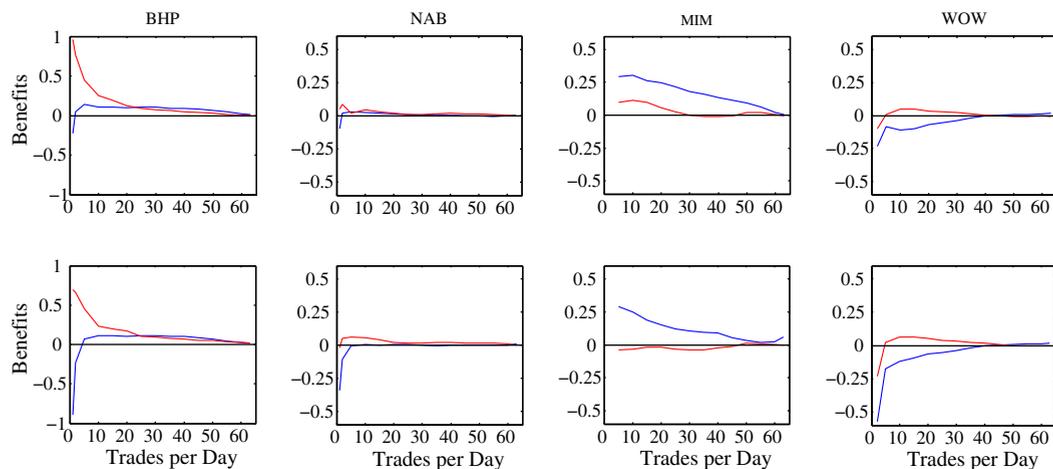


Fig. 10. Average percentage gains by reduced transaction costs compared to an equal-splitting strategy when buying (blue) and selling (red) shares based on m DSFM-predicted time points per day. Upper panel: Daily volumes corresponding to 5 (2) times the average first level market depth for BHP, NAB, WOW (MIM). Lower panel: Daily volumes corresponding to 10 (5) times the average first level market depth for BHP, NAB, WOW (MIM). Prediction period: July 22 to August 16, 2002 (20 trading days).

Example 2. (Demand and Supply Elasticity)

A straightforward dimension-less measure for the order book slope is the curve's elasticity which we compute at best bid ($\tilde{S}_{t,101}^b$) and best ask prices ($\tilde{S}_{t,1}^a$) as

$$\hat{E}_{t+h}^d = \frac{\hat{Y}_{t+h,1}^b - \hat{Y}_{t+h,101}^b}{\hat{Y}_{t+h,101}^b} / \frac{\tilde{S}_{t,1}^b - \tilde{S}_{t,101}^b}{\tilde{S}_{t,101}^b}, \quad (15)$$

$$\hat{E}_{t+h}^s = \frac{\hat{Y}_{t+h,101}^a - \hat{Y}_{t+h,1}^a}{\hat{Y}_{t+h,1}^a} / \frac{\tilde{S}_{t,101}^a - \tilde{S}_{t,1}^a}{\tilde{S}_{t,1}^a}, \quad (16)$$

for the demand (bid) and supply (ask) side, respectively. The elasticity is a measure for the marginal trading costs reflecting the curve's curvature.

Suppose at 10:30 an investor aims to predict the demand and supply elasticity at best bid and best ask prices for all 5-min intervals during the trading day covering the forecast horizons $h = 1, \dots, 66$. As above, the forecasts are computed using the last 10 trading days. Since we are not forecasting the price process, the last observed ask and bid quotes are used for prediction. Fig. 11 shows the 10:30 predictions of demand and supply elasticities at best bid and best ask prices during all trading days. We observe that marginal trading costs exhibit significant variations over time. The fact that predicted elasticities reveal temporarily trending patterns might be used for improving trading strategies.

Consider the case of NAB on July 24 and July 30, 2002. We observe that the demand elasticities (in absolute terms) are increasing on the first day, and decreasing on the second day. Practically, it would be better to sell shares late on July 24, and early on July 30, under the assumption that the price does not change significantly over both trading days. The supply elasticities show converse patterns across the days. Consequently, it would be advisable to buy shares early on July 24, and late on July 30, provided that the prices remain unchanged. While this section aims to illustrate possible applications of the DSFM approach, more detailed elaborations of dynamic strategies are beyond the scope of the paper.

6. Conclusions

In this paper, we propose a dynamic semiparametric factor model (DSFM) for modelling and forecasting liquidity supply. The main idea of the DSFM as proposed by Brüggemann et al. (2008), Cao et al. (2009), Fenger et al. (2007) and Park et al. (2009) is to capture the order curve's spatial structure across various relative distances to the best quotes using a factor decomposition which is estimated nonparametrically. To capture the order book's time variations, the corresponding factor loadings are modelled using a multivariate time series model. The framework is flexible though parsimonious and turns out to provide a powerful way to reduce the high dimension of the book and to extract the relevant underlying information regarding order book dynamics.

The model is applied to four stocks traded at the Australian Stock Exchange (ASX). It is shown that two underlying factors can explain up to 95% of in-sample variations of ask and bid liquidity supply. While the first factor captures the overall order curve's slope, the second factor is associated with the curve's curvature. The extracted factor loadings reveal highly persistent though weakly stationary dynamics which are successfully captured by a vector error correction specification. We find relatively weak spill-over effects between both sides of the limit order book sides that are more pronounced for less liquid stocks compared to high frequently traded ones. It is shown that order book shapes have short-term prediction power for quote changes. Furthermore, we show that the

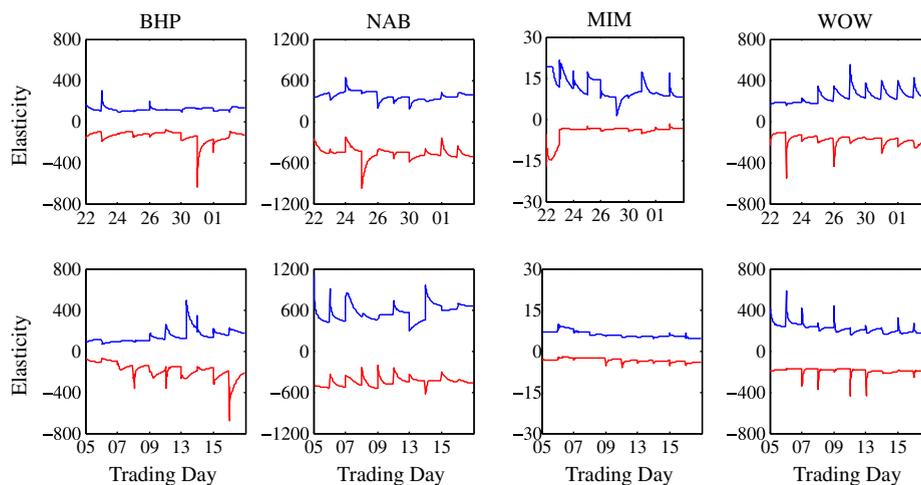


Fig. 11. Predicted demand and supply elasticities at best bid (red) and best ask prices (blue) using the DSFM-separated approach with two factors for selected stocks traded at the ASX from July 22 to August 2, 2002 (upper panels, 10 trading days) and from August 5 to August 16, 2002 (lower panels, 10 trading days).

order curves' shapes are driven by explanatory variables reflecting the recent liquidity demand, volatility as well as mid-quote returns.

Employing the DSFM approach in an extensive and realistic out-of-sample forecasting exercise we show that the model successfully predicts the liquidity supply over various forecasting horizons during a trading day and outperforms a naive approach. Using the forecasting results in a trading strategy it is shown that order execution costs can be reduced if orders are optimally placed according to predictions of liquidity supply. In particular, it turns out that optimal order placement in periods of high liquidity results in smaller transaction costs than in the case of a proportional splitting over time. Finally, our flexible approach allows us to estimate and to predict future (excess) demand and supply elasticities.

These results show that the DSFM approach is suitable for modelling and forecasting liquidity supply. Since it is computationally tractable, it can serve as a valuable building block for automated trading models.

Acknowledgements

We are very grateful to Anthony Hall for providing us the data. For helpful comments and discussions we thank Jean-Philippe Bouchaud, Joachim Grammig, Jeffrey Russell and the participants of the 2009 Humboldt-Copenhagen Conference on Financial Econometrics in Berlin, the 2009 annual conference of the Society for Financial Econometrics (SoFiE) in Geneva, the 2009 European Meeting of the Econometric Society in Barcelona, the International Conference on Price, Liquidity and Credit Risk in Konstanz, 2008, as well as the 4th World Congress of the International Association for Statistical Computing in Yokohama, 2008. Furthermore, we are grateful to Szymon Borak for helping us with the implementation of the Dynamic Semiparametric Factor Model in MATLAB. Finally, we thank the reviewers for their constructive comments and helpful suggestions.

References

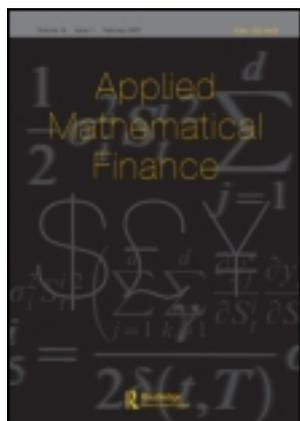
- Ahn, H.J., Bae, K.H., Chan, K., 2001. Limit orders, depth, and volatility: evidence from the stock exchange of Hong Kong. *Journal of Finance* 56, 767–788.
- Alfonsi, A., Fruth, A., Schied, A., 2010. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance* 10, 143–157.
- Almgren, R., Chriss, N., 2000. Optimal execution of portfolio transactions. *Journal of Risk* 3, 5–39.
- Bertsimas, D., Lo, A.W., 1998. Optimal control of execution costs. *Journal of Financial Markets* 1, 1–50.
- Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris bourse. *Journal of Finance* 50, 1655–1689.
- Bloomfield, R., O'Hara, M., Saar, G., 2005. The "make or take" decision in an electronic market: evidence on the evolution of liquidity. *Journal of Financial Economics* 75, 165–200.
- Brownlees, C.T., Cipollini, F., Giampiero, M.G., 2009. Intra-daily Volume Modeling and Prediction for Algorithmic Trading. Discussion paper, Stern School of Business.
- Brüggemann, R., Härdle, W., Mungo, J., Trenkler, C., 2008. VAR modelling for dynamic semiparametric factors of volatility strings. *Journal of Financial Econometrics* 5 (2), 189–218.
- Cao, J., Härdle, W., Mungo, J., 2009. A Joint Analysis of the KOSPI 200 Option and ODAX Option Markets Dynamics. Discussion Paper 019, Collaborative Research Center 649 "Economic Risk". Humboldt-Universität zu Berlin.
- Chacko, G.C., Jurek, J.W., Stafford, E., 2008. The price of immediacy. *Journal of Finance* 63, 1253–1290.
- Degryse, H., Jong, F., Ravenswaaij, M., Wuyts, G., 2005. Aggressive orders and the resiliency of a limit order market. *Review of Finance* 9, 201–242.
- Engle, R.F., Ferstenberg, R., 2007. Execution risk. *Journal of Portfolio Management* 33, 34–45.
- Engle, R.F., Patton, A.J., 2004. Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets* 7, 1–25.
- Fengler, M.R., Härdle, W., Mammen, E., 2007. A dynamic semiparametric factor model for implied volatility string dynamics. *Journal of Financial Econometrics* 5 (2), 189–218.
- Gallant, A.R., 1981. On the bias of flexible functional forms and an essentially unbiased form. *Journal of Econometrics* 15, 211–245.
- Garvey, R., Wu, F., 2009. Intraday time and order execution quality dimensions. *Journal of Financial Markets* 12, 203–228.
- Glosten, L., 1994. Is the electronic Limit Order Book inevitable. *Journal of Finance* 49 (4), 1127–1161.
- Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? *Journal of Financial Economics* 92, 153–181.
- Griffiths, M.D., Smith, B.F., Turnbull, D.A.S., White, R.W., 2000. The costs and determinants of order aggressiveness. *Journal of Financial Economics* 56, 65–88.
- Hall, A.D., Hautsch, N., 2006. Order aggressiveness and order book dynamics. *Empirical Economics* 30, 973–1005.
- Hall, A.D., Hautsch, N., 2007. Modelling the buy and sell intensity in a limit order book market. *Journal of Financial Markets* 10 (3), 249–286.
- Hasbrouck, J., 2009. Trading costs and returns for U.S. equities: estimating effective costs from daily data. *Journal of Finance* 64, 1445–1477.
- Hasbrouck, J., Saar, G., 2009. Technology and liquidity provision: the blurring of traditional definitions. *Journal of Financial Markets* 12, 143–172.
- Hautsch, N., Huang, R., 2012. The market impact of a limit order. *Journal of Economic Dynamics & Control* 36, 501–522.
- Hollifield, B., Miller, R.A., Sandás, P., 2004. Empirical analysis of limit order markets. *Review of Economic Studies* 71, 1027–1063.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59 (6), 1551–1580.
- Johnson, T.C., 2008. Volume, liquidity and liquidity risk. *Journal of Financial Economics* 87, 388–417.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.
- Large, J., 2007. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets* 10, 1–25.
- Liu, W.-M., 2009. Monitoring and limit order submission risks. *Journal of Financial Markets* 12, 107–141.
- Nelson, C.R., Siegel, A.F., 1987. Parsimonious modelling of yield curves. *Journal of Business* 60, 473–489.
- Obizhaeva, A., Wang, J., 2005. Optimal Trading Strategy and Supply/Demand Dynamics. Working Paper 11444, NBER.
- Park, B., Mammen, E., Härdle, W., Borak, S., 2009. Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association* 104 (485), 284–298.
- Rinaldo, A., 2004. Order aggressiveness in limit order book markets. *Journal of Financial Markets* 7, 53–74.
- Schmidt, P., Phillips, P.C.B., 1992. LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.

This article was downloaded by: [Humboldt-Universität zu Berlin Universitätsbibliothek]

On: 25 April 2012, At: 06:53

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Applied Mathematical Finance

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ramf20>

The Implied Market Price of Weather Risk

Wolfgang Karl Härdle^a & Brenda López Cabrera^a

^a Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin, Berlin, Germany

Available online: 17 Oct 2011

To cite this article: Wolfgang Karl Härdle & Brenda López Cabrera (2012): The Implied Market Price of Weather Risk, Applied Mathematical Finance, 19:1, 59-95

To link to this article: <http://dx.doi.org/10.1080/1350486X.2011.591170>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

The Implied Market Price of Weather Risk

WOLFGANG KARL HÄRDLE & BRENDA LÓPEZ CABRERA

Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin, Berlin, Germany

(Received 7 May 2010; in revised form 28 February 2011)

ABSTRACT *Weather derivatives (WD) are end-products of a process known as securitization that transforms non-tradable risk factors (weather) into tradable financial assets. For pricing and hedging non-tradable assets, one essentially needs to incorporate the market price of risk (MPR), which is an important parameter of the associated equivalent martingale measure (EMM). The majority of papers so far has priced non-tradable assets assuming zero or constant MPR, but this assumption yields biased prices and has never been quantified earlier under the EMM framework. Given that liquid-derivative contracts based on daily temperature are traded on the Chicago Mercantile Exchange (CME), we infer the MPR from traded futures-type contracts (CAT, CDD, HDD and AAT). The results show how the MPR significantly differs from 0, how it varies in time and changes in sign. It can be parameterized, given its dependencies on time and temperature seasonal variation. We establish connections between the market risk premium (RP) and the MPR.*

KEY WORDS: CAR process, CME, HDD, seasonal volatility, risk premium

1. Introduction

In the 1990s weather derivatives (WD) were developed to hedge against the random nature of temperature variations that constitute weather risk. WD are financial contracts with payments based on weather-related measurements. WD cover against volatility caused by temperature, rainfall wind, snow, and frost. The key factor in efficient usage of WD is a reliable valuation procedure. However, due to their specific nature one encounters several difficulties. First, because the underlying weather (and indices) is not tradable and second, the WD market is incomplete, meaning that the WD cannot be cost-efficiently replicated by other WD.

Since the largest portion of WD traded at Chicago Mercantile Exchange (CME) is written on temperature indices, we concentrate our research on temperature derivatives. There have been basically four branches of temperature derivative pricing: actuarial approach, indifference pricing, general equilibrium theory or pricing via no arbitrage arguments. While the actuarial approach considers the conditional expectation of the pay-off under the real physical measure discounted at the riskless rate (Brix *et al.*, 2005), the indifference pricing relies on the equivalent utility principle (Barrieu and El Karoui, 2002; Brockett *et al.*, 2010) and the general equilibrium theory assumes

Correspondence Address: Brenda López Cabrera, Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin, Berlin, Germany. Tel: +49(0)30 2093 1457 Email: lopezcab@wiwi.hu-berlin.de

investors' preferences and rules of Pareto optimal risk allocation (Cao and Wei, 2004; Horst and Mueller, 2007; Richards *et al.*, 2004). The Martingale approach, although less demanding in terms of assumptions, concentrates on the econometric modelling of the underlying dynamics and requires the selection of an adequate equivalent martingale measure (EMM) to value the pay-offs by taking expectations (Alaton *et al.*, 2002; Benth, 2003; Benth and Saltyte-Benth, 2007; Benth *et al.*, 2007; Brody *et al.*, 2002; Huang-Hsi *et al.*, 2008; Mraoua and Bari, 2007).

Here we prefer the latter approach. First, since the underlying (temperature) we consider is of local nature, our analysis aims at understanding the pricing at different locations around the world. Second, the EMM approach helps identify the market price of risk (MPR), which is an important parameter of the associated EMM, and it is indispensable for pricing and hedging non-tradable assets. The MPR can be extracted from traded securities and one can use this value to price other derivatives, though any inference about the MPR requires an assumption about its specification.

The MPR is of high scientific interest, not only for financial risk analysis, but also for better economic modelling of fair valuation of risk. Constantinides (1987) and Landskroner (1977) studied the MPR of tradable assets in the Capital Asset Pricing Model (CAPM) framework. For pricing interest rate derivatives, Vasicek (1977) assumed a constant market price of interest rate, while Hull and White (1990) used the specification of Cox *et al.* (1985). In the oil market, Gibson and Schwartz (1990) supposed an intertemporal constant market price of crude oil conveniences yield risk. Benth *et al.* (2008) introduced a parameterization of the MPR to price electricity derivatives. In the WD framework, Cao and Wei (2004) and Richards *et al.* (2004) studied the MPR as an implicit parameter in a generalization of the Lucas' (1978) equilibrium framework. They showed that the MPR is not only statistically significant for temperature derivatives, but also economically large as well. However, calibration problems arise with the methodology suggested by Cao and Wei (2004), since it deals with a global model like the Lucas' (1978) approach while weather is locally specified. Benth and Saltyte-Benth (2007) introduced theoretical ideas of equivalent changes of measure to get no arbitrage futures/option prices written on different temperature indices. Huang-Hsi *et al.* (2008) examined the MPR of the Taiwan Stock Exchange Capitalization-Weighted Stock Index ((the mean of stock returns – risk-free interest rate)/SD of stock returns) and used it as a proxy for the MPR on temperature option prices. The majority of temperature pricing papers so far has priced temperature derivatives assuming 0 or constant MPR (Alaton *et al.*, 2002; Cao and Wei, 2004; Dorfleitner and Wimmer, 2010; Huang-Hsi *et al.*, 2008; Richards *et al.*, 2004), but this assumption yields biased prices and has never been quantified earlier using the EMM framework. This article deals exactly with the differences between 'historical' and 'risk neutral' behaviours of temperature.

The contribution of this article is threefold. First, in contrast to Campbell and Diebold (2005), Benth and Saltyte-Benth (2007) and Benth *et al.* (2007), we correct for seasonality and seasonal variation of temperature with a local smoothing approach to get, independently of the chosen location, the driving stochastics close to a Gaussian Process and with that being able to apply pricing technique tools of financial mathematics (Karatzas and Shreve, 2001). Second and the main contribution, using statistical modelling and given that liquid derivative contracts based on daily

temperature are traded on the CME, we imply the MPR from traded futures-type contracts (CAT/HDD/CDD/AAT) based on a well-known pricing model developed by Benth *et al.* (2007). We have chosen this methodology because it is a stationary model that fits the stylized characteristics of temperature well; it nests a number of previous models (Alaton *et al.*, 2002; Benth, 2003; Benth and Saltyte-Benth, 2005, 2007; Brody *et al.*, 2002; Dornier and Querel, 2007); it provides closed-form pricing formulas; and it computes, after deriving the MPR, non-arbitrage prices based on a continuous-time hedging strategy. Moreover, the price dynamics of futures are easy to compute and require only a one-time estimation. Our implied MPR approach is a calibration procedure for financial engineering purposes. In the calibration exercise, a single date (but different time horizons and calibrated instruments are used) is required, since the model is recalibrated daily to detect intertemporal effects. Moreover, we use an economic and statistical testing approach, where we start from a specification of the MPR and check consistency with the data. The aim of this analysis is to study the effect of different MPR specifications (a constant, a (two) piecewise linear function, a time-deterministic function and a ‘financial-bootstrapping’) on the temperature futures prices. The statistical point of view is to beat this as an inverse problem with different degrees of smoothness expressed through the penalty parameter of a smoothing spline. The degrees of smoothness will allow for a term structure of risk. Since smoothing estimates are fundamentally different from estimating a deterministic function, we also assure our results by fitting a parametric function to all available contract prices (calendar year estimation). The economic point of view is to detect possible time dependencies that can be explained by investor’s preferences in order to hedge weather risk. Our findings that the MPR differs significantly from 0 confirm the results found in Cao and Wei (2004), Huang-Hsi *et al.* (2008), Richards *et al.* (2004) and Alaton *et al.* (2002), but we differ from them, by showing that it varies in time and changes in sign. It is not a reflection of bad model specification, but data-extracted MPR. This contradicts the assumption made earlier in the literature that the MPR is 0 or constant and rules out the ‘burn-in’ analysis, which is popular among practitioners since it uses the historical average index value as the price for the futures (Brix *et al.*, 2005). This brings significant challenges to the statistical branch of the pricing literature. We also establish connections between the market risk premium (RP) (a Girsanov-type change of probability) and the MPR. As a third contribution, we discuss how to price over-the-counter (OTC) temperature derivatives with the information extracted.

Our article is structured as follows. Section 2 presents the fundamentals of temperature derivatives (futures and options) and describes the temperature data and the temperature futures traded at CME, the biggest market offering this kind of product. Section 3 is devoted to explaining the dynamics of temperature data – the econometric part. The temperature model captures linear trend, seasonality, mean reversion, intertemporal correlations and seasonal volatility effects. Section 4 – the financial mathematics part – connects the weather dynamics with the pricing methodology. Section 5 solves the inverse problem of determining the MPR of CME temperature futures using different specifications. Section 1 introduces the estimation results and test procedures of our specifications applied into temperature-derivative data. Here we give (statistical and economic) interpretations of the estimated MPR. The pricing of

OTC temperature products is discussed at the end of this section. Section 6 concludes the article. All computations in this article were carried out in Matlab version 7.6 (The MathWorks, Inc., Natick, MA, USA). To simplify notation, dates are denoted with yyyyymmdd format.

2. Temperature Derivatives

The largest portion of futures and options written on temperature indices is traded on the CME. Most of the temperature derivatives are written on daily average temperature indices, rather than on the underlying temperature by itself. A call option written on futures $F_{(t, \tau_1, \tau_2)}$ with exercise time $t \leq \tau_1$ and delivery over a period $[\tau_1, \tau_2]$ will pay $\max\{F_{(t, \tau_1, \tau_2)} - K, 0\}$ at the end of the measurement period $[\tau_1, \tau_2]$. The most common weather indices on temperature are Heating Degree Day (HDD), Cooling Degree Day (CDD) and Cumulative Averages (CAT). The HDD index measures the temperature over a period $[\tau_1, \tau_2]$, usually between October and April:

$$HDD(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} \max(c - T_u, 0) du, \quad (1)$$

where c is the baseline temperature (typically 18°C or 65°F) and $T_u = (T_{u, \max} + T_{u, \min})/2$ is the average temperature on day u . Similarly, the CDD index measures the temperature over a period $[\tau_1, \tau_2]$, usually between April and October:

$$CDD(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} \max(T_u - c, 0) du. \quad (2)$$

The HDD and the CDD index are used to trade futures and options in 24 US cities, 6 Canadian cities and 3 Australian cities. The CAT index accounts the accumulated average temperature over $[\tau_1, \tau_2]$:

$$CAT(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} T_u du. \quad (3)$$

The CAT index is the substitution of the CDD index for 11 European cities. Since $\max(T_u - c, 0) - \max(c - T_u, 0) = T_u - c$, we get the HDD–CDD parity:

$$CDD(\tau_1, \tau_2) - HDD(\tau_1, \tau_2) = CAT(\tau_1, \tau_2) - c(\tau_2 - \tau_1). \quad (4)$$

Therefore, it is sufficient to analyse only HDD and CAT indices. An index similar to the CAT index is the Pacific Rim Index, which measures the accumulated total of 24-hr average temperature (C24AT) over a period $[\tau_1, \tau_2]$ days for Japanese cities:

$$C24AT(\tau_1, \tau_2) = \int_{\tau_1}^{\tau_2} \tilde{T}_u du, \quad (5)$$

where $\tilde{T}_u = \frac{1}{24} \int_1^{24} T_{u_i} du_i$ and T_{u_i} denotes the temperature of hour u_i . A difference of the CAT and the C24AT index is that the latter is traded over the whole year. Note that

temperature is a continuous-time process even though the indices used as underlying for temperature futures contracts are discretely monitored.

As temperature is not a marketable asset, the replication arguments for any temperature futures contract do not hold and incompleteness of the market follows. In this context, any probability measure Q equivalent to the objective P is also an EMM and a risk neutral probability turns all the tradable assets into martingales after discounting. However, since temperature futures/option prices dynamics are indeed tradable assets, they must be free of arbitrage. Thanks to the Girsanov theorem, equivalent changes of measures are simply associated with changes of drift. Hence, under a probability space (Ω, \mathcal{F}, Q) with a filtration $\{\mathcal{F}_t\}_{0 \leq t \leq \tau_{\max}}$, where τ_{\max} denotes a maximal time covering all times of interest in the market, we choose a parameterized equivalent pricing measure $Q = Q_\theta$ that completes the market and pin it down to compute the arbitrage-free temperature futures price:

$$F_{(t, \tau_1, \tau_2)} = E^{Q_\theta} [Y | \mathcal{F}_t], \tag{6}$$

where Y refers to the pay-off from the temperature index in Equations (2)–(5). The MPR θ is assumed to be a real-valued, bounded and piecewise continuous function. We later relax that assumption, by considering the time-dependent MPR θ_t . In fact, the MPR can depend on anything that can affect investors’ attitudes. The MPR can be inferred from market data.

The choice of Q determines the RP demanded for investors for holding the temperature derivative, and opposite, having knowledge of the RP determines the choice of the risk-neutral probability. The RP is defined as a drift of the spot dynamics or a Girsanov-type change of probability. In Equation (6), the futures price is set under a risk-neutral probability $Q = Q_\theta$, thereby the RP measures exactly the differences between the risk-neutral $F_{(t, \tau_1^i, \tau_2^i, Q)}$ (market prices) and the temperature market probability predictions $\hat{F}_{(t, \tau_1^i, \tau_2^i, P)}$ (under P):

$$RP = F_{(t, \tau_1^i, \tau_2^i, Q)} - \hat{F}_{(t, \tau_1^i, \tau_2^i, P)}. \tag{7}$$

Using the ‘burn-in’ approach of Brix *et al.* (2005), the futures price is only the historical average index value, therefore there is no RP since $Q = P$.

2.1 Data

We have temperature data available from 35 US, 30 German, 159 Chinese and 9 European weather stations. The temperature data were obtained from the National Climatic Data Center (NCDC), the Deutscher Wetterdienst (DWD), Bloomberg Professional Service, the Japanese Meteorological Agency (JMA) and the China Meteorological Administration (CMA). The temperature data contain the minimum, maximum and average daily temperatures measured in degree Fahrenheit for US cities and degree Celsius for other cities. The data set period is, in most of the cities, from 19470101 to 20091231.

The WD data traded at CME were provided by Bloomberg Professional Service. We use daily closing prices from 20000101 to 20091231. The measurement periods for the

different temperature indices are standardized to be as each month of the year or as seasonal strips (minimum of 2 and maximum of 7 consecutive calendar months). The futures and options at the CME are cash settled, that is, the owner of a futures contract receives 20 times the index at the end of the measurement period, in return for a fixed price. The currency is British pounds for the European futures contracts, US dollars for the US contracts and Japanese Yen for the Asian cities. The minimum price increment is 1 index point. The degree day metric is Celsius or Fahrenheit and the termination of the trading is two calendar days following the expiration of the contract month. The accumulation period of each CAT/CDD/HDD/C24AT index futures contract begins with the first calendar day of the contract month and ends with the calendar day of the contract month. Earth Satellite Corporation (ESC) reports to CME the daily average temperature. Traders bet that the temperature will not exceed the estimates from ESC.

3. Temperature Dynamics

In order to derive explicitly no arbitrage prices for temperature derivatives, we need first to describe the dynamics of the underlying under the physical measure. This article studies the average daily temperature data (because most of the temperature derivative trading is based on this quantity) for US, European and Asian cities. In particular, we analyse the weather dynamics for Atlanta, Portland, Houston, New York, Berlin, Essen, Tokyo, Osaka, Beijing and Taipei (Table 1). Our interest in these cities is because all of them with the exception of the latter two are traded at CME and because a casual examination of the trading statistics on the CME website reveals that the Atlanta HDD, Houston CDD and Portland CDD temperature contracts have relatively more liquidity.

Most of the literature that discuss models for daily average temperature and capture a linear trend (due to global warming and urbanization), seasonality (peaks in cooler winter and warmer summers), mean reversion, seasonal volatility (a variation that varies seasonally) and strong correlations (long memory); see, for example, Alaton *et al.* (2002), Cao and Wei (2004), Campbell and Diebold (2005) and Benth *et al.* (2007). They differ from their definition of temperature variations, which is exactly the component that characterizes weather risk. Here we show that an autoregressive (AR) model AR of high order (p) for the detrended daily average temperatures (rather than the underlying temperature itself) is enough to capture the stylized facts of temperature.

We first need to remove the seasonality in mean Λ_t from the daily temperature series T_t , check intertemporal correlations and remove the seasonality in variance to deal with a stationary process. The deterministic seasonal mean component can be approximated with Fourier-truncated series (FTS):

$$\Lambda_t = a + bt + \sum_{l=1}^L c_l \cos \left\{ \frac{2\pi(t - d_l)}{l \cdot 365} \right\}, \quad (8)$$

where the coefficients a and b indicate the average daily temperature and global warming, respectively. We observe low temperatures in winter times and high temperatures in summer for different locations. The temperature data sets do not deviate from its

Table 1. Coefficients of the Fourier-truncated seasonal series of average daily temperatures in different cities.

City	Period	\hat{a} (CI)	\hat{b} (CI)	\hat{c}_1 (CI)	\hat{d}_1 (CI)
Atlanta	19480101–20081204	61.95 (61.95, 61.96)	-0.0025 (-0.0081, 0.0031)	18.32 (18.31, 18.33)	-165.02 (-165.03, -165.02)
Beijing	19730101–20090831	12.72 (12.71, 12.73)	0.0001 (-0.0070, 0.0073)	14.93 (14.92, 14.94)	-169.59 (-169.59, -169.58)
Berlin	19480101–20080527	9.72 (9.71, 9.74)	-0.0004 (-0.0147, 0.0139)	9.75 (9.74, 9.77)	-164.79 (-164.81, -164.78)
Essen	19700101–20090731	10.80 (10.79, 10.81)	-0.0020 (-0.0134, 0.0093)	8.02 (8.01, 8.03)	-161.72 (-161.73, -161.71)
Houston	19700101–20081204	68.52 (68.51, 68.52)	-0.0006 (-0.0052, 0.0039)	15.62 (15.62, 15.63)	-165.78 (-165.79, -165.78)
New York	19490101–20081204	53.86 (53.86, 53.87)	-0.0004 (-0.0079, 0.0071)	21.43 (21.42, 21.44)	-156.27 (-156.27, -156.26)
Osaka	19730101–20090604	16.78 (16.77, 16.79)	-0.0021 (-0.0109, 0.0067)	11.61 (11.60, 11.62)	-153.57 (-153.58, -153.56)
Portland	19480101–20081204	55.35 (55.35, 55.36)	-0.0116 (-0.0166, -0.0065)	14.36 (14.36, 14.37)	-155.58 (-155.58, -155.57)
Taipei	19920101–20090806	23.32 (23.31, 23.33)	0.0023 (-0.0086, 0.0133)	6.67 (6.66, 6.68)	-158.67 (-158.68, -158.66)
Tokyo	19730101–20090831	16.32 (16.31, 16.33)	-0.0003 (-0.0085, 0.0079)	10.38 (10.37, 10.38)	-153.52 (-153.53, -153.52)

Notes: CI, Confidence interval.

All coefficients are non-zero at 1% significance level. CIs are given in parentheses. Dates given in yyyyymmdd format. The daily temperature is measured in degree Celsius, except for American cities measured in degree Fahrenheit.

mean level and in most of the cases a linear trend at 1% significance level is detectable as it is displayed in Table 1.

Our findings are similar to Alaton *et al.* (2002) and Benth *et al.* (2007) for Sweden; Benth *et al.* (2007) for Lithuania; Campbell and Diebold (2005) for the United States; and Papazian and Skiadopoulos (2010) for Barcelona, London, Paris and Rome. In our empirical results, the number of periodic terms of the FTS series varies from city to city, sometimes from 4 to 21 or more terms. We notice that the series expansion in Equation (8) with more and more periodic terms provides a fine tuning, but this will increase the number of parameters. Here we propose a different way to correct for seasonality. We show that a local smoothing approach does that job instead, but with less technical expression. Asymptotically they can be approximated by FTS estimators. For a fixed time point $s \in [1, 365]$, we smooth Λ_s with a Local Linear Regression (LLR) estimator:

$$\Lambda_s = \arg \min_{e, f} \sum_{t=1}^{365} \{ \bar{T}_t - e_s - f_s(t-s) \}^2 K\left(\frac{t-s}{h}\right), \quad (9)$$

where \bar{T}_t is the mean of average daily temperature in J years, h is the smoothing parameter and $K(\cdot)$ denotes a kernel. This estimator, using Epanechnikov Kernel, incorporates an asymmetry term since high temperatures in winter are more pronounced than in summer as Figure 1 displays in a stretch of 8 years plot of the average daily temperatures over the FTS estimates.

After removing the LLR-seasonal mean (Equation (9)) from the daily average temperatures ($X_t = T_t - \Lambda_t$), we apply the Augmented Dickey–Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests to check whether X_t is a stationary process. We then plot the Partial Autocorrelation Function (PACF) of X_t to detect possible intertemporal correlations. This suggests that persistence of daily average is captured by AR processes of higher order p :

$$X_{t+p} = \sum_{i=1}^p \beta_i X_{t-i} + \varepsilon_t, \varepsilon_t = \sigma_t e_t, e_t \sim N(0, 1), \quad (10)$$

with residuals ε_t . Under the stationarity hypothesis of the coefficients β s and the mean zero of residuals ε_t , the mean temperature $E[T_t] = \Lambda_t$. This is different to the approach of Campbell and Diebold (2005), who suggested to regress deseasonalized temperatures on original temperatures. The analysis of the PACFs and Akaike's information criterion (AIC) suggests that the $AR(3)$ model in Benth *et al.* (2007) explains the temperature evolution well and holds for many cities. The results of the stationarity tests and the coefficients of the fitted $AR(3)$ are given in Table 2. Figure 2 illustrates that the ACFs of the residuals ε_t are close to 0 and according to Box-Ljung statistic the first few lags are insignificant, whereas the ACFs of the squared residuals ε_t^2 show a high seasonal pattern.

We calibrate the deterministic seasonal variance function σ_t^2 with FTS and an additional generalized autoregressive conditional heteroskedasticity (GARCH) (p, q) term:

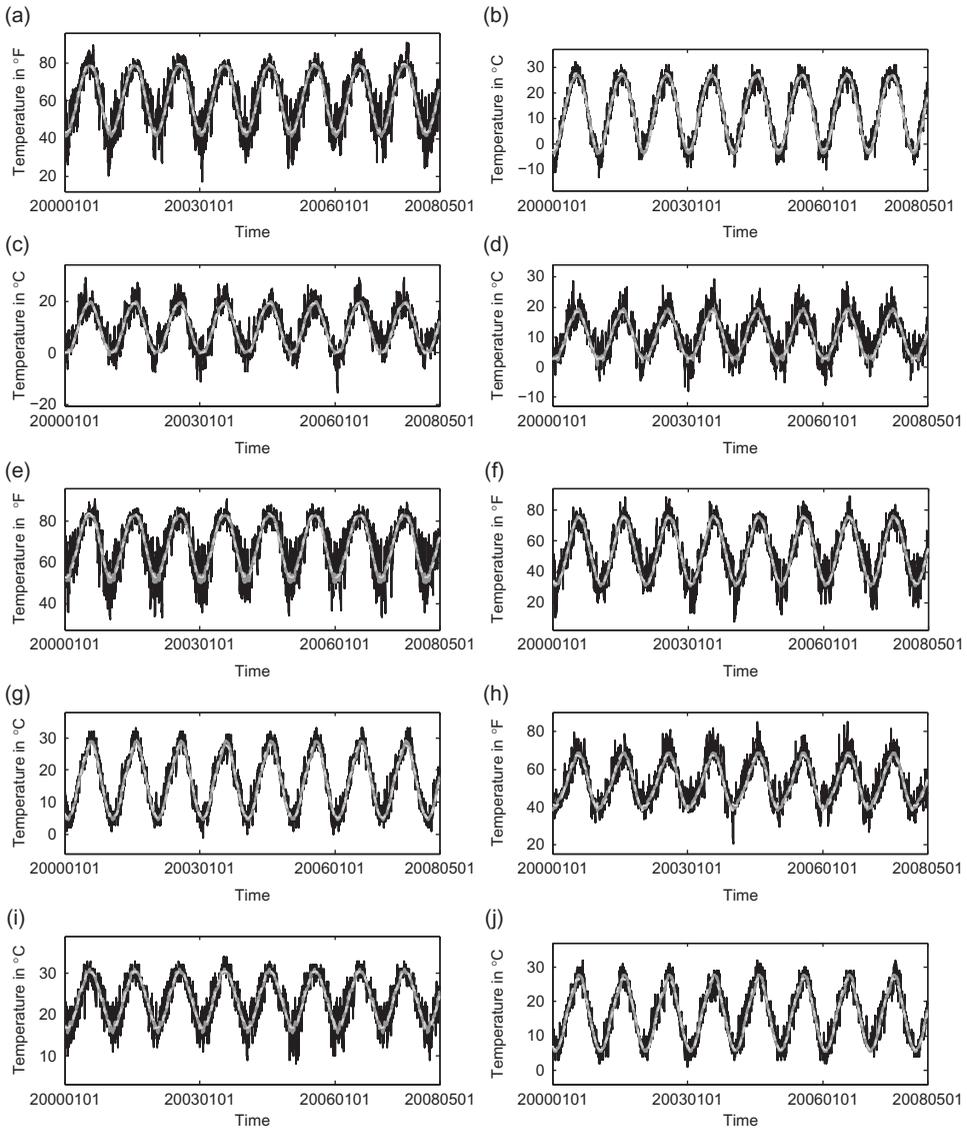


Figure 1. A stretch of 8 years plot of the average daily temperatures (black line), the seasonal component modelled with a Fourier-truncated series (dashed line) and the local linear regression (grey line) using Epanechnikov Kernel. (a) Atlanta, (b) Beijing, (c) Berlin, (d) Essen, (e) Houston, (f) New York, (g) Osaka, (h) Portland, (i) Taipei and (j) Tokyo.

$$\hat{\sigma}_t^2 = c + \sum_{l=1}^L \left\{ c_{2l} \cos\left(\frac{2l\pi t}{365}\right) + c_{2l+1} \sin\left(\frac{2l\pi t}{365}\right) \right\} + \alpha_1(\sigma_{t-1}\eta_{t-1})^2 + \beta_1\sigma_{t-1}^2\eta_t \quad (11)$$

$$\sim iid N(0, 1).$$

Table 2. Result of the stationary tests and the coefficients of the fitted $AR(3)$.

City	ADF–KPSS		$AR(3)$			$CAR(3)$				
	$\hat{\tau}$	\hat{k}	β_1	β_2	β_3	α_1	α_2	α_3	λ_1	$\lambda_{2,3}$
Atlanta	–55.55 ⁺	0.21 ^{***}	0.96	–0.38	0.13	2.03	1.46	0.28	–0.30	–0.86
Beijing	–30.75 ⁺	0.16 ^{***}	0.72	–0.07	0.05	2.27	1.63	0.29	–0.27	–1.00
Berlin	–40.94 ⁺	0.13 ^{**}	0.91	–0.20	0.07	2.08	1.37	0.20	–0.21	–0.93
Essen	–23.87 ⁺	0.11 [*]	0.93	–0.21	0.11	2.06	1.34	0.16	–0.16	–0.95
Houston	–38.17 ⁺	0.05 [*]	0.90	–0.39	0.15	2.09	1.57	0.33	–0.33	–0.87
New York	–56.88 ⁺	0.08 [*]	0.76	–0.23	0.11	2.23	1.69	0.34	–0.32	–0.95
Osaka	–18.65 ⁺	0.09 [*]	0.73	–0.14	0.06	2.26	1.68	0.34	–0.33	–0.96
Portland	–45.13 ⁺	0.05 [*]	0.86	–0.22	0.08	2.13	1.48	0.26	–0.27	–0.93
Taipei	–32.82 ⁺	0.09 [*]	0.79	–0.22	0.06	2.20	1.63	0.36	–0.40	–0.90
Tokyo	–25.93 ⁺	0.06 [*]	0.64	–0.07	0.06	2.35	1.79	0.37	–0.33	–1.01

Notes: ADF, augmented Dickey–Fuller; KPSS, Kwiatkowski–Phillips–Schmidt–Shin; AR, autoregressive process; CAR, continuous autoregressive model.

ADF and KPSS statistics, coefficients of the $AR(3)$, $CAR(3)$ and eigenvalues $\lambda_{1,2,3}$, of the matrix \mathbf{A} of the $CAR(3)$ model for the detrended daily average temperatures for different cities.

⁺0.01 critical values, ^{*}0.1 critical value, ^{**}0.05 critical value, ^{***}0.01 critical value.

The Fourier part in Equation (11) captures the seasonality in volatility, whereas the GARCH part captures the remaining non-seasonal volatility. Note again that more and more periodic terms in Equation (11) provide a good fitting but this will increase the number of parameters. To avoid this and in order to achieve positivity of the variance, Gaussian risk factors and volatility model flexibility in a continuous time, we propose the calibration of the seasonal variance in terms of an LLR:

$$\arg \min_{g,h} \sum_{t=1}^{365} \{\hat{\varepsilon}_t^2 - g_s - h_s(t-s)\}^2 K\left(\frac{t-s}{h}\right), \quad (12)$$

where $\hat{\varepsilon}_t^2$ is the mean of squared residuals in J years and $K(\cdot)$ is a kernel. Figure 3 shows the daily empirical variance (the average of squared residuals for each day of the year), the fittings using the FTS-GARCH(1,1) and the LLR (with Epanechnikov kernel) estimators. Here we obtain the Campbell and Diebold’s (2005) effect for different temperature data, high variance in winter to earlier summer and low variance in spring to late summer. The effects of non-seasonal GARCH volatility component are small.

Figure 4 displays the ACFs of temperature residuals ε_t and squared residuals ε_t^2 after dividing out the deterministic LLR seasonal variance. The ACF plots of the standardized residuals remain unchanged but now the squared residuals presents a non-seasonal pattern. The LLR seasonal variance creates almost normal residuals and captures the peak seasons as Figure 5 in a log Kernel smoothing density plot shows against a Normal Kernel evaluated at 100 equally spaced points. Table 3 presents the calibrated coefficients of the FTS-GARCH seasonal variance estimates and the

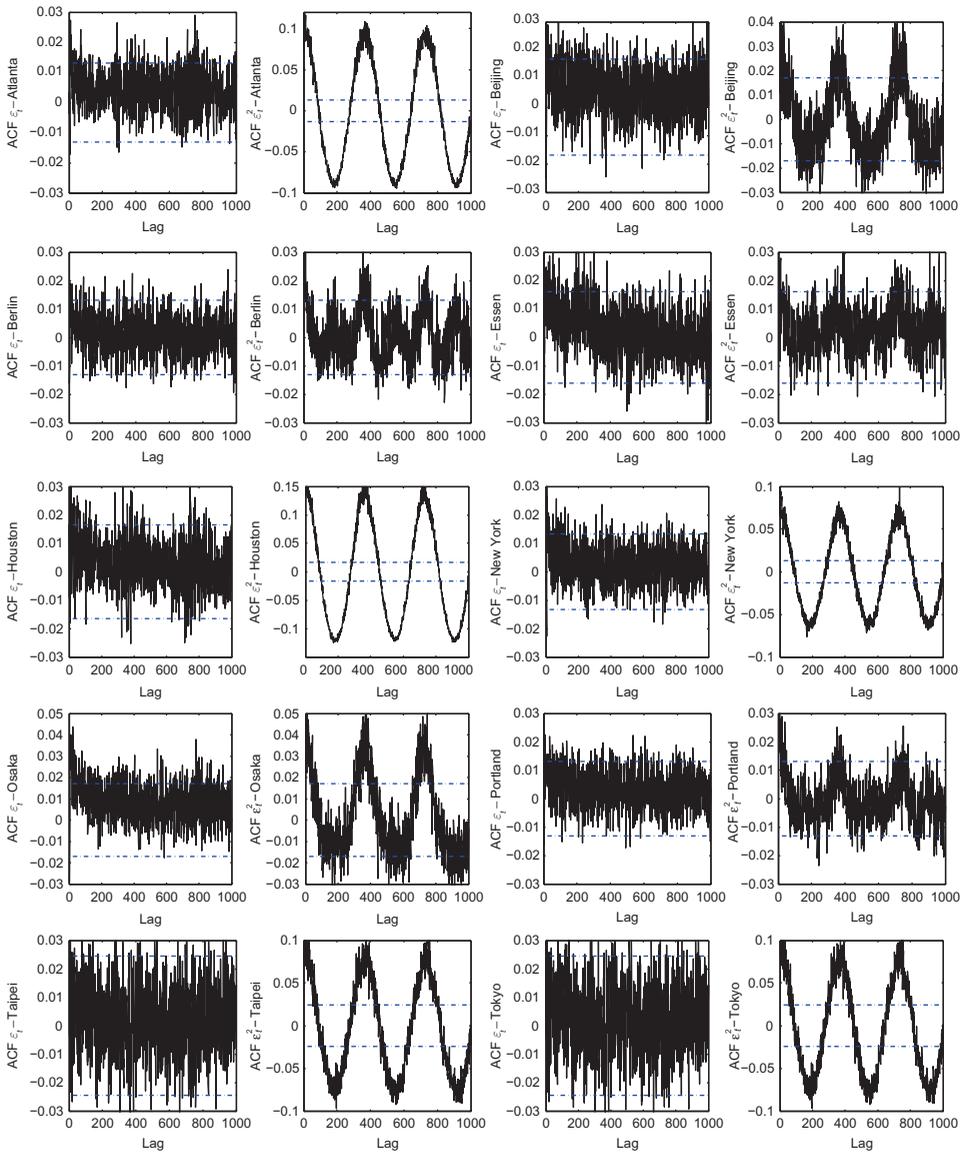


Figure 2. The ACF of residuals ε_t (left panels) and squared residuals ε_t^2 (right panels) of detrended daily temperatures for different cities.

descriptive statistics for the residuals after correcting by the FTS-GARCH and LLR seasonal variance. We observe that independently of the chosen location, the driving stochastics are close to a Wiener process. This will allow us to apply the pricing tools of financial mathematics.

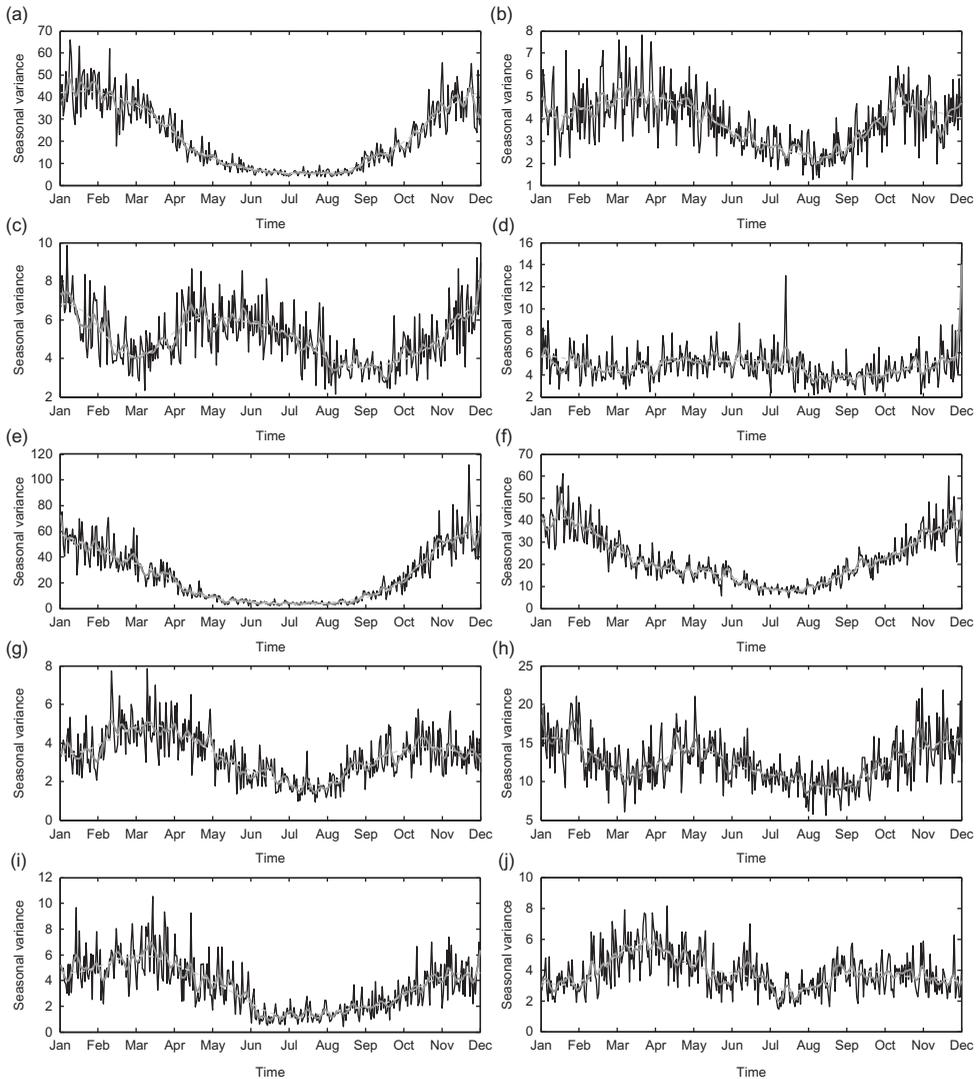


Figure 3. The daily empirical variance (black line), the Fourier-truncated (dashed line) and the local linear smoother seasonal variation using Epanechnikov kernel (grey line) for different cities. (a) Atlanta, (b) Beijing, (c) Berlin, (d) Essen, (e) Houston, (f) New York, (g) Osaka, (h) Portland, (i) Taipei and (j) Tokyo.

4. Stochastic Pricing Model

Temperatures are naturally evolving continuously over time, so it is very convenient to model the dynamics of temperature with continuous-time stochastic processes, although the data may be on a daily scale. We therefore need the reformulation of the underlying process in continuous time to be more convenient with market definitions.

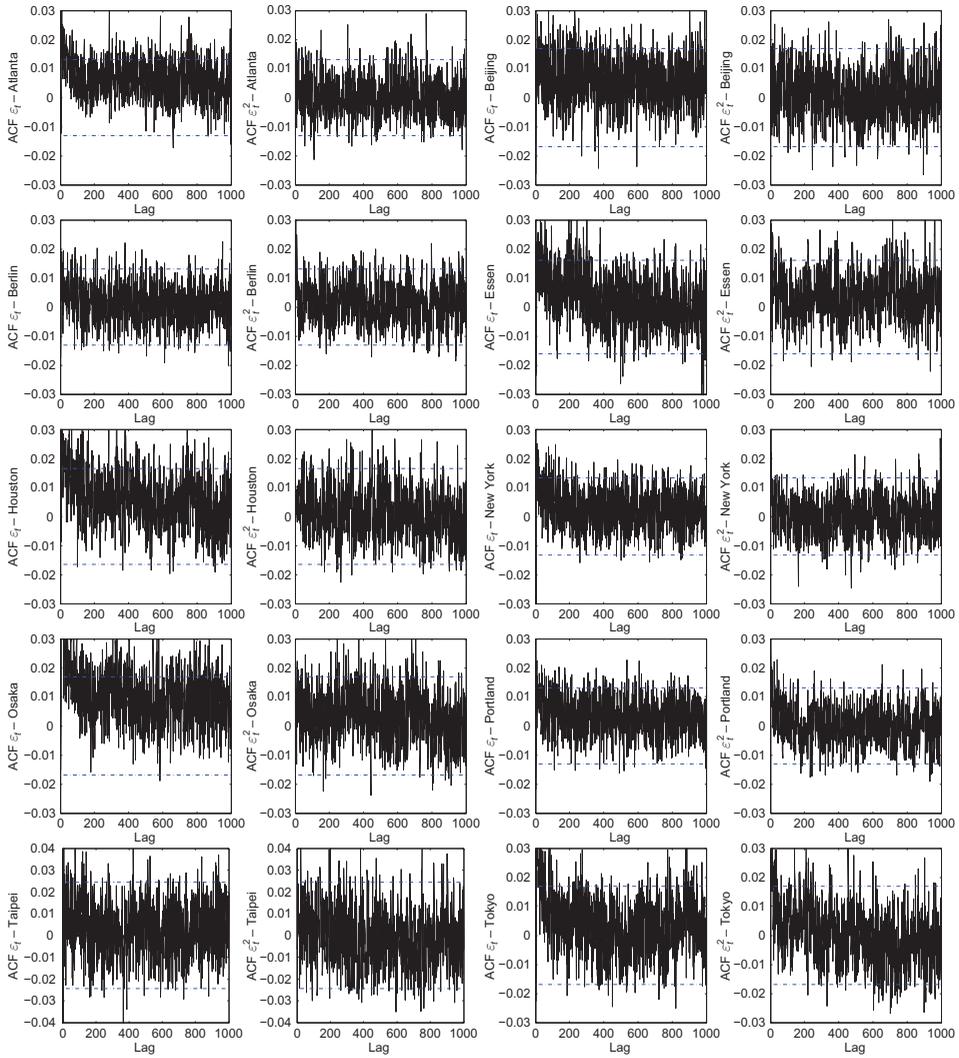


Figure 4. The ACF of residuals ε_t (left panels) and squared residuals ε_t^2 (right panels) of detrended daily temperatures after dividing out the local linear seasonal variance for different cities.

We show that the $AR(p)$ (Equation (10)) estimated in Section 3 for the detrended temperature can be therefore seen as a discretely sampled continuous-time autoregressive (CAR) process ($CAR(p)$) driven by a one-dimensional Brownian motion B_t (though the continuous-time process is Markov in higher dimension) (Benth *et al.*, 2007):

$$dX_t = \mathbf{A}X_t dt + \mathbf{e}_p \sigma_t dB_t, \tag{13}$$

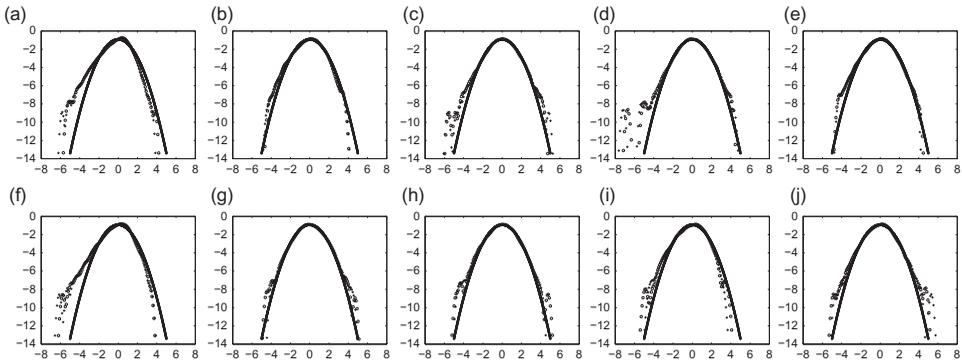


Figure 5. The log of Normal Kernel (*) and log of Kernel smoothing density estimate of residuals after correcting FTS (+) and local linear (o) seasonal variance for different cities (a) Atlanta, (b) Beijing, (c) Berlin, (d) Essen, (e) Houston, (f) New York, (g) Osaka, (h) Portland, (i) Taipei and (j) Tokyo.

where the state vector $X_t \in \mathbb{R}^p$ for $p \geq 1$ is a vectorial Ornstein–Uhlenbeck process, namely, the temperatures after removing seasonality at times $t, t-1, t-2, t-3, \dots$; e_k denotes the k th unit vector in \mathbb{R}^p for $k = 1, \dots, p$; $\sigma_t > 0$ is a deterministic volatility (real-valued and square integrable function); and \mathbf{A} is a $p \times p$ matrix:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \dots & \dots & & 0 & 0 & 1 \\ -\alpha_p & -\alpha_{p-1} & \dots & & 0 & -\alpha_1 \end{pmatrix}, \quad (14)$$

with positive constants α_k . Following this nomenclature, $X_{q(t)}$ with $q = 1, \dots, p$ is the q th coordinate of X_t and by setting $q = 1$ is equivalent to the detrended temperature time series $X_{1(t)} = T_t - \Delta_t$. The proof is derived by an analytical link between $X_{1(t)}, X_{2(t)}$ and $X_{3(t)}$ and the lagged temperatures up to time $t-3$. $X_{1(t+3)}$ is approximated by Euler discretization. Thus for $p = 1, X_t = X_{1(t)}$ and Equation (13) becomes

$$dX_{1(t)} = -\alpha_1 X_{1(t)} dt + \sigma_t dB_t, \quad (15)$$

which is the continuous version of an $AR(1)$ process. Similarly for $p = 2$, assume a time step of length 1 $dt = 1$ and substitute $X_{2(t)}$ iteratively to get

$$X_{1(t+2)} \approx (2 - \alpha_1)X_{1(t+1)} + (\alpha_1 - \alpha_2 - 1)X_{1(t)} + \sigma_t e_t, \quad (16)$$

Table 3. Coefficients of the FTS-GARCH seasonal variance.

City	Coefficients of the FTS										Corrected residuals $\frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$ with					
											FTS			LLR		
	\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5	\hat{c}_6	\hat{c}_7	\hat{c}_8	\hat{c}_9	\hat{c}_{10}	JB	Kurt	Skew	JB	Kurt	Skew
Atlanta	21.51	18.10	7.09	2.35	1.69	-0.39	-0.68	0.24	-0.45	272.01	3.98	-0.70	253.24	3.91	-0.68	
Beijing	3.89	0.70	0.84	-0.22	-0.49	-0.20	-0.14	-0.11	0.08	219.67	3.27	-0.28	212.46	3.24	-0.28	
Berlin	5.07	0.10	0.72	0.98	-0.43	0.45	0.06	0.16	0.22	224.55	3.48	-0.05	274.83	3.51	-0.08	
Essen	4.78	0.00	0.42	0.63	-0.20	0.17	-0.06	0.05	0.17	273.90	3.65	-0.05	251.89	3.61	-0.08	
Houston	23.61	25.47	4.49	6.65	-0.38	1.00	-2.67	0.68	-1.56	140.97	3.96	-0.60	122.83	3.87	-0.57	
New York	22.29	13.80	3.16	3.30	-0.47	0.80	2.04	0.11	0.01	367.38	3.43	-0.23	355.03	3.43	-0.22	
Osaka	3.34	0.80	0.80	-0.57	-0.27	-0.18	-0.07	0.01	-0.03	105.32	3.37	-0.11	101.50	3.36	-0.11	
Portland	12.48	1.55	1.05	1.42	-1.19	0.46	0.34	-0.40	0.45	67.10	3.24	0.06	75.01	3.27	0.02	
Taipei	3.50	1.49	1.59	-0.38	-0.16	0.03	-0.17	-0.09	-0.18	181.90	3.26	-0.39	169.41	3.24	-0.37	
Tokyo	3.80	0.01	0.73	-0.69	-0.33	-0.14	-0.14	0.26	-0.13	137.93	3.45	-0.10	156.58	3.46	-0.13	

Notes: FTS, Fourier-truncated series; JB, Jarque Bera; LLR, local linear regression; GARCH, generalized autoregressive conditional heteroskedasticity. Seasonal variance estimate of $\{c_{ij}\}_{j=1}^9$, fitted with an FTS and statistics – Skewness (Skew), kurtosis (Kurt) and JB test statistics – of the standardized residuals with seasonal variances fitted with FTS-GARCH and with LLR. Coefficients are significant at 1% level.

where $e_t = B_{t+1} - B_t$. For $p = 3$, we have:

$$\begin{aligned}
 X_{1(t+1)} - X_{1(t)} &= X_{2(t)}dt, \\
 X_{2(t+1)} - X_{2(t)} &= X_{3(t)}dt, \\
 X_{3(t+1)} - X_{3(t)} &= -\alpha_3 X_{1(t)}dt - \alpha_2 X_{2(t)}dt - \alpha_1 X_{3(t)}dt + \sigma_t e_t, \\
 &\dots, \\
 X_{1(t+3)} - X_{1(t+2)} &= X_{2(t+2)}dt, \\
 X_{2(t+3)} - X_{2(t+2)} &= X_{3(t+2)}dt, \\
 X_{3(t+3)} - X_{3(t+2)} &= -\alpha_3 X_{1(t+2)}dt - \alpha_2 X_{2(t+2)}dt - \alpha_1 X_{3(t+2)}dt + \sigma_t e_t,
 \end{aligned} \tag{17}$$

substituting into the $X_{1(t+3)}$ dynamics and setting $dt = 1$:

$$X_{1(t+3)} \approx \underbrace{(3 - \alpha_1)}_{\beta_1} X_{1(t+2)} + \underbrace{(2\alpha_1 - \alpha_2 - 3)}_{\beta_2} X_{1(t+1)} + \underbrace{(-\alpha_1 + \alpha_2 - \alpha_3 + 1)}_{\beta_3} X_{1(t)}. \tag{18}$$

Please note that this corrects the derivation in Benth *et al.* (2007) and Equation (18) leads to Equation (10) (with $p = 3$). The approximation of Equation (18) is required to compute the eigenvalues of matrix \mathbf{A} . The last columns of Table 2 display the *CAR*(3)-parameters and the eigenvalues of the matrix \mathbf{A} for the studied temperature data. The stationarity condition is fulfilled since the eigenvalues of \mathbf{A} have negative real parts and the variance matrix $\int_0^t \sigma_{t-s}^2 \exp\{\mathbf{A}(s)\} \mathbf{e}_p \mathbf{e}_p^\top \exp\{\mathbf{A}^\top(s)\} ds$ converges as $t \rightarrow \infty$.

By applying the multidimensional *Itô Formula*, the process in Equation (13) with $X_t = x \in \mathbb{R}^p$ has the explicit form $X_s = \exp\{\mathbf{A}(s-t)\}x + \int_t^s \exp\{\mathbf{A}(s-u)\} \mathbf{e}_p \sigma_u dB_u$ for $s \geq t \geq 0$.

Since dynamics of temperature futures prices must be free of arbitrage under the pricing equivalent measure Q_θ , the temperature dynamics of Equation (13) becomes for $s \geq t \geq 0$:

$$\begin{aligned}
 dX_t &= (\mathbf{A}X_t + \mathbf{e}_p \sigma_t \theta_t)dt + \mathbf{e}_p \sigma_t dB_t^\theta, \\
 X_s &= \exp\{\mathbf{A}(s-t)\}x + \int_t^s \exp\{\mathbf{A}(s-u)\} \mathbf{e}_p \sigma_u \theta_u du + \int_t^s \exp\{\mathbf{A}(s-u)\} \mathbf{e}_p \sigma_u dB_u^\theta.
 \end{aligned} \tag{19}$$

By inserting Equations (1)–(3) into Equation (6), Benth *et al.* (2007) explicitly calculated the risk neutral prices for HDD/CDD/CAT futures (and options) for contracts traded before the temperature measurement period, that is $0 \leq t \leq \tau_1 < \tau_2$:

$$\begin{aligned}
 F_{HDD}(t, \tau_1, \tau_2) &= \int_{\tau_1}^{\tau_2} v_{t,s} \psi \left[\frac{c - m_{\{t,s, \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{X}_t\}}}{v_{t,s}} \right] ds, \\
 F_{CDD}(t, \tau_1, \tau_2) &= \int_{\tau_1}^{\tau_2} v_{t,s} \psi \left[\frac{m_{\{t,s, \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{X}_t\}} - c}{v_{t,s}} \right] ds, \\
 F_{CAT}(t, \tau_1, \tau_2) &= \int_{\tau_1}^{\tau_2} \Lambda_u du + \mathbf{a}_{t, \tau_1, \tau_2} \mathbf{X}_t + \int_t^{\tau_1} \theta_u \sigma_u \mathbf{a}_{t, \tau_1, \tau_2} \mathbf{e}_p du \\
 &\quad + \int_{\tau_1}^{\tau_2} \theta_u \sigma_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2 - u)\} - I_p] \mathbf{e}_p du,
 \end{aligned}
 \tag{20}$$

with $\mathbf{a}_{t, \tau_1, \tau_2} = \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2 - t)\} - \exp\{\mathbf{A}(\tau_1 - t)\}]$; I_p is a $p \times p$ identity matrix; $\psi(x) = x\Phi(x) + \varphi(x)$ (Φ denotes the standard normal cumulative distribution function (cdf)) with $x = \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{X}_t$; $v_{t,s}^2 = \int_t^s \sigma_u^2 [\mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p]^2 du$; and $m_{\{t,s,x\}} = \Lambda_s + \int_t^s \sigma_u \theta_u \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p du + x$. The solution to Equation (20) depends on the assumed specification for the MPR θ . In the next section, it is shown that different assumed risk specifications can lead into different derivative prices.

The model in Benth *et al.* (2007) nests a number of previous models (Alaton *et al.*, 2002; Benth, 2003; Benth and Saltyte-Benth, 2005; Brody *et al.*, 2002); it generalizes the Benth and Saltyte-Benth (2007) and Dornier and Querel (2007) approaches and is a very well studied methodology in the literature (Benth *et al.*, 2011; Papazian and Skiadopoulos, 2010; Zapranis and Alexandridis, 2008). Besides it gives a clear connection between the discrete- and continuous-time versions, it provides closed-form non-arbitrage pricing formulas and it requires only a one-time estimation for the price dynamics. With the time series approach (Campbell and Diebold, 2005), the continuous-time approaches (Alaton *et al.*, 2002; Huang-Hsi *et al.*, 2008), neural networks (Zapranis and Alexandridis, 2008, 2009) or the principal component analysis approach (Papazian and Skiadopoulos, 2010) are not easy to compute price dynamics of CAT/CDD/HDD futures and one needs to use numerical approaches or simulations in order to calculate conditional expectations in Equation (6). In that case, partial differential equations or Monte Carlo simulations are being used. For option pricing, this would mean to simulate scenarios from futures prices. This translates into intensive computer simulation procedures.

5. The Implied Market Price of Weather Risk

For pricing and hedging non-tradable assets, one essentially needs to incorporate the MPR θ which is an important parameter of the associated EMM and it measures the additional return for bearing more risk. This section deals exactly with the differences between ‘historical’ (P) and ‘risk-neutral’ (Q) behaviours of temperature. Using statistical modelling and given that liquid-derivative contracts based on daily temperatures are traded on the CME, one might infer the MPR (the change of drift) from traded (CAT/CDD/HDD/C24AT) futures–options-type contracts.

Our study is a calibration procedure for financial engineering purposes. In the calibration exercise, a single date (but different time horizons and calibrated instruments are used) is required, since the model is recalibrated daily to detect intertemporal effects. Moreover, we use an economic and statistical testing approach, where we start from a specification of the MPR and check consistency with the data. By giving assumptions about the MPR, we implicitly make an assumption about the aggregate risk aversion of the market. The risk parameter θ can then be inferred by finding the value that satisfies Equation (20) for each specification. Once we know the MPR for temperature futures, then we know the MPR for options and thus one can price new ‘non-standard maturities’ or OTC derivatives. The concept of implied MPR is similar to that used in extracting implied volatilities (Fengler *et al.*, 2007) or the market price of oil risk (Gibson and Schwartz, 1990).

To value temperature derivatives, the following specifications of the MPR are investigated: a constant, a piecewise linear function, a two-piecewise linear function, a time-deterministic function and a ‘financial-bootstrapping’ MPR. The statistical point of view is to beat this as an inverse problem with different degrees of smoothness expressed through the penalty parameter of a smoothing spline. The economic point of view is to detect possible time dependencies that can be explained by investor’s preferences in order to hedge weather risk.

In this article we concentrate on contracts with monthly measurement length periods, but similar implications apply for seasonal strip contracts. We observe different temperature futures contracts $i = 1, \dots, I$ with measurement periods $t \leq \tau_1^i < \tau_2^i$ and $\tau_2^i \leq \tau_1^{i+1}$ traded at time t , meaning that contracts expire at some point in time and roll over to another contract. Therefore, $i = 1$ denotes contract types with measurement period in 30 days, $i = 2$ denotes contract types in 60 days and so on. For example, a contract with $i = 7$ is six months ahead from the trading day t . For United States and Europe, the number of temperature futures contracts is $I = 7$ (April–October or October–April), while for Asia $I = 12$ (January–December). The details of the temperature futures data are displayed in Table 4. To simplify notation, dates are written in `yyyymmdd` format.

5.1 Constant MPR for Each Contract per Trading Date

Given observed temperature futures market prices and by inverting Equation (20), we imply the MPR θ_u for $i = 1, \dots, I$ futures contracts with different measurement time horizon periods $[\tau_1^i, \tau_2^i]$, $t \leq \tau_1^i < \tau_2^i$ and $\tau_2^i \leq \tau_1^{i+1}$ traded at date t . Our first assumption is to set, for the i th contract, a constant MPR over $[t, \tau_2^i]$, that is, we have that $\theta_u = \theta_t^i$:

$$\hat{\theta}_{i,CAT}^i = \arg \min_{\theta_t^i} \left(F_{CAT}(t, \tau_1^i, \tau_2^i) - \int_{\tau_1^i}^{\tau_2^i} \hat{\Lambda}_u du - \mathbf{a}_{t, \tau_1^i, \tau_2^i} \mathbf{X}_t - \theta_t^i \left\{ \int_t^{\tau_1} \hat{\sigma}_u \mathbf{a}_{t, \tau_1^i, \tau_2^i} \mathbf{e}_p du + \int_{\tau_1^i}^{\tau_2^i} \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp \{ \mathbf{A}(\tau_2^i - u) \} - I_p] \mathbf{e}_p du \right\} \right)^2,$$

Table 4. Weather futures and futures prices listed on date (yyyymmdd) at CME.

Contract type	Trading date		Measurement period		Futures prices $F_{(t, \tau_1, \tau_2, \beta)}$			Realized T_t $I_{(\tau_1, \tau_2)}$
	t	τ_1	τ_2	CME	MPR = 0	Constant MPR		
Berlin-CAT	20070316	20070401	20070430	288.00	363.00	291.06	362.90	
Berlin-CAT	20070316	20070501	20070531	457.00	502.11	454.91	494.20	
Berlin-CAT	20070316	20070601	20070630	529.00	571.78	630.76	574.30	
Berlin-CAT	20070316	20070701	20070731	616.00	591.56	626.76	583.00	
Berlin-CAT	20070316	20070801	20070831	610.00	566.14	636.22	580.70	
Berlin-CAT	20070316	20070901	20070930	472.00	414.33	472.00	414.80	
Berlin-CAT	20070427	20070501	20070531	457.00	506.18	457.52	494.20	
Berlin-CAT	20070427	20070601	20070630	529.00	571.78	534.76	574.30	
Berlin-CAT	20070427	20070701	20070731	616.00	591.56	656.76	583.00	
Berlin-CAT	20070427	20070801	20070831	610.00	566.14	636.22	580.70	
Berlin-CAT	20070427	20070901	20070930	472.00	414.33	472.00	414.80	
Tokyo-C24AT	20081027	20090301	20090331	450.00	118.32	488.90	305.00	
Tokyo-C24AT	20081027	20090401	20090430	592.00	283.18	563.27	479.00	
Tokyo-C24AT	20081027	20090501	20090531	682.00	511.07	696.31	623.00	
Tokyo-C24AT	20081027	20090601	20090630	818.00	628.24	835.50	679.00	
Tokyo-C24AT	20081027	20090701	20090731	855.00	731.30	706.14	812.00	

Notes: CME, Chicago Mercantile Exchange; MPR, market price of risk.
 Weather futures at CME; futures prices $F_{(t, \tau_1, \tau_2, \beta)}$ from CME; estimated prices with MPR = 0; constant MPR for different controls per trading date (constant MPR).
 Source: Bloomberg Professional Service (Weather futures).

$$\hat{\theta}_{i,HDD}^i = \arg \min_{\theta_i^i} \left(F_{HDD(t,\tau_1^i,\tau_2^i)} - \int_{\tau_1^i}^{\tau_2^i} v_{t,s} \psi \left[\frac{c - \hat{m}_{\{t,s,\mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\}\mathbf{X}_t\}}^1}{v_{t,s}} \right] ds \right)^2, \quad (21)$$

with $\hat{m}_{\{t,s,x\}}^1 = \Lambda_s + \theta_t^i \int_t^s \sigma_u \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p du + x$, $v_{t,s}^2$, $\psi(x)$ and x defined as in Equation (20). The MPR for CDD futures $\hat{\theta}_{i,CDD}^i$ is equivalent to the HDD case in Equation (21) and we will therefore omit CDD parameterizations. Note that this specification can be seen as a deterministic time-varying MPR θ_t^i that varies with date for any given contract i , but it is constant over $[t, \tau_2^i]$.

5.2 One Piecewise Constant MPR

A simpler MPR parameterization is to assume that it is constant across all time horizon contracts priced in a particular date (θ_t). We therefore estimate this constant MPR for all contract types traded at $t \leq \tau_1^i < \tau_2^i$, $i = 1, \dots, I$ as follows:

$$\hat{\theta}_{i,CAT} = \arg \min_{\theta_t} \sum_{i=1}^I \left(F_{CAT(t,\tau_1^i,\tau_2^i)} - \int_{\tau_1^i}^{\tau_2^i} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{t,\tau_1^i,\tau_2^i} \mathbf{X}_t - \theta_t \left\{ \int_t^{\tau_1^i} \hat{\sigma}_u \hat{\mathbf{a}}_{t,\tau_1^i,\tau_2^i} \mathbf{e}_p du + \int_{\tau_1^i}^{\tau_2^i} \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2^i - u)\} - I_p] \mathbf{e}_p du \right\} \right)^2,$$

$$\hat{\theta}_{i,HDD} = \arg \min_{\theta_t} \sum_{i=1}^I \left(F_{HDD(t,\tau_1^i,\tau_2^i)} - \int_{\tau_1^i}^{\tau_2^i} v_{t,s} \psi \left[\frac{c - \hat{m}_{\{t,s,\mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\}\mathbf{X}_t\}}^2}{v_{t,s}} \right] ds \right)^2, \quad (22)$$

with $\hat{m}_{\{t,s,x\}}^2 = \Lambda_s + \theta_t \int_t^s \sigma_u \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p du + x$ and $v_{t,s}^2$, $\psi(x)$ and x as defined in Equation (20). This ‘one piecewise constant’ MPR specification (θ_t) is solved by means of the ordinary least squares (OLS) minimization procedure and differs from θ_t^i in Equation (21) because for all traded contracts at date t , we get only one MPR estimate (instead of i estimates) at time t , that is, θ_t is constant over $[t, \tau_2^I]$.

5.3 Two Piecewise Constant MPR

Assuming now that, instead of one constant MPR per trading day, we have a step function with a given jump point ξ (take e.g. the first 150 days before the beginning of the measurement period), so we have that $\hat{\theta}_t = I(u \leq \xi) \theta_t^1 + I(u > \xi) \theta_t^2$. The two piecewise constant function $\hat{\theta}_t$ with $t \leq \tau_1^i < \tau_2^i$ is estimated with the OLS minimization procedure as follows:

$$\begin{aligned}
 f_{CAT}(\xi) = \arg \min_{\theta_{t,CAT}^1, \theta_{t,CAT}^2} \sum_{i=1}^I & \left(F_{CAT(t, \tau_1^i, \tau_2^i)} - \int_{\tau_1^i}^{\tau_2^i} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{t, \tau_1^i, \tau_2^i} \mathbf{X}_t \right. \\
 & - \theta_{t,CAT}^1 \left\{ \int_t^{\tau_1^i} I(u \leq \xi) \hat{\sigma}_u \hat{\mathbf{a}}_{t, \tau_1^i, \tau_2^i} \mathbf{e}_p du \right. \\
 & + \left. \int_{\tau_1^i}^{\tau_2^i} I(u \leq \xi) \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2^i - u)\} - I_p] \mathbf{e}_p du \right\} \\
 & - \theta_{t,CAT}^2 \left\{ \int_t^{\tau_1^i} I(u > \xi) \hat{\sigma}_u \hat{\mathbf{a}}_{t, \tau_1^i, \tau_2^i} \mathbf{e}_p du \right. \\
 & \left. \left. + \int_{\tau_1^i}^{\tau_2^i} I(u > \xi) \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2^i - u)\} - I_p] \mathbf{e}_p du \right\} \right)^2, \tag{23}
 \end{aligned}$$

$$f_{HDD}(\xi) = \arg \min_{\theta_{t,HDD}^1, \theta_{t,HDD}^2} \sum_{i=1}^I \left(F_{HDD(t, \tau_1^i, \tau_2^i)} - \int_{\tau_1^i}^{\tau_2^i} v_{t,s} \psi \left[\frac{c - \hat{m}_{\{t,s, \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{X}_t\}}^3}{v_{t,s}} \right] ds \right)^2,$$

$$\begin{aligned}
 \hat{m}_{\{t,s,x\}}^3 = \Lambda_s + \theta_{t,HDD}^1 & \left\{ \int_t^s I(u \leq \xi) \sigma_u \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p du + x \right\} \\
 + \theta_{t,HDD}^2 & \left\{ \int_t^s I(u > \xi) \sigma_u \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p du + x \right\},
 \end{aligned}$$

and $v_{t,s}^2$, $\psi(x)$ and x as defined in Equation (20). In the next step, we optimized the value of ξ such as $f_{CAT}(\xi)$ or $f_{HDD}(\xi)$ is minimized. This MPR specification will vary according to the unknown ξ . This would mean that the market does a risk adjustment for contracts traded close or far from the measurement period.

5.4 General Form of the MPR per Trading Day

Generalizing the piecewise continuous function given in the previous subsection, the (inverse) problem of determining θ_t with $t \leq \tau_1^i < \tau_2^i$, $i = 1, \dots, I$, can be formulated via a series expansion for θ_t :

$$\begin{aligned}
 \arg \min_{\gamma_k} \sum_{i=1}^I & \left(F_{CAT(t, \tau_1^i, \tau_2^i)} - \int_{\tau_1^i}^{\tau_2^i} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{t, \tau_1^i, \tau_2^i} \hat{\mathbf{X}}_t - \int_t^{\tau_1^i} \sum_{k=1}^K \gamma_k h_k(u_i) \hat{\sigma}_u \hat{\mathbf{a}}_{t, \tau_1^i, \tau_2^i} \mathbf{e}_p du_i \right. \\
 & \left. - \int_{\tau_1^i}^{\tau_2^i} \sum_{k=1}^K \gamma_k h_k(u_i) \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2^i - u_i)\} - I_p] \mathbf{e}_p du_i \right)^2,
 \end{aligned}$$

$$\arg \min_{a_k} \sum_{i=1}^I \left(F_{HDD}(t, \tau_1^i, \tau_2^i) - \int_{\tau_1^i}^{\tau_2^i} v_{t,s} \psi \left[\frac{c - \hat{m}_{\{t,s, \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} X_t\}}^4}{v_{t,s}} \right] ds \right)^2, \quad (24)$$

with $\hat{m}_{\{t,s,x\}}^4 = \Lambda_s + \int_t^s \sum_{k=1}^K a_k l_k(u_i) \hat{\sigma}_{u_i} \mathbf{e}_1^\top \exp\{\mathbf{A}(s-t)\} \mathbf{e}_p du_i + x$ and $v_{t,s}^2$, $\psi(x)$ and x as defined in Equation (20). $h_k(u_i)$ and $l_k(u_i)$ are vectors of known basis functions and may denote a B-spline basis for example. γ_k and a_k define the coefficients and K is the number of knots. This means that the inferred MPR is going to be a solution for an inverse problem with different degrees of smoothness expressed through the penalty parameter of a smoothing spline. The degrees of smoothness will allow for a term structure of risk. In other words, a time-dependent risk factor offers the possibility to have different risk adjustments for different times of the year.

5.5 Bootstrapping the MPR

In this section we propose a bootstrapping technique to detect possible MPR time-dependent paths of temperature futures contracts. More importantly, since these futures contract types have different measurement periods $[\tau_1^i, \tau_2^i]$ with $\tau_1^i < \tau_1^{i+1} \leq \tau_2^i < \tau_2^{i+1}$, $i = 1, \dots, I$, and they roll over to another contracts when they expire at some point in time, it makes sense to construct MPR estimates from which we can price contracts with any maturity, without the need of external information. This ‘financial’ bootstrapping idea consists of estimating by forward substitution the MPR θ_t^i of the futures price contracts with the closest measurement period and placing it into the estimation for the next MPR θ_t^{i+1} . We implement the estimation for CAT contracts, but the idea applies also for HDD/CDD contract types. First, for the first contract $i = 1$ and $t \in [\tau_1^1, \tau_2^1]$, $\hat{\theta}_{t,CAT}^1$ is estimated from Equation (21):

$$\begin{aligned} \hat{\theta}_{t,CAT}^1 = \arg \min_{\theta_t^1} & \left(F_{CAT}(t, \tau_1^1, \tau_2^1) - \int_{\tau_1^1}^{\tau_2^1} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{t, \tau_1^1, \tau_2^1} \hat{\mathbf{X}}_t - \theta_t^1 \left\{ \int_t^{\tau_1^1} \hat{\sigma}_u \hat{\mathbf{a}}_{t, \tau_1^1, \tau_2^1} \mathbf{e}_p du \right. \right. \\ & \left. \left. + \int_{\tau_1^1}^{\tau_2^1} \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2^1 - u)\} - I_p] \mathbf{e}_p du \right\} \right)^2. \end{aligned} \quad (25)$$

Second, the estimated $\hat{\theta}_{t,CAT}^1$ is substituted in the period $[\tau_1^1, \tau_2^1]$ to get an estimate of $\hat{\theta}_{t,CAT}^2$:

$$\begin{aligned} \hat{\theta}_{t,CAT}^2 = \arg \min_{\theta_{t,CAT}^2} & \left(F_{CAT}(t, \tau_1^2, \tau_2^2) - \int_{\tau_1^2}^{\tau_2^2} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{t, \tau_1^2, \tau_2^2} \hat{\mathbf{X}}_t - \int_t^{\tau_1^1} \hat{\theta}_{t,CAT}^1 \hat{\sigma}_u \hat{\mathbf{a}}_{t, \tau_1^1, \tau_2^1} \mathbf{e}_p du \right. \\ & \left. - \int_{\tau_1^1}^{\tau_2^2} \theta_{t,CAT}^2 \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp\{\mathbf{A}(\tau_2^2 - u)\} - I_p] \mathbf{e}_p du \right)^2. \end{aligned} \quad (26)$$

Then substitute $\hat{\theta}_{i,CAT}^1$ in the period $[\tau_1^1, \tau_2^1]$ and $\hat{\theta}_{i,CAT}^2$ in the period $[\tau_1^2, \tau_2^2]$ to estimate $\hat{\theta}_{i,CAT}^3$:

$$\hat{\theta}_{i,CAT}^3 = \arg \min_{\hat{\theta}_{i,CAT}^3} \left(F_{CAT}(t, \tau_1^3, \tau_2^3) - \int_{\tau_1^3}^{\tau_2^3} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{i, \tau_1^3, \tau_2^3} \mathbf{X}_t - \int_t^{\tau_1^1} \hat{\theta}_{i,CAT}^1 \hat{\sigma}_u \hat{\mathbf{a}}_{i, \tau_1^3, \tau_2^3} \mathbf{e}_p du - \int_{\tau_1^2}^{\tau_2^2} \hat{\theta}_{i,CAT}^2 \hat{\sigma}_u \hat{\mathbf{a}}_{i, \tau_1^3, \tau_2^3} \mathbf{e}_p du - \int_{\tau_1^3}^{\tau_2^3} \theta_{i,CAT}^3 \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp \{ \mathbf{A}(\tau_2^3 - u) \} - I_p] \mathbf{e}_p du \right)^2.$$

In a similar way, one obtains the estimation of $\hat{\theta}_{i,CAT}^4, \dots, \hat{\theta}_{i,CAT}^I$.

5.6 Smoothing the MPR over Time

Since smoothing individual estimates is different from estimating a deterministic function, we also assure our results by fitting a parametric function to all available contract prices (calendar year estimation). After computing the MPR $\hat{\theta}_{i,CAT}$, $\hat{\theta}_{i,HDD}$ and $\hat{\theta}_{i,CDD}$ for each of the previous specification and for each of the n th trading days t for different i th contracts, the MPR time series can be smoothed with the inverse problem points to find an MPR $\hat{\theta}_u$ for every calendar day u and with that being able to price temperature derivatives for any date:

$$\arg \min_{f \in F_j} \sum_{t=1}^n \left\{ \hat{\theta}_t - f(u_t) \right\}^2 = \arg \min_{\alpha_j} \sum_{t=1}^n \left\{ \hat{\theta}_t - \sum_{j=1}^J \alpha_j \Psi_j(u_t) \right\}^2, \quad (27)$$

where $\Psi_j(u_t)$ is a vector of known basis functions, α_j defines the coefficients, J is the number of knots, $u_t = t - \Delta + 1$ with increment Δ and n is the number of days to be smoothed. In our case, $u_t = 1$ day and $\Psi_j(u_t)$ is estimated using cubic splines.

Alternatively, one can first do the smoothing with basis functions of all available futures contracts:

$$\arg \min_{\beta_j} \sum_{t=1}^n \sum_{i=1}^I \left\{ F_{(t, \tau_1^i, \tau_2^i)} - \sum_{j=1}^J \beta_j \Psi_j(u_t) \right\}^2, \quad (28)$$

and then estimate the time series of $\hat{\theta}_t^s$ s with the obtained smoothed futures prices $F_{(t, \tau_1^i, \tau_2^i)}^s$.

For example, for a constant MPR for all CAT futures contracts type traded over all t s with $t \leq \tau_1^i < \tau_2^i$ and $\tau_2^i \leq \tau_1^{i+1}$, we have:

$$\hat{\theta}_{t,CAT}^s = \arg \min_{\theta_{t,CAT}^s} \left(F_{CAT(t,\tau_1^1,\tau_2^1)}^s - \int_{\tau_1^1}^{\tau_2^1} \hat{\Lambda}_u du - \hat{\mathbf{a}}_{t,\tau_1^1,\tau_2^1} \mathbf{X}_t - \theta_{t,CAT}^s \left\{ \int_t^{\tau_1^1} \hat{\sigma}_u \hat{\mathbf{a}}_{t,\tau_1^1,\tau_2^1} \mathbf{e}_p du + \int_{\tau_1^1}^{\tau_2^1} \hat{\sigma}_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp \{ \mathbf{A}(\tau_2^1 - u) \} - I_p] \mathbf{e}_p du \right\} \right)^2. \quad (29)$$

5.7 Statistical and Economical Insights of the Implied MPR

In this section, using the previous specifications, we imply the MPR (the change of drift) for CME (CAT/CDD/HDD/C24AT) futures contracts traded for different cities. Note that one might also infer the MPR from options data and compare the findings with prices in the futures market.

Table 5 presents the descriptive statistics of different MPR specifications for Berlin-CAT, Essen-CAT and Tokyo-C24AT daily futures contracts with $t \leq \tau_1^i < \tau_2^i$ traded during 20031006–20080527 (5102 contracts in 1067 trading days with 29 different measurement periods), 20050617–20090731 (3530 contracts in 926 trading dates with 28 measurement periods) and 20040723–20090831 (2611 contracts in 640 trading dates with 27 measurement periods). The MPR ranges vary between $[-10.71, 10.25]$, $[31.05, 5.73]$ and $[-82.62, 52.17]$ for Berlin-CAT, Essen-CAT and Tokyo-C24AT futures contracts, respectively, whereas the MPR averages are 0.04, 0.00 and -3.08 for constant MPR for different contracts; -0.08 , -0.38 and 0.73 for one piecewise constant; -0.22 , -0.43 and -3.50 for two piecewise constant; 0.04 , 0.00 and -3.08 for spline; and 0.07 , 0.00 and -0.11 when bootstrapping the MPR. We observe that the two piecewise constant MPR function is a robust least square estimation, since its values are sensitive to the choice of ξ . Figure 6 shows the MPR estimates for Berlin-CAT futures prices traded on 20060530 with $\xi = 62, 93, 123$ and 154 and sum of squared errors equal to 2759, 14,794, 15,191 and 15,526. The line displays a discontinuity indicating that trading was not taking place (CAT futures are only traded from April to November and MPR estimates cannot be computed since there are no market prices). When the jump ξ is getting far from the measurement period, the value of the MPR $\hat{\theta}_t^1$ decreases and $\hat{\theta}_t^2$ increases, yielding a $\hat{\theta}_t$ around 0. Table 5 also displays the estimates of the time-dependent MPR (or spline MPR) from the bootstrapping technique. The spline MPR smooths the estimates over time and it is estimated using cubic polynomials with k equal to the number of traded contracts I at date t . The performance of the bootstrapped MPR is similar to the constant MPR for different contracts per trading date estimates, suggesting that the only risk which the statistical model might imply is that the MPR will be equal at any trading date across all temperature contract types.

The first panel in Figure 7 displays the Berlin-CAT, Essen-CAT and Tokyo-C24AT futures contracts traded at 20060530, 20060530 and 20050531, respectively. The second, third and fourth panels of Figure 7 show the MPR when it is assumed to be constant for different contracts per trading date, a two piecewise constant and the spline MPR. In the case of the constant MPR for different contracts per trading date, the lines overlap because the MPR for every contract $i = 1, \dots, 12$ is supposed to be constant over the period $[t, \tau_2^i]$ at trading date t . The two piecewise constant function adjusts the risk according to the choice of ξ

Table 5. Statistics of MPR specifications for Berlin-CAT, Essen-CAT and Tokyo-C24AAT.

Type	No. of contracts	Statistic	Constant	1 piecewise	2 piecewise	Bootstrap	Spline
Berlin-CAT							
30 days	487	WS (Prob)	0.93 (0.66)	0.01 (0.08)	0.00 (0.04)	0.93 (0.04)	0.91 (0.66)
(<i>i</i> = 1)		Min (Max)	-0.28 (0.12)	-0.65 (0.11)	-1.00 (0.20)	-0.28 (0.12)	0.02 (0.21)
60 days	874	Med (SD)	0.05 (0.07)	-0.10 (0.12)	-0.17 (0.31)	0.05 (0.07)	0.14 (0.06)
(<i>i</i> = 2)		Min (Max)	-0.54 (0.84)	-4.95 (8.39)	-10.71 (10.25)	-0.54 (0.84)	0.00 (0.21)
90 days	858	Med (SD)	0.03 (0.13)	-0.09 (1.06)	-0.17 (1.44)	0.03 (0.13)	0.05 (0.05)
(<i>i</i> = 3)		Min (Max)	-0.54 (0.82)	-4.95 (8.39)	-10.71 (10.25)	-0.54 (0.82)	0.00 (0.21)
120 days	815	Med (SD)	0.02 (0.09)	-0.10 (1.07)	-0.17 (1.46)	0.02 (0.09)	0.11 (0.07)
(<i>i</i> = 4)		Min (Max)	-0.53 (0.84)	-4.95 (4.56)	-7.71 (6.88)	-0.53 (0.84)	0.00 (0.21)
150 days	752	Med (SD)	0.03 (0.14)	-0.10 (0.94)	-0.19 (1.38)	0.03 (0.14)	0.11 (0.09)
(<i>i</i> = 5)		Min (Max)	-0.54 (0.84)	-4.53 (4.56)	-8.24 (6.88)	-0.54 (0.84)	-0.00 (0.20)
180 days	711	Med (SD)	0.13 (0.09)	-0.10 (0.95)	-0.20 (1.47)	0.13 (0.09)	0.01 (0.07)
(<i>i</i> = 6)		Min (Max)	-0.54 (0.82)	-4.53 (4.56)	-8.24 (6.88)	-0.54 (0.82)	-0.03 (0.12)
		Med (SD)	0.02 (0.12)	-0.10 (0.95)	-0.14 (1.48)	0.02 (0.12)	0.00 (0.03)
Essen-CAT							
30 days	384	WS (Prob)	0.02 (0.11)	0.20 (0.34)	0.01 (0.09)	0.02 (0.11)	1.63 (0.79)
(<i>i</i> = 1)		Min (Max)	-0.98 (0.52)	-31.05 (5.73)	-1.83 (1.66)	-0.98 (0.52)	-0.00 (0.00)
60 days	796	Med (SD)	0.01 (0.12)	-0.39 (1.51)	-0.43 (9.62)	0.01 (0.12)	0.00 (0.00)
(<i>i</i> = 2)		Min (Max)	-1.35 (0.62)	-31.05 (5.73)	-1.83 (1.66)	-1.35 (0.62)	0.00 (0.00)
90 days	738	Med (SD)	0.02 (0.14)	-0.40 (1.56)	-0.46 (9.99)	0.02 (0.14)	0.00 (0.00)
(<i>i</i> = 3)		Min (Max)	-1.56 (0.59)	-6.68 (5.14)	-5.40 (1.71)	-1.56 (0.59)	0.00 (0.00)
120 days	551	Med (SD)	-0.02 (0.19)	-0.40 (0.88)	-0.40 (0.85)	-0.02 (0.19)	0.00 (0.00)
(<i>i</i> = 4)		Min (Max)	-0.29 (0.51)	-4.61 (1.44)	-6.60 (1.43)	-0.29 (0.51)	0.00 (0.00)
150 days	468	Med (SD)	0.03 (0.05)	-0.37 (0.51)	-0.41 (0.85)	0.03 (0.05)	0.00 (0.00)
(<i>i</i> = 5)		Min (Max)	-0.44 (0.13)	-4.61 (1.44)	-6.60 (1.43)	-0.44 (0.13)	0.00 (0.00)
180 days	405	Med (SD)	0.00 (0.07)	-0.35 (0.49)	-0.30 (0.81)	0.00 (0.07)	0.00 (0.00)
(<i>i</i> = 6)		Min (Max)	-0.10 (0.57)	-4.61 (1.44)	-4.25 (0.52)	-0.10 (0.57)	0.00 (0.00)
		Med (SD)	-0.02 (0.06)	-0.12 (0.45)	-0.21 (0.52)	-0.02 (0.06)	0.00 (0.00)
Tokyo-C24AAT							
30 days	419	WS (Prob)	0.76 (0.61)	0.02 (0.13)	0.01 (0.11)	0.76 (0.61)	4.34 (0.96)
		Min (Max)	-7.55 (0.17)	-69.74 (52.17)	-69.74 (52.17)	-7.55 (0.17)	-0.23 (-0.18)

(Continued)

Table 5. (Continued).

Type	No. of contracts	Statistic	Constant	1 piecewise	2 piecewise	Bootstrap	Spline
($i = 2$)		Med (SD)	-3.87 (2.37)	-0.33 (19.68)	-0.48 (20.46)	-3.87 (2.37)	-0.20 (0.01)
60 days	416	Min (Max)	-7.56 (0.14)	-69.74 (52.17)	-69.74 (52.17)	-7.56 (0.14)	-0.18 (-0.13)
($i = 3$)		Med (SD)	-3.49 (2.47)	-0.23 (21.34)	-0.41 (21.63)	-3.49 (2.47)	-0.15 (0.01)
90 days	393	Min (Max)	-7.55 (1.02)	-69.74 (26.82)	-69.74 (38.53)	-7.55 (1.02)	-0.13 (-0.09)
($i = 4$)		Med (SD)	-2.96 (2.65)	0.04 (20.84)	-0.33 (20.22)	-2.96 (2.65)	-0.11 (0.01)
120 days	350	Min (Max)	-7.55 (1.02)	-69.74 (26.82)	-69.74 (48.32)	-7.55 (1.02)	-0.10 (-0.06)
($i = 5$)		Med (SD)	-2.08 (2.74)	1.26 (19.54)	-0.11 (19.69)	-2.08 (2.74)	-0.08 (0.01)
150 days	305	Min (Max)	-7.55 (1.02)	-51.18 (26.82)	-51.18 (48.32)	-7.55 (1.02)	-0.06 (-0.04)
($i = 6$)		Med (SD)	-2.08 (2.71)	1.26 (16.03)	7.17 (17.02)	-2.08 (2.71)	-0.05 (0.00)
180 days	243	Min (Max)	-7.39 (1.26)	-51.18 (19.10)	-51.18 (48.32)	-7.39 (1.26)	-0.05 (-0.04)
($i = 7$)		Med (SD)	-2.08 (2.74)	3.66 (15.50)	7.63 (17.69)	-2.08 (2.74)	-0.04 (0.00)
210 days	184	Min (Max)	-7.39 (1.26)	-24.88 (26.16)	-54.14 (39.65)	-7.39 (1.26)	-0.07 (-0.05)
($i = 8$)		Med (SD)	-3.00 (2.86)	13.69 (10.05)	10.61 (14.63)	-3.00 (2.86)	-0.06 (0.00)
240 days	167	Min (Max)	-7.39 (1.26)	-24.88 (21.23)	-82.62 (42.14)	-7.39 (1.26)	-0.07 (-0.07)
($i = 9$)		Med (SD)	-3.00 (2.74)	3.46 (11.73)	-4.24 (37.25)	-3.00 (2.74)	-0.07 (0.00)
270 days	134	Min (Max)	-7.39 (0.44)	-24.88 (26.16)	-82.62 (42.14)	-7.39 (0.44)	-0.07 (-0.03)
($i = 10$)		Med (SD)	-4.24 (2.39)	8.48 (13.88)	-7.39 (40.76)	-4.24 (2.39)	-0.05 (0.00)

Notes: WS, Wald statistics; SD, standard deviation.

Futures contracts traded during (20031006-20080527), (20050617-20090731), and (20040723-20090630) respectively, with trading date before measurement period $t \leq t_i^j < t_i^j$, $i = 1, \dots, I$ (where $i = 1$ (30 days), $i = 2$ (60 days), $\dots, i = I$ (210 days)); the WS, the WS probabilities (Prob), Minimum (Min), Maximum (Max), Median (Med) and SD. MPR specifications: Constant for different contracts per trading date (Constant), 1 piecewise constant, 2 piecewise constant ($\xi = 150$ days), bootstrap and spline.

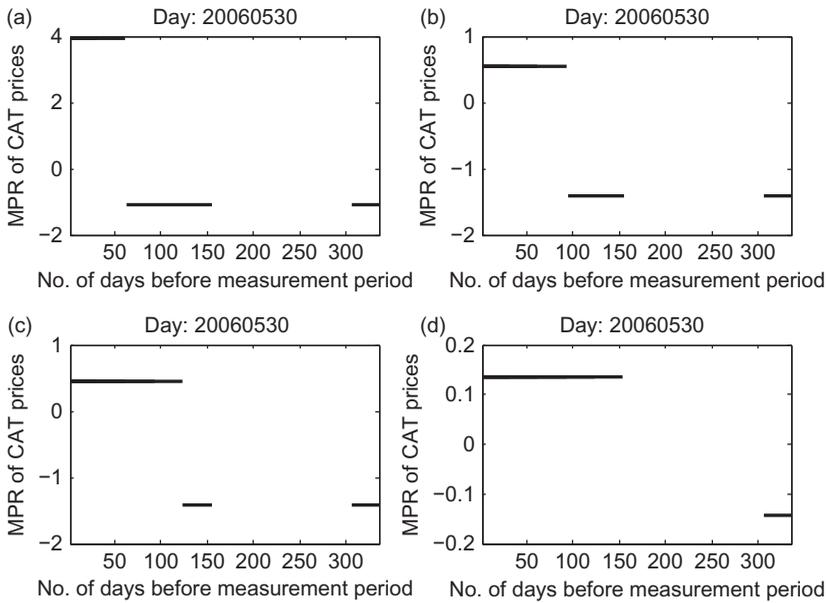


Figure 6. Two piecewise constant MPR with jumps $\xi =$ (a) 62, (b) 93, (c) 123 and (d) 154 days for Berlin-CAT contracts traded on 20060530. The corresponding sum of squared errors are 2759, 14794, 15191 and 15526. When the jump ξ is getting far from the measurement period, the value of the MPR $\hat{\theta}_t^1$ decreases and $\hat{\theta}_t^2$ increases, yielding a $\hat{\theta}_t$ around 0.

(in this case $\xi = 150$ days). The spline MPR smooths over time and for days without trading (see the case of Berlin-CAT or Essen-CAT futures), it displays a maximum, for example, in winter. A penalizing term in Equation (24) might correct for this.

In all the specifications, we verified the discussion that MPR is different from 0 (as Cao and Wei (2004), Huang-Hsi *et al.* (2008), Richards *et al.* (2004) and Alaton *et al.* (2002) do) varies in time and moves from a negative to a positive domain according to the changes in the seasonal variation. The MPR specifications change signs when a contract expires and rolls over to another contract (e.g. from 210 to 180, 150, 120, 90, 60, 30 days before measurement period); they react negatively to the fast changes in seasonal variance σ_t within the measurement period (Figure 3) and to the changes in CAT futures volatility $\sigma_t \mathbf{a}_{t, \tau_1, \tau_2} \mathbf{e}_p$. Figure 8 shows the Berlin-CAT volatility paths for contracts issued before and within the measurement periods 2004–2008. We observed the Samuelson effect for mean-reverting futures: for contracts traded within the measurement period, CAT volatility is close to 0 when the time to measurement is large and it decreases up to the end of the measurement period. For contracts traded before the measurement period, CAT volatility is also close to 0 when the time to measurement is large, but increases up to the start of the measurement period. In Figure 9, two Berlin-CAT contracts issued on 20060517 but with different measurement periods are plotted: the longest the measurement period, the largest the volatility. Besides this, one observes the effect of the *CAR*(3) in both contracts when the volatility decays just before maturity of the contracts. These two effects are comparable with the study for Stockholm CAT futures in Benth *et al.* (2007); however, the deviations are less smoothed for Berlin.

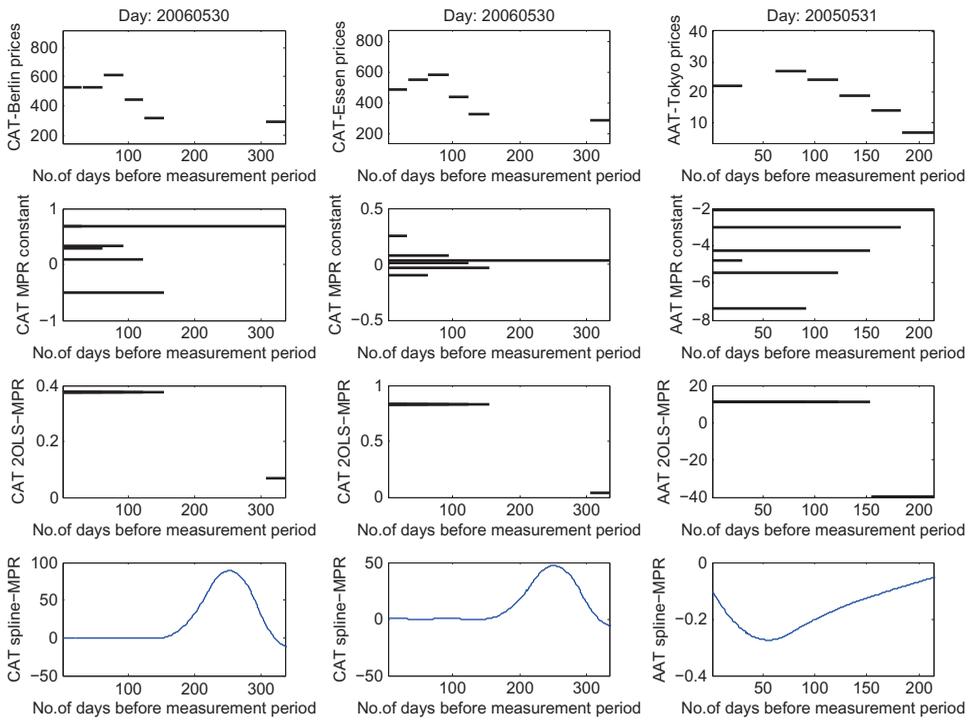


Figure 7. Futures CAT prices (1 row panel) and MPR specifications: constant MPR for different contracts per trading day, two piecewise constant and spline (2, 3 and 4 row panel) for Berlin-CAT (left), Essen-CAT (middle), Tokyo-AAT (right) of futures traded on 20060530, 20060530 and 20060531, respectively.

We investigate the proposition that the MPR derived from CAT/HDD/CDD futures is different from 0. We conduct the Wald statistical test to check whether this effect exists by testing the true value of the parameter based on the sample estimate. In the multivariate case, the Wald statistic for $\{\theta_t \in \mathbb{R}^i\}_{t=1}^n$ is

$$(\hat{\theta}_t - \theta_0)^\top \Sigma (\hat{\theta}_t - \theta_0) \sim \chi_p^2, \Sigma^{\frac{1}{2}} (\hat{\theta}_t - \theta_0) \sim N(0, I_i),$$

where Σ is the variance matrix and the estimate $\hat{\theta}_t$ is compared with the proposed value $\theta_0 = 0$. Using a sample size of n trading dates of contracts with $t \leq \tau_1^i < \tau_2^i$, $i = 1, \dots, I$, we illustrate in Table 5 the Wald statistics for all previous MPR specifications. We reject $H_0 : \hat{\theta}_t = 0$ under the Wald statistic $\{\theta_t \in \mathbb{R}^i\}_{t=1}^n$ for all cases. Although the constant per trading day and general MPR specifications smooth deviations over time, the Wald statistic confirms that the MPR differs significantly from 0. Our results are robust to all specifications.

Figure 10 shows the smoothing of MPR individuals (Equation (27)) for different specifications in 1 (20060530), 5 (20060522–20060530) and 30 trading days (20060417–20060530) of Berlin-CAT futures, while the last panel in Figure 10 gives the results

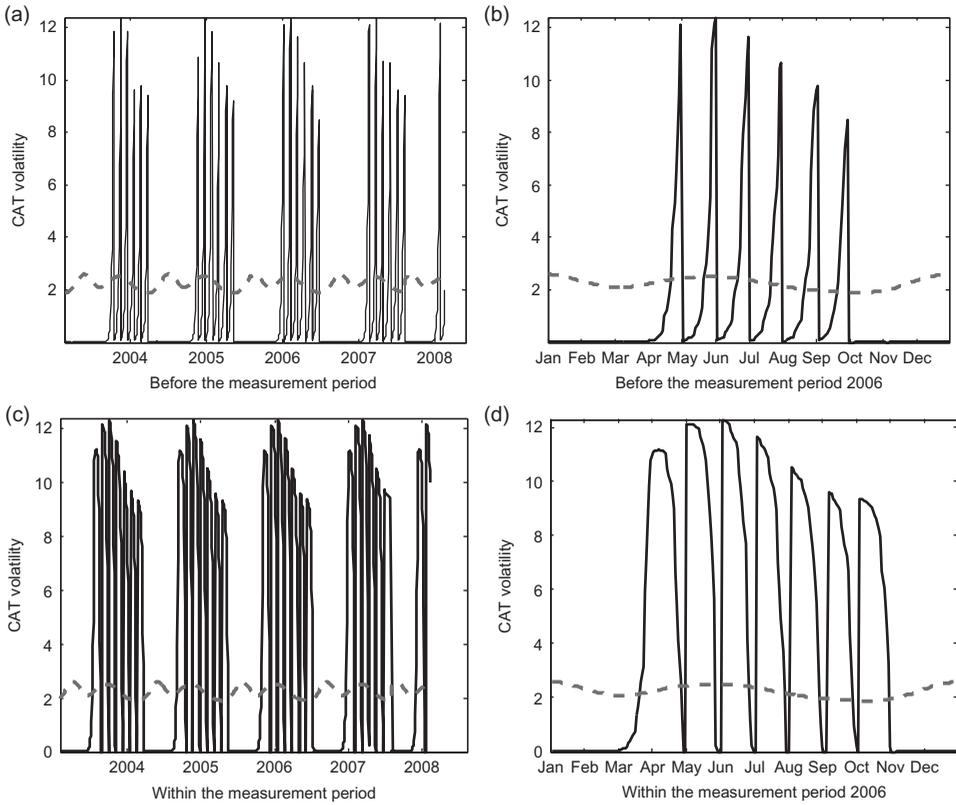


Figure 8. The Samuelson effect for Berlin-CAT futures explained by the CAT volatility $\sigma_t \mathbf{a}_{t, \tau_1, \tau_2} \mathbf{e}_p$ (black line) and the volatility σ_t of Berlin-CAT futures (dash line) from 2004 to 2008 and 2006 for contracts traded before (a) and (b) and within (c) and (d) the measurement period.

when MPR estimates are obtained from smoothed prices using the calendar year estimation (Equation (29)). Both smoothing procedures lead to similar outcomes: notable changes in sign, MPR deviations are smoothed over time and the higher the number of calendar days, the closer the fit of Equations (27) and (29). This indicates that sample size does not influence the stochastic behaviour of the MPR.

To interpret the economic meaning of the previous MPR results, recall, for example, the relationship between the RP (the market price minus the implied futures price with MPR equal to 0) and the MPR for CAT temperature futures:

$$RP_{CAT} = \int_t^{\tau_1^i} \theta_u \sigma_u \mathbf{a}_{t, \tau_1^i, \tau_2^i} \mathbf{e}_p du + \int_{\tau_1^i}^{\tau_2^i} \theta_u \sigma_u \mathbf{e}_1^\top \mathbf{A}^{-1} [\exp \{ \mathbf{A}(\tau_2^i - u) \} - I_p] \mathbf{e}_p du, \quad (30)$$

which can be interpreted as the aggregated MPR times the amount of temperature risk σ_t over $[t, \tau_1^i]$ (first integral) and $[\tau_1^i, \tau_2^i]$ (second integral). By adjusting the MPR value, these two terms contribute to the CAT futures price. For temperature futures with values that are positive related to weather changes in the short term, this implies a

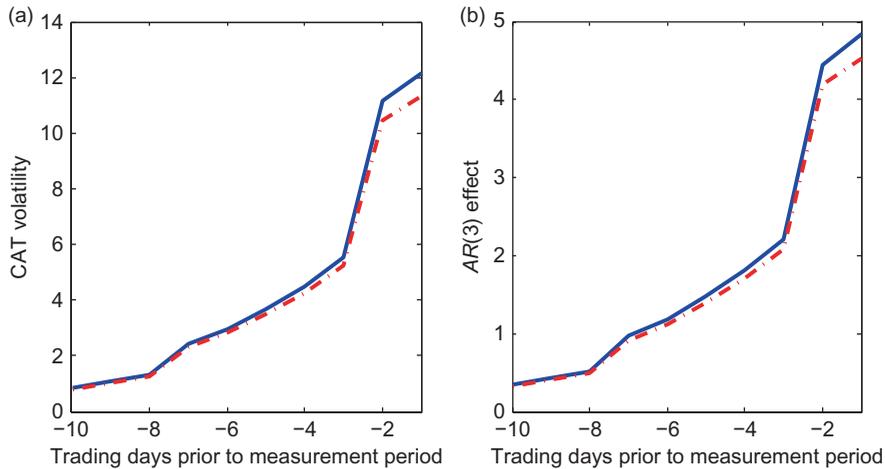


Figure 9. (a) The CAT term structure of volatility and (b) the autoregressive effect of two contracts issued on 20060517: one with whole June as measurement period (straight line) and the other one with only the 1st week of June (dotted line).

negative RP meaning that buyers of temperature derivatives expect to pay lower prices to hedge weather risk (insurance RP). In this case, θ_t must be negative for CAT futures, since σ_t and X_t are both positive. Negative MPRs translate into premiums for bearing risk, implying that investor will accept a reduction in the return of the derivative equal to the right-hand side of Equation (30) in exchange for eliminating the effects of the seasonal variance on pay-offs. On the other side, positive RP indicates the existence of consumers, who consider temperature derivatives for speculation purposes. In this case, θ_t must be positive and implies discounts for taking additional (weather) risk. This rules out the ‘burn-in’ analysis of Brix *et al.* (2005), which seems to popular among practitioners since it uses the historical average index value as the price for the futures. The sign of MPR–RP reflects the risk attitude and time horizon perspectives of market participants in the diversification process to hedge weather risk in peak seasons. By understanding the MPR, market participants might earn money (by shorting or longing, according to the sign). The investors impute value to the weather products, although they are non-marketable. This might suggest some possible relationships between risk aversion and the MPR.

The non-stationarity behaviour of the MPR (sign changes) is also possible because it is capturing all the non-fundamental information affecting the futures pricing: investors preferences, transaction costs, market illiquidity or other fractions like effects on the demand function. When the trading is illiquid the observed prices may contain some liquidity premium, which can contaminate the estimation of the MPR.

Figure 11 illustrates the RP of Berlin-CAT futures for monthly contracts traded on 20031006–20080527. We observe RPs different from 0, time dependent, where positive (negative) MPR contributes positively (negatively) to futures prices. The mean for the constant MPR for the $i = 1, \dots, 7$ th Berlin-CAT futures contracts per trading date is of size 0.02, 0.05, 0.02, 0.01, 0.10, 0.02 and 0.04, thus the terms in Equation (30)

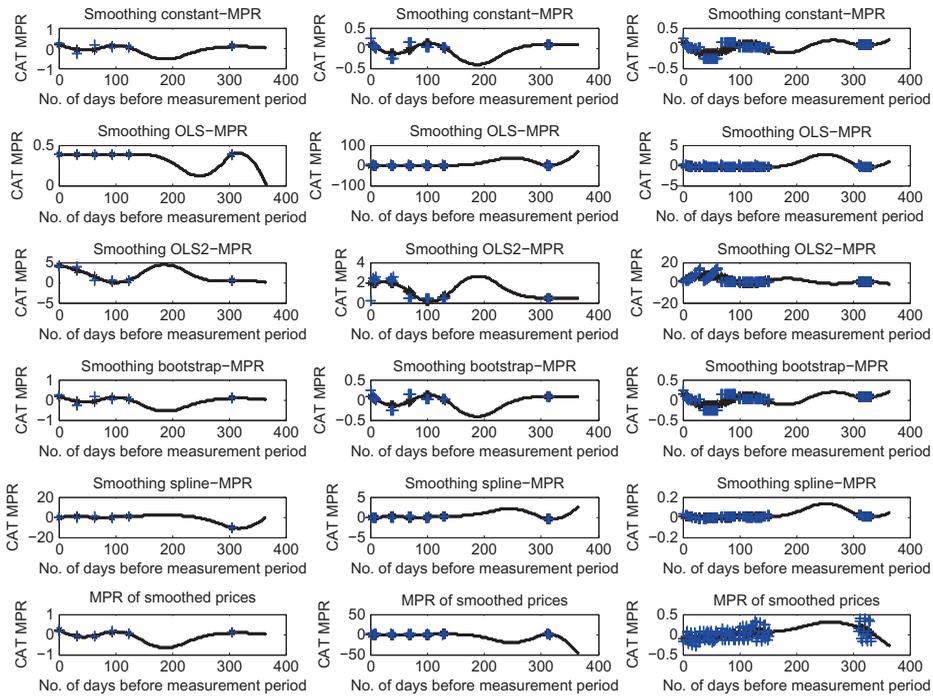


Figure 10. Smoothing the MPR parameterization for Berlin-CAT futures traded on 20060530: the calendar year smoothing (black line) for 1 day (left), 5 days (middle) and 30 days (right). The last row gives MPR estimates obtained from smoothed prices.

contribute little to the prices compared to the seasonal mean Λ_t . The RPs are very small for all contract types, and they behave constant within the measurement month but fluctuate with σ_t and θ_t , leading to higher RPs during volatile months (winters or early summers). This suggests that the temperature market does the risk adjustment according to the seasonal effect, where low levels of mean reversion mean that volatility plays a greater role in determining the prices.

Our data extracted MPR results can be comparable with Cao and Wei (2004), Richards *et al.* (2004) and Huang-Hsi *et al.* (2008), who showed that the MPR is not only different from 0 for temperature derivatives, but also significant and economically large as well. However, the results in Cao and Wei (2004) and Richards *et al.* (2004) rely on the specification of the dividend process and the risk aversion level, while the approach of Huang-Hsi *et al.* (2008) depends on the studied Stock index to compute the proxy estimate of the MPR. Alaton *et al.* (2002) concluded that the MPR impact is likely to be small. Our findings can also be compared with the MPR of other non-tradable assets, for example, in commodities markets; the MPR may be either positive or negative depending on the time horizon considered. In Schwartz (1997), the calibration of futures prices of oil and copper delivered negative MPR in both cases. For electricity, Cartea and Figueroa (2005) estimated a negative MPR. Cartea and Williams (2008) found a positive MPR for gas long-term contracts and for short-term

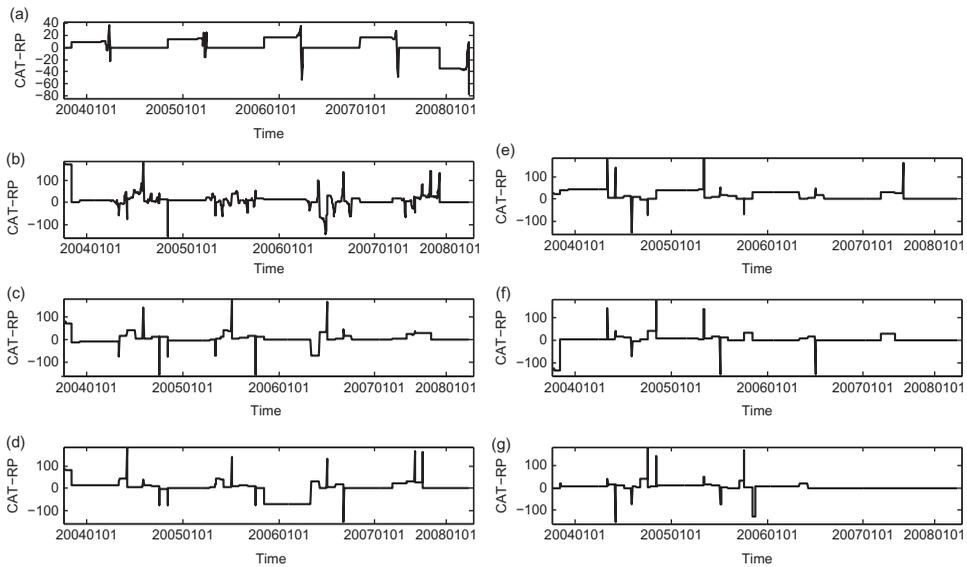


Figure 11. Risk premiums (RPs) of Berlin-CAT monthly futures prices traded during (20031006–20080527) with $t \leq \tau_1^i < \tau_2^i$ and contracts $i = 1$ (30 days), $i = 2$ (60 days), . . . , $i = I$ (210 days) traded before measurement period. RPs of Berlin CAT futures for (a) 30 days, (b) 60 days, (c) 90 days, (d) 120 days, (e) 150 days, (f) 180 days and (g) 210 days.

contracts the MPR changes signs across time. Doran and Ronn (2008) demonstrated the need of a negative market price of volatility risk in both equity and commodity-energy markets (gas, heating oil and crude oil). Similar to weather, electricity, natural gas and heating oil markets show seasonal patterns, where winter months have higher RP. The only difference is that in temperature markets, the spot–futures relation is not clear since the underlying is not storable (Benth *et al.*, 2008).

5.8 Pricing CAT–HDD–CDD and OTC Futures

Once that market prices of traded derivatives are used to back out the MPR for temperature futures, the MPR for options is also known and thus one can price other temperature contract types with different maturity (weekly, daily or seasonal contracts) and over the counter OTC derivatives (e.g. Berlin-CDD futures or for cities without formal WD market). This method seems to be popular among practitioners in other markets.

This section tests the MPR specifications to fit market prices in sample. The implied MPR (under multiple specifications) from monthly CAT futures in Section 5.7 are used to calculate theoretical CDD prices Equation (20) for Berlin, Essen and Tokyo. We then compute HDD futures prices from the HDD–CDD parity in Equation (4) and compare them with market data (in sample performance). Table 4 shows the CME futures prices (Column 5), the estimated risk-neutral prices with $P = Q$ (MPR = 0), the estimated futures prices with constant MPR for different contracts per trading date and the index values computed from the realized temperature data $I_{(\tau_1, \tau_2)}$. While

the inferred prices with constant MPR replicate market prices, the estimated prices with $P = Q$ are close to the realized temperatures, meaning that the history is likely a good prediction of the future. Table 6 describes the root mean squared errors (RMSEs) of the differences between the market prices and the estimated futures prices, with MPR values implied directly from specific futures contract types and with MPR values extracted from the HDD/CDD/CAT parity method, over different periods and cities. The RMSE is defined as

$$\text{RMSE} = \sqrt{n^{-1} \sum_{t=1}^n (F_{t,\tau_1^i,\tau_2^i} - \hat{F}_{t,\tau_1^i,\tau_2^i})^2},$$

where $\hat{F}_{t,\tau_1^i,\tau_2^i}$ are the estimated futures prices and small RMSE values denote good measure of precision. The RMSE estimates in the case of the constant MPR for different CAT futures contracts are statistically significant enough to know CAT futures prices, but fail for HDD futures. Since temperature futures are written on different indices, the implied MPR will be then contract-specific hence requiring a separate estimation procedure. We argue that this inequality in prices results from additional premiums that the market incorporates to the HDD estimation, due to possible temperature market probability predictions operating under a more general equilibrium rather than non-arbitrage conditions (Horst and Mueller, 2007) or due to the incorporation of weather forecast models in the pricing model that influence the risk attitude of market participants in the diversification process of hedging weather risk (Benth and Meyer-Brandis, 2009; Dorfleitner and Wimmer, 2010; Papazian and Skiadopoulos, 2010).

We investigate the pricing algorithm for cities without formal WD market. In this context, the stylized facts of temperature data (Δ_t, σ_t) are the only risk factors. Hence, a natural way to infer the MPR for emerging regions is by knowing the MPR dependency on seasonal variation of the closest geographical location with formal WD market. For example, for pricing Taipei weather futures derivatives, one could take the WD market in Tokyo and learn the dependence structure by simply regressing the average MPR of Tokyo-C24AT futures contracts i over the trading period against the seasonal variation in period $[\tau_1, \tau_2]$:

$$\hat{\theta}_{\tau_1,\tau_2}^i = \frac{1}{\tau_1 - t} \sum_t^{\tau_1} \hat{\theta}_t^i,$$

$$\hat{\sigma}_{\tau_1,\tau_2}^2 = \frac{1}{\tau_2 - \tau_1} \sum_{t=\tau_1}^{\tau_2} \hat{\sigma}_t^2.$$

In this case, the quadratic function that parameterizes the dependence is $\theta_t = 4.08 - 2.19\hat{\sigma}_{\tau_1,\tau_2}^2 + 0.28\hat{\sigma}_{\tau_1,\tau_2}^4$, with $R_{\text{adj}}^2 = 0.71$ and MPR increases by increasing the drift and volatility values (Figure 12). The dependencies of the MPR on time and temperature seasonal variation indicate that for regions with homogeneous weather risk there is some common market price of weather risk (as we expect in equilibrium).

Table 6. RMSE of the differences between observed CAT/HDD/CDD.

Contract type	Measurement period		No. of contracts	RMSE between estimated with MPR (θ_t) and CME prices					
	τ_1	τ_2		MPR = 0	Constant	1 piecewise	2 piecewise	Bootstrap	Spline
Atlanta-CDD ⁺	20070401	20070430	230	15.12	20.12	150.54	150.54	20.15	27.34
Atlanta-CDD ⁺	20070501	20070531	228	20.56	53.51	107.86	107.86	53.52	28.56
Atlanta-CDD ⁺	20070601	20070630	230	18.52	43.58	97.86	97.86	44.54	35.56
Atlanta-CDD ⁺	20070701	20070731	229	11.56	39.58	77.78	77.78	39.59	38.56
Atlanta-CDD ⁺	20070801	20070831	229	21.56	33.58	47.86	47.86	33.59	38.56
Atlanta-CDD ⁺	20070901	20070930	230	17.56	53.58	77.86	77.86	53.54	18.56
Berlin-HDD*	20061101	20061130	22	129.94	164.52	199.59	199.59	180.00	169.76
Berlin-HDD*	20061201	20061231	43	147.89	138.45	169.11	169.11	140.00	167.49
Berlin-HDD ⁺	20061101	20061130	22	39.98	74.73	89.59	89.59	74.74	79.86
Berlin-HDD ⁺	20061201	20061231	43	57.89	58.45	99.11	99.11	58.45	88.49
Berlin-CAT ⁺	20070401	20070430	230	18.47	40.26	134.83	134.83	40.26	18.44
Berlin-CAT ⁺	20070501	20070531	38	40.38	47.03	107.342	107.34	47.03	40.38
Berlin-CAT ⁺	20070601	20070630	58	10.02	26.19	78.18	78.18	26.20	10.02
Berlin-CAT ⁺	20070701	20070731	79	26.55	16.41	100.22	100.22	16.41	26.55
Berlin-CAT ⁺	20070801	20070831	101	34.31	12.22	99.59	99.59	12.22	34.31
Berlin-CAT ⁺	20070901	20070930	122	32.48	17.96	70.45	70.45	17.96	32.48
Essen-CAT ⁺	20070401	20070430	230	13.88	33.94	195.98	195.98	33.94	13.87
Essen-CAT ⁺	20070501	20070531	39	52.66	52.95	198.18	198.188	52.95	52.66
Essen-CAT ⁺	20070601	20070630	59	15.86	21.35	189.45	189.45	21.38	15.86
Essen-CAT ⁺	20070701	20070731	80	16.71	44.14	155.82	155.82	44.14	16.71
Essen-CAT ⁺	20070801	20070831	102	31.84	22.66	56.93	56.92	22.66	31.84
Essen-CAT ⁺	20070901	20070930	123	36.93	14.28	111.58	111.58	14.28	33.93
Tokyo-C24AT ⁺	20090301	20090331	57	161.81	148.21	218.99	218.99	148.21	158.16
Tokyo-C24AT ⁺	20090401	20090430	116	112.65	99.55	156.15	156.15	99.55	109.78
Tokyo-C24AT ⁺	20090501	20090531	141	81.64	70.81	111.21	111.21	70.81	79.68
Tokyo-C24AT ⁺	20090601	20090630	141	113.12	92.66	104.75	110.68	92.66	111.20
Tokyo-C24AT ⁺	20090701	20090731	141	78.65	74.95	116.34	3658.39	74.95	77.07

Notes: RMSE, root mean squared error; MPR, market price of risk; CME, Chicago Mercantile Exchange.

Futures prices with $t \leq \tau_1^i < \tau_2^i$ and the estimated futures with implied MPR under different MPR parameterizations (MPR = 0, constant MPR for different contracts (Constant), 1 piecewise constant MPR, 2 piecewise constant MPR, bootstrap MPR and spline MPR).

+ Computations with MPR implied directly from specific futures contract types (+) and * through the parity HDD/CDD/CAT parity method(*).

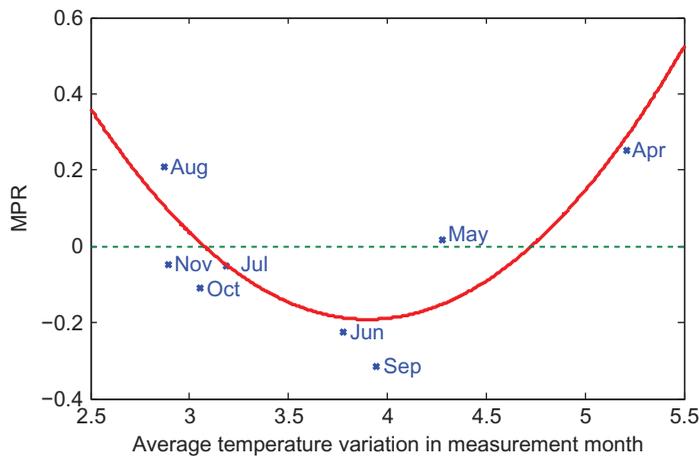


Figure 12. The calibrated MPR as a deterministic function of the monthly temperature variation of Tokyo-C24AT futures from November 2008 to November 2009 (prices for 8 contracts were available).

6. Conclusions and Further Research

This article deals with the differences between ‘historical’ and ‘risk-neutral’ behaviours of temperature and gives insights into the MPR, a drift adjustment in the dynamics of the temperature process to reflect how investors are compensated for bearing risk when holding the derivative. Our empirical work shows that independently of the chosen location, the temperature-driving stochastics are close to the Gaussian risk factors that allow us to work under the financial mathematical context.

Using statistical modelling, we imply the MPR from daily temperature futures-type contracts (CAT, CDD, HDD, C24AT) traded at the CME under the EMM framework. Different specifications of the MPR are investigated. It can be parameterized, given its dependencies on time and seasonal variation. We also establish connections between the RP and the MPR. The results show that the MPRs–RPs are significantly different from 0, changing over time. This contradicts with the assumption made earlier in the literature that MPR is 0 or constant and rules out the ‘burn-in’ analysis, which is popular among practitioners. This brings significant challenges to the statistical branch of the pricing literature, suggesting that for regions with homogeneous weather risk there is a common market price of weather risk. In particular, using a relationship of the MPR with a utility function, one may link the sign changes of the MPR with risk attitude and time horizon perspectives of market participants in the diversification process to hedge weather risk.

A further research on the explicit relationship between the RP and the MPR should be carried out to explain possible connections between modelled futures prices and their deviations from the futures market. An important issue for our results is that the econometric part in Section 2 is carried out with estimates rather than true values. One thus deals with noisy observations, which are likely to alter the subsequent estimations and test procedures. An alternative to this is to use an adaptive local parametric estimation procedure, for example, in Mercurio and Spokoiny (2004) or Härdle *et al.* (2011).

Finally, a different methodology, but related to this article, would be to imply the pricing kernel of option prices.

Acknowledgement

We thank Fred Espen Benth and two anonymous referees for several constructive and insightful suggestions on how to improve the article.

References

- Alaton, P., Djehiche, B. and Stillberger, D. (2002) On modelling and pricing weather derivatives, *Applied Mathematical Finance*, 9(1), pp. 1–20.
- Barrieu, P. and El Karoui, N. (2002) Optimal design of weather derivatives, *ALGO Research*, 5(1), pp. 79–92.
- Benth, F. (2003) On arbitrage-free pricing of weather derivatives based on fractional Brownian motion, *Applied Mathematical Finance*, 10(4), pp. 303–324.
- Benth, F., Cartea, A. and Kiesel, R. (2008) Pricing forward contracts in power markets by the certainty equivalence principle: explaining the sign of the market risk premium, *Journal of Finance and Banking*, 32(10), pp. 2006–2021.
- Benth, F., Härdle, W. K. and López Cabrera, B. (2011) Pricing Asian temperature risk. In: P. Cizek, W. Härdle and R. Weron (Eds.), *Statistical Tools for Finance and Insurance*, pp. 163–199 (Heidelberg: Springer-Verlag).
- Benth, F. and Meyer-Brandis, T. (2009) The information premium for non-storable commodities, *Journal of Energy Markets*, 2(3), pp. 111–140.
- Benth, F. and Saltyte-Benth, J. (2005) Stochastic modelling of temperature variations with a view towards weather derivatives, *Applied Mathematical Finance*, 12(1), pp. 53–85.
- Benth, F. and Saltyte-Benth, J. (2007) The volatility of temperature and pricing of weather derivatives, *Quantitative Finance*, 7(5), pp. 553–561.
- Benth, F., Saltyte-Benth, J. and Jalinska, P. (2007) A spatial-temporal model for temperature with seasonal variance, *Applied Statistics*, 34(7), pp. 823–841.
- Benth, F., Saltyte-Benth, J. and Koekebakker, S. (2007) Putting a price on temperature, *Scandinavian Journal of Statistics*, 34, pp. 746–767.
- Benth, F., Saltyte-Benth, J. and Koekebakker, S. (2008) *Stochastic Modelling of Electricity and Related Markets*, Advanced Series on Statistical Science and Applied Probability, 2nd ed. (Singapore: World Scientific).
- Brix, A., Jewson, S. and Ziehmann, C. (2005) *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations* (Cambridge: Cambridge University Press).
- Brockett, P., Golden, L. L., Wen, M. and Yang, C. (2010) Pricing weather derivatives using the indifference pricing approach, *North American Actuarial Journal*, 13(3), pp. 303–315.
- Brody, D., Syroka, J. and Zervos, M. (2002) Dynamical pricing of weather derivatives, *Quantitative Finance*, 2(3), pp. 189–198.
- Campbell, S. and Diebold, F. (2005) Weather forecasting for weather derivatives, *Journal of American Statistical Association*, 100(469), pp. 6–16.
- Cao, M. and Wei, J. (2004) Weather derivatives valuation and market price of weather risk, *The Journal of Future Markets*, 24(11), pp. 1065–1089.
- Cartea, A. and Figueroa, M. (2005) Pricing in electricity markets: a mean reverting jump diffusion model with seasonality, *Applied Mathematical Finance*, 12(4), pp. 313–335.
- Cartea, A. and Williams, T. (2008) UK gas markets: the market price of risk and applications to multiple interruptible supply contracts, *Energy Economics*, 30(3), pp. 829–846.
- Constantinides, G. (1987) Market risk adjustment in project valuation, *The Journal of Finance*, 33(2), pp. 603–616.
- Cox, J. C., Ingersoll, J. and Ross, S. (1985) A theory of the term structure of interest rates, *Econometrica*, 59, pp. 385–407.
- Doran, J. S. and Ronn, E. (2008) Computing the market price of volatility risk in the energy commodity markets, *Journal of Banking and Finance*, 32, pp. 2541–2552.

- Dorflleitner, G. and Wimmer, M. (2010) The pricing of temperature futures at the Chicago Mercantile Exchange, *Journal of Banking and Finance*, 34(6), pp. 1360–1370.
- Dornier, F. and Querel, M. (2007) Caution to the wind: energy power risk management, *Weather Risk Special Report*, August, pp. 30–32.
- Fengler, M., Härdle, W. and Mammen, E. (2007) A dynamic semiparametric factor model for implied volatility string dynamics, *Financial Econometrics*, 5(2), pp. 189–218.
- Gibson, R. and Schwartz, E. (1990) Stochastic convenience yield and the pricing of oil contingent claims, *Journal of Finance*, 45, pp. 959–976.
- Härdle, W. K., López Cabrera, B., Okhrin, O. and Wang, W. (2011) Localizing temperature risk, Working Paper, Humboldt-Universität zu Berlin.
- Horst, U. and Mueller, M. (2007) On the spanning property of risk bonds priced by equilibrium, *Mathematics of Operation Research*, 32(4), pp. 784–807.
- Huang-Hsi, H., Yung-Ming, S. and Pei-Syun, L. (2008) HDD and CDD option pricing with market price of weather risk for Taiwan, *The Journal of Future Markets*, 28(8), pp. 790–814.
- Hull, J. and White, A. (1990) Valuing derivative securities using the explicit finite difference method, *Journal of Financial and Quantitative Analysis*, 28, pp. 87–100.
- Karatzas, I. and Shreve, S. (2001) *Methods of Mathematical Finance* (New York: Springer-Verlag).
- Landskroner, Y. (1977) Intertemporal determination of the market price of risk, *The Journal of Finance*, 32(5), pp. 1671–1681.
- Lucas, R. E. (1978) Asset prices in an exchange economy, *Econometrica*, 46(6), pp. 1429–1445.
- Mercurio, D. and Spokoiny, V. (2004) Statistical inference for time-inhomogeneous volatility models, *The Annals of Statistics*, 32(2), pp. 577–602.
- Mraoua, M. and Bari, D. (2007) Temperature stochastic modelling and weather derivatives pricing: empirical study with Moroccan data, *Afrika Statistika*, 2(1), pp. 22–43.
- Papazian, G. and Skiadopoulos, G. (2010) Modeling the dynamics of temperature with a view to weather derivatives, Working Paper, University of Piraeus.
- Richards, T., Manfredo, M. and Sanders, D. (2004) Pricing weather derivatives, *American Journal of Agricultural Economics*, 86(4), pp. 1005–1017.
- Schwartz, E. (1997) The stochastic behaviour of commodity prices: implications for valuation and hedging, *Journal of Finance*, LII(3), pp. 923–973.
- Vasicek, O. (1977) An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, pp. 177–188.
- Zapranis, A. and Alexandridis, A. (2008) Modeling the temperature time-dependent speed of mean reversion in the context of weather derivative pricing, *Applied Mathematical Finance*, 15, pp. 355–386.
- Zapranis, A. and Alexandridis, A. (2009) Weather derivatives pricing: modeling the seasonal residual variance of an Ornstein-Uhlenbeck temperature process with neural networks, *Neurocomputing*, 73(1–3), pp. 37–48.

Using wiki to build an e-learning system in statistics in the Arabic language

Taleb Ahmad · Wolfgang Härdle ·
Sigbert Klinke · Shafiqah Alawadhi

Received: 1 June 2007 / Accepted: 4 February 2012
© Springer-Verlag 2012

Abstract E-learning plays an important role in education as it supports online teaching via computer networks and provides educational services by utilising information technologies. This paper presents a case study describing the development of an Arabic language e-learning course in statistics. Under discussion are problems concerning e-learning in Arab countries with special focus on the difficulties of the application of e-learning in the Arabic world as well as designing an Arabic platform with its language and technical challenges. For the platform we have chosen a wiki that supports $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ for formulas and R to generate tables and figures as well as some interactivity. Our system, Arabic MM*Stat, can be found at http://mars.wiwi.hu-berlin.de/mediawiki/mmstat_ar.

Keywords E-learning · MM*Stat · Wiki

T. Ahmad (✉)
Tishreen University, Latakia, Syria
e-mail: taleb_ahmad1976@yahoo.de

W. Härdle
Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin,
Unter den Linden 6, 10099 Berlin, Germany
e-mail: haerdle@wiwi.hu-berlin.de

S. Klinke
Ladislaus von Bortkiewicz Chair of Statistics, Humboldt-Universität zu Berlin,
Unter den Linden 6, 10099 Berlin, Germany
e-mail: sigbert@wiwi.hu-berlin.de

S. Alawadhi
Department of Statistics and Operations Research, Faculty of Science,
Kuwait University, P. O. Box 5969, 13060 Safat, Kuwait
e-mail: alawadi@kuc01.kuniv.edu.kw

1 Introduction

Due to the proliferation of the Internet, e-learning has become a significant aspect of education and many universities and educational institutions have created their own web sites and e-learning systems. Future trends predict that e-learning will significantly complement classic learning. Statistics show that the size of the worldwide e-learning market is estimated to be 52.6 billion US dollars yearly, with the ratio at 65–75% for the United States and Europe. Statistics also indicate that 30% of the education was delivered electronically. In comparison the e-learning market in Arab countries with a size around 15 million US dollars yearly is very weak. The gap between Europe and the United States and the Arab countries is very large.

The reasons for this gap is briefly summarised below:

- According to the latest figures available on Internet World Statistics 2010 (de Argaez 2011), Internet world usage still varies widely across the world and across languages as shown in Table 1. The diffusion of Internet services in the most Arab countries is weak compared to other regions of the world. This is mainly due to the government monopolies over the telecommunications sector, resulting in higher prices. As a consequence only 3.3% of Internet users come from the Arabic region, even though the Arabic population is 5% of world population. Another example for this gap is that the percentage of web users in the Arabic world is 18.8% compared with 58.4% in Europe, 77.4% in the USA and 28.7% on average in the whole world. Arabic users have much less experience with e-learning platforms, telecourses and educational courses.
- English is the most common language in the e-learning platforms, but most Arabic users have difficulties in understanding and speaking English.
- General educational problems: A high level of illiteracy can be found in the Arabic world which varies between 25 and 45% (Clayton 2007; Al-Fadhli 2008).
- There is only a limited number of specialised cadres and scientific expertise in the area of e-learning in Arab countries (Maegaard et al. 2005).

Due to the above mentioned problems Arab countries need more time to acquire the advantage of e-learning. The dissemination of the culture of e-learning in schools and universities needs a new generation of qualified professionals who can deal successfully with modern technology and the experiences of e-learning.

In fact, our Internet research showed that only a few Arabic e-learning platforms exists, especially for statistics we could not find a single one. For this reason we find the creation of a platform that would aid Arabic students in learning statistics highly necessary. The platform should cover the basic statistical topics, and is supported by multiple examples and ease-of-use will be adapted for Arabic students.

From the perspective described above we developed an Arabic e-learning platform in statistics (Arabic MM*Stat), which might become an important reference point in the study of statistics in Arabic through the Internet.

Around about 2000 a system known as MM*Stat was developed at the School for Business and Economics of Humboldt-Universität zu Berlin (Müller et al. 2000).

Table 1 World internet users for 10 languages by June 2010 (de Argaez 2011)

Top 10 languages in the internet	Internet users (Mio.)	Internet penetration, (%)	Growth in internet (2000–2010), (%)	Internet users % of total, (%)	World population (2010 estimate)
English	537	42.0	281	27.3	1,278
Chinese	445	32.6	1,277	22.6	1,278
Spanish	153	36.5	742	7.8	420
Japanese	99	78.2	111	5.0	127
Portuguese	83	33.0	990	4.2	250
German	75	78.6	173	3.8	96
Arabic	65	18.8	2,501	3.3	347
French	60	17.2	389	3.0	348
Russian	60	42.8	1,826	3.0	139
Korean	39	55.2	107	2.0	71
Top 10 languages	1,615	36.4	421	82.2	4,442
Other languages	351	14.6	588	17.8	2,403
World total	1,966	28.7	444	100.0	6,846

MM*Stat is a platform for e-learning statistics and is an HTML based multimedia environment to support teaching and learning statistics via CD or Internet.

A MM*Stat course consists of lectures of specific topics in basic statistics, see Fig. 1 for the *hypergeometric distribution*. Each lecture gives the basic concepts of general statistical theory, definitions, formulae and mathematical proofs. At the bottom is a set of buttons, on the left-hand side three buttons for navigation (go to the previous lecture, jump to the table of contents, go to the next lecture) and on the right-hand side a number of buttons which link to pages with additional information. Four types of additional information are provided, these are:

Explained examples which require only knowledge of the current lecture to understand them.

Enhanced examples which require knowledge from different lectures than the current one to understand them.

Interactive examples which allow the user, via an embedded statistical software, to run them. For example, to plot the probability density function or the cumulative distribution function for different parameters of n and p) or apply tests.

More information which contain for example historical information or mathematical derivations which are not necessary for first-hand understanding.

Each chapter with lectures is finished with a lecture containing multiple-choice questions such that a user can evaluate his/her learning progress.

Students or anyone interested in statistics can interactively learn about the basic concepts of statistics at anytime and anywhere and consequently we based Arabic

The screenshot shows a web browser window displaying the MM*Stat interface. At the top, there are navigation tabs for 'lecture', 'explained', 'interactive', and 'lecture'. The main content area is titled '6.4 Hypergeometric Distribution'. It contains the following text:

The Hypergeometric distribution is based on a **random event** with the following characteristics:

- total number of elements is N
- from the N elements, M elements have the property $N-M$ elements do not have this property, i.e. only two **events**, A and \bar{A} are possible
- we randomly choose n elements out of the N

This means the **probability** $P(A)$ is not constant and the draws (events) are not independent in this sort of experiment.

The **random variable** X , which contains number of successes A after n repetitions of the experiment has a Hypergeometric distribution with parameters N, M , and n , with **probability density function**:

$$f_H(x; N, M, n) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{for } x = \max\{0, n - (N - M)\}, \dots, \min\{n, M\} \\ 0 & \text{otherwise} \end{cases}$$

Shorthand notation is: $X \sim H(N, M, n)$

The **expected value** and the **variance** of the Hypergeometric distribution $H(N, M, n)$:

$$E(X) = n \cdot \frac{M}{N}$$

$$\text{Var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

At the bottom of the interface, there are navigation buttons for 'information', 'examples', 'enhanced', and 'interactive'.

Fig. 1 The graphical user interface of MM*Stat, as an example the lecture entitled *hypergeometric distribution*. Note the navigation button (*bottom left*) and buttons to examples and more information (*bottom right*). The tabs at the top reflect the user history and allow for a fast change between lectures

MM*Stat on the existing MM*Stat, which already existed in various languages: Czech, German, English, Spanish, French, Indonesian, Italian, Polish and Portuguese.

2 Difficulties to design Arabic platforms

There are some problems, however, associated with the making of an Arabic platform, these relate to language as well as technology. We summarise these problems below:

Language problems

There are some items related to translation, some words and scientific terms are similar in Arabic and could create a problem when translated. For example, see Table 2. The Arabic language makes no distinction between “administration” and “management” or “calculate” and “compute”. The reader must recognise from the context which meaning is correct. This makes a text more difficult to understand.

Technical problems

1. User interface

The different language versions of MM*Stat were based on two different systems:

Table 2 Some similar words in Arabic

Arabic	English
إدارة	Administration
إدارة	Management
حسب	Calculate
حسب	Compute

- The German version was written in HTML and the user interface was developed with JavaScript for Internet Explorer 5. The problem was that neither later Internet Explorer versions nor browsers other than the Internet Explorer were able to run the JavaScript code.
- The English version was written in L^AT_EX for a variety of reasons, for example, translating MM*Stat into a new language just required a change to the L^AT_EX text which is much easier to handle than translating from a HTML page with a lot of embedded JavaScript codes. We used a software based on LaTeX2HTML to create the HTML/JavaScript version of MM*Stat with the same user interface as before (Witzel and Klinke 2002).

2. *Writing from left to right*

Arabic script runs from right to left as opposed to most other languages, therefore all lists, paragraphs, statistical forms, tables and graphics also run from right to left. In some cases however Arabic text may contain information that needs to run in the opposite direction (from left to right) such as numbers and Latin texts. Any program that supports the Arabic language should provide the possibility of changing the direction when needed. A solution would be to use ArabTeX (Lagally 2004), but with ArabTeX the Arabic texts are written in English with special character combinations and not in Arabic, see Fig. 2. Obviously this is unfamiliar to most Arabic speaking people. Additionally LaTeX2HTML supports neither text from right to left, Arabic or ArabTeX.

3. *Interactive examples*

MM*Stat contained a set of interactive examples, which are important since they allow the user to practice repeatedly with various variables or data sets, and with alternate sample sizes or parameters of the statistical methods applied. In this manner, the student obtains a better understanding of how the statistical method works. However, the client-server technology implemented by Lehmann (2004) for MM*Stat worked only with the statistical software XploRe. The development and support of the XploRe software has unfortunately ceased, so the question arises how one should include the interactive examples in Arabic MM*Stat such that they will be runnable in future.

The language problem can only be solved by adapting the texts. To solve the technical problems we decided to use another technology, the so called “wiki technology”.

```

\documentstyle[12pt,arabtex,atrans,nashbf]{article}
\begin{document}
...
\begin{arabtext}
i^starY ^gu.hA 'a^saraTa .hamIriN.
fari.ha bihA wa-sAqahA 'amAmahu,
_tumma rakiba wA.hidaN minhA.
wa-fI al-.t.tarIqi 'adda .hamIrahu wa-huwa rAkibuN,
fa-wa^gadahA tis'aTaN.
_tumma nazala wa-'addahA fa-ra'AhA 'a^saraTuN fa-qAla:

'am^sI wa-'aksibu .himAraN,
'af.dalu min 'an 'arkaba wa-'a_hsara .himAraN.
\end{arabtext}
...
\end{document}

```

Fig. 2 Sample Arabtex input in L^AT_EX, see `examples/ghuA.tex` in Arabtex

3 Wiki technology

3.1 What is a wiki?

Wiki is a system that allows users to collaborate in forming the content of a web site. The first wiki web site, “WikiWikiWeb”, was designed by Cunningham and Leuf in 1995 (Leuf and Cunningham 2001). They describe the wiki system as a simple database that can operate on the World Wide Web. The goal is to simplify the process of participation and cooperation in the development of web content with maximum flexibility. The main advantages of a wiki are:

- Wiki simplifies the process of content editing. Each web page contains a link to change content within the web browser. After saving a modified page it can be viewed immediately.
- It uses simple markup to coordinate content, and it is suitable for users with little experience with computers or web site development, as no HTML language knowledge is required.
- Wiki sites keep a record of the page history and therefore makes the comparison of older and newer web pages an easy task. If a mistake is made, one can revert to the older version of the page.
- Wiki sites can be publicly open and therefore allow any user to improve the content.
- Wiki simplifies the organisation of a site: Wiki sites create hypertext databases and can regulate the content in any manner desired; many content management systems require the planning of the organization of the content before anything is written. This allows for flexibility which is not available in content management systems.

3.2 Application of wiki

The flexibility of the wiki concept makes it an ideal knowledge transfer tool, at universities, educational institutes, in companies and with specialised web sites. For example,

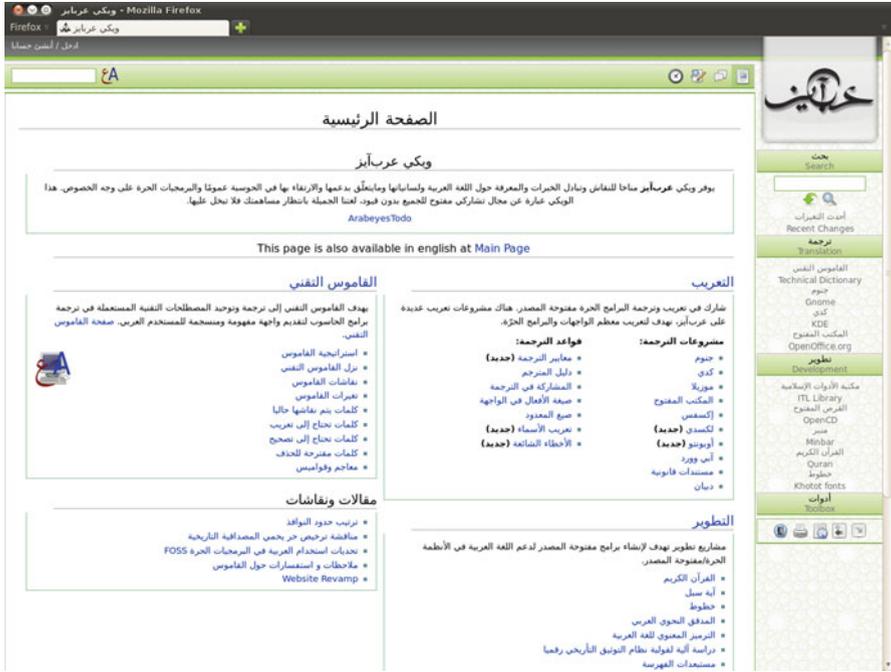


Fig. 3 Entry page of Arabeyes wiki (<http://wiki.arabeyes.org>)

a teacher could write his course using a wiki and offer it to his students as useful material for study, see e.g., [Klinke \(2011\)](#).

Nowadays, we have many more examples of web sites using wikis as a tool for the development of content, like [Wikipedia \(2011\)](#). The Wikipedia project started 15 January 2001 and today there are more than 10 million articles in the encyclopedia in all languages, more than 3.7 million articles in the English encyclopedia alone. Millions of volunteers around the world modify and add to the contents daily and new articles are created. The Arabic version of the free encyclopedia was launched in July 2003 and currently contains approximately 160 thousand articles as the Arabic encyclopedia is in the content-building phase.

The Arabic wiki platform “Arabeyes” ([Afifi et al. 2011](#)) provides a good environment for discussion and exchange of experience and knowledge about the Arabic language. Arabeyes offers the translation into Arabic for free open-source programs. In addition Arabeyes provides a technical dictionary that aims to translate and standardise the technical terms used in translating the software to the Arabic user. Arabeyes is a solution for the *language problem*, see [Fig. 3](#).

3.3 Implementation of Arabic MM*Stat

Arabic MM*Stat is directed at students and Arabic users that serve the e-learning issues in the Arabic region. The content of Arabic MM*Stat is a translation of the content of the former CD’s into Arabic.

المعلومات الإضافية لتطبيقات طريقة الأمكانية العظمى

المعلومات الإضافية لتطبيقات طريقة الأمكانية العظمى

تقدير الأمكانية العظمى للـ μ و σ^2 لأجل توزيع غاوسيان

نحرض المتغير العشوائي X الموزع بتوزيع غاوسيان مع المعاملات المجهولة μ و σ^2

نحرض أن X_1, \dots, X_n هي عينة عشوائية مستقلة من هذا التوزيع عندئذ لكل $i = 1, \dots, n$ لدينا

$$f(x_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

يعطى تابع الأمكانية العظمى بواسطة

$$L(\mu, \sigma^2|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2\right)$$

أحد التفاضلات بينج لدينا:

$$\log L(\mu, \sigma^2|x_1, \dots, x_n) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log \sigma^2 - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^n (x_i-\mu)^2$$

تقدير الأمكانية العظمى للـ μ و σ^2

تنظيم $L(\mu; \sigma^2)$ لأجل (x_1, \dots, x_n)

$\hat{\mu}$ و $\hat{\sigma}^2$ متحدره تعظم التابع التفاضلي للأمكانية العظمى بأحد الاشتقاقات الجزئية بالتفاضل إلى μ و σ^2

يجمع المعادلات الناتجة معاوية للتفاضل بينج:

$$\frac{\partial \log L}{\partial \mu} = -\frac{2 \cdot \sum_{i=1}^n (x_i - \hat{\mu}) \cdot (-1)}{2\sigma^2} = 0$$

Fig. 4 Graphical user interface (GUI) of Arabic MM*Stat. Note that the interface language (English) is different from the content language (Arabic) due to user settings

Wikimatrix (2011) offers an overview of the available wiki software and their capabilities. A useful wiki should support:

- $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ to provide the possibility to write a statistical formula in “mathematical” language rather than integrate it as a graphic, generated for example by $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}2\text{HTML}$.
- Arabic as a language for the content and the interface.
- Integration of statistical software, preferably R, to recreate interactive examples.
- Multiple choice questions to test students knowledge.

As wiki software we finally decided to use the Mediawiki, the software behind the (Arabic) Wikipedia. It solves all possible technical problems (see Figs. 4 and 5):

- *User interface*
It is able to have the content and the user interface in the Arabic language as the Arabic Wikipedia shows.
- *Writing from left to right*
To some extent, it can change the writing direction for formulas, list etc.
- *Interactive examples*
Through Mediawiki extensions we are able to transfer the functionality of the MM*Stat CD to the new system:
 - The R extension allows to embed (interactive) tables and graphics generated by R into wiki page as well as interactive examples.

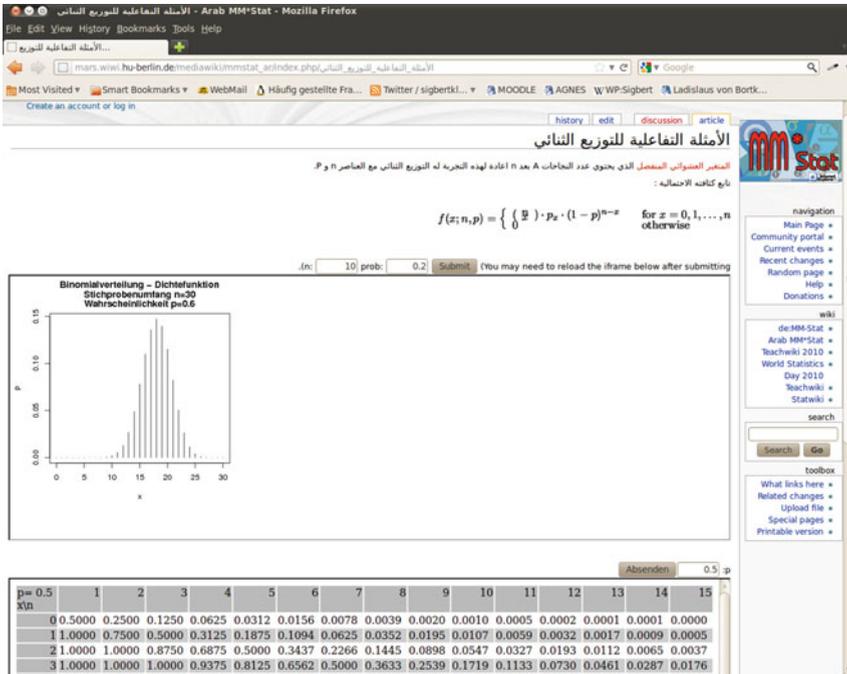


Fig. 5 An interactive example of a graphic of a probability density function and a table of a cdf of a binomial distribution ($p = 0.6$)

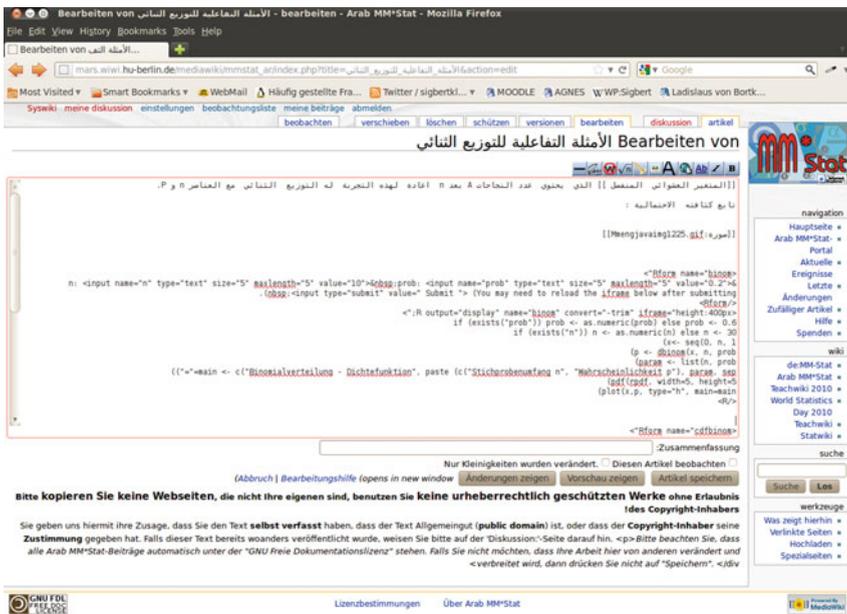


Fig. 6 The wiki source code for the page shown in Fig. 5. On top the Arabic text and within the RForm tags the input parameters and within the R tags the R program

- The Quiz extension provided multiple choice questions (Babé 2007).
- The Math extension allows formulas written in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ to be embedded into the wiki page (Wegrzanowski and Vibber 2011).

3.4 Integration of R program into Arabic MM*Stat

R is a language and environment for statistical computing and graphics (R Development Core Team 2011). Arabic MM*Stat uses R programs to create tables and graphics which can be incorporated in courses notes. For the Mediawiki software an extension to embed R into the wiki page exists.

They enable the students and learners, for example to visualise statistics distributions and probability tables via the Internet. See Fig. 5 as an example of a graphic of a probability density function and a table of a cdf function of a binomial distribution. Choosing other input values will lead to different tables or graphics.

Figure 6 shows the wiki source code for the example shown in Fig. 5. The interactive example consists of two tags `Rform` and `R` which share a common attribute name.

```
<Rform name='binom'>
... Input parameters...
</Rform>

<R output='display' name='binom'>
... R program...
</R>
```

Between the opening and closing `Rform` tags are the input parameters as defined in an HTML form. The following opening and closing `R` tag contain the R program which produces a graphic. For more detail see Klinke and Zlatkin-Troitschanskaia (2007).

There are in Arabic MM*Stat other examples, e.g., for other distributions like normal, Poisson and exponential distribution.

4 Conclusion

Using E-learning/e-teaching tools to offer effective learning of statistics is a necessity for students. There is the possibility of creating an e-learning system with Arabic MM*Stat through the application of wiki technology. Some of the specific characteristics we have discussed earlier for developing an Arabic platform already exist in the wiki. We see that embedding of R is an solution for the interactive examples in Arabic MM*Stat. We hope that the Arabic MM*Stat platform for e-learning of statistics will be a significant help for the Arabic user as it clearly overcomes weaknesses in developing such electronic platforms in Arabic.

This research was supported by the Deutsche Forschungsgemeinschaft through the CRC 649 ‘Economic Risk’.

References

- Afifi D, Chahibi Y, Hosny K, Trojette MA (2011) Arabeyes.org—The Arabic Unix Project. <http://wiki.arabeyes.org>. Accessed 18 Aug 2011
- Al-Fadhli S (2008) Students' perceptions of e-learning in Arab society: Kuwait University as a case study. *E-Learning* 5(4):418–428
- Babé LR (2007) Extension: Quiz. <http://www.mediawiki.org/wiki/Extension:Quiz>, Accessed 26 Aug 2011
- CosmoCode (2011) Wikimatrix. <http://www.wikimarix.org>, Accessed 24 Aug 2011
- Clayton PM (2007) Women, violence and the internet. *E-Learning* 4(1):79–92
- de Argaez E (2011) Internet world stats, Columbia 2011. <http://www.internetworldstats.com>, Accessed 18 Aug 2011
- Klinke S (2011) Developing web-based tools for the teaching of statistics: Our Wikis and the German Wikipedia. ISI Conference Proceedings, Dublin/Ireland (forthcoming)
- Klinke S, Zlatkin-Troitschanskaia O (2007) Embedding R in the Mediawiki, Sonderforschungsbereich 649 Discussion Papers 61, Humboldt University, Berlin/Germany 2007. <http://edoc.hu-berlin.de/docviews/abstract.php?id=28421>, Accessed 18 Aug 2011
- Lagally K (2004) ArabTeX : a system for typesetting Arabic and Hebrew. Report Nr. 2004/03, User Manual Version 4.00, Fakultät Informatik, University Stuttgart, Germany. http://www2.informatik.uni-stuttgart.de/ivi/bs/research/arab_e.htm, Accessed 18 Aug 2011
- Lehmann H (2004) Client/server based statistical computing. Dissertation, Humboldt-Universität zu Berlin, Germany. <http://edoc.hu-berlin.de/docviews/abstract.php?id=20803>, Accessed 18 Aug 2011
- Leuf B, Cunningham W (2001) The wiki way: quick collaboration on the web. Addison-Wesley, Reading, MA
- Maegaard B, Choukri K, Mokbel C, Yaseen M (2005) Language technology for Arabic. Report of The NEMLAR project, Center for Sprogteknologi. University of Copenhagen, Denmark. http://www.medar.info/The_Nemlar_Project/Publications/Arabic_LT.pdf, Accessed 18 Aug 2011
- Müller M, Rönz B, Ziegenhagen U (2006) The Multimedia Project MM*Stat for Teaching Statistics. In: Proceedings of the COMPSTAT 2000. pp 409–414
- R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna/Austria. <http://www.r-project.org>, Accessed 18 Aug 2011
- Wegrzanowski T, Vibber B et al. (2011) Extension: Math. <http://www.mediawiki.org/wiki/Extension:Math>, Accessed 26 Aug 2011
- Wikimedia Foundation Inc. (2011) Wikipedia: The free encyclopedia. <http://www.wikipedia.org>, Accessed 18 Aug 2011
- Witzel R, Klinke S (2002) MD*Book online & e-stat: Generating e-stat Modules from Latex. In: Proceedings of the COMPSTAT 2002. pp 449–454

Shape Invariant Modeling of Pricing Kernels and Risk Aversion

MARIA GRITH

*CASE—Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin*

WOLFGANG HÄRDLE

*CASE—Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin*

JUHYUN PARK

Department of Mathematics and Statistics, Lancaster University

ABSTRACT

Several empirical studies reported that pricing kernels exhibit a common pattern across different markets. The main interest in pricing kernels lies in validating the presence of the peaks and their variability in location among curves. Motivated by this observation we investigate the problem of estimating pricing kernels based on the shape invariant model, a semi-parametric approach used for multiple curves with shape-related nonlinear variation. This approach allows us to capture the common features contained in the shape of the functions and at the same time characterize the nonlinear variability with a few interpretable parameters. These parameters provide an informative summary of the curves and can be used to make a further analysis with macroeconomic variables. Implied risk aversion function and utility function can also be derived. The method is demonstrated with the European options and returns values of the German stock index DAX. (JEL: C14, C32, G12)

KEYWORDS: pricing kernels, risk aversion, risk neutral density

The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin is gratefully acknowledged. Address correspondence to Juhyun Park, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK, or email: juhyun.park@lancaster.ac.uk

doi:10.1093/jffinec/nbs019 Advanced Access publication November 18, 2012

© The Author, 2012. Published by Oxford University Press. All rights reserved.

For Permissions, please email: journals.permissions@oxfordjournals.org

1 METHODOLOGY

1.1 Pricing Kernel and Risk Aversion

Risk analysis and management drew much attention in quantitative finance recently. Understanding the basic principles of financial economics is a challenging task in particular in a dynamic context. With the formulation of utility maximization theory, individuals' preferences are explained through the shape of the underlying utility functions. Namely, a concave, convex, or linear utility function is associated with risk averse, risk seeking, or risk neutral behavior, respectively. The comparison is often made through the Arrow-Pratt measure of absolute risk aversion (ARA), as a summary of aggregate investor's risk-averseness. The quantity is originated from the expected utility theory and is defined by

$$ARA(u) = -\frac{U''(u)}{U'(u)},$$

where U is the individual utility as a function of wealth.

With an economic consideration that one unit gain and loss does not carry the same value for every individual, understanding state-dependent risk behavior becomes an increasingly important issue. The fundamental problem is that individual agents are not directly observable but it is assumed that the prices of goods traded in the market reflect the dynamics of their risk behavior. Several efforts have been made to relate the price processes of assets and options traded in a market to risk behavior of investors, since options are securities guarding against losses in risky assets.

A standard option pricing model in a complete market assumes a *risk neutral* distribution of returns, which gives the fair price under no arbitrage assumptions. If markets are not complete, there are more risk neutral distributions and the fair price depends on the hedging problem. The *subjective or historical* distribution of observed returns reflects a risk-adaptive behavior of investors based on subjective assessment of the future market. Then the equilibrium price is the arbitrage free price and the transition from risk neutral pricing to subjective rule is achieved through the pricing kernel. Assuming those densities exist, write q for the risk neutral density and p for the historical density. The pricing kernel \mathcal{K} is defined by the ratio of those densities:

$$\mathcal{K}(u) = \frac{q(u)}{p(u)}.$$

Through the intermediation of these densities, there exists a link between the pricing kernel and ARA, see for example Leland (1980)

$$ARA(u) = \frac{p'(u)}{p(u)} - \frac{q'(u)}{q(u)} = -\frac{d \log \mathcal{K}(u)}{du}.$$

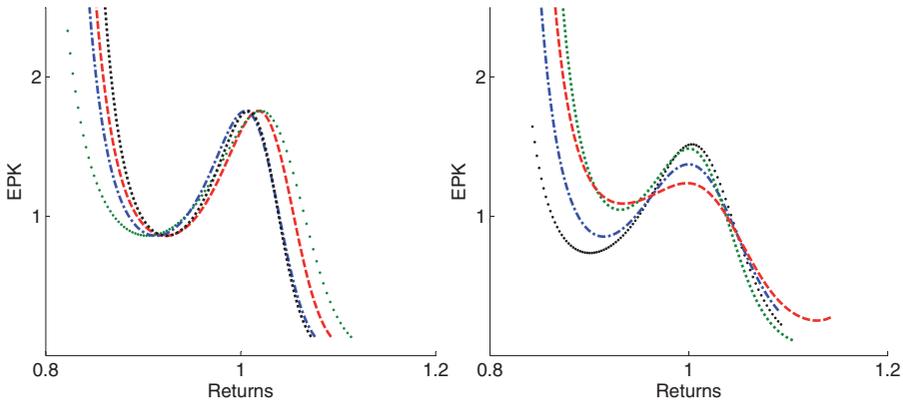


Figure 1 Examples of inter-temporal pricing kernels for various maturities in January–February 2006 (left) and monthly pricing kernels from the first six months in 2006 for maturity one month (right).

In this way, rather than specifying a priori preferences of agents (risk neutral, averse, or risk seeking) and implicitly the monotonicity of the pricing kernel, we can infer the risk patterns from the shape of the pricing kernel.

1.2 Dynamics of Empirical Pricing Kernels (EPKs)

With increasing availability of large market data, several approaches to recovering pricing kernels from empirical data have been proposed. As many of them estimate p and q separately to recover \mathcal{K} , potentially relevant are studies focusing on recovering risk neutral density, see e.g. Jackwerth (1999), and Bondarenko (2003) for comparison of different approaches. For the estimation of p nonparametric kernel methods or parametric models such as GARCH or Heston models are popular choices.

Examples of empirical pricing kernels obtained from European options data on the German stock index DAX (Deutscher Aktien index) in 2006 are shown in Figure 1, based on separate estimation of p and q . A detailed account of estimation is given in Section 3.4. To make these comparable, they are shown on a continuously compounded returns scale. Throughout the article, the pricing kernel is considered as a function of this common scale of returns. Figure 1 depicts inter-temporal pricing kernels with various maturities in January–February 2006 (left), and monthly pricing kernels with fixed maturity one month in 2006 (right). The sample of curves appears to have a bump around 1 and has convexity followed by concavity in all cases. The location as well as the magnitude of the bump vary among curves, which reflects individual variability on different dates or under different investment horizons. Some features that are of particular economic interest include

the maximum of the bump, the spread or duration of the bump and the location of the bump.

From a statistical perspective, properties of the pricing kernel are intrinsically related to assumptions about the data generation process. A very restrictive model, with normal marginal distributions, is the Black–Scholes model. This results in an overall decreasing pricing kernel in wealth, which is consistent with overall risk-averse behavior. These preferences are often assumed in the classical economic theory of utility-maximizing agent and correspond to a concave indirect von Neumann and Morgenstern utility function. However, under richer parametric specifications or nonparametric models monotonicity of the pricing kernel has been rejected in practice (Rosenberg and Engle, 2002; Giacomini and Härdle, 2008). The phenomenon of locally nondecreasing pricing kernel is referred to as *the pricing kernel puzzle* in the literature. There have been many attempts to reconcile the underlying economic theory with the empirical findings. A recent solution suggested by Hens and Reichlin (2012) relates the puzzle to the violation of the fundamental assumptions in the equilibrium model framework.

Most of earlier works adopt a static viewpoint, showing a snapshot of markets on selected dates but report that there is a common pattern across different markets. The first dynamic viewpoint appears in Jackwerth (2000), who recovers a series of pricing kernels in consecutive times and claims that these do not correspond to the basic assumptions of asset pricing theory. In a similar framework Giacomini and Härdle (2008) perform a factor analysis based on the so-called dynamic semiparametric factor models, while Giacomini, Härdle and Handel (2008) introduce time series analysis of daily summary measures of pricing kernels to examine variability between curves.

Chabi-Yo, Garcia, and Renault (2008) explain the observed dynamics or the puzzles by means of latent variables in the asset pricing models. Effectively, they propose to build conditional models of the pricing kernels given the state variables reflecting preferences, economic fundamentals, or beliefs. Within this framework they are able to reproduce the puzzles, in conjunction with some joint parametric specifications for the pricing kernel and the asset return processes.

Due to evolution of markets over time under different circumstances, the pricing kernels are intrinsically time varying. Thus, approaches that do not take into account the changing market make limited use of information available in the current data. On the other hand, changes over time may not be completely arbitrary, as there are common rules and underlying laws that assure some consistency across different market system. Moreover, variability observed in pricing kernels, as shown in Figures 1, is not necessarily linear, and thus factors constructed from a linear combination of observations are only meaningful for explaining aggregated effects.

Considering the pricing kernels as an object of curves, we approach the problem of estimating the pricing kernels and implied risk aversion functions from a functional data analysis viewpoint (Ramsay and Silverman, 2002). The main interest in pricing kernels lies in validating the presence of the peaks and their

variability in location among curves. Motivated by this observation we investigate the estimation method based on the *shape invariant model*, which will be formally introduced in Section 2. This is chosen over the commonly adopted functional principal component analysis to accommodate the nonlinear features such as variation of peak locations, which encapsulate quantities amenable to economic interpretation. The shape invariant model allows us to capture the common characteristics, reported across different studies on different markets. We then explain individual variability as a deviation from the common curve or a reference.

Our contribution is three-fold. Firstly, we analyze the phenomenon of pricing kernel puzzle from a dynamic viewpoint using shape invariant modeling approach. The starting question was how to compare the empirical evidence. By taking into account variability among curves, we quantify a *trend of the puzzle* in the series of the pricing kernels by a few interpretable parameters. Secondly, we provide a unified framework for estimation and interpretation of ARA and utility functions consistent with the underlying pricing kernels with the same set of parameters. The ARA corresponding to the reference pricing kernel could be viewed as a typical pattern of risk behavior for the period under consideration. Due to nonlinear transformation involved in deriving ARA from the pricing kernel function, this common ARA function does not necessarily coincide with the simple average ARA functions. Thirdly, the output of the analysis provides a summary measure to study the relationship with macroeconomic variables. Through real data example we have related the changes in risk behavior to some macroeconomic variables of interest and found that local risk loving behavior is procyclical. We acknowledge that we do not provide an economic explanation to the puzzle but rather try to understand the nature of the phenomenon by means of statistical analysis.

The paper is organized as follows. Section 2 motivates common shape modeling approach and Section 3 reviews the shape invariant model and describes it in detail in the context of pricing kernel estimation. This section serves the basis of our analysis. Numerical studies based on simulation are found in Section 4. An application to real data example is summarized in Section 5.

2 COMMON SHAPE MODELING

2.1 Shape Invariant Model for Pricing Kernel

We consider a common shape modeling approach for the series of pricing kernels with explicit components of location and scale. To represent varying pricing kernels, we introduce the time index t in the pricing kernel as \mathcal{K}_t and consider a general regression model:

$$Y_t = \mathcal{K}_t + \varepsilon_t,$$

where ε_t represents an error with mean 0 and variance σ_t^2 . We begin with a working assumption of independent error as in Kneip and Engel (1995). The effect of

dependent error is investigated in simulation studies in Section 4.2. The relationship among K_t s is specified as

$$\mathcal{K}_t(u) = \theta_{t1} g\left(\frac{u - \theta_{t3}}{\theta_{t2}}\right) + \theta_{t4}, \quad (1)$$

with some unknown constants $\theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4})$ and an unknown function g . The common shape function g can be interpreted as a reference curve. Deviation from the reference curve is described by four parameters $\theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4})$ that capture scale changes and a shift in horizontal and vertical direction. This parametrization in (1) is commonly known as a shape invariant model (SIM), originally introduced by Lawton, Sylvestre and Maggio (1972) and studied by Kneip and Engel (1995). Note that the model includes as a special case complete parametric models with known g .

In contrast to standard applications of SIM as a regression model, the SIM application to pricing kernel estimation does not, strictly speaking, satisfy the model assumption. There is no realization of the pricing kernels available and thus our formulation of regression model should be viewed as an approximation. The original data used would be intraday options data and daily returns data, which are collected from separate sources with sample sizes of different orders of magnitude but estimation of p and q can be effectively done independently of each other. It may be possible to elaborate our approach to incorporate simultaneous estimation with a two-step state-dependent dynamic model formulation whereby the dynamics of the observed return processes are specified and the unobserved pricing kernel processes enter as a state variable. However, with current advancement in the methodology, this is only possible with limited parametric model choices, see for example Chabi-Yo *et al.* (2008), and extension to a flexible shape invariant model is left for future work.

Instead we exploit the fact that preliminary estimates of pricing kernels based on separate estimation of p and q are readily available from market data and this can easily substitute Y . From now on, we treat the estimates as something observable and denote by Y_t , similar to the regression formulation with direct measurements Y_t and state the asymptotic result without further complication of pre-processing steps. After all, these estimates of curves are available from the beginning and the SIM aims to characterize a structural relationship among these curves. This however may impact the parametric rate of convergence attainable (Kneip and Engel, 1995) because our observations are already contaminated by a nonparametric error of estimation. As is shown in Section 3.6, the dominating error comes from the estimation of q , which involves second derivative estimation. The optimal rate of convergence for estimating second derivative is known to be $\mathcal{O}(N^{-2/9})$, where N is the sample size used (Stone, 1982). This implies that $\sigma_t^2 = \alpha_{N,t} v^2$ where $\alpha_{N,t}$ is a constant of order $\mathcal{O}(N^{-2/9})$, which should be understood as the multiplication factor for the parametric rate of convergence.

A particular choice of estimates of individual pricing kernels is not part of the model formulation but affects the starting values for the estimation of shape

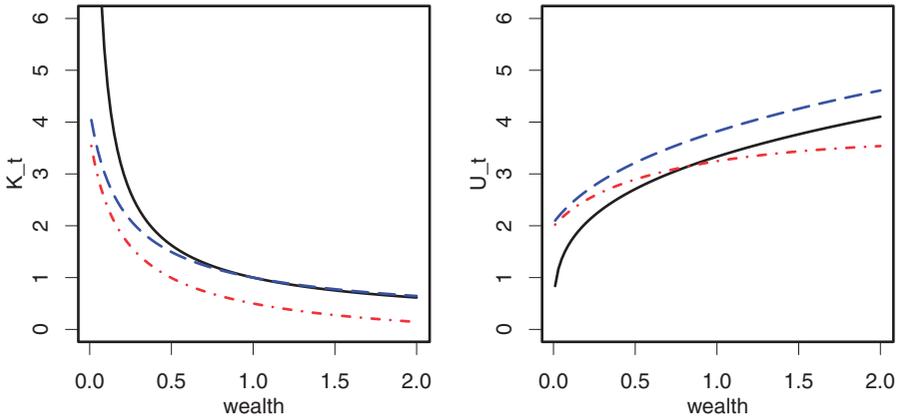


Figure 2 Example of location and scale shift family of pricing kernels (left) and corresponding utility functions (right). Solid line in each plot represents reference curves of $g(u) = u^{-\gamma}$ and $U_0(u) = u^{1-\gamma} / (1-\gamma)$ with $\gamma = 0.7$, respectively. Parameters are $\theta_{t1} = 1.1, \theta_{t2} = 1, \theta_{t3} = 1 - \theta_{t1}^{(1/\gamma)}$, and $\theta_{t4} = 0$ for dot-dashed (red) and $\theta_{t4} = -0.5$ for dashed (blue) lines.

invariant model. Our choice of initial estimates will be explained in Section 3.4. Our main interest lies in quantifying the variation among the pricing kernels given those estimates.

The new message here is an analysis of a sequence of pricing kernels through shape invariant models. Although we start with different motivation, our approach is in line with that of Chabi-Yo *et al.* (2008). In contrast to their approach, we impose a structural constraint that is related to the shape of the function. This way we strike a balance between flexibility much desired in parametric model specification and interpretability of the results lacking in full nonparametric models.

2.2 SIM and Black–Scholes Model

To appreciate the model formulation, given in the Equation (1), it is instructive to consider utility functions implied by this family of pricing kernels together. The utility function can be derived from

$$U_t(u) = \alpha \int_0^u \mathcal{K}_t(x) dx,$$

for a constant α . Figure 2 shows an example based on a power utility function, which corresponds to risk averse behavior. Pricing kernels \mathcal{K}_t are shown on the left and the corresponding utility functions U_t are on the right. The solid lines represent reference curves and the dashed and dot-dashed lines represent \mathcal{K}_t and U_t with appropriate parameters θ_t in the Equation (1). Depending on the choice of

parameters, the utility function can increase quickly or slowly. As an illustration, we consider the Black–Scholes model with power utility function. The Black–Scholes model assumes that the stock price follows a geometric Brownian motion

$$dS_t/S_t = \mu dt + \sigma dW_t,$$

which gives rise to a log normal distribution for the historical density p . Under the risk neutral measure, the drift μ is replaced by the riskless rate r but the density q is still log normal. The pricing kernel can be written as a power function

$$\mathcal{K}(u) = \lambda u^{-\gamma}, 0 < \gamma < 1,$$

with appropriate constants λ and γ . The corresponding utility function is a power utility

$$U(u) = \lambda \frac{u^{1-\gamma}}{1-\gamma}.$$

Assume that $\lambda = 1$ and suppose that g is a power function, say $u^{-\gamma}$. Then the class of pricing kernels implied by (1) is given by

$$\begin{aligned} \mathcal{K}_t(u) &= \theta_{t1} \left(\frac{u - \theta_{t3}}{\theta_{t2}} \right)^{-\gamma} + \theta_{t4} \\ &= \theta_{t1}^* (u - \theta_{t3})^{-\gamma} + \theta_{t4}, \end{aligned}$$

where $\theta_{t1}^* = \theta_{t1} \theta_{t2}^\gamma$. Notice that with this family of functions θ_{t1} and θ_{t2} are not identifiable and \mathcal{K}_t is defined for $u > \theta_{t3}$. For the sake of argument we set $\theta_{t2} = 1$ for the moment. The corresponding utility function is

$$\begin{aligned} U_t(u) &= \int_{\theta_{t3}}^u \mathcal{K}_t(x) dx \\ &= \frac{\theta_{t1}}{1-\gamma} (u - \theta_{t3})^{1-\gamma} + \theta_{t4} (u - \theta_{t3}) \\ &\stackrel{\text{def}}{=} \theta_{t1}^{**} (u - \theta_{t3})^{1-\gamma} + \theta_{t4} (u - \theta_{t3}). \end{aligned}$$

When $\theta_{t4} = 0$, this produces again a transformed power utility. When $\theta_{t4} \neq 0$, there is additional linear term in the function. See Figure 2 for comparison.

2.3 Identifiability Condition for SIM

The previous section illustrates two aspects of applicability of the shape invariant models. The class of functions that can be generated by the relation (1) is rich, but in order to uniquely identify the model parameters, some restriction is necessary.

For example, we have seen that the two scale parameters in the pricing kernel functions corresponding to the Black–Scholes model are not separable. Basically, unless there exist some qualitatively distinct common characteristics for each curve, the model is not identifiable (Kneip and Gasser, 1988). In the case of no prior structural information available as in the case of pricing kernels, it is sufficient to consider a few landmarks such as peaks and inflection points.

Even with a unique g , some translation and scaling of parameters lead to multiple representations of the models. For uniqueness of parameters, we will impose normalizing conditions suggested in Kneip and Engel (1995):

$$T^{-1} \sum_{t=1}^T \theta_{t1} = 1, \quad T^{-1} \sum_{t=1}^T \theta_{t2} = 1, \quad T^{-1} \sum_{t=1}^T \theta_{t3} = 0, \quad T^{-1} \sum_{t=1}^T \theta_{t4} = 0$$

in the sense that there exists an *average curve*. These conditions are not restriction at all and can be replaced by any appropriate combination of parameters. Alternatively, we could consider the first curve as a reference, as done in Härdle and Marron (1990), which implies the restriction $\theta_1 = (1, 1, 0, 0)$. Generally, an application-driven normalization scheme can be devised and examples are found in Lawton, Sylvestre and Maggio (1972).

2.4 SIM Implied Risk Aversion and Utility Function

The utility function corresponding to \mathcal{K}_t is given by

$$\begin{aligned} U_t(u) &= \theta_{t1}\theta_{t2} \left\{ G\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right) - G\left(-\frac{\theta_{t3}}{\theta_{t2}}\right) \right\} + \theta_{t4}u \\ &\equiv \theta_{t1}^* G\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right) + \theta_{t4}^* + \theta_{t4}u, \end{aligned}$$

where $G(t) = \int_0^t g(u) du$. The utility function U_t is a combination of a SIM class of the common utility function and a linear utility function.

The ARA measure is given by

$$ARA_t(u) = \frac{-\frac{\theta_{t1}}{\theta_{t2}} g'\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right)}{\theta_{t1} g\left(\frac{u-\theta_{t3}}{\theta_{t2}}\right) + \theta_{t4}}. \quad (2)$$

For example, assuming $g(u) = u^{-\gamma}$ with $\theta_{t2} = 1$ gives

$$ARA_t(u) = \gamma \left\{ (u - \theta_{t3}) + (\theta_{t4}/\theta_{t1})(u - \theta_{t3})^{\gamma+1} \right\}^{-1}.$$

When $\theta_{t4} = 0$, this function is monotonically decreasing but in general this is not the case. Note the common ARA function corresponding to g is γu^{-1} compared to the mean ARA function computed by taking the sample average $T^{-1} \sum_{t=1}^T ARA_t(u)$.

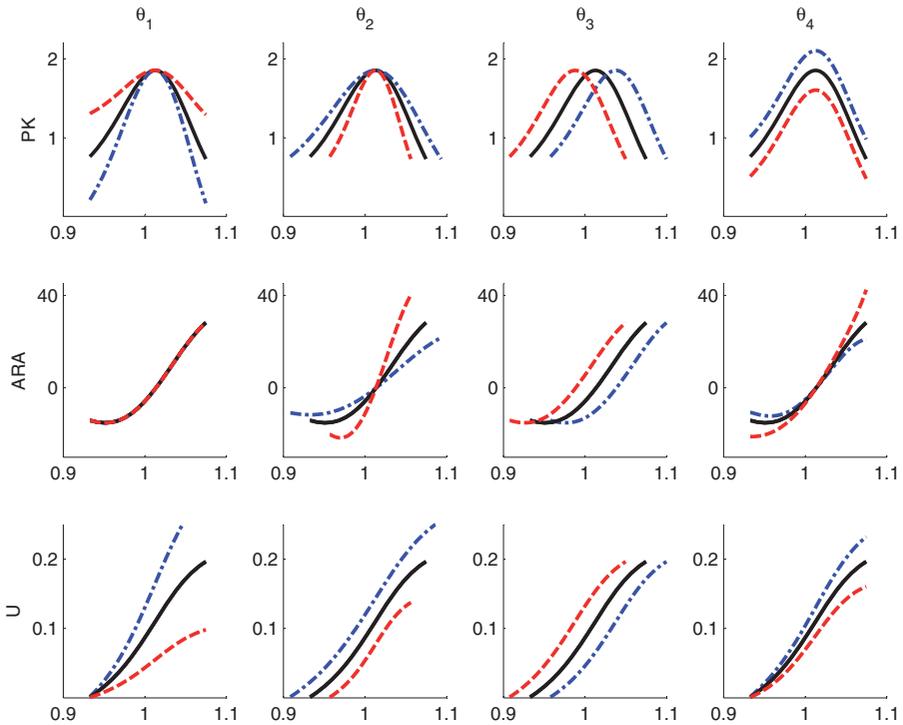


Figure 3 Effect of parameters on pricing kernel (top), ARA (middle), and utility function (bottom) compared to the baseline model $\theta_0 = (1, 1, 0, 0)$ (black). Dot-dashed lines are used for increasing direction and dashed lines for decreasing direction.

In order to gain some insights, we take a closer look at the changes in relation to individual scale and shift parameters. These individual effects are demonstrated in Figure 3. We vary each θ_i with respect to a baseline model and then we show how these modifications translate into changes of the risk attitudes and the corresponding utility functions. The parameters used in Figure 3 are $\theta = (0.5, 0.7, -0.025, -0.25)$ in dashed line and $\theta = (1.5, 1.3, 0.025, 0.25)$ in dot-dashed line.

For this exercise we first standardize the common curve that we have estimated via the shape invariant model so that the peak occurs at the value 0 on the abscissa and the effect of the scale and shift parameters is separately captured. But we added the peak coordinates back for visualization so that they are comparable to other figures shown on returns scale. We observe that an increase in θ_1 marks the bump of the pricing kernel more distinctive while the shape of ARA remains unchanged compared to the baseline model because, as we can see from (2), ARA does not depend on θ_1 when $\theta_4 = 0$. Yet, the effect of θ_1 on ARA can be analyzed by considering two distinct cases: $\theta_4 > 0$ and $\theta_4 < 0$. These specifications are important

because the direction of change in the slope of ARA is dictated by the sign of θ_4 . In the present case—after normalization— θ_1 varies around 0 and its effect on ARA is almost nil.

A larger value in the parameter θ_2 as compared to a benchmark value stretches the x -axis, which implies larger spread of the bump. When we vary θ_2 alone the slope of $ARA(\theta_2 u)$ is $1/\theta_2^2 \left[\left\{ g'^2(u) - g''(u)/g(u) \right\} / g^2(u) \right]$. The term in brackets does not depend on θ_2 ; it is equal to the slope of $ARA(u)$. Therefore, there is an inverse relationship between the direction of change in the parameter and that of the absolute value of the slope. These changes in slope occur around an inflection point that corresponds to the peak of the pricing kernel.

A positive increment in θ_3 shifts both curves to the left without any modification in the shape. θ_4 simply translates pricing kernel curves above or below the reference curve following a sign rule. Similarly to θ_2 , the shape of ARA modifies around the fixed inflection point that marks the change from risk proclivity (negative ARA) to risk aversion (positive ARA). The effect of θ_4 on the values of ARA is straightforward: since θ_4 adds to the g in the denominator its increase will diminish the absolute ARA level and the other way around. Insulating the effects of a change in θ_4 on the slope of $ARA(u)$ analytically proves to be a more complicated task than in the case of θ_2 because the change in the slope depends jointly on the change in θ_4 and on the pricing kernel values and its first two derivatives. In our case, the slope around the inflection point increases when θ_4 decreases.

As for the utility function, positive changes in θ_1 and θ_4 increases its absolute slope. In the horizontal direction, θ_3 translates the curve to the left or right similarly to the pricing kernel and ARA while θ_2 shrinks or expands its domain.

With this information at hand we can characterize the changes in risk patterns in relation to economic variables of interest, see Section 5.4.

3 FITTING SHAPE INVARIANT MODELS

3.1 Estimation of SIM

The model in (1) is equivalently written as

$$\mathcal{K}_t(\theta_{t2}u + \theta_{t3}) = \theta_{t1}g(u) + \theta_{t4}, \quad \theta_{t1} > 0, \quad \theta_{t2} > 0. \quad (3)$$

The estimation procedure is developed using the least squares criterion based on nonparametric estimates of individual curves. If there are only two curves, parameter estimates are obtained by minimizing

$$\int \{ \hat{\mathcal{K}}_2(\theta_2 u + \theta_3) - \theta_1 \hat{\mathcal{K}}_1(u) - \theta_4 \}^2 w(u) du, \quad (4)$$

where $\hat{\mathcal{K}}_i$ are nonparametric estimates of the curves. Härdle and Marron (1990) studied comparison of two curves and Kneip and Engel (1995) extended to multiple curves with an iterative algorithm. We consider an adaption of such algorithm here.

The weight function w is introduced to ensure that the functions are compared in a domain where the common features are defined. We assume that there is an interval $[a, b] \in J$ where boundary effects are eliminated and then define

$$w(u) = \prod_t 1_{[a, b]} \left\{ (u - \theta_{t3}) / \theta_{t2} \right\}.$$

The parameter estimates are compared only in the common region defined by w but the individual curve estimates are defined on the whole interval. Weights can be extended to account for additional variability.

The normalization leads to:

$$T^{-1} \sum_{t=1}^T \mathcal{K}_t(\theta_{t2}u + \theta_{t3}) = g(u). \quad (5)$$

Formula (5) was exploited in the algorithm proposed by Kneip and Engel (1995). We adopt a similar strategy here.

- Initialize

- Let $\hat{K}_t = Y_t$ and set starting values $(\theta_{t2}^{(0)}, \theta_{t3}^{(0)})$ for $t = 1, 2, \dots, T$.
- Construct an initial estimate $g^{(0)}$ by

$$g^{(0)}(u) = T^{-1} \sum_{t=1}^T \hat{\mathcal{K}}_t(\theta_{t2}^{(0)}u + \theta_{t3}^{(0)}).$$

- For r -th step, $r = 1, 2, \dots, R$,

- Determine parameters $\theta^{(r)}$ separately for $t = 1, 2, \dots, T$ by minimizing

$$\int \left\{ \hat{\mathcal{K}}_t(\theta_{t2}u + \theta_{t3}) - \theta_{t1}g^{(r-1)}(u) - \theta_{t4} \right\}^2 w(u) du.$$

- Normalize parameters: for $j = (1, 2)$ and $k = (3, 4)$

$$\theta_{tj}^{(r)} \leftarrow \frac{\theta_{tj}^{(r)}}{\sum_t \theta_{tj}^{(r)}}, \quad \theta_{tk}^{(r)} \leftarrow \theta_{tk}^{(r)} - T^{-1} \sum_t \theta_{tk}^{(r)}.$$

- Update $g^{(r-1)}$ to

$$g^{(r)}(u) = T^{-1} \sum_{t=1}^T \hat{\mathcal{K}}_t(\theta_{t2}^{(r)}u + \theta_{t3}^{(r)}).$$

- Determine final estimates:

$$\tilde{\theta}_t = \theta_t^{(R)},$$

$$\tilde{g}(u) = T^{-1} \sum_{t=1}^T \hat{\mathcal{K}}_t(\tilde{\theta}_{t2}u + \tilde{\theta}_{t3}).$$

Kneip and Engel (1995) proved consistency of the estimator. In particular despite nonparametric initial curve estimates, the parameters are shown to be \sqrt{T} consistent. In their analysis it is noted that the initial estimates of the curves are of minor importance compared to the final estimate of g . So the original algorithm includes the final updating of each curve. This improves precision of the estimates because the pooled sample estimate reduces the variance of \tilde{g} , which allows undersmoothing at the final stage to reduce bias. However, this final updating step is not practical for our situation with indirect measurements and is not implemented here for pricing kernel estimation. On the other hand, we can take advantage of having smooth curves evaluated at finite grid points as data. It is easier to improve the initialization step, explained in Section 3.2. This leads to simplification of the estimating procedure with little compromise of the quality of the fit. In fact, the number of iterations required is very small and often 3 or 4 is sufficient in practical terms. We found that when the initial estimates are determined sufficiently accurate, the iteration is not necessary.

As a working model we have assumed an independent error. If there is a reasonable dependence structure available, this could be incorporated easily in the estimation algorithm with weighted least squares estimation in (4). The effect of independence assumption mainly appears in the standard error estimation and a correction can be made with a sandwich variance–covariance estimator. To assess the effect of model misspecification, we also carried out some simulation studies with dependent errors and reported the results in Section 4.

3.2 Starting Values

If there is no scale change in horizontal direction, due to prominent peaks in each curve, the parameter θ_3 can be identified easily by the location of the individual peak. If the models hold true, and there are two unique landmarks identifiable for each curve, simple linear regression between the individual mark and the average mark provides an estimate of the slope parameter θ_2 . Suppose that the peak is identified by u satisfying $K'_t(u) = 0$. Then we have

$$0 = \mathcal{K}'_t(u) = \frac{\theta_{t1}}{\theta_{t2}} g' \left(\frac{u - \theta_{t3}}{\theta_{t2}} \right).$$

Writing u_t^* for \mathcal{K}'_t and u_0^* for g' leads to a simple linear relation:

$$u_t^* = \theta_{t2}u_0^* + \theta_{t3}. \tag{6}$$

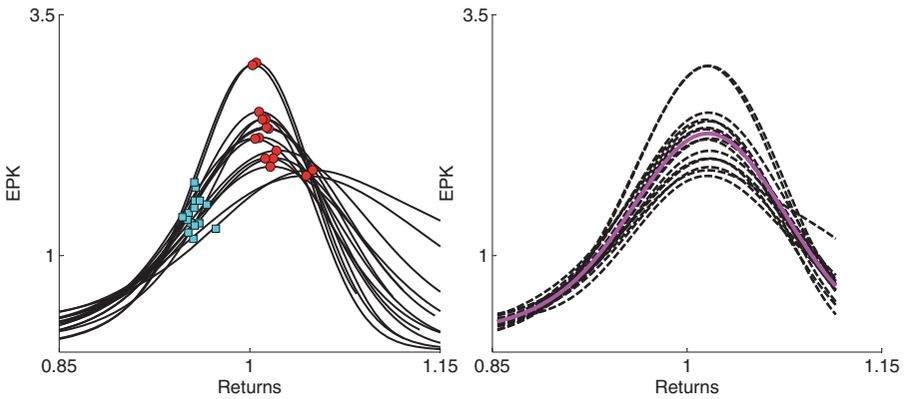


Figure 4 Initial estimates $\mathcal{K}_t(u)$ (left) and final estimates $\mathcal{K}_t(\theta_{12}u + \theta_{13})$ from SIM (right) with g overlaid. Marked in the left plot are two landmarks identified for estimation of the starting values of $(\theta_{12}, \theta_{13})$.

If an inflection point is used, we would have

$$0 = \mathcal{K}_t''(u) = \frac{\theta_{11}}{\theta_{12}^2} g''\left(\frac{u - \theta_{13}}{\theta_{12}}\right),$$

which gives rise to the same relation as (6), with the corresponding u_t^{**} and u_0^{**} substituted. The coefficients of intercept and slope estimates are used for starting values of θ_{13} and θ_{12} , respectively.

We used the peak and the inflection points around 1 as landmarks, marked in Figure 4. The location of the landmarks is defined by the zero crossings of the first and second derivatives. Because the initial observations \mathcal{K}_t are a smoothed curve, we find that additional smoothing procedure is not required at this stage: a finite difference operation is sufficient to apply mean value theorem with linear interpolation.

The slope between any two points did not vary much, which is consistent with the model specification. This step is also used as an informal check and should there be any nonlinearity detected, the model needs to be extended to include a nonlinear transformation. With our example, this was not the case.

3.3 Nonlinear Optimization

Given the estimates of $(\theta_{12}, \theta_{13})$, the nonlinear least squares optimization uses (4), which is approximated by

$$\sum_j \left\{ \hat{\mathcal{K}}_t(\theta_{12}u_j + \theta_{13}) - \theta_{11}\hat{g}(u_j) - \theta_{14} \right\}^2 w(u_j). \tag{7}$$

When the initial values of $(\theta_{t2}, \theta_{t3})$ are sufficiently accurate, this step is simplified to a linear regression. Conditional on θ_{t2}, θ_{t3} and \hat{g} , the solutions to the least square regression with response variable $\hat{K}_t(\theta_{t2}u_j + \theta_{t3})$ and explanatory variable $\hat{g}(u_j)$ provide $(\theta_{t1}, \theta_{t4})$. When a further optimization routine is employed to improve the estimates, these numbers serve as initial values for $(\theta_{t1}, \theta_{t4})$.

3.4 Initial Estimates of \mathcal{K}

To start the algorithm the initial estimates of \mathcal{K} should be supplied. An example of initial estimates of \mathcal{K} is shown in Figure 4 on the scale of continuously compounded returns. These are obtained from separate estimation of p and q , which are described below. Individual smoothing parameter choice is discussed in Section 5 with real data example.

3.4.1 Estimation of the historical density p . We use the nonparametric kernel density estimates similar to Ait-Sahalia and Lo (2000) based on the past observations of returns for a fixed maturity τ . With this approach the returns of the stock prices are assumed to vary slowly and thus the process can be assumed stationary for a short period of time. Alternatively, if additional modeling assumption is made for the evolution of the stock price such as GARCH, a simulation-based approach could be employed.

At given time t and $T = t + \tau$ we obtain realizations of future return values from a window of historical return values of length J :

$$r_T^k = \log\left(S_{t-(k-1)} / S_{t-\tau-k+1}\right) \quad \text{and} \quad S_T^k = S_t e^{r_T^k}, \quad k = 1, \dots, J.$$

The probability density of r_T is obtained by the kernel density estimator

$$\hat{p}_{h_p}(r) = \frac{1}{Jh_p} \sum_{k=1}^J K\left(\frac{r_T^k - r}{h_p}\right),$$

where K is a kernel weight function and h_p is the bandwidth. Some variations are also explored such as overlapping and nonoverlapping windows with a real data example in Section 5.

3.5 Estimation of the Risk Neutral Density q

We begin with the call price option formula that links the call prices to the risk neutral density estimation. The European call price option formula is given

by (Ait-Sahalia and Duarte, 2003)

$$C(X, \tau, r_{t,\tau}, \delta_{t,\tau}, S_t) = e^{-r_{t,\tau}\tau} \int_0^\infty \max(S_T - X, 0) q(S_T | \tau, r_{t,\tau}, \delta_{t,\tau}, S_t) dS_T$$

where

- S_t : the underlying asset price at time t ,
- X : the strike price,
- τ : the time to maturity,
- $T = t + \tau$: the expiration date,
- $r_{t,\tau}$: the deterministic risk free interest rate for that maturity,
- $\delta_{t,\tau}$: the corresponding dividend yield of the asset.

Write $q(S_T)$ for $q(S_T | \tau, r_{t,\tau}, \delta_{t,\tau}, S_t)$. For fixed t and τ , assume $r_{t,\tau} = r$ and $\delta_{t,\tau} = \delta$, the risk neutral density is expressed as

$$q(u) = e^{r\tau} \frac{\partial^2 C}{\partial X^2} \Big|_{X=u}.$$

The relation is due to Breeden and Litzenberger (1978) and serves the basis of many current semi-parametric and nonparametric approaches. We employ the semiparametric estimates of Rookley (1997), where the parametric Black–Scholes formula is assumed except that the volatility parameter σ is a function of the option's moneyness and the time to maturity τ . In this work, we fix the maturity and consider it as one dimensional regression problem.

Define $F = S_t e^{(r-\delta)\tau}$ and $m = X/F$ is moneyness. Write Φ and ϕ for the cumulative distribution function and its density of standard normal random variable, respectively. The Black–Scholes model assumes

$$\begin{aligned} C_{BS}(X, \tau) &= S_t e^{-\delta\tau} \Phi(d_1) - e^{-r\tau} X \Phi(d_2) \\ &= e^{-r\tau} F \{ \Phi(d_1) - m \Phi(d_2) \}. \end{aligned}$$

In a semiparametric call price function, the volatility parameter σ is expressed as a function of the option's moneyness and the time to maturity τ :

$$C(X, \tau, r, \delta, S_t) = C_{BS}(X, \tau, F, \sigma(m, \tau)).$$

To derive the second derivative of C , it is simpler to work with a standardized call price function $c(m, \tau) = e^{r\tau} C(X, \tau, r, \delta, \sigma) / F = \Phi(d_1) - m \Phi(d_2)$. The derivatives of C and c are related as

$$\begin{aligned} \frac{\partial C}{\partial X} &= e^{-r\tau} F \frac{\partial c}{\partial m} \frac{\partial m}{\partial X} = e^{-r\tau} \frac{\partial c}{\partial m}, \\ \frac{\partial^2 C}{\partial X^2} &= e^{-r\tau} \frac{\partial c^2}{\partial m^2} \frac{\partial m}{\partial X} = \frac{e^{-r\tau}}{F} \frac{\partial c^2}{\partial m^2}. \end{aligned}$$

With some manipulation we obtain the following expressions, which are only functions of $(\sigma, \sigma', \sigma'')$:

$$\begin{aligned} \frac{\partial c}{\partial m} &= \phi(d_1) \frac{\partial d_1}{\partial m} - \Phi(d_2) - m\phi(d_2) \frac{\partial d_2}{\partial m} \\ \frac{\partial^2 c}{\partial m^2} &= -d_1\phi(d_1) \left(\frac{\partial d_1}{\partial m}\right)^2 + \phi(d_1) \frac{\partial^2 d_1}{\partial m^2} - \phi(d_2) \frac{\partial d_2}{\partial m} - \phi(d_2) \frac{\partial d_2}{\partial m} \\ &\quad + md_2\phi(d_2) \left(\frac{\partial d_2}{\partial m}\right)^2 - m\phi(d_2) \frac{\partial^2 d_2}{\partial m^2}, \end{aligned}$$

where

$$\begin{aligned} \frac{\partial d_1}{\partial m} &= -\frac{1}{\sqrt{\tau}} \frac{1}{m\sigma(m, \tau)} + \frac{1}{\sqrt{\tau}} \ln(m) \frac{\sigma'(m, \tau)}{\sigma^2(m, \tau)} + \frac{\sqrt{\tau}}{2} \sigma'(m, \tau) \\ \frac{\partial d_2}{\partial m} &= \frac{\partial d_1}{\partial m} - \sqrt{\tau} \sigma'(m, \tau) \\ \frac{\partial^2 d_1}{\partial m^2} &= \frac{1}{m^2 \sqrt{\tau} \sigma(m, \tau)} + \frac{2}{\sqrt{\tau}} \frac{\sigma'(m, \tau)}{\sigma^2(m, \tau)} \left\{ \frac{1}{m} - \ln(m) \frac{\sigma'(m, \tau)}{\sigma(m, \tau)} \right\} \\ &\quad + \sigma''(m, \tau) \left\{ \frac{\ln(m)}{\sigma^2(m, \tau) \sqrt{\tau}} + \frac{\sqrt{\tau}}{2} \right\} \\ \frac{\partial^2 d_2}{\partial m^2} &= \frac{\partial^2 d_1}{\partial m^2} - \sqrt{\tau} \sigma''(m, \tau). \end{aligned}$$

Note that this leads to a slightly different derivation from Rookley (1997), albeit using the same principle.

In order to compute the derivatives of σ , we used the local polynomial smoothing on implied volatility. Let σ_i be the implied volatility corresponding to the call price C_i with moneyness m_i . The local polynomial smoothing estimates are obtained by minimizing

$$\sum_i \left\{ \sigma_i - \sum_{j=0}^3 \beta_j(m) (m_i - m)^j \right\}^2 W((m_i - m)/h_q),$$

where $W(\cdot)$ is a weight function. The estimates are computed as $\hat{\sigma}(m) = \hat{\beta}_0(m)$, $\hat{\sigma}'(m) = \hat{\beta}_1(m)$ and $\hat{\sigma}''(m) = 2\hat{\beta}_2(m)$. Substituting the estimates to the above expressions gives an estimate of q . The density estimates are defined on the scale of S_T . To define the density on the same returns scale $r_T = \log(S_T/S_t)$ as p , a simple transformation can be applied:

$$q(r_T) = q(S_T) S_T.$$

Notice that all results are shown on a continuously compounded 1-month period returns $R_T = 1 + r_T = 1 + \log(S_T/S_t)$.

3.6 Word on Asymptotics

There are two layers of estimation involved. The first step deals with individual curve estimation and the second step introduces shape invariant modeling. The shape invariant modeling is largely robust to how the data are prepared before entering the iterative algorithm and the resulting estimates are interpreted as conditional on the individual curves. Therefore, the main estimation error arises in the first stage where p and q are separately estimated with possibly different sample sizes and separately chosen bandwidths.

In practical terms, the sample size used in estimating p is normally of smaller order, say n compared to $N = nM$ for q for a constant M . This is due to the difference between the daily observations available for estimating p and the intraday observations available for estimating q . Thus it might be expected that the estimation error will be dominated by the estimation error of p . On the other hand, the underlying function p for which simple kernel estimation is used is much simpler and more stable compared to q for which nonparametric second derivative estimation is required.

Because the estimates of ratios are constructed from the ratio of the estimates, we can decompose the error as

$$\begin{aligned}\hat{\mathcal{K}}(u) - \mathcal{K}(u) &= \frac{\hat{q}(u)}{\hat{p}(u)} - \frac{q(u)}{p(u)} \\ &\simeq \frac{\hat{q}(u) - q(u)}{p(u)} - \frac{q(u)}{p(u)} \frac{\hat{p}(u) - p(u)}{p(u)}.\end{aligned}$$

Numerical instability might occur in the region where $\hat{p} \approx 0$ however this is not of theoretical concern. In fact, the pricing kernel is the Radon-Nikodym derivative of an absolutely continuous measure, and thus p and q are equivalent measures, that is, the null set of p is the same as the null set of q . So we can limit our attention to the case where $p(u) > \epsilon$ for some constant ϵ . Provided that $p(u) > \epsilon$ and $q(u) > \epsilon$, the asymptotic approximation is straightforward and asymptotic bias and variance can be computed from

$$\begin{aligned}\mathbb{E}[\hat{\mathcal{K}}(u) - \mathcal{K}(u)] &\simeq \frac{\mathbb{E}[\hat{q}(u) - q(u)]}{p(u)} - \frac{q(u)}{p(u)} \frac{\mathbb{E}[\hat{p}(u) - p(u)]}{p(u)} \\ &= \mathcal{O}(h_q^4) + \mathcal{O}(h_p^2) + \mathcal{O}(h_p^2 + h_q^4), \\ \text{Var}[\hat{\mathcal{K}}(u) - \mathcal{K}(u)] &\simeq \mathcal{K}^2(u) \left\{ \frac{\text{Var}[\hat{q}(u)]}{q^2(u)} + \frac{\text{Var}[\hat{p}(u)]}{p^2(u)} \right\} \\ &= \mathcal{O}\{(Nh_q)^{-1}\} + \mathcal{O}\{(nh_p)^{-1}\} + \mathcal{O}\{(Nh_q)^{-1} + (nh_p)^{-1}\}.\end{aligned}$$

Since \hat{q} involves estimation of second derivative of a regression function, the error is dominated by the estimation of q . The optimal rate of convergence for q is $\mathcal{O}(N^{-2/9})$

while that for p is $\mathcal{O}(n^{-2/5})$. These will be equivalent when $M = \mathcal{O}(n^{39/15}) > \mathcal{O}(n^2)$. In practice M is of much smaller order and therefore the leading error terms come from the estimation of q . Ait-Sahalia and Lo (2000) showed in a similar framework that the error is dominated by the estimation of q and for the purpose of asymptotics p can be regarded as a fixed quantity. For this reason we actually implement a semiparametric estimator for q to stabilize the estimator.

Consistency and asymptotic normality of the parameter estimates are shown in Härdle and Marron (1990) for two curves and in Kneip and Engel (1995) for multiple curves. We write the approximate distribution for $\hat{\theta}_t$ as

$$\hat{\theta}_t \approx N(\theta_t, \Sigma_t).$$

Due to the iterative algorithm, the asymptotic covariance matrix is more complicated for multiple curves but Kneip and Engel (1995) show that, as the number of curves increases, the additional terms arising in the covariance matrix is of lower order than the standard error term due to nonlinear least square methods. There is no suggested estimate for the asymptotic covariance matrix but a consistent estimate can be constructed as in standard nonlinear least square methods. Define the residual $\hat{\epsilon}_{tj} = \hat{\mathcal{K}}_t(u_j) - \tilde{\mathcal{K}}_t(u_j)$ where $\hat{\mathcal{K}}$ is the initial estimates and $\tilde{\mathcal{K}}$ is the SIM estimates and let

$$\hat{\sigma}_t^2 = \frac{1}{n} \sum_{j=1}^n \hat{\epsilon}_{tj}^2.$$

The covariance matrix can be estimated as

$$\hat{\Sigma}_t = \hat{\sigma}_t^2 \left[n^{-1} \sum_{j=1}^n \left\{ \nabla_{\theta} \tilde{\mathcal{K}}_t(u_j; \tilde{\theta}) \right\} \left\{ \nabla_{\theta} \tilde{\mathcal{K}}_t(u_j; \tilde{\theta}) \right\}^{\top} \right]^{-1},$$

where $\nabla_{\theta} \mathcal{K}(u; \theta)$ is the first derivative of the function, given by

$$\begin{aligned} \frac{\partial \mathcal{K}(u)}{\partial \theta_1} &= g\left(\frac{u - \theta_3}{\theta_2}\right), \\ \frac{\partial \mathcal{K}(u)}{\partial \theta_2} &= -\frac{\theta_1}{\theta_2} \left(\frac{u - \theta_3}{\theta_2}\right) g'\left(\frac{u - \theta_3}{\theta_2}\right), \\ \frac{\partial \mathcal{K}(u)}{\partial \theta_3} &= -\frac{\theta_1}{\theta_2} g'\left(\frac{u - \theta_3}{\theta_2}\right), \\ \frac{\partial \mathcal{K}(u)}{\partial \theta_4} &= 1. \end{aligned}$$

To see whether the location or scale parameters are different between any pair of curves, we can compute the standard errors of the estimates to make a comparison. A formal hypothesis testing also appears in Härdle and Marron (1990) for kernel-based estimates and in Ke and Wang (2001) for spline-based estimates. For example

Table 1 Parameter values of θ

	Distribution	Mean	Standard deviation
θ_1	Log-normal	1	0.33
θ_2	Log-normal	1	0.28
θ_3	Normal	0	0.27
θ_4	Normal	0	0.35

we might be interested in testing whether a location or a scale parameter can be removed.

Although these results are practically relevant, we note that the methods mentioned all assume direct observations of the underlying function of interest, with one smoothing parameter selection involved. Obtaining comparable rigorous results for our estimator is complicated in the present situation due to the nonstandard nature of the estimator being a ratio of two separate nonparametric estimates with independent bandwidths. We consider this out of scope of this paper and leave it for separate work.

4 NUMERICAL STUDIES OF SIM ESTIMATION

Applying the SIM to pricing kernels involves two rather separate estimation steps, the initial estimation of the pricing kernels and the SIM estimation given the pre-estimates. The former has been studied extensively and in particular the properties of the nonparametric methods that we have used are well established in the literature. This section mainly concerns the latter.

We identify the two main factors that could affect the performance of SIM estimation to be error misspecification and smoothing parameter selection for the individual curves. Their effects are evaluated in the following simulation studies. The effects on pricing kernel estimation are separately studied in Section 5.4, in comparison to the standard nonparametric approach used in Jackwerth (2000).

4.1 Generating Curves

In each simulation 50 curves are generated at 50 (random uniform) grid points. In order to mimic the common shape of the observed pricing kernel, we generated the common curve by a ratio of two densities

$$g(u) = q_0(u)/p_0(u),$$

where p_0 is density of Gamma(0.8,1) distribution and q_0 is density of mixture $w * \text{Gamma}(0.2, 1) + (1-w) * N(0.91, 0.3^2)$ distribution with $w=0.3$. In accordance with the normalization scheme, the θ values are set as in Table 1. The values of the

Table 2 Parameter values for error specification

		Error 1	Error 2	Error3
Case 1	σ	0.02	0.05	0.10
Case 2	ϕ	0.75	0.75	0.75
	σ_u	0.02	0.03	0.09
Case 3	α	-3.69	-2.99	-2.30
	β	0.75	0.52	0.53
	σ_v	0.01	0.02	0.02
Case 4	α	-2.41	-1.89	-1.39
	β	0.45	0.40	0.42
	ϕ	0.75	0.45	0.45
	σ_v	0.10	0.25	0.25

standard deviation were chosen to be similar to the observed ones in the real data example.

4.2 Error Specification

For the error specification, we have included dependent errors in time as well as in moneyness as following.

- Case 1: Independent error: $\varepsilon_{t,j} \sim N(0, \sigma^2)$
- Case 2: Dependent error in moneyness:

$$\varepsilon_{t,j} = \phi \varepsilon_{t,j-1} + u_{t,j}, \quad \text{the set of the } u_{t,j} \sim N(0, \sigma_u^2)$$

- Case 3: Dependent error in time: $\varepsilon_{t,j} \sim N(0, \sigma_t^2)$

$$\log(\sigma_t) = \alpha + \beta \log(\sigma_{t-1}) + v_t, \quad v_t \sim N(0, \sigma_v^2)$$

- Case 4: Dependent error in moneyness and time:

$$\varepsilon_{t,j} = \phi \varepsilon_{t,j-1} + u_{t,j}, \quad u_{t,j} \sim N(0, \sigma_{ut}^2),$$

$$\log(\sigma_{ut}) = \alpha + \beta \log(\sigma_{u,t-1}) + v_t, \quad v_t \sim N(0, \sigma_v^2)$$

Cases 1 and 2 are commonly assumed but Cases 3 and 4 were rarely used in the literature with SIM estimation. Table 2 lists the parameter values used for simulation. These values are chosen to be comparable in terms of overall variability among cases.

4.3 Smoothing Parameter Selection

We consider three versions of the least squares cross-validation (CV) based criteria for bandwidth selection:

$$CV_t(h) = \sum_{i=1}^n \left\{ Y_{t,i} - \widehat{K}_{t,h}^{-(i)}(u_i) \right\}^2,$$

where $\widehat{K}_{t,h}^{-(i)}$ is the local linear fit without using the i -th observation. For each observed curve we find the optimal bandwidth $h_t^* = \operatorname{argmin} CV_t(h)$. Due to considerable variability in the x -dimension we standardize the optimal bandwidths ($\tilde{h}_t^* = h_t^*/s_t$), where s_t is the empirical standard deviation of u_i , and we choose the common bandwidth as follows:

$$h_{opt,1} = \max(\tilde{h}_t^*) \quad h_{opt,2} = \operatorname{average}(\tilde{h}_t^*) \quad \text{or} \quad h_{opt,3} = \operatorname{argmin}_t \sum_t CV_t(h).$$

Finally, we multiply h_{opt} by s_t and use these values to perform smoothing of each curve.

4.4 Results of Simulation

We considered various simulation scenarios based on the combinations of the case of errors and bandwidth selection methods. Table 3 summarizes the results of the goodness of fit measured by MSE for the case $\sigma = 0.05$. For comparison we added in the last row the MSE for the standard nonparametric estimates based on individual optimal bandwidths to their advantage. For larger error ($\sigma = 0.1$, not shown) we also observed some dramatic deterioration with Case 4. Nevertheless, the simulation studies suggest that the overall error is in the same order of magnitude and we suspect that the impact of these factors is limited. The fit was however best with smoothing parameters selected by h_1 .

5 REAL DATA EXAMPLE

We use intraday European options data on the Deutscher Aktien index (DAX), provided by European Exchange EUREX and maintained by the CASE, RDC SFB 649 (<http://sfb649.wiwi.hu-berlin.de>) in Berlin. We have identified options data with maturity one month (31 working days/23 trading days) from June 2003 to June 2006 from DAX 30 Index European options, which adds up to 37 days.

We obtain the initial estimates for p and q as described in Section 3.4. For the choice of kernel functions, we have used quartic function for both p and q .

Table 3 Comparison of SIM estimation with respect to error misspecification and smoothing parameter selection

		$\sigma = 0.05$			
methods	parms.	case 1	case 2	case 3	case 4
h_1	θ_1	31	32	67	65
	θ_2	60	70	84	77
	θ_3	54	62	81	76
	θ_4	32	32	77	75
	\mathcal{K}_i s	1.2	1.6	1.5	1.5
h_2	θ_1	67	68	80	69
	θ_2	115	115	110	99
	θ_3	111	110	105	103
	θ_4	70	72	99	85
	\mathcal{K}_i s	1.1	1.6	1.9	1.9
h_3	θ_1	67	71	67	73
	θ_2	115	108	91	82
	θ_3	111	100	88	84
	θ_4	70	74	83	88
	\mathcal{K}_i s	1.1	1.6	1.8	1.8
	np \mathcal{K}	3.5	2.0	4.2	3.6

Numbers are MSE multiplied by 10000. \mathcal{K}_i s computes the average MSE for all curves from SIM and np \mathcal{K} without SIM but using individual optimal bandwidths for each curve.

5.1 Estimation of the Risk Neutral Density q

The stock index price varies within one day and we would need to identify the price at which a certain transaction has taken place. However, several authors (e.g. Jackwerth, 2000) report that the intraday change of the index price is stale and we use instead the prices of futures contracts closest to the time of the registered pair of the option and strike prices to derive the corresponding stock price, corrected for dividends and difference in taxation following a methodology described in Fengler (2005).

The data contains the actually traded call prices, the implied stock index price corrected for the dividends from the futures derivatives on the DAX, the strike prices, the interest rates (linearly interpolated based on EURIBOR to approximate a *riskless* interest rate for the specific option's time to maturity), the maturity, the type of the options, calculated moneyness, calculated Black and Scholes implied volatility, the volume, and the date. For each day, we use only at-the-money and out-of-the-money call options and in-the-money puts to compute the Black–Scholes implied volatilities. This guarantees that unreliable observations (high volatility) will be removed from our estimation samples. Since the intraday stock price varies, we use its median (S_t) to compute the risk neutral density and correct the strike

price to preserve the ratio relative to the underlying stock price. For this price, we verify if our observations satisfy the no arbitrage condition:

$$S_t \geq C_i \geq \max(S_t - X_i e^{-r\tau}, 0),$$

where X_i is the adjusted strike price and C_i is the corresponding call price. For the remaining observations (X_i, C_i) we compute the (m_i, σ_i) counterparts for the fixed S_t by implicitly assuming that the volatility does not depend on the changes in the intraday stock price. The estimates are computed based on these pairs (m_i, σ_i) .

5.2 Estimation of the Historical Density p

We compute the nonparametric kernel density estimates as described earlier. Jackwerth (2000) argues that some discrepancies between the nonparametric estimates are attributed to overlapping and nonoverlapping windows of the past observations selected. For comparison to the earlier works, we also experimented with a choice of time varying equity premium and constant equity premium (we demean the densities and supplant it with the risk free rate on the estimation day plus 8% equity premium per annum as in Jackwerth (2000) adjusted for the corresponding maturity), overlapping and nonoverlapping returns, window lengths of 2, 4, and 6 years, respectively. The estimates for different choices of parameters are then compared subsequently in terms of pricing kernel, implied risk aversion and implied utility function estimation. We find that with varying degrees of assumptions on the model, common characteristics such as peaks and skewness are reportedly observed in a wide range of estimates.

5.3 Smoothing Parameter Selection

In contrast to the simulation studies, the effect of smoothing parameter is less transparent with real data when we estimate p and q separately. At first glance, the bandwidth selection for q seems more influential than that of p in gauging performance of the estimates, as it involves derivative estimation. Figure 5 examines the effect of the bandwidth choices on \hat{q} . Top left panel shows the implied volatility estimates overlayed, the top right shows the first derivative estimates and bottom left shows the second derivative estimates, respectively, which are used as inputs to create the estimates of q on bottom right panel. The bandwidths used are (0.05, 0.10, 0.15, 0.20). With the apparent undersmoothing at the smallest bandwidth, there is notable variability in terms of smoothness in estimation of implied volatility and its derivatives, however the resulting density estimates demonstrate robustness. Similar observations are made to other dates. However by smoothing on implied volatility domain, we find that the estimates are stable with relatively a wide range of bandwidth choices.

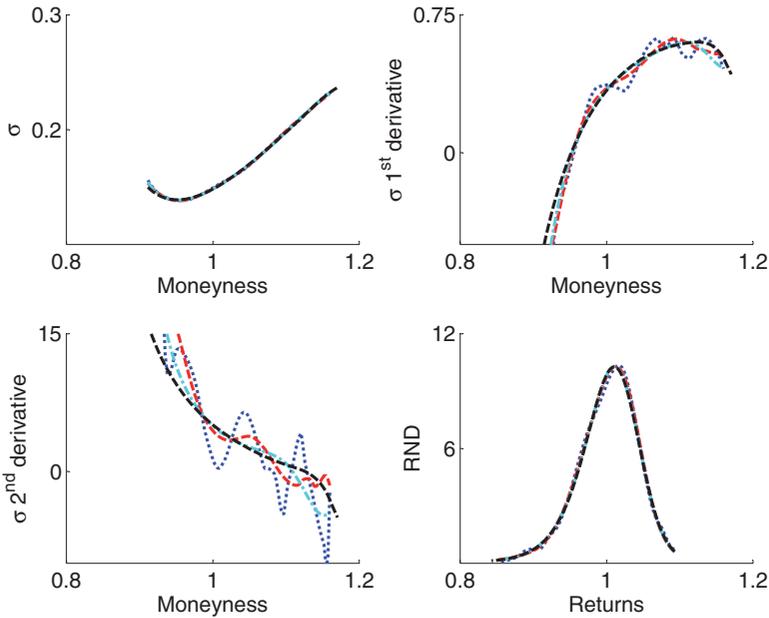


Figure 5 Example of q estimates with varying bandwidths (0.05, 0.1, 0.15, 0.20). The first three panels show estimates of implied volatility, its first and second derivative. The corresponding densities are shown in lower right panel. Estimates are stable for a wide range of bandwidths choices.

For a systematic choice, we employed a version of CV criteria ($h_{opt,1}$ defined in Section 4.3) for p and q estimation. For estimation of q , we have used the least squares CV for local cubic estimation to include the second derivative of σ :

$$CV(h_q) = \sum_{i=1}^n \sum_{j \neq i}^n \left\{ \sigma_i - \hat{\sigma}_{h_q, -i}^{(0)}(m_i) - \hat{\sigma}_{h_q, -i}^{(1)}(m_i)(m_j - m_i) - \frac{1}{2} \hat{\sigma}_{h_q, -i}^{(2)}(m_i)(m_j - m_i)^2 \right\}^2 w(m_i),$$

where $\hat{\sigma}_{h_q, -i}^{(k)}$ is the k -th derivative estimate without the i -th observation (m_i, σ_i) and $0 \leq w(m_i) \leq 1$ is a weight function. The h_1 -optimal bandwidth in implied volatility space turns out to be $h_q = 0.2$.

For estimation of p , we have used the likelihood CV for each curve on returns scale:

$$\log L(h_p) = \sum_{i=1}^n \log \hat{p}_{h_p}^{-i}(r_i),$$

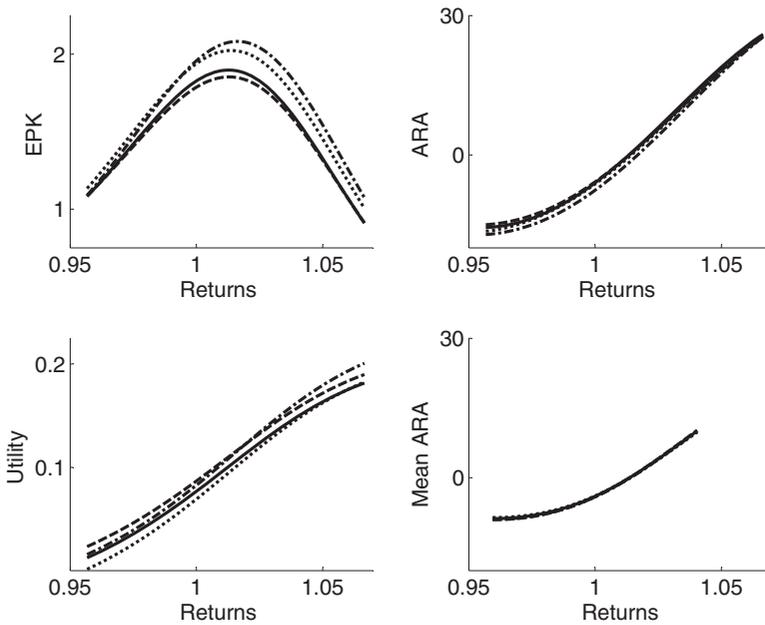


Figure 6 Illustration of SIM with common EPK, ARA, utility function, and mean ARA.

where $\hat{p}_{h_p}^{-i}(r_i)$ is the leave-one-out kernel estimator for $p_{h_p}(r_i)$. However, we found that the optimal bandwidth selected tends to systematically oversmooth and thus we chose a smaller value close to the maximum of individually optimal bandwidths, which is in our case $h_p = 0.05$.

5.4 Estimation of Pricing Kernels, ARA and Utility Function

We have considered in Section 5.2 various options for the parameter choice in estimating p and have ended up with 12 series of pre-estimates of pricing kernel. We are interested in seeing how these choices influence the estimated common curves and θ_i parameters by SIM. Since, as it turns out, the results are very similar among specifications we depict graphically only four of them in Figures 6 and 7: those based on nonoverlapping (solid) and overlapping (dashed) returns over the last two years, nonoverlapping returns over the last four (dot-dashed) and six (dotted) years, respectively with varying equity premium. The added lines in Figure 7 are 95% pointwise confidence band for the first series of pre-estimates.

The common curves are represented in Figure 6. All estimates display a *paradoxical* feature: pricing kernel has a bump, ARA has a region of negative values that correspond to the increasing region in the pricing kernel, utility function has a convex region in the domain around the peak of the pricing kernel. The variability

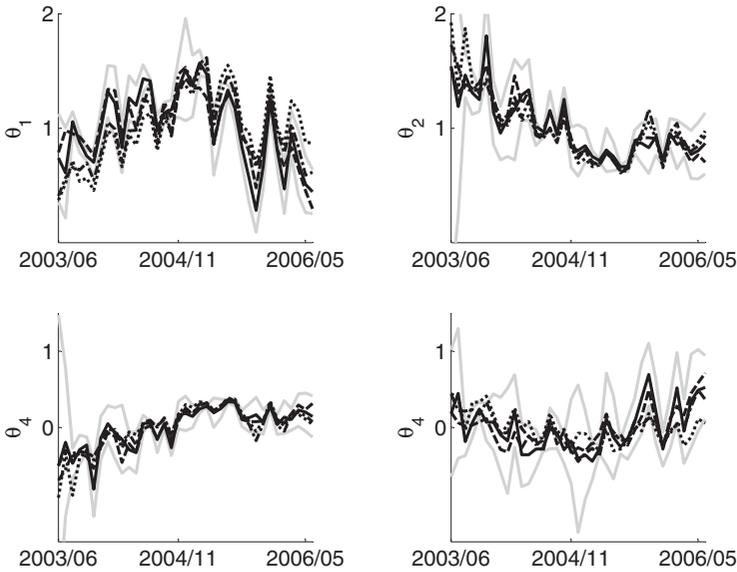


Figure 7 Estimated SIM parameters under variations in the choice of the window lengths of returns values.

among curves is expressed by θ_t -s. In Figure 7, we observe that the main difference in the dynamics of different series has to do with the magnitude but less with the direction of change. In addition, we computed the mean of implied ARA corresponding to our estimation period by computing the sample average and found that it was similar to the the mean ARA for S&P500 appearing in Figure 3C—19 March 1991 to 19 August 1993 in Jackwerth (2000), and to a certain extent to the yearly average from 2003 and 2005 shown in Figure 4 in Chabi-Yo *et al.* (2008). It is worth noting that the mean ARA and the common ARA curves differ a great deal due to the nonlinear transformation involved in deriving ARA from the pricing kernel, e.g. see Equation (2) in Section 2.4. This is not surprising since the interpretation of common curve is different from the average curve, in particular the common curve and the mean curves have different scales of the x -domains—by means of registration.

5.5 Relation to Macroeconomic Variables

With an aid of the SIM model for EPK, we wish to characterize changes in risk patterns in relation to economic variables of interest. Before doing this, we should mention that in the case of nonstandard common curves—in our empirical example the peak does not occur at 0—both θ_1 and θ_2 introduce a shift effect in EPK together with its shape effect. In order to disentangle these effects and improve interpretation we first standardize the EPK curves by the location of the peak before applying SIM.

Table 4 Correlation table for the first difference of SIM parameters and the selected macro economic variables

	θ_1	θ_2	θ_3	θ_4	CS	DAX	YT
θ_1	1.00	0.55*	0.02	0.78*	-0.25	0.38**	-0.26
θ_2		1.00	0.38*	-0.04	0.06	-0.12	-0.39**
θ_3			1.00	-0.18	0.07	-0.21	-0.28***
θ_4				1.00	-0.37**	0.62*	-0.04

*, **, and *** significant at 1%, 5%, and 10% levels, respectively.

This introduces two more parameters, the horizontal and vertical coordinates of the peaks in the analysis. Since their shift effect is comprised by parameters θ_3 and θ_4 we will not treat them here separately.

Previous studies trying to link the parameters describing risk attitudes to the business conditions include Rosenberg and Engle (2002). Based on power pricing kernel specifications they show that risk aversion is counter-cyclical. Other related work investigates the relation between equity premiums, (e.g., Fama and French, 1989), smile asymmetry of volatility (Bekaert and Wu, 2000; Drechsler and Yaron, 2010), or market efficiency (Marshall, Cahan, and Cahan, 2008). The advantage of our approach over Rosenberg and Engle (2002) is that it allows us to identify how the change in economic variables relates to the shape of a nonparametrically estimated pricing kernel. Due to limited sample size—37 observations—it is impossible to estimate a structural model that correctly deals with the simultaneity of our set of dependent variables. Further research will involve the estimation of a (S)VAR specification, in order to account for the aforementioned endogeneity. We instead evaluate the potential univariate correlations between the estimated θ_i parameters and macroeconomic variables associated with the business cycle and interpret our results from the perspective of local EPK and risk aversion functions. We use the following variables that have a revealed relation with the state of the economy: credit spread (CS) is the difference between the yield on the corporate bond, based on the German CORPTOP Bond maturing in 3–5 years, and the government bond maturing in 5 years; the yield curve slope (YT) refers to the difference between the 30-year government bond yield and three-months interbank rate; short-term interest rate (IR) is the three-months interbank rate; and DAX 30 Performance index as a proxy for consumption. Depending on data availability we collect daily or monthly data. Tests on unit roots failed to reject stationarity in all parameter series and economic variables; we therefore work with their first difference. For conciseness we present only the correlation table for nonoverlapping returns over the past two years with varying equity premium and interpret the results below in relation to Figure 3.

In Table 4, we read significant positive correlation between changes in θ_1 and DAX and negative one with the credit spread, indicating that the EPK becomes more pronounced when the economic indicators suggest an expanding economy; changes in θ_2 and YT are negatively correlated, suggesting that risk aversion slope

becomes locally steeper during economic boom. The same interpretation holds for the negative correlation between changes in θ_3 and YT. The height of the peak varies with the returns on the index, pointing to an increasing local risk proclivity in periods of economic expansion. We have not found any significant correlation between changes in θ_t and in the short-term interest rate. Finally, we observe a positive correlation between the increments in θ_1 and θ_2 that suggests that over periods of concerted negative evolution of the economic indicators the EPK bump will shrink in both horizontal and vertical direction, possibly leading to an overall decreasing EPK.

In summary, the sense of the relations between the indicators of the business cycle and the parameters that summarize risk preferences indicates that locally risk loving behavior is procyclical. These findings are also in line with the results found in Rosenberg and Engle (2002).

REFERENCES

- Ait-Sahalia, Y., and J. Duarte. 2003. Nonparametric Option Pricing under Shape Restrictions. *Journal of Econometrics* 16: 9–47.
- Ait-Sahalia, Y., and A. Lo. 2000. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94: 9–51.
- Bekaert, G., and G. Wu. 2000. Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies* 13: 1–42.
- Bondarenko, O. 2003. Estimation of Risk-Neutral Densities using Positive Convolution Approximation. *Journal of Econometrics* 116: 85–112.
- Breedon, D., and R. Litzenberger. 1978. Prices of State-Contingent Claims Implicit in Option Prices. *The Journal of Business* 51: 621–651.
- Chabi-Yo, Y., R. Garcia, and E. Renault. 2008. State Dependence can Explain the Risk Aversion Puzzle. *Review of Financial Studies* 21: 973–1011.
- Drechsler, I., and A. Yaron. 2010. What's Vol Got to Do with it. *Review of Financial Studies* 20: 1–45.
- Fama, E. F., and K. R. French. 1989. Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics* 25: 23–49.
- Fengler, M. R. 2005. *Semiparametric Modeling of Implied Volatility*. Springer-Verlag Berlin and Heidelberg.
- Giacomini, E., and W. Härdle. 2008. "Dynamic semiparametric factor models in pricing kernel estimation." In S. Dabo-Niang and F. Ferraty (eds.), *Functional and Operational Statistics, Contributions to Statistics*, pp. 181–187. Physica-Verlag HD.
- Giacomini, E., W. Härdle, and M. Handel. 2008. "Time dependent relative risk aversion." In G. Bol, S. Rachev, and R. Wrth (eds.), *Risk Assessment: Decisions in Banking and Finance, Contributions to Economics*, pp. 15–46. Physica-Verlag HD.

- Härdle, W., and J. S. Marron. 1990. Semiparametric Comparison of Regression Curves. *The Annals of Statistics* 18: 63–89.
- Hens, T., and C. Reichlin. 2012. “Three Solutions to the Pricing Kernel Puzzle.” Technical report, Swiss Finance Institute Research Paper No. 10-14. Available at SSRN: <http://ssrn.com/abstract=1582888> or <http://dx.doi.org/10.2139/ssrn.1582888>
- Jackwerth, J. C. 1999. Option-Implied Risk-Neutral Distributions and Implied Binomial Trees: a Literature Review. *Journal of Derivatives* 2: 66–82.
- Jackwerth, J. C. 2000. Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies* 13: 433–451.
- Ke, C., and Y. Wang. 2001. Semiparametric Nonlinear Mixed-Effects Models and their Applications. *Journal of the American Statistical Association* 96: 1272–1298.
- Kneip, A., and J. Engel. 1995. Model Estimation in Nonlinear Regression under Shape Invariance. *The Annals of Statistics* 23: 551–570.
- Kneip, A., and T. Gasser. 1988. Convergence and Consistency Results for Self-Modeling Nonlinear Regression. *The Annals of Statistics* 16: 82–112.
- Lawton, W. H., E. A. Sylvestre, and M. S. Maggio. 1972. Self Modeling Nonlinear Regression. *Technometrics* 14: 513–532.
- Leland, H. E. 1980. Who Should Buy Portfolio Insurance? *Journal of Finance* 35: 581–594.
- Marshall, B. R., R. H. Cahan, and J. Cahan. 2008. “Technical Analysis Around the World: Does it Ever Add Value?” Technical report, SSRN eLibrary.
- Ramsay, J. O., and B. W. Silverman. 2002. *Applied Functional Data Analysis*. Springer Series in Statistics. New York: Springer. Methods and case studies.
- Rookley, C. 1997. “Fully Exploiting the Information Content of Intra day Option Quotes: Applications in Option Pricing and Risk Management.” Technical report, University of Arizona.
- Rosenberg, J. V., and R. F. Engle. 2002. Empirical Pricing Kernels. *Journal of Financial Economics* 64: 341–372.
- Stone, C. J. 1982. Optimal Global Rates of Convergence for Nonparametric Regression. *The Annals of Statistics* 10: 1040–1053.

Variance swap dynamics

K. DETLEFSEN* and W. K. HÄRDLE

Center for Applied Statistics and Economics, Humboldt Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany

(Received 23 September 2007; in final form 8 November 2012)

We compare several parametric and non-parametric approaches for modelling variance swap curves by conducting an in-sample and an out-of-sample analysis using market prices. The forecasted Heston model gives the best overall performance. Moreover, the static Heston model highlights some problems of stochastic volatility models in option pricing of forward starting products.

Keywords: Empirical finance; Equities; Volatility modelling; Term structure

JEL Classification: C5, D4, G1

1. Introduction

Variance swap markets have become liquid and this has led to the trading of options on realized variance. In this way, variance swap markets have become important for investments and hedging. Because of this we analyse the dynamics of variance swap curves in order to find a model that gives a good overall performance—in- and out-of-sample. On the one hand, this may help build models for direct investment. On the other hand, we look at a popular stochastic volatility model and analyse its forecasting performance, which is essential for forward starting products.

For implied volatility surfaces, the dynamics have often been analysed and modeled by factor approaches (see, e.g. Cont and da Fonseca (2002) and Fengler (2005)). As variance swap curves are basically the term structure of implied volatility surfaces, these studies also analyse, in principal, the dynamics of variance swaps. General studies on variance swaps that do not focus on the dynamics are the survey of Demeterfi *et al.* (1999) and the hedging and trading analysis of Carr and Madan (1998). More recently, Bühler (2006) considered a modeling approach for the joint plain vanilla and variance swap market. For yield curves, forecasting questions have been analysed by, for example, Duffie and Kan (1996) and Diebold and Li (2003). Amengual (2009) found a significant improvement of the fit of variance swap prices when using higher-dimensional models with two stochastic factors for the variance. We focus on lower-dimensional models as they are easier to use and give more stable fits.

In section 2 we describe the modeling framework for variance swaps and the analysed models, which comprise a

Heston model, a Nelson–Siegel parametrization and a semiparametric approach. In section 3 we conduct an empirical analysis, describing the data, estimating the models and forecasting the variance swap curves. In section 4 we summarize our results.

2. Modeling the term structure

In this section we introduce variance swaps, explain the construction of variance swap curves and describe the approach that we use for fitting and forecasting the variance swap curves. We start with the stochastic volatility model of Heston (1993) and derive the corresponding model for the variance curves. In addition to this two-parameter model we consider a generalization with three parameters. Moreover, we describe a semiparametric factor model. Finally, we see how good some stylized facts are replicable in these models.

2.1. Constructing variance swap curves

Variance swaps are forward contracts on future realized volatility. They exchange at expiration the realized annualized variance of the log returns of an underlying against a predefined strike. These contracts vary in several respects: they may or may not assume zero mean of the log returns, they differ with respect to the annualization factor and it must be specified when the underlying is sampled. We assume a zero mean of the returns, use $c = 252$ trading days for annualization with daily sampling and focus on zero strikes.

Given an underlying S , the price of such a variance swap for the period $[0, T]$ with business days $0 = t_0 < \dots < t_n = T$ is given by

*Corresponding author. Email: kai.detlefsen@gmx.de

$$\sigma_R^2(T) \stackrel{\text{def}}{=} \frac{c}{n} \sum_{i=1}^n \left(\log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2.$$

At time $t \in (0, T)$, the first part of the variance is already realized while the second is still uncertain. Hence, the prices are composed of the value of the realized variance and the price of the uncertain variance. In our analysis, we will focus on the uncertain part and denote the price for the non-annualized variance that still has to be realized by $V_t(T)$.

At a point in time we observe the prices of variance swaps $V(x_i)/x_i$ with times-to-expiry x_1, \dots, x_n . The variance swap curve at this time is then given by the mapping $T \mapsto V(T)/T$. We call V the variance curve and V' the forward variance curve. The variance swap curve quoted in volatility strikes is given by $T \mapsto \sqrt{V(T)/T}$. Several approaches for modeling variance swap curves are based on forward variance curves, but variance curves or forward variance curves are not observed. Instead, they must be estimated from a discrete set of observed variance swap prices which is often done via (piecewise) polynomial functions for interpolation between the observations. Such an approach often makes the forward variance curves vary significantly from day to day.

Hence we choose a non-parametric method for constructing smooth curves (see, e.g. Härdle *et al.* (2004)). We apply a local quadratic regression to the variance prices $V(x_{ji})$, leading to the following minimization problem:

$$\min_{\beta} \sum_{i=1}^n \{V(x_i) - \beta_0 - \beta_1(x_i - x) - \beta_2(x_i - x)^2\} K_h(x_i - x),$$

where the vector $\beta = (\beta_0, \beta_1, \beta_2)$ depends on x . The result $\hat{\beta}(x)$ is a weighted least-squares estimator where the variance curve is given by $\hat{\beta}_0$ and its first derivative by $\hat{\beta}_1$. We use the quartic kernel K and choose the bandwidth h by a rule of thumb described by Fan and Gijbels (1996). We do not consider higher-order kernels because of their inferior finite sample bias.

2.2. Modeling variance swap curves

On each day, we construct a variance curve to which we fit two parametric models and a semiparametric model. The resulting time series of factor loadings are the basis for the forecasting. We use the functional form derived from the Heston model and, in addition, we consider an extension that leads to the form of Nelson and Siegel (1987). These models are convenient and parsimonious exponential factor approximations. Moreover, we analyse a semiparametric approach.

In the Heston model,

$$\frac{dS_t}{S_t} = rdt + \sqrt{\zeta_t} dW_t^{(1)},$$

$$d\zeta_t = \kappa(\theta - \zeta_t)dt + v\sqrt{\zeta_t}dW_t^{(2)}$$

with correlated Wiener processes $W^{(1)}$ and $W^{(2)}$, the prices of (annualized) variance swaps $V(T)/T$ are given by

$$\theta + (\zeta_0 - \theta) \frac{1 - \exp(-\kappa T)}{\kappa T}.$$

Hence, only the short variance ζ_0 , the long variance θ and the mean-reversion speed κ determine the variance swap prices. In the Heston model, the smile of the implied volatility surfaces is controlled by the other two parameters: the correlation between the Brownian motions and the volatility of variance v . These two parameters do not enter the formula for the variance swap price.

The corresponding model for the forward variance curve is given by

$$v(T) \stackrel{\text{def}}{=} V'(T) = \theta + (\zeta_0 - \theta)\exp(-\kappa T).$$

This forward variance curve model implies exactly the above variance swap prices because of the constraint $V(0) = 0$. Reparametrizing this model and writing it in factor notation we obtain for the prices of variance swaps $V(T)/T$

$$z_1 + z_2 \frac{1 - \exp(-\kappa T)}{\kappa T}, \tag{1}$$

where $z = (z_1, z_2)$ are the two factor loadings. They correspond to the model parameters $(\theta, \zeta_0 - \theta)$ for the volatility. The reparametrization can be described formally in terms of a reparametrization matrix by

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \theta \\ \zeta \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

This model for forward variance curves is also called the linearly mean-reverting (forward) variance curve model (Bühler 2006).

As Diebold and Li (2003) obtained good results with the Nelson–Siegel parametrization, we generalize the above model in such a way that the resulting variance swap prices have a Nelson–Siegel parametrization:

$$v(T) = z_1 + z_2 \exp(-\kappa T) + z_3 \kappa T \exp(-\kappa T).$$

This model is called the double mean-reverting (forward) variance curve model. The variance swap prices $V(T)/T$ are given in this model by

$$z_1 + z_2 \frac{1 - \exp(-\kappa T)}{\kappa T} + z_3 \left\{ \frac{1 - \exp(-\kappa T)}{\kappa T} - \exp(-\kappa T) \right\}. \tag{2}$$

Thus, the generalized Heston model leads exactly to a Nelson–Siegel parametrization for the prices of variance swaps.

While the linearly mean-reverting model is basically the Heston model, the second approach was considered by Bühler (2006), who analysed the conditions for an arbitrage-free joint market of variance swaps and stock. His considerations imply that the mean-reversion speed κ should be constant. In practice, a constant mean-reversion speed is important for the stability of the parameters. For these theoretical and practical reasons we fix this parameter.

We interpret z_{1t} , z_{2t} and z_{3t} as latent dynamic factor loadings for the prices of variances swaps $V(T)/T$. The factor on z_{1t} is the constant 1. As this factor does not decay to zero in the long run it can be interpreted as a long-term factor. The factor on z_{2t} is $\{1 - \exp(\kappa T)\}/(\kappa T)$. This function is monotonically decreasing from 1 to 0. As it influences only the short end of the curve it can be interpreted as a short-term factor. Besides these two factors the generalized model also controls the medium term. The factor on z_{3t} is $\{1 - \exp(-\kappa T)\}/(\kappa T) - \exp(-\kappa T)$. This mapping increases monotonically from 0 to a peak and then decreases to zero in the long term in a similar way as the second factor. This form explains the interpretation as a medium-term factor. These three factors are presented in figure 1.

The interpretation of these factors corresponds to their meaning in the Heston model: z_1 is the long variance and $z_1 + z_2$ is the short variance. Moreover, these quantities can be recovered from the variance swap curve: from the limits $\lim_{T \rightarrow 0} V(T)/T = z_1 + z_2$ and $\lim_{T \rightarrow \infty} V(T)/T = z_1$ we see that z_1 is the long variance (i.e. θ) and $z_1 + z_2$ is the short variance (i.e. ζ_0). Hence, we have used the parametrization $\zeta_0 = z_1 + z_2$ and $\theta = z_1$. Thus, the original parameters of the Heston model (θ, ζ) can be recovered by multiplying the inverse of the reparametrization matrix by the factor loadings (z_1, z_2).

Moreover, the parameters have interpretations as level, slope and curvature. As an increase in z_1 increases the whole curve by the same amount, the factor on z_1 represents the level of the curve. An increase of the short-term factor increases the curve more at the short end than at the

long end. Hence it controls the slope of the curve. Finally, the third factor moves the middle of the curve while keeping the ends almost fixed. In this way it changes the curvature of the curve. Hence, the difference between the Heston model and its three-factor generalization is the capability to control the curvature.

Besides these parametric approaches we analyse a semi-parametric model described by Fessler (2005). It offers a low-dimensional representation of variance swap curves that are approximated by basis functions. These basis functions are unknown and have to be estimated from the data. The dynamics of the curves are described by the time series of the corresponding factor loadings.

Let $Y_{i,j}$ be an observed price of a variance swap on day i with maturity $T_j \in \{0.12, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0\}$. Let $X_{i,j}$ be a one-dimensional variable representing the time-to-maturity. Then the model regresses $Y_{i,j}$ on $X_{i,j}$ by

$$Y_{i,j} = m_0(X_{i,j}) + \sum_{l=1}^L \beta_{i,l} m_l(X_{i,j}),$$

where m_0 is an invariant basis function, m_l ($l = 1, \dots, L$) are the ‘dynamic’ basis functions and $\beta_{i,l}$ are the factor weights depending on time i . We describe the estimation procedure and the obtained basis functions in section 3.2 where we use actual data.

2.3. Stylized facts of variance swap curves

A model of variance swap curves should in principle have a reasonable in-sample fit and be able to reproduce the variety of observed shapes of variance swap curves. A good model should moreover reflect the dynamics of the curves by a reasonable out-of-sample performance.

We consider briefly some stylized facts of variance swap curves and see how the described approaches can model these characteristic shapes. The average variance swap curve is increasing and concave. The slope factor can readily replicate the increasing structure. The concavity can be modeled in the generalized model by the third factor. Variance swap curves show many different shapes in different markets over time. They can be upward- or downward-sloping and some have a hump. These shapes can be replicated, in principle, by varying the three factors accordingly. The short end of the curves is more volatile than the long end. This is reflected in the models because two factors (z_1 and z_2) control the short end while only one factor (z_1) models the long end. The Heston model can replicate many patterns, but not the humps (i.e. the curvature), while the generalized model has, in principle, the capability of replicating all stylized facts. The semiparametric model should be able to replicate these stylized facts because of its more flexible structure.

3. Forecasting the term structure

In this section, we describe the data used, estimate the factor loadings, model them and compare the forecasted variance swap curves.

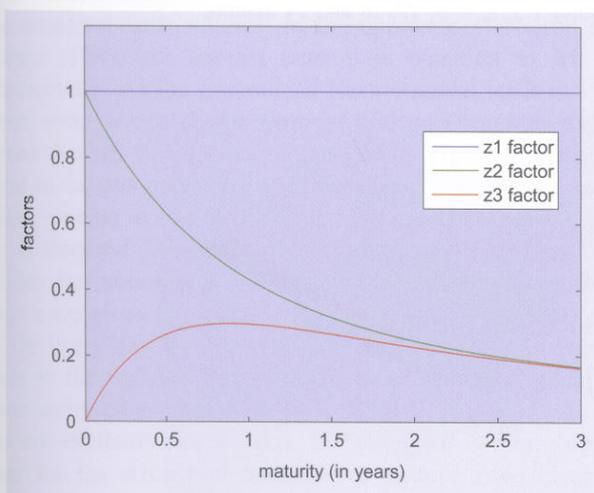


Figure 1. Factors for variance curves in the Heston model and its generalization (for $\kappa = 2$).

3.1. The data

The data set studied contains prices of variance swaps on the S&P 500 index between 1 October 2003 and 30 September 2005. These swaps use daily closing prices of the index, have 252 business days as annualization factor and assume a zero mean for the calculation of the variance of the returns. The prices are quoted in volatility strikes and represent the mid-market prices. Hence, we observe on every trading day prices $\sqrt{V(x_i)/x_i}$, $i = 1, \dots, n$, of variance swaps with times-to-maturity x_1, \dots, x_n . We have around $n = 7$ observations per day.

Our analysis does not require the use of fixed maturities because we always model the entire variance swap curve. But we use fixed maturities in order to simplify the following variance swap curve forecasts. Hence we first create from the discrete data curves by local quadratic smoothing as described in section 2.1. Then we extract the data for the fixed maturities 1.5, 3, 6, 9, 12, 18 and 24 months.

The variance swap prices (not quoted in volatility strikes) and level, slope and curvature of these curves are the basis for the following. In figure 2 we present the smoothed variance swap curves. The figure also shows the variance swap curves quoted in volatility strikes. Although these prices are often quoted in volatility strikes, we

estimate and forecast the variance curves because the integrated variance is normally the essential quantity in option pricing. In figure 3 we show the corresponding variance and forward variance curves. The variation of the level is clearly visible for the variance swap curves; the changes in the slope and curvature are less apparent. The changes are more readily observed from the forward variance curves.

We provide some descriptive statistics of the variance swap curves in table 1. Here, we also present the level (defined as the 24 month price), the slope (defined as the 24 month price minus the 1.5 month price) and the curvature (defined as twice the 6 month price minus the 1.5 month price minus the 24 month price). We will see below that these empirical factor loadings are highly correlated with the loadings of the parametric models. In figure 4 we show the median variance swap curve, with pointwise interquartile ranges. The mentioned upward-sloping and concave form is visible.

3.2. Fitting the variance swap curves

We estimate the Heston model and its generalization by minimizing the difference between the observed variance swap curves and the model prices. In the Heston model, these prices are given by

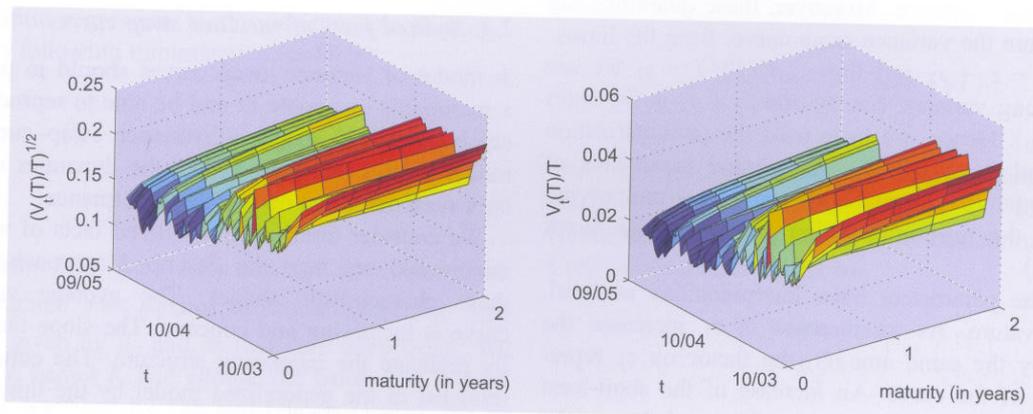


Figure 2. Variance swap curves quoted in volatility strikes (left) and variance swap curves (right), 01/10/03–30/09/05. The sample consists of weekly curves from October 2003 to September 2005 at maturities 1.5, 3, 6, 9, 12, 18 and 24 months.

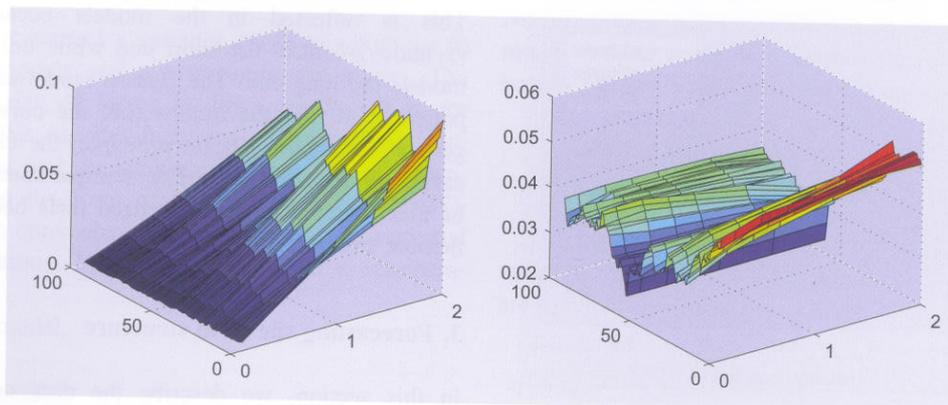


Figure 3. Variance curves (left) and forward variance curves (right), 01/10/03–30/09/05. The sample consists of weekly curves from October 2003 to September 2005 at maturities 1.5, 3, 6, 9, 12, 18 and 24 months.

M
1.
3
6
9
12
18
24
Sl
Cu

Figure
plot th

and in

$z_1 + z_2$

The fac
by non
Siegel
paramet
ance sw
sense th
tical to
eling, pr
as in Be
plifies th
ings z ar
On ev
tion to t
time seri
do not e
tant for
tions with
We est
the first y

Table 1. Descriptive statistics of variance swap curves [E^{-2}].

Mat. (months)	Mean	Std. dev.	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(4)$	$\hat{\rho}(12)$
1.5	2.03	0.65	0.91	4.36	80.1	64.5	51.1
3	2.48	0.70	1.39	4.39	90.1	79.3	65.4
6	2.75	0.73	1.51	4.41	93.3	82.6	64.7
9	2.89	0.73	1.60	4.45	94.1	82.5	61.3
12	2.98	0.72	1.69	4.54	94.5	81.8	57.1
18	3.14	0.70	1.85	4.68	94.9	79.6	47.3
24	3.27	0.69	2.02	4.79	95.0	77.1	36.8
Slope	1.24	0.44	-0.24	2.21	74.1	38.8	-16.6
Curvature	0.20	0.31	-0.50	0.98	83.6	62.0	49.5

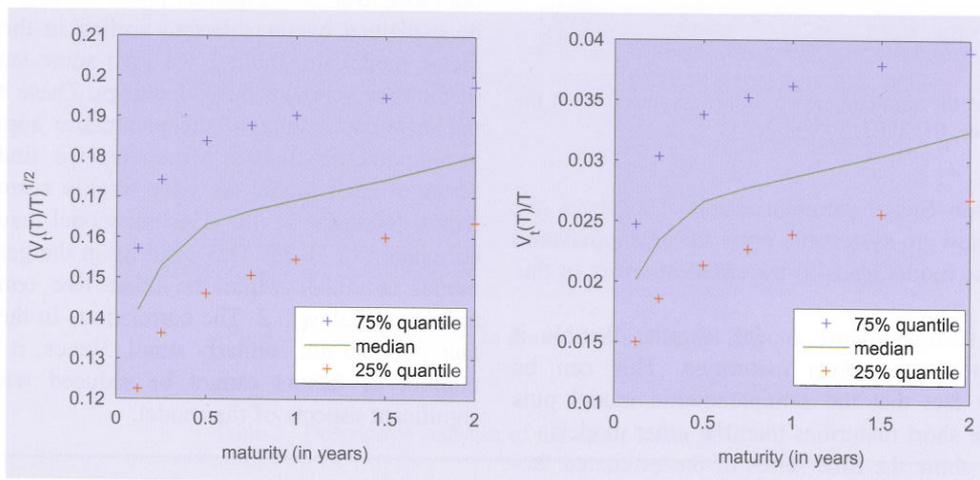


Figure 4. Median data-based forward variance curve quoted in volatility strikes with pointwise interquartile range. For each maturity, we plot the median along with the 25th and 75th quantiles.

$$z_1 + z_2 \frac{1 - \exp(-\kappa T)}{\kappa T},$$

and in the generalized Heston model by

$$z_1 + z_2 \frac{1 - \exp(-\kappa T)}{\kappa T} + z_3 \left\{ \frac{1 - \exp(-\kappa T)}{\kappa T} - \exp(-\kappa T) \right\}.$$

The factor loadings z and the parameter κ can be estimated by nonlinear least squares. In the approach of Nelson and Siegel (1987) for interest rates it is common to fix the parameter κ . As the generalized Heston model leads to variance swap prices in the form of Nelson–Siegel, it makes sense that we also fix the parameter κ . Moreover, it is practical to fix this parameter in the Heston model for the modeling, pricing and hedging of options. Hence, we use $\kappa = 2$ as in Bergomi (2004). Keeping this parameter constant simplifies the numerics considerably because the factor loadings z are given by OLS.

On every day we apply an ordinary least-squares estimation to the variance swap curves. In this way we obtain a time series of estimated factor loadings $(\hat{z}_1, \hat{z}_2, \hat{z}_3)^T$. As we do not explicitly use weights, the short end is more important for the estimation because we sample more observations with short maturities.

We estimate the factors in the semiparametric model from the first year of our time series. The factors or basis func-

tions \hat{m}_l and the factor loadings $\hat{\beta}_{i,l}$ are estimated by minimizing the following least-squares criterion ($\beta_{i,0} = 1$):

$$\sum_{i=1}^I \sum_{j=1}^{J_i} \int \left\{ Y_{i,j} - \sum_{l=0}^L \hat{\beta}_{i,l} \hat{m}_l(u) \right\}^2 K_h(u - X_{i,j}) du,$$

where K_h denotes a kernel function. The minimization procedure searches through all functions $\hat{m}_l : \mathbb{R} \rightarrow \mathbb{R}$, $l = 0, \dots, L$ and time series $\hat{\beta}_{i,l} \in \mathbb{R}$ ($i = 1, \dots, I, l = 1, \dots, L$) by an iterative procedure. The estimates are then orthogonalized and normalized (see Fengler (2005) for details). The estimated factors are plotted in figure 5. They can be interpreted as in the parametric models as level, slope and curvature (see section 2.2). A comparison with the parametric factors in figure 1 reveals that the estimated factors have a different scaling and become negative. The positivity of the parametric factors allows us to ensure the positivity of the resulting curve by imposing simple constraints on the loadings. After the factors have been estimated the factor loadings can be estimated by ordinary least squares as in the parametric models.

Information about the in-sample fit of the models is presented in figure 6. It shows that the Heston and the semiparametric model have problems fitting the short end of the curves. This systematic deviation holds in principle

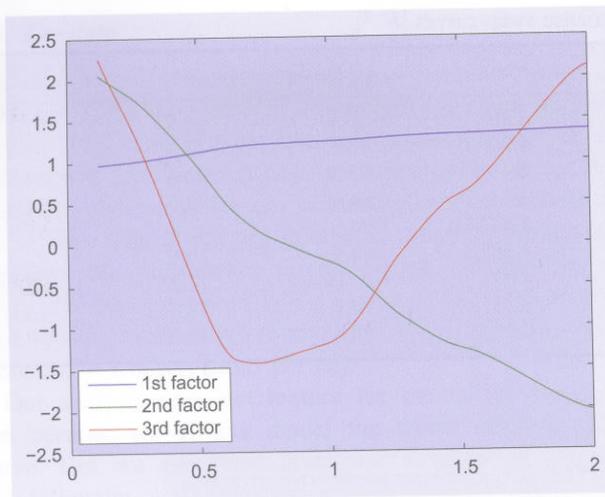


Figure 5. Factors in the semiparametric model estimated from the variance swap curves, 01/10/03–30/09/04.

also for the Nelson–Siegel parametrization. The prices for long maturities show no systematic error for all models and the semiparametric model leads to the smallest errors in this region.

Although the semiparametric model is quite flexible it has problems fitting the short maturities. This can be explained by the fact that the semiparametric model puts less weight on the short maturities than the other models.

In figure 7 we show the time series of the estimated factor loadings for all models. In this figure we have plotted the negative slope loadings of the models and we have scaled the curvature loadings of the models by 0.3. As the factors have interpretations as level, slope and curvature, we compare them with the empirical level, slope and curva-

ture as defined in section 3.1. As the empirical and model quantities are highly correlated we see that the definitions and interpretations of the empirical quantities are appropriate. The empirical level factor is very similar to the level loadings in the generalized Heston model. The corresponding loadings of the Heston model lie above, and the loadings of the semiparametric model lie below the empirical levels. The empirical slope is a good estimator for slope loadings in the Heston and in the generalized Heston model. The empirical curvature differs from the loadings of the models. In all three cases the model and the empirical values are highly correlated.

The loadings of the semiparametric factor model differ the most from the empirical factor loadings. This can partly be explained by the different scaling in the semiparametric factor model. In Table 2 we give some summary statistics of the time series of factor loadings. These statistics confirm the different scaling of the parametric approaches and the semiparametric model. Moreover, we find that the time series of each model are only weakly correlated. The two-factor loadings in the Heston model have an empirical correlation of -0.39 . This quantity in the generalized Heston model is -0.33 , while the other two correlations of the model are below 0.2. The correlations in the semiparametric factor model are similarly small. Hence, it appears that the number of factors cannot be reduced without sacrificing significant aspects of the model.

3.3. Modeling the factor loadings and forecasting the variance swap curves

Diebold and Li (2003) model the factor loadings of the Nelson–Siegel framework by univariate AR(1) processes for

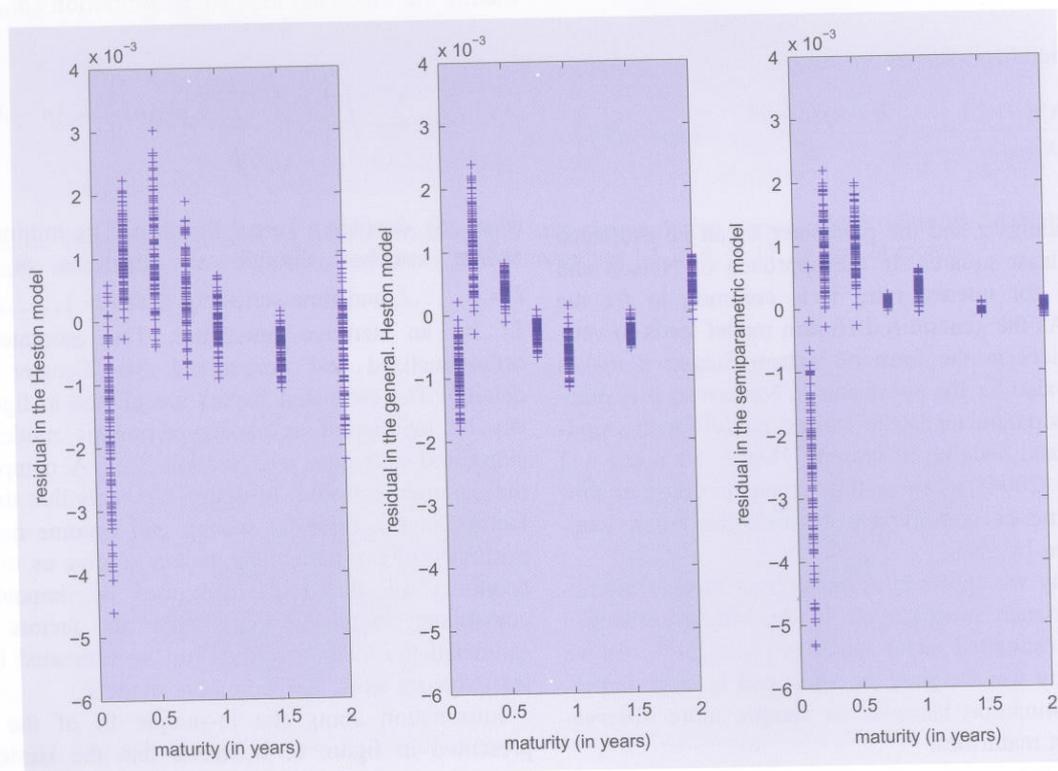


Figure 6. Variance swap curve residuals, 01/10/03–30/09/05. left, Heston; middle, generalized Heston; right, semiparametric model.

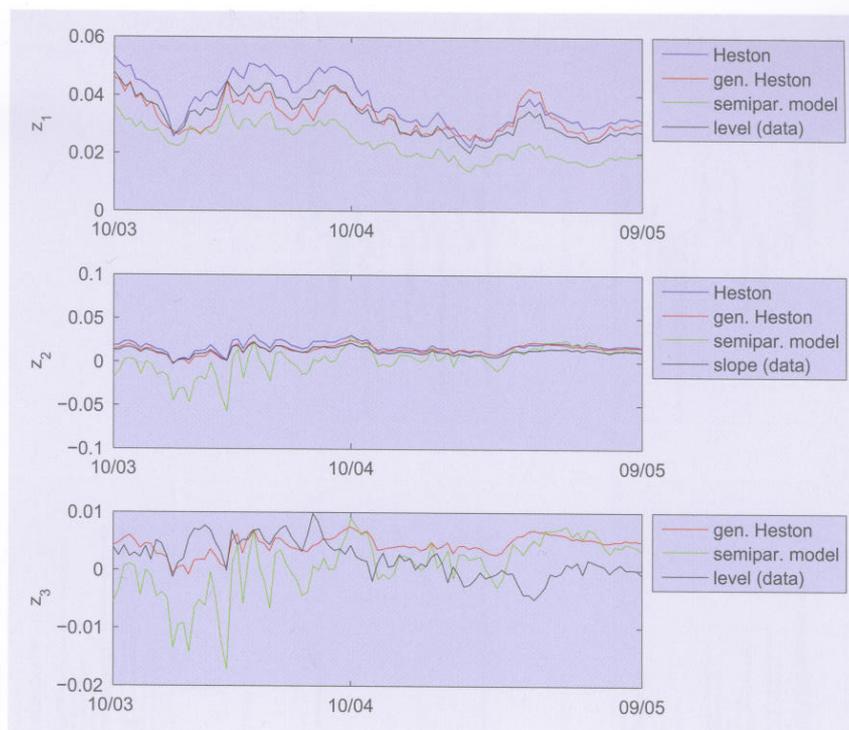


Figure 7. Factor loadings in the models and in the data.

Table 2. Descriptive statistics of the factor loadings [E^{-2}].

Model	Factor	Mean	Std. dev.	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(4)$	$\hat{\rho}(12)$
Heston	z_1	3.74	0.80	2.21	5.32	95.4	75.2	31.1
	z_2	-1.74	0.61	-3.08	0.33	74.1	38.1	-17.9
Generalized Heston	z_1	3.27	0.58	2.45	4.60	87.9	51.9	-26.6
	z_2	-1.46	0.56	-2.50	0.27	79.8	49.0	-11.2
	z_3	1.34	1.30	-1.37	4.61	81.3	57.6	45.7
Semiparam. model	z_1	2.37	0.58	1.36	3.67	93.5	81.1	58.1
	z_2	-0.01	0.10	-0.17	0.33	73.1	50.0	23.1
	z_3	0.00	0.03	-0.06	0.06	70.0	36.9	24.8

yield curves. Cont and da Fonseca (2002) also use these models for the factor loadings in their principal components analysis of implied volatility surfaces. Hence, we follow this approach. Moreover, more complex ARMA models did not improve the forecasting results in tests. We do not consider multivariate AR processes because there is only little correlation between the factor loadings (see section 3.2). In addition, the use of AR(1) processes allows us to more easily compare our results with the findings of Diebold and Li (2003).

The resulting forecasts of the variance swap curves τ weeks ahead at time t are given by

$$V_{t+\tau}(T)/T = \hat{z}_{1,t/t+\tau}f_1(T) + \hat{z}_{2,t/t+\tau}f_2(T) + \hat{z}_{3,t/t+\tau}f_3(T),$$

where $\hat{z}_{i,t/t+\tau}$ are the forecasts of the i th factor loading and f_1, f_2 and f_3 are the factors. These factor loading forecasts can be computed by regressing the loadings at $t+h$ on the loadings at t . As we obtained better results using repeated 1-day forecasts we used this second approach instead. In figure 8 we show the autocorrelation functions of the

residuals of the estimated AR(1) models. In general, only a few autocorrelations lie slightly outside the 95% confidence interval for all models. Hence, the models seem to be in line with the data.

As the models describe the variance swap curves by the factor loadings z , we forecast the loadings $(\hat{z}_{1t}, \hat{z}_{2t}, \hat{z}_{3t})$. Our data set consists of observations from 10/2003 to 9/2005. We use the first part of the data for the estimation of the factor loadings and forecast the variance swap curves of the second year. In the semiparametric model, the factors are estimated from the data of the first year. Then we keep these factors fixed for the forecasting. Actually, these factors differ only slightly from the factors estimated from the whole data set. If we want to forecast at time t the variance curve at time $t + \tau$, then we use the whole history of the factor loadings up to time t , i.e. z_1, \dots, z_t .

We made forecasts for 1 week, 1 month, 3 months and 6 months. In Tables 3 and 4 we show the results for 1 week and 6 months forecasts. As before, we consider the variance swap curves at maturities 1.5, 3, 6, 9, 12, 18 and 24 months. In addition to the three models described above we

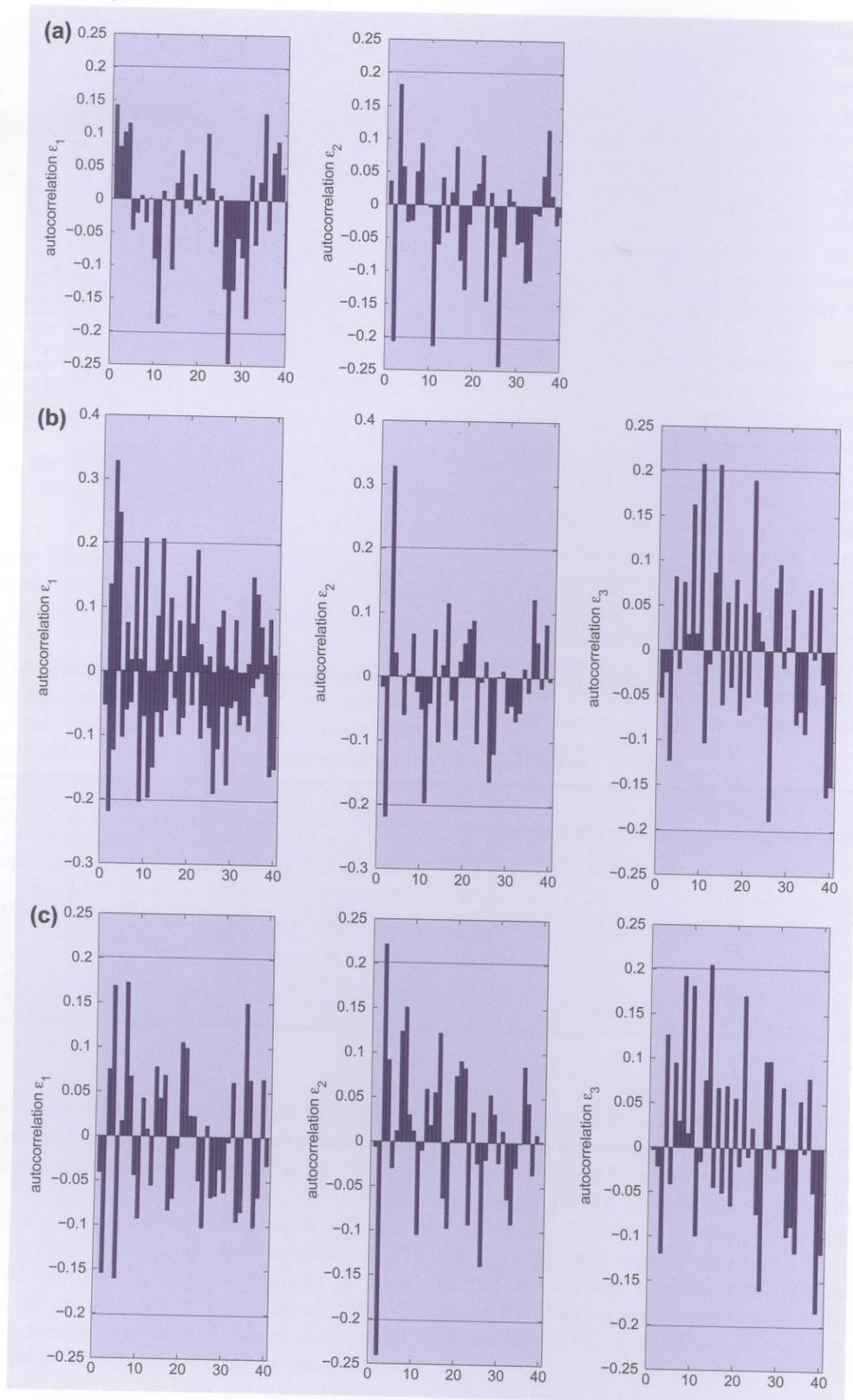


Figure 8. Factor loadings in the Heston model (top), in the generalized Heston model (middle) and in the semiparametric factor model (bottom).

consider two simple alternatives, the static Heston model and the random walk.

- *The static Heston model:* The Heston model (or, in general, stochastic volatility models) is often used when the dynamics of the volatility surfaces

are important. For example, forward starting call spreads depend on the skew of the volatility surface at the start date. For such products, the model is normally calibrated at the valuation date and the price is computed without any parameter forecast to the start date. Thus the model is used

Table 3. Out-of-sample 1-week-ahead forecasting results [E^{-2}].

Model	Mat. (months)	Mean	Std. dev.	MAE	MARE
Heston	1.5	0.14	0.27	0.24	16.8
	3	-0.04	0.21	0.16	8.4
	6	-0.02	0.21	0.15	7.2
	9	0.03	0.21	0.15	6.8
	12	0.07	0.21	0.16	7.0
	18	0.08	0.21	0.17	6.6
Generalized Heston	24	0.04	0.20	0.16	5.7
	1.5	0.19	0.23	0.24	17.1
	3	0.06	0.20	0.16	8.6
	6	0.13	0.20	0.18	9.0
	9	0.20	0.20	0.23	10.4
	12	0.22	0.20	0.25	10.7
Static Heston	18	0.19	0.19	0.22	8.8
	24	0.10	0.19	0.17	6.2
	1.5	0.17	0.25	0.25	17.0
	3	-0.01	0.21	0.15	8.1
	6	-0.00	0.20	0.15	6.9
	9	0.04	0.20	0.15	6.7
Semiparam. model	12	0.07	0.20	0.16	6.8
	18	0.08	0.20	0.16	6.4
	24	0.03	0.20	0.15	5.5
	1.5	0.32	0.26	0.35	24.4
	3	0.04	0.20	0.16	8.6
	6	0.03	0.20	0.15	7.3
Random walk	9	0.10	0.21	0.18	7.9
	12	0.06	0.20	0.16	6.8
	18	0.07	0.20	0.16	6.3
	24	0.04	0.19	0.15	5.4
	1.5	0.00	0.24	0.18	12.0
	3	0.01	0.21	0.15	8.2
	6	0.02	0.20	0.15	7.2
	9	0.02	0.20	0.15	6.7
	12	0.02	0.20	0.15	6.3
	18	0.02	0.19	0.15	5.7
	24	0.02	0.19	0.14	5.2

in a static way. For the Heston model, this method yields, for variance swaps that start at time τ , the price

$$V_{t+\tau}(\widehat{T})/T = \frac{V_t(T + \tau) - V_t(\tau)}{T}, \tag{3}$$

where V_t denotes the variance curve at time t . This price can be interpreted as the forecasted variance swap price in the static Heston model. Considering this model, we can see if this static approach or the forecast of factor loadings gives better price forecasts for variance swaps, i.e. better dynamics.

- *The random walk:* This natural benchmark 'model' forecasts that the variance swap curves do not change:

$$V_{t+\tau}(\widehat{T})/T = V_t(T)/T.$$

Duffie and Kan (1996) found that the analysed yield curve models seem to have problems in giving better forecasts than this model. Diebold and Li (2003) found that the reparametrized Nelson–Siegel

approach seems to outperform the random walk for yield curves. The Nelson–Siegel model corresponds to the generalized Heston model.

The forecast errors at time t are defined as the difference between the variance swap curve observed at $t + \tau$ and the forecasted curve:

$$V_{t+\tau}(\widehat{T})/T - V_{t+\tau}(T)/T,$$

for $T = 1.5, 3, 6, 9, 12, 18$ or 24 months. We examine a number of descriptive statistics of the errors, including the mean absolute error,

$$MAE \stackrel{\text{def}}{=} \frac{1}{n} \sum_t \|V_{t+\tau}(\widehat{T})/T - V_{t+\tau}(T)/T\|,$$

and the corresponding relative error,

$$MARE \stackrel{\text{def}}{=} \frac{1}{n} \sum_t \left\| \frac{V_{t+\tau}(\widehat{T})/T - V_{t+\tau}(T)/T}{V_{t+\tau}(T)/T} \right\|,$$

actor model

starting call
atility sur-
ducts, the
uation date
parameter
del is used

Table 4. Out-of-sample 6-months-ahead forecasting results [E^{-2}].

Model	Mat. (months)	Mean	Std. dev	MAE	MARE
Heston	1.5	0.52	0.40	0.54	39.8
	3	0.37	0.44	0.44	26.0
	6	0.40	0.48	0.46	24.6
	9	0.45	0.51	0.50	24.9
	12	0.48	0.53	0.53	24.8
	18	0.48	0.56	0.53	23.1
	24	0.42	0.58	0.50	20.6
Generalized Heston	1.5	0.82	0.30	0.82	62.3
	3	0.79	0.27	0.79	46.1
	6	0.92	0.31	0.92	46.9
	9	0.98	0.33	0.98	47.0
	12	0.99	0.35	0.99	44.8
	18	0.87	0.38	0.87	36.9
	24	0.71	0.40	0.71	28.6
Static Heston	1.5	1.16	0.48	1.16	84.4
	3	0.90	0.53	0.90	52.5
	6	0.76	0.59	0.76	39.5
	9	0.68	0.63	0.69	34.0
	12	0.62	0.66	0.64	30.3
	18	0.49	0.71	0.57	25.1
	24	0.36	0.75	0.54	22.2
Semiparam. model	1.5	0.97	0.31	0.97	71.8
	3	0.74	0.31	0.74	43.2
	6	0.77	0.36	0.77	39.7
	9	0.84	0.39	0.84	40.5
	12	0.79	0.41	0.79	36.1
	18	0.74	0.44	0.74	31.3
	24	0.64	0.46	0.64	26.0
Random walk	1.5	0.18	0.35	0.33	24.0
	3	0.21	0.38	0.32	18.9
	6	0.21	0.47	0.36	18.9
	9	0.20	0.52	0.38	18.9
	12	0.18	0.55	0.40	18.8
	18	0.14	0.62	0.44	19.1
	24	0.11	0.67	0.48	19.2

where the index t runs over all forecast days and n is the number of forecast days. All these errors are measured in variance.

The results of the 1-week forecasts are presented in table 3. All models except the generalized Heston model show a similar forecasting performance over this short time horizon with an average absolute error of around 0.15%. This seems reasonably small because the average variance swap curve has a level around 2.7%. We see that the random walk model has the smallest errors of all models. The static Heston model and the Heston model with parameter forecasts have similar errors. The three-factor models have the worst performance, and the Nelson–Siegel framework leads to larger errors for long maturities. Moreover, the errors tend to be larger for short maturities. Hence, the short ends of the variance swap curves are harder to forecast. This corresponds to the in-sample fit problems at the short end.

Forecasts for 1 month and 3 months show qualitatively similar results. The errors in the dynamic Heston model become smaller than in the static Heston model. Both Heston models produce better forecasts than the three-fac-

tor model. The semiparametric factor model, as before, outperforms the Nelson–Siegel parametrization.

In table 4, we present the forecast results for half a year ahead. For these long periods the dynamic Heston model gives forecasts of similar quality as the random walk. The static Heston model leads to rather large forecast errors and the semiparametric model produces similar errors as the Nelson–Siegel approach.

Variance swap curve forecasts are hard to forecast over short time horizons with models because the models already exhibit in-sample fit problems. For longer time horizons, the dynamic Heston model performs well and particularly better than the static Heston model.

4. Conclusion

We analyse the modeling and forecasting of variance swap curves in a Heston model, a three-factor Nelson–Siegel parametrization and a three-factor semiparametric model. The in-sample fit gives good results for long maturities, but

all models have problems in fitting the short end of the variance swap curves. The Nelson–Siegel parametrization naturally outperforms the Heston model, and the flexible semiparametric factor model leads to the best fit for long maturities, but also has problems with short maturities. In the forecasting analysis, all models give similar results for short forecasting horizons. But for longer time horizons the Heston model clearly leads to the smallest forecasting errors. In option pricing, model parameters are generally not forecasted. When we consider the Heston model in such a static way its performance is worse than the forecasted Heston model. The semiparametric approach seems problematic for forecasting.

Combining the in-sample and out-of-sample results the Heston model gives the best overall performance due to its good forecast results for long time horizons. This can be interpreted in a way that convexity is either difficult to model or not significant for long time horizons. In yield curve modelling, convexity seems more important, and the Nelson–Siegel model appears to be superior.

In addition, we show that the static Heston model can lead to bad variance swap curve dynamics. Hence the usual way of pricing forward starting products could be problematic in standard stochastic volatility models. Recent stochastic volatility approaches such as, for example, that of Bergomi (2005) appear necessary in order to capture the variance swap dynamics better.

References

- Amengual, D., The term structure of variance risk premia. Technical report, Princeton University, March 2009.
- Bergomi, L., Smile dynamics. *RISK*, 2004, 17(9), 117–123.
- Bergomi, L., Smile dynamics II. *RISK*, 2005, 18(10), 67–73.
- Bühler, H., Consistent variance curve models. *Finance Stochast.*, 2006, 10(2), 178–203.
- Carr, P. and Madan, D., Towards a theory of volatility trading. Reprinted in *Option Pricing, Interest Rates, and Risk Management*, edited by Musiella, Jouini, Cvitanic, pp. 417–427, 1998 (University Press).
- Cont, R. and da Fonseca, J., Dynamics of implied volatility surfaces. *Quant. Finance*, 2002, 2(1), 45–60.
- Demeterfi, K., Derman, E., Kamal, M. and Zou, J., More than you ever wanted to know about volatility swaps. Quantitative Strategies Research Notes, Goldman Sachs, 1999.
- Diebold, F.X. and Li, C., Forecasting the term structure of government bond yields. NBER Working Papers 10048, National Bureau of Economic Research, October 2003.
- Duffie, D. and Kan, R., A yield factor model of interest rates. *Math. Finance*, 1996, 6(4), 379–406.
- Fan, J. and Gijbels, I., *Local Polynomial Modelling and Its Applications*, 1996 (Chapman and Hall: London).
- Fengler, M., *Semiparametric Modeling of Implied Volatility*, 2005 (Springer: Berlin).
- Härdle, W., Müller, M., Sperlich, S. and Werwatz, A., *Nonparametric and Semiparametric Models*, 2004 (Springer: Heidelberg).
- Heston, S., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, 6(2), 327–343.
- Nelson, C. and Siegel, A., Parsimonious modeling of the yield curve. *J. Business*, 1987, 60, 473–489.