Research Article

Local Likelihood Density Estimation and Value-at-Risk

Christian Gourieroux[1] and Joann Jasiak[2]

[1] *CREST and University of Toronto, Canada*
[2] *York University, Canada*

Correspondence should be addressed to Joann Jasiak, jasiakj@econ.yorku.ca

This paper presents a new nonparametric method for computing the conditional Value-at-Risk, based on a local approximation of the conditional density function in a neighborhood of a predetermined extreme value for univariate and multivariate series of portfolio returns. For illustration, the method is applied to intraday VaR estimation on portfolios of two stocks traded on the Toronto Stock Exchange. The performance of the new VaR computation method is compared to the historical simulation, variance-covariance, and J. P. Morgan methods.

1. Introduction

The Value-at-Risk (VaR) is a measure of market risk exposure for portfolios of assets. It has been introduced by the Basle Committee on Banking Supervision (BCBS) and implemented in the financial sector worldwide in the late nineties. By definition, the VaR equals the Dollar loss on a portfolio that will not be exceeded by the end of a holding time with a given probability. Initially, the BCBS has recommended a 10-day holding time (and allowed for computing the VaR at horizon 10 days by rescaling the VaR at a shorter horizon) and loss probability 1%; (see, [1], page 3), Banks use the VaR to determine the required capital to be put aside for coverage of potential losses. (The required capital reserve is defined as $RC_t = \text{Max}[\text{VaR}_t, (M + m)1/60 \sum_{h=1}^{60} \text{VaR}_{t-h}]$, (see, [1], page 14 and [2], page 2), where M is a multiplier set equal to 3, and m takes a value between 0 and 1 depending on the predictive quality of the internal model used by the bank.) The VaR is also used in portfolio management and internal risk control. Therefore, some banks compute intradaily VaRs, at horizons of one or two hours, and risk levels of 0.5%, or less.

Formally, the conditional Value-at-Risk is the lower-tail conditional quantile and satisfies the following expression:

$$P_t[x_{t+1} < -\mathrm{VaR}_t(\alpha)] = \alpha, \tag{1.1}$$

where x_t is the portfolio return between $t-1$ and t, α denotes the loss probability, and P_t represents the conditional distribution of x_{t+1} given the information available at time t. Usually, the information set contains the lagged values $x_t, x_{t-1}, \ldots$ of portfolio returns. It can also contain lagged returns on individual assets, or on the market portfolio.

While the definition of the VaR as a market risk measure is common to all banks, the VaR computation method is not. In practice, there exist a variety of parametric, semiparametric, and nonparametric methods, which differ with respect to the assumptions on the dynamics of portfolio returns. They can be summarized as follows (see, e.g., [3]).

(a) Marginal VaR Estimation

The approach relies on the assumption of i.i.d. returns and comprises the following methods.

(1) Gaussian Approach

The VaR is the α-quantile, obtained by inverting the Gaussian cumulative distribution function

$$\mathrm{VaR}(\alpha) = -Ex_{t+1} - \Phi^{-1}(\alpha)(Vx_{t+1})^{1/2}, \tag{1.2}$$

where Ex_{t+1} is the expected return on a portfolio, Vx_{t+1} is the variance of portfolio returns, and $\Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution. This method assumes the normality of returns and generally underestimates the VaR. The reason is that the tails of the normal distribution are much thinner than the tails of an empirical marginal distribution of portfolio returns.

(2) Historical Simulation (see [1])

$\mathrm{VaR}(\alpha)$ is approximated from a sample quantile at probability α, obtained from historical data collected over an observation period not shorter than one year. The advantage of this method is that it relaxes the normality assumption. Its major drawback is that it provides poor approximation of small quantiles at α's such as 1%, for example, as extreme values are very infrequent. Therefore, a very large sample is required to collect enough information about the true shape of the tail. (According to the asymptotic properties of the empirical quantile, at least 200–300 observations, that is, one year, approximately, are needed for $\alpha = 5\%$ and at least 1000, that is, 4 years are needed for $\alpha = 1\%$, both for a Gaussian tail. For fatter tails, even more observations can be required (see, e.g., the discussion in [3]).

(3) Tail Model Building

The marginal quantile at a small risk level α is computed from a parametric model of the tail and from the sample quantile(s) at a larger α. For example, McKinsey Inc. suggests to

infer the 99th quantile from the 95th quantile by multiplying the latter one by 1.5, which is the weight based on a zero-mean Gaussian model of the tail. This method is improved by considering two tail quantiles. If a Gaussian model with mean μ and variance σ is assumed to fit the tail for $\alpha < 10\%$, then the $\widehat{\text{VaR}}(\alpha)$, for any $\alpha < 10\%$, can be calculated as follows, Let $\widehat{\text{VaR}}(10\%)$ and $\widehat{\text{VaR}}(5\%)$ denote the sample quantiles at risk levels 5% and 10%. From (1.2), the estimated mean and variance in the tail arise as the solutions of the system

$$\widehat{\text{VaR}}(10\%) = -\widehat{m} - \Phi^{-1}(10\%)\widehat{\sigma},$$
$$\widehat{\text{VaR}}(5\%) = -\widehat{m} - \Phi^{-1}(5\%)\widehat{\sigma}. \tag{1.3}$$

The marginal VaR at any loss probability α less than 10% is calculated as

$$\widehat{\text{VaR}}(\alpha) = -\widehat{m} - \Phi^{-1}(\alpha)\widehat{\sigma}, \tag{1.4}$$

where $\widehat{m}, \widehat{\sigma}$ are solutions of the above system. Equivalently, we get

$$\frac{\widehat{\text{VaR}}(\alpha) - \widehat{\text{VaR}}(10\%)}{\widehat{\text{VaR}}(5\%) - \widehat{\text{VaR}}(10\%)} = \frac{\Phi^{-1}(\alpha) - \Phi^{-1}(10\%)}{\Phi^{-1}(5\%) - \Phi^{-1}(10\%)}. \tag{1.5}$$

Thus, $\widehat{\text{VaR}}(\alpha)$ is a linear combination of sample quantiles $\widehat{\text{VaR}}(10\%)$ and $\widehat{\text{VaR}}(5\%)$ with the weights determined by the Gaussian model of the tail.

This method is parametric as far as the tail is concerned and nonparametric for the central part of the distribution, which is left unspecified.

The marginal VaR estimation methods discussed so far do not account for serial dependence in financial returns, evidenced in the literature. (These methods are often applied by rolling, that is, by averaging observations over a window of fixed length, which implicitly assumes independent returns, with time dependent distributions.)

(b) Conditional VaR Estimation

These methods accommodate serial dependence in financial returns.

(1) J. P. Morgan

The VaR at 5% is computed by inverting a Gaussian distribution with conditional mean zero and variance equal to an estimated conditional variance of returns. The conditional variance is estimated from a conditionally Gaussian IGARCH-type model of volatility σ_t^2, called the Exponentially Weighted Moving Average, where $\sigma_t^2 = \theta\sigma_{t-1}^2 + (1-\theta)x_{t-1}^2$, and parameter θ is arbitrarily fixed at 0.94 for any portfolio [4].

(2) CaViar [5]

The CaViar model is an autoregressive specification of the conditional quantile. The model is estimated independently for each value of α, and is nonparametric in that respect.

Table 1: Computation of the VaR.

	Parametric	Semi-parametric	Nonparametric
i.i.d.	Gaussian approach	tail model building (advisory firms)	historical simulation (regulators)
Serial Dependence	IGARCH by J. P. Morgan DAQ		CaViar

(3) Dynamic Additive Quantile (DAQ) [6]

This is a parametric, dynamic factor model of the conditional quantile function.

Table 1 summarizes all the aforementioned methods.

This paper is intended to fill in the empty cell in Table 1 by extending the tail model building method to the conditional Value-at-Risk. To do that, we introduce a parametric pseudomodel of the conditional portfolio return distribution that is assumed valid in a neighbourhood of the VaR of interest. Next, we estimate locally the pseudodensity, and use this result for calculating the conditional VaRs in the tail.

The local nonparametric approach appears preferable to the fully parametric approaches for two reasons. First, the nonparametric methods are too sensitive to specification errors. Second, even if the theoretical rate of convergence appears to be smaller than that of a fully parametric method (under the assumption of no specification error in the latter one), the estimator proposed in this paper is based on a local approximation of the density in a neighborhood where more observations are available than at the quantile of interest.

The paper is organized as follows. Section 2 presents the local estimation of a probability density function from a misspecified parametric model. By applying this technique to a Gaussian pseudomodel, we derive the local drift and local volatility, which can be used as inputs in expression (1.2). In Section 3, the new method is used to compute the intraday conditional Value-at-Risk for portfolios of two stocks traded on the Toronto Stock Exchange. Next, the performance of the new method of VaR computation is compared to other methods, such as the historical simulation, Gaussian variance-covariance method, J. P. Morgan IGARCH, and ARCH-based VaR estimation in Monte Carlo experiments. Section 4 discusses the asymptotic properties of the new nonparametric estimator of the log-density derivatives. Section 5 concludes the paper. The proofs are gathered in Appendices.

2. Local Analysis of the Marginal Density Function

The local analysis of a marginal density function is based on a family of pseudodensities. Among these, we define the pseudodensity, which is locally the closest to the true density. Next, we define the estimators of the local pseudodensity, and show the specific results obtained for a Gaussian family of pseudodensities.

2.1. Local Pseudodensity

Let us consider a univariate or multivariate random variable Y, with unknown density f_0, and a parametric multivariate family of densities $\mathcal{F} = \{f(y, \theta), \ \theta \in \Theta\}$, called the family of pseudodensities where the parameter set $\Theta \subset \mathbb{R}^p$. This family is generally misspecified. Our

method consists in finding the pseudodensity $f(y; \theta_0^*)$, which is locally the closest to the true density. To do that we look for the local pseudo-true value of parameter θ.

In the first step, let us assume that variable Y is univariate and consider an approximation on an interval $A = [c - h, c + h]$, centered at some value c of variable Y. The pseudodensity is derived by optimizing the Kullback-Leibler criterion evaluated from the pseudo and true densities truncated over A. The pseudo-true value of θ is

$$
\begin{aligned}
\tilde{\theta}_{c,h} = \underset{\theta}{\mathrm{Argmax}} \Big[& E_0 \big[1_{c-h<Y<c+h} \log f(Y; \theta) \big] \\
& - E_0 [1_{c-h<Y<c+h}] \log \int 1_{c-h<y<c+h} \, f(y; \theta) dy \Big] \\
= \underset{\theta}{\mathrm{Argmax}} \ & E_0 \Big[\frac{1}{2h} 1_{c-h<Y<c+h} \log f(Y; \theta) \Big] \\
& - E_0 \Big[\frac{1}{2h} 1_{c-h<Y<c+h} \Big] \log \int \frac{1}{2h} 1_{c-h<y<c+h} f(y; \theta) dy,
\end{aligned}
\tag{2.1}
$$

where E_0 denotes the expectation taken with respect to the true probability density function (pdf, henceforth) f_0. The pseudo-true value depends on the pseudofamily, the true pdf, the bandwidth, and the location c. The above formula can be equivalently rewritten in terms of the uniform kernel $K(u) = (1/2)1_{[-1,1]}(u)$. This leads to the following extended definition of the pseudo-true value of the parameter which is valid for vector Y of any dimension d, kernel K, bandwidth h, and location c:

$$
\begin{aligned}
\tilde{\theta}_{c,h} = \underset{\theta}{\mathrm{Argmax}} \ & E_0 \Big[\frac{1}{h^d} K \Big(\frac{Y-c}{h} \Big) \log f(Y; \theta) \Big] \\
& - E_0 \Big[\frac{1}{h^d} K \Big(\frac{Y-c}{h} \Big) \Big] \log \int \frac{1}{h^d} K \Big(\frac{y-c}{h} \Big) f(y; \theta) dy.
\end{aligned}
\tag{2.2}
$$

Let us examine the behavior of the pseudo-true value when the bandwidth tends to zero.

Definition 2.1. (i) The *local parameter function (l.p.f.)* is the limit of $\tilde{\theta}_{c,h}$ when h tends to zero, given by

$$
\tilde{\theta}(c, f_0) = \lim_{h \to 0} \tilde{\theta}_{c,h},
\tag{2.3}
$$

when this limit exists.

(ii) The local pseudodensity is $f[y; \tilde{\theta}(c, f_0)]$.

The local parameter function provides the set of local pseudo-true values indexed by c, while the local pseudodensity approximates the true pdf in a neighborhood of c. Let us now discuss some properties of the l.p.f.

Proposition 2.2. *Let one assume the following:*

(A.1) *There exists a unique solution to the objective function maximized in (2.2) for any h, and the limit $\tilde{\theta}(c, f_0)$ exists.*

(A.2) *The kernel K is continuous on $\mathbb{R}^d$, of order 2, such that $\int K(u)du = 1$, $\int uK(u)du = 0$, $\int uu'K(u)du = \eta^2$, positive definite.*

(A.3) *The density functions $f(y, \theta)$ and $f_0(y)$ are positive and third-order differentiable with respect to y.*

(A.4) $\dim \theta = p \geq d$, *and, for any c in the support of f_0,*

$$\left\{ \frac{\partial \log f(c; \theta)}{\partial y}, \theta \in \Theta \right\} = \mathbb{R}^d. \tag{2.4}$$

(A.5) *For h small and any c, the following integrals exist: $\int K(u) \log f(c + uh; \theta) f_0(c + uh)du$, $\int K(u)f_0(c+uh)du$, $\int K(u)f(c+uh; \theta)du$, and are twice differentiable under the integral sign with respect to h.*

Then, the local parameter function is a solution of the following system of equations:

$$\frac{\partial \log f\left[c; \tilde{\theta}(c; f_0)\right]}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}, \quad \forall c. \tag{2.5}$$

Proof. See Appendix A.   □

The first-order conditions in Proposition 2.2 show that functions $f[y, \tilde{\theta}(c, f_0)]$ and $f_0(c)$ have the same derivatives at c. When p is strictly larger than d, the first-order conditions are not sufficient to characterize the l.p.f.

Assumption (A.1) is a local identification condition of parameter θ. As shown in the application given later in the text, it is verified to hold for standard pseudofamilies of densities such as the Gaussian, where $\tilde{\theta}(c, f_0)$ has a closed form. (The existence of a limit $\tilde{\theta}(c, f_0)$ is assumed for expository purpose. However, the main result concerning the first-order conditions is easily extended to the case when $\tilde{\theta}_{c,h}$ exists, with a compact parameter set Θ. The proof in Appendix A shows that, even if the $\lim_{h \to 0} \tilde{\theta}(c, f_0)$ does not exist, we get $\lim_{h \to 0}(\partial \log f[c, \tilde{\theta}_{c,h}]/\partial y) = \partial \log f_0(c)/\partial y$, $\forall c$. This condition would be sufficient to define a local approximation to the log-derivative of the density.)

It is known that a distribution is characterized by the log-derivative of its density due to the unit mass restriction. This implies the following corollary.

Corollary 2.3. *The local parameter function characterizes the true distribution.*

2.2. Estimation of the Local Parameter Function and of the Log-Density Derivative

Suppose that $y_1, \ldots, y_T$ are observations on a strictly stationary process (Y_t) of dimension d. Let us denote by f_0 the true marginal density of Y_t and by $\{f(y : \theta), \theta \in \Theta\}$ a (misspecified)

pseudoparametric family used to approximate f_0. We now consider the l.p.f. characterization of f_0, and introduce nonparametric estimators of the l.p.f. and of the marginal density.

The estimator of the l.p.f. is obtained from formula (2.2), where the theoretical expectations are replaced by their empirical counterparts:

$$\tilde{\theta}_T(c) = \underset{\theta}{\text{Argmax}} \left[\sum_{t=1}^T \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \log f(y_t; \theta) \right.$$

$$\left. - \sum_{t=1}^T \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \log \int \frac{1}{h^d} K\left(\frac{y - c}{h}\right) f(y; \theta) dy \right]. \tag{2.6}$$

The above estimator depends on the selected kernel and bandwidth. This estimator allows us to derive from Proposition 2.2 a new nonparametric consistent estimator of the log-density derivative defined as

$$\frac{\partial \log \hat{f}_T(c)}{\partial y} = \frac{\partial \log f\left[c; \tilde{\theta}_T(c)\right]}{\partial y}. \tag{2.7}$$

The asymptotic properties of the estimators of the l.p.f. and log-density derivatives are discussed in Section 4, for the exactly identified case $p = d$. In that case, $\tilde{\theta}_T(c)$ is characterized by the system of first-order conditions (2.7).

The quantity $f[c, \tilde{\theta}_T(c)]$ is generally a nonconsistent estimator of the density $f_0(c)$ at c (see, e.g., [7] for a discussion of such a bias in an analogous framework). However, a consistent estimator of the log-density (and thus of the density itself, obtained as the exponential function of the log-density) is derived directly by integrating the estimated log-density derivatives under the unit mass restriction. This offers a correction for the bias, and is an alternative to including additional terms in the objective function (see, e.g., [7, 8]).

2.3. Gaussian Pseudomodel

A Gaussian family is a natural choice of pseudomodel for local analysis, as the true density is locally characterized by a local mean and a local variance-covariance matrix. Below, we provide an interpretation of the Gaussian local density approximation. Next, we consider a Gaussian pseudomodel parametrized by the mean only, and show the relationship between the l.p.f. estimator and two well-known nonparametric estimators of regression and density, respectively.

(i) Interpretation

For a Gaussian pseudomodel indexed by mean m and variance Σ, we have

$$\frac{\partial \log f(y; m, \Sigma)}{\partial y} = \Sigma^{-1}(y - m). \tag{2.8}$$

Thus, the approximation associated with a Gaussian pseudofamily is the standard one, where the partial derivatives of the log-density are replaced by a family of hyperplanes

parallel to the tangent hyperplanes. These tangent hyperplanes are not independently defined, due to the Schwartz equality

$$\frac{\partial^2 \log f(y)}{\partial y_i \partial y_j} = \frac{\partial^2 \log f(y)}{\partial y_j \partial y_i}; \quad \forall i \neq j. \tag{2.9}$$

The Schwartz equalities are automatically satisfied by the approximated densities because of the symmetry of matrix Σ^{-1}.

(ii) Gaussian Pseudomodel Parametrized by the Mean and Gaussian Kernel

Let us consider a Gaussian kernel: $K(\cdot) = \phi(\cdot)$ of dimension d, where ϕ denotes the pdf of the standard Normal $N(0, Id)$.

Proposition 2.4. *The l.p.f. estimator for a Gaussian pseudomodel parametrized by the mean and with a Gaussian kernel can be written as*

$$\widetilde{\theta}_T(c) = c + \frac{1+h^2}{h^2}[\widetilde{m}_T(c) - c] = c + \left(1 + h^2\right)\frac{\partial \log}{\partial c}\widetilde{f}_T(c), \tag{2.10}$$

where

$$\widetilde{m}_T(c) = \frac{\left(\sum_{t=1}^{T}(1/h^d)\phi((y_t - c)/h)y_t\right)}{\left(\sum_{t=1}^{T}(1/h^d)\phi((y_t - c)/h)\right)} \tag{2.11}$$

is the Nadaraya-Watson estimator of the conditional mean $m(c) = E[Y \mid Y = c] = c$, and

$$\widetilde{f}_T(c) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}\,\phi\left(\frac{y_t - c}{h}\right) \tag{2.12}$$

is the Gaussian kernel estimator of the unknown value of the true marginal pdf at c.

Proof. See Appendix B. □

In this special case, the asymptotic properties of $\widetilde{\theta}_T(c)$ follow directly from the asymptotic properties of $\widetilde{f}_T(c)$ and $\partial\widetilde{f}_T(c)/\partial c$ [9]. In particular, $\widetilde{\theta}_T(c)$ converges to $c + \partial \log f_0(c)/\partial y$, when T and h tend to infinity and zero, respectively, with $Th^{d+2} \to 0$.

Alternatively, the asymptotic behavior can be inferred from the Nadaraya-Watson estimator [10, 11] in the degenerate case when the regressor and the regressand are identical. Section 5 will show that similar relationships are asymptotically valid for non-Gaussian pseudofamilies.

2.4. Pseudodensity over a Tail Interval

Instead of using the local parameter function and calibrating the pseudodensity locally about a value, one could calibrate the pseudodensity over an interval in the tail. (We thank

an anonymous referee for this suggestion.) More precisely, we could define a pseudo-true parameter value

$$\theta^*(c, f_0) = \underset{\theta}{\text{Argmax}}\, E_0\left\{1_{Y>c}\log\left[\frac{f(y;\theta)}{S(c,\theta)}\right]\right\}, \tag{2.13}$$

where S denotes the survival function, and consider an approximation of the true distribution over a tail interval $f[y;\theta^*(c, f_0)]$, for $y > c$. From a theoretical point of view, this approach can be criticized as it provides different approximations of $f_0(y)$ depending on the selected value of c, $c < y$.

3. From Marginal to Conditional Analysis

Section 2 described the local approach to marginal density estimation. Let us now show the passage from the marginal to conditional density analysis and the application to the conditional VaR.

3.1. General Approach to VaR Computation

The VaR analysis concerns the future return on a given portfolio. Let x_t denote the return on that portfolio at date t. In practice, the prediction of x_t is based on a few summary statistics computed from past observations, such as a lagged portfolio return, realized market volatility, or realized idiosyncratic volatility in a previous period. The application of our method consists in approximating locally the joint density of series $y_t = (y'_{1t}, y'_{2t})'$, whose component y_{1t} is x_t, and component y_{2t} contains the summary statistics, denoted by z_{t-1}. Next, from the marginal density of y_t, that is, the joint density of y_{1t} and y_{2t}, we derive the conditional density of y_{1t} given y_{2t}, and the conditional VaR.

The joint density is approximated locally about c which is a vector of two components, $c = (c'_1, c'_2)'$. The first component c_1 is a tail value of portfolio returns, such as the 5% quantile of the historical distribution of portfolio returns, for example, if the conditional VaR at $\alpha < 5\%$ needs to be found. The second component c_2 is the value of the conditioning set, which is fixed, for example, at the last observed value of the summary statistics in $y_{2t} = z_{t-1}$. Due to the difference in interpretation, the bandwidths for c_1 and c_2 need to be different.

The approach above does not suffer from the curse of dimensionality. Indeed, in practice, y_1 is univariate, and the number of summary statistics is small (often less than 3), while the number of observations is sufficiently large (250 per year) for a daily VaR.

3.2. Gaussian Pseudofamily

When the pseudofamily is Gaussian, the local approximation of the density of y_t is characterized by the local mean and variance-covariance matrix. For $y_t = (y'_{1t}, y'_{2t})'$, these moments are decomposed by blocks as follows:

$$\mu(c) = \begin{pmatrix} \mu_1(c) \\ \mu_2(c) \end{pmatrix}, \qquad \Sigma(c) = \begin{pmatrix} \Sigma_{11}(c) & \Sigma_{12}(c) \\ \Sigma_{21}(c) & \Sigma_{22}(c) \end{pmatrix}. \tag{3.1}$$

The local conditional first and second-order moments are functions of these joint moments:

$$\mu_{1|2}(c) = \mu_1(c) - \Sigma_{12}(c)\Sigma_{22}^{-1}(c)\mu_2(c), \tag{3.2}$$

$$\Sigma_{1|2}(c) = \Sigma_{11}(c) - \Sigma_{12}(c)\Sigma_{22}^{-1}(c)\Sigma_{21}(c). \tag{3.3}$$

When $y_{1t} = x_t$ is univariate, these local conditional moments can be used as inputs in the basic Gaussian VaR formula (1.2).

The method is convenient for practitioners, as it suggests them to keep using the misspecified Gaussian VaR formula. The only modifications are the inputs, which become the local conditional mean and variance in the tail that are easy to calculate given the closed-form expressions given above.

Even though the theoretical approach is nonparametric, its practical implementation is semi-parametric. This is because, once an appropriate location c has been selected, the local pseudodensity estimated at c is used to calculate any VaR in the tail. Therefore, the procedure can be viewed as a model building method, in which the two benchmark loss probabilities are arbitrarily close. As compared with other model building approaches, it allows for choosing a location c with more data-points in its neighborhood than the quantile of interest.

4. Application to Value-at-Risk

The nonparametric feature of our localized approach requires the availability of a sufficient number of observations in a neighborhood of the selected c. This requirement is easily satisfied when high-frequency data are used and an intraday VaR is computed. We first consider an application of this type. It is followed by a Monte-Carlo study, which provides information on the properties of the estimator when the number of observations is about 200, which is the sample size used in practice for computing the daily VaR.

4.1. Comparative Study of Portfolios

We apply the local conditional mean and variance approach to intraday data on financial returns and calculate the intraday Value-at-Risk. The financial motivation for intraday risk analysis is that internal control of the trading desks and portfolio management is carried out continuously by banks, due to the use of algorithmic trading that implements automatic portfolio management, based on high-frequency data. Also, the BCBS in [2, page 3], suggests that a weakness of the current (daily) risk measure is that it is based on the end-of-day positions, and disregards the intraday trading risk. It is known that intraday stock price variation can be often as high as the variation of the market closure prices over 5 to 6 consecutive days.

Our analysis concerns two stocks traded on the Toronto Stock Exchange: the Bank of Montreal (BMO) and the Royal Bank (ROY) from October 1st to October 31, 1998, and all portfolios with nonnegative allocations in these two stocks. This approach under the no-short-sell constraint will suffice to show that allocations of the least risky portfolios differ, depending on the method of VaR computation.

From the tick-by-tick data, we select stock prices at a sampling interval of two minutes, and compute the two minute returns $x_t = (x_{1t}, x_{2t})'$. The data contain a large proportion of

zero price movements, which are not deleted from the sample, because the current portfolio values have to be computed from the most recent trading prices.

The BMO and ROY sample consists of 5220 observations on both returns from October 1 to October 31, 1998. The series have equal means of zero. The standard deviations are 0.0015 and 0.0012 for BMO and ROY, respectively. To detect the presence of fat tails, we calculate the kurtosis, which is 5.98 for BMO and 3.91 for ROY, and total range, which is 0.0207 for BMO and 0.0162 for ROY. The total range is approximately 50 (for BMO) and 20 (for ROY) times greater than the interquartile range, equal to 0.0007 in both samples.

The objective is to compute the VaR for any portfolio that contains these two assets. Therefore, $y_t = (y_{1t}, y_{2t})$ has two components; each of which is a bivariate vector. We are interested in finding a local Gaussian approximation of the conditional distribution of $y_{1t} = x_t$ given $y_{2t} = x_{t-1}$ in a neighborhood of values $c_1 = (c_{11}, c_{12})$ of x_t and $c_2 = (c_{21}, c_{22})$ of x_{t-1} (which does not mean that the conditional distribution itself is Gaussian) . We fix $c_{21} = c_{22} = 0$. Because a zero return is generally due to nontrading, by conditioning on zero past returns, we investigate the occurrence of extreme price variations after a non-trading period. As a significant proportion of returns is equal to zero, we eliminate smoothing with respect to these conditioning values in our application.

The local conditional mean and variance estimators were computed from formulae (3.2)-(3.3) for $c_{11} = 0.00188$ and $c_{12} = 0.00154$, which are the 90% upper percentiles of the sample distribution of each return on the dates preceded by zero returns. The bandwidth for x_t was fixed at $h = 0.001$, proportionally to the difference between the 10% and 1% quantiles. The estimates are

$$\mu_1 = -6.54 \, 10^{-3}, \qquad \mu_2 = -0.48 \, 10^{-3},$$

$$\sigma_{11} = 10.2 \, 10^{-6}, \qquad \sigma_{22} = 1.33 \, 10^{-6}, \qquad \rho = \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} = -0.0034. \tag{4.1}$$

They can be compared to the global conditional moments of the returns, which are the moments computed from the whole sample, $\overline{\mu} = E(x_t \mid x_{t-1} = 0)$, $\overline{\Sigma} = V(x_t \mid x_{t-1} = 0)$. Their estimates are

$$\overline{\mu}_1 = -2.057 \, 10^{-5}, \qquad \overline{\mu}_2 = -1.359 \, 10^{-4},$$

$$\overline{\sigma}_{11} = 2.347 \, 10^{-6}, \qquad \overline{\sigma}_{22} = 1.846 \, 10^{-6}, \qquad \overline{\rho} = \frac{\overline{\sigma}_{12}}{\overline{\sigma}_{11}\overline{\sigma}_{22}} = 0.0976. \tag{4.2}$$

As the conditional distribution of x_t given $x_{t-1} = 0$ has a sharp peak at zero, it comes as no surprise that the global conditional moments estimators based on the whole sample lead to smaller Values-at-Risk than the localized ones. More precisely, for loss probability 5% and a portfolio with allocations $a, 1 - a, 0 \le a \le 1$, in the two assets, the Gaussian VaR is given by

$$\text{VaR}(5\%, a) = -(a, 1-a)\mu + 1.64\big[(a, 1-a)\Sigma(a, 1-a)'\big]^{1/2}, \tag{4.3}$$

and determines the required capital reserve for loss probability 5%. Figure 1 presents the Values-at-Risk computed from the localized and unlocalized conditional moments, for any admissible portfolios of nonnegative allocations. The proportion a invested in the BMO is measured on the horizontal axis.
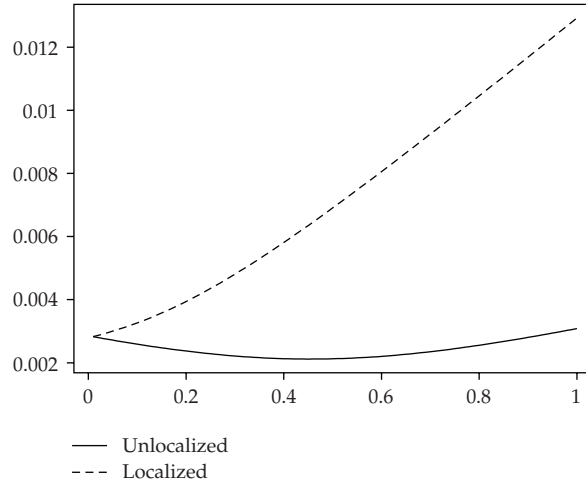
Figure 1: Localized and Unlocalized VaRs.

As expected, the localized VaR lies far above the unlocalized one. This means that the localized VaR implies a larger required capital reserve. We also note that, under the unlocalized VaR, the least risky portfolio contains equal allocations in both assets. In contrast, the localized measure suggests to invest the whole portfolio in a single asset to avoid extreme risks (under the no-short-sell constraint).

4.2. Monte-Carlo Study

The previous application was based on a quite large number of data (more than 5000) on trades in October 1998 and risk level of 5%. It is natural to assess the performance of the new method in comparison to other methods of VaR computation, for smaller samples, such as 200 (resp. 400) observations that correspond to one year (resp., two years) of daily returns and for a smaller risk level of 1%.

A univariate series of 1000 simulated portfolio returns is generated from an ARCH(1) model, with a double exponential (Laplace) error distribution. More precisely, the model is

$$x_t = (0.4 + 0.95x_{t-1}^2)^{1/2} u_t, \tag{4.4}$$

where the errors u_t are i.i.d. with pdf

$$g(u) = \frac{1}{2} \exp(-|u|). \tag{4.5}$$

The error distribution has exponential tails that are slightly heavier than the tails of a Gaussian distribution. The data generating process are assumed to be unknown to the person who estimates the VaR. In practice, that person will apply a method based on a misspecified model (such as the i.i.d. Gaussian model of returns in the Gaussian variance-covariance method or the IGARCH model of squared returns by J. P. Morgan with an ad-hoc fixed

parameter 0.94). Such a procedure leads to either biased, or inefficient estimators of the VaR level.

The following methods of VaR computation at risk level of 1% are compared. Methods 1 to 4 are based on standard routines used in banks, while method 5 is the one proposed in this paper.

(1) The historical simulation based on a rolling window of 200 observations. We will see later (Figure 2) that this approach results in heavy smoothing with respect to time. A larger bandwidth would entail even more smoothing.

(2) The Gaussian variance-covariance approach based on the same window.

(3) The IGARCH-based method by J. P. Morgan:

$$\widehat{\text{VaR}}_t = -\Phi^{-1}(1\%)0.06 \sum_{h=0}^{\infty} (0.94)^h x_{t-h}^2. \tag{4.6}$$

(4) Two conditional ARCH-based procedures that consist of the following steps. First, we consider a subset of observations to estimate an ARCH(1) model:

$$x_t = (a_0 + a_1 x_{t-1}^2)^{1/2} v_t, \tag{4.7}$$

where v_t are i.i.d. with an unknown distribution. First, the parameters a_0 and a_1 are estimated by the quasi-maximum likelihood, and the residuals are computed. From the residuals we infer the empirical 1% quantile $\widehat{q}$, say. The VaR is computed as $\widehat{\text{VaR}}_t = -(\widehat{a}_0 + \widehat{a}_1 x_t^2)^{1/2} \widehat{q}$. We observe that the ARCH parameter estimators are very inaccurate, which is due to the exponential tails of the error distribution. Two subsets of data were used to estimate the ARCH parameters and the 1%-quantile. The estimator values based on a sample of 200 observations are $\widehat{a}_0 = 8.01, \widehat{a}_1 = 0.17,$ and $\widehat{q} = -3.85$. The estimator values based on a sample of 800 observations are $\widehat{a}_0 = 4.12, \widehat{a}_1 = 0.56,$ and $\widehat{q} = -2.78$. We find that the ratios $\widehat{a}_1/\widehat{a}_0$ are quite far from the true value 0.95/0.4 used to generate the data, which is likely due to fat tails.

(5) Localized VaR.

We use a Gaussian pseudofamily, a Gaussian kernel, and two different bandwidths for the current and lagged value of returns, respectively. The bandwidths were set proportional to the difference between the 10% and 1% quantiles, and the bandwidth for the lagged return is 4 times the bandwidth for the current return. Their values are 1.16 and 4.64, respectively. We use a Gaussian kernel (resp., a simple bandwidth) instead of an optimal kernel (resp., an optimal bandwidth) for the sake of robustness. Indeed, an optimal approach may not be sufficiently robust for fixing the required capital. Threshold c is set equal to the 3%-quantile of the marginal empirical distribution. The localized VaR's are computed by rolling with a window of 400 observations.

For each method, Figures 2, 3, 4, 5, 6 and 7 report the evolution of the true VaR corresponding to the data generating model along with the evolution of the estimated VaR. For clarity, only 200 data points are plotted.

The true VaR features persistence and admits extreme values. The rolling methods such as the historical simulation and variance-covariance method produce stepwise patterns of VaR, as already noted, for example, by Hull and White [12]. These patterns result from the i.i.d. assumption that underlies the computations. The J. P. Morgan IGARCH approach creates spurious long memory in the estimated VaR and is not capable to recover the dynamics of the true VaR series. The comparison of the two ARCH-based VaR's shows that the estimated
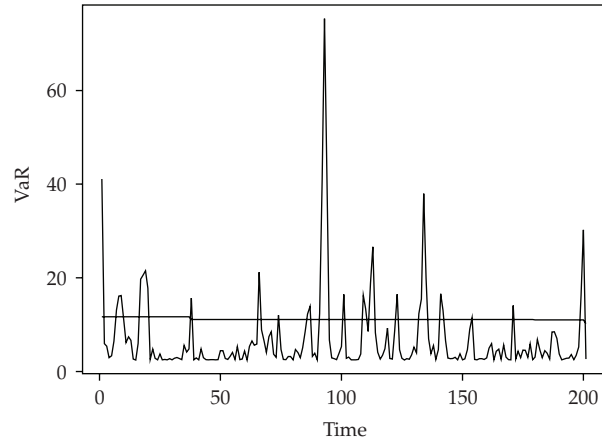
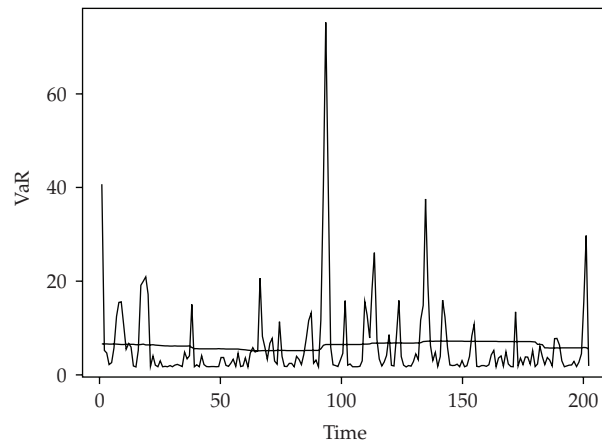**Figure 2:** True and Historical Simulation-Based VaRs.
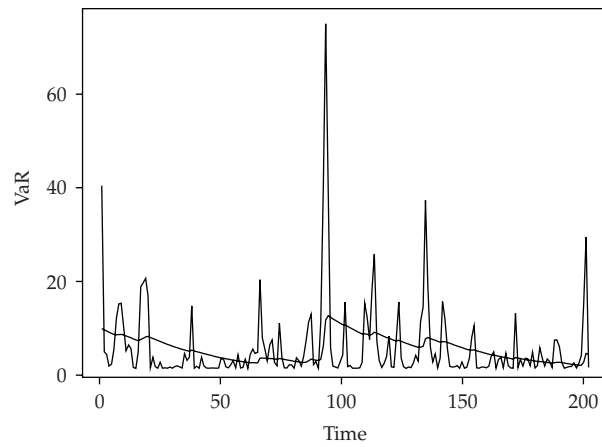

Figure 3: True and Variance-Covariance Based VaRs.

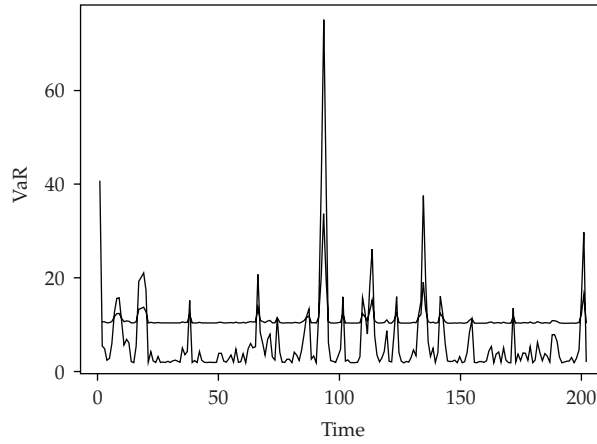**Figure 4:** True and IGARCH-Based VaR.

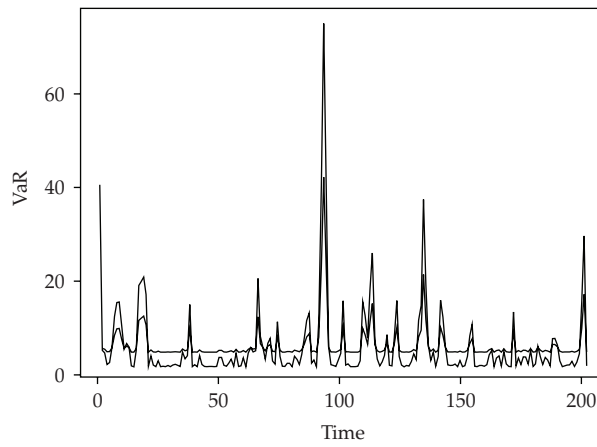**Figure 5:** True and ARCH-Based VaR (200 obs).

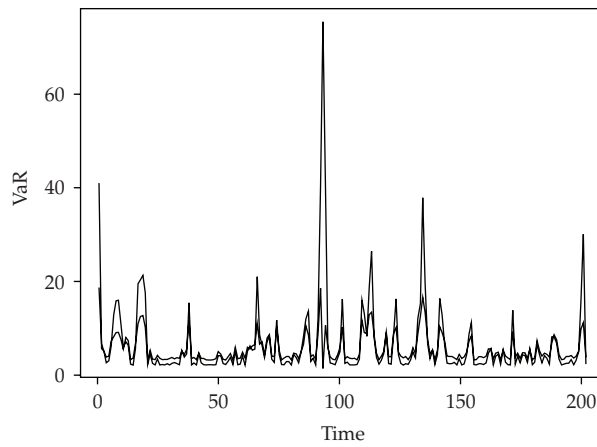**Figure 6:** True and ARCH-Based VaR (800 obs).

Figure 7: True and Localized VaR.

paths strongly depend on the estimated ARCH coefficients. When the estimators are based on 200 observations, we observe excess smoothing. When the estimators are based on 800 observations, the model is able to recover the general pattern, but overestimates the VaR when it is small and underestimates the VaR when it is large. The outcomes of the localized VaR method are similar to the second ARCH model, with a weaker tendency to overestimate the VaR when it is small.

The comparison of the different approaches shows the good mimicking properties of the ARCH-based methods and of the localized VaR. However, these methods need also to be compared with respect to their tractability. It is important to note that the ARCH parameters were estimated only once and were kept fixed for future VaR computations. The approach would become very time consuming if the ARCH model was reestimated at each point in time. In contrast, it is very easy to regularly update the localized VaR.

5. Properties of the Estimator of the Local Parameter Function

5.1. Asymptotic Properties

In this section, we discuss the asymptotic properties of the local pseudomaximum likelihood estimator under the following strict stationarity assumption.

Assumption 5.1. The process $Y = (Y_t)$ is strictly stationary, with marginal pdf f_0.

Let us note that the strict stationarity assumption is compatible with nonlinear dynamics, such as in the ARCH-GARCH models, stochastic volatility models, and so forth, All proofs are gathered in Appendices.

The asymptotic properties of the local P. M. L. estimator of θ are derived along the following lines. First, we find the asymptotic equivalents of the objective function and estimator, that depend only on a limited number of kernel estimators. Next, we derive the properties of the local P. M. L. estimator from the properties of these basic kernel estimators. As the set of assumptions for the existence and asymptotic normality of the basic kernel estimators for multivariates dependent observations can be found in the literature (see the study by Bosq in [13]), we only list in detail the additional assumptions that are necessary to satisfy the asymptotic equivalence. The results are derived under the assumption that θ is exactly identified (see Assumptions 5.2 and 5.3). (In the overidentified case $p > d$, the asymptotic analysis can be performed by considering the terms of order h^3, h^4 in the expansion of the objective function (see Appendix A), which is out of the scope of this paper.)

Let us introduce the additional assumptions.

Assumption 5.2. The parameter set Θ is a compact set and $p = d$.

Assumption 5.3. (i) There exists a unique solution $\tilde{\theta}(c; f_0)$ of the system of equations:

$$\frac{\partial \log f(c; \theta)}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}, \tag{5.1}$$

and this solution belongs to the interior of Θ.

(ii) The matrix $\partial^2 \log f[c, \tilde{\theta}(c, f_0)]/\partial\theta\partial y'$ is nonsingular.

Assumption 5.4. The following kernel estimators are strongly consistent:

(i) $(1/T) \sum_{t=1}^{T} (1/h^d) K((y_t - c)/h) \overset{a.s.}{\rightarrow} f_0(c)$,

(ii) $(1/T) \sum_{t=1}^{T} (1/h^d) K((y_t - c)/h)((y_t - c)/h)((y_t - c)/h)' \overset{a.s.}{\rightarrow} \eta^2 f_0(c)$,

(iii) $(1/Th) \sum_{t=1}^{T} (1/h^d) K((y_t - c)/h)((y_t - c)/h) \overset{a.s.}{\rightarrow} f_0(c) \eta^2 (\partial \log f_0(c)/\partial y)$.

Assumption 5.5. In any neighbourhood of θ, the third-order derivatives $\partial^3 \log f(y; \theta)/ \partial y_i \partial y_j \partial y_k, i, j, k$ varying, are dominated by a function $a(y)$ such that $\|y\|^3 a(y)$ is integrable.

Proposition 5.6. *The local pseudomaximum likelihood estimator $\tilde{\theta}_T(c)$ exists and is strongly consistent for the local parameter function $\tilde{\theta}(c; f_0)$ under Assumptions 5.1–5.5.*

Proof. See Appendix C. □

It is possible to replace the set of Assumptions 5.4 by sufficient assumptions concerning directly the kernel, the true density function f_0, the bandwidth h, and the Y process. In particular it is common to assume that the process Y is geometrically strong mixing, and that $h \rightarrow 0, Th^d/(\log T)^2 \rightarrow +\infty$, when T tends to infinity (see [13–15]).

Proposition 5.7. *Under Assumptions 5.1–5.5 the local pseudomaximum likelihood estimator is asymptotically equivalent to the solution $\tilde{\tilde{\theta}}_T(c)$ of the equation:*

$$\frac{\partial \log f\left[c; \tilde{\tilde{\theta}}_T(c)\right]}{\partial y} = \left(\eta^2\right)^{-1} \frac{1}{h^2} (\tilde{m}_T(c) - c), \tag{5.2}$$

where:

$$\tilde{m}_T(c) = \frac{\sum_{t=1}^{T} K((y_t - c)/h) y_t}{\sum_{t=1}^{T} K((y_t - c)/h)} \tag{5.3}$$

is the Nadaraya-Watson estimator of $m(c) = E(Y \mid Y = c) = c$ based on the kernel K.

Proof. See Appendix D. □

Therefore the asymptotic distribution of $\tilde{\theta}_T(c)$ may be derived from the properties of $\tilde{m}_T(c) - c$, which are the properties of the Nadaraya-Watson estimator in the degenerate case when the regressand and the regressor are identical. Under standard regularity conditions [13], the numerator and denominator of $1/h^2(\tilde{m}_T(c) - c)$ have the following asymptotic properties.

Assumption 5.8. If $T \to \infty$, $h \to 0$, $Th^{d+2} \to \infty$, and $Th^{d+4} \to 0$, we have the limiting distribution

$$
\begin{pmatrix}
\sqrt{Th^{d+2}}\left[\dfrac{1}{Th^{d+2}} \displaystyle\sum_{t=1}^{T} K\left(\dfrac{y_t - c}{h}\right)(y_t - c) - \eta^2 \dfrac{\partial f_0(c)}{\partial y} \right] \\[2em]
\sqrt{Th^{d}}\left[\dfrac{1}{Th^{d}} \displaystyle\sum_{t=1}^{T} K\left(\dfrac{y_t - c}{h}\right) - f_0(c) \right]
\end{pmatrix}
$$

$$
\xrightarrow{d} N\left(0, f_0(c) \begin{bmatrix} \int uu' K^2(u)\,du & \int u K^2(u)\,du \\[1em] \int u' K^2(u)\,du & \int K^2(u)\,du \end{bmatrix} \right). \tag{5.4}
$$

The formulas of the first- and second-order asymptotic moments are easy to verify (see Appendix E). (Assumption 5.8 is implied by sufficient conditions concerning the kernel, the process... (see, [13]). In particular it requires some conditions on the multivariate distribution of the process such as $\sup_{t_1 < t_2} \| f_{t_1,t_2} - f \otimes f \|_\infty < \infty$, where f_{t_1,t_2} denotes the joint p.d.f. of (Y_{t_1}, Y_{t_2}) and $f \otimes f$ the associated product of marginal distributions, and $\sup_{t_1 < t_2 < t_3 < t_4} \| f_{t_1,t_2,t_3,t_4} \|_\infty < \infty$, where f_{t_1,t_2,t_3,t_4} denotes the joint p.d.f of $(Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_{t_4})$.) Note that the rate of convergence of the numerator is slower than the rate of convergence of the denominator since we study a degenerate case, when the Nadaraya-Watson estimator is applied to a regression with the regressor equal to the regressand.

We deduce that the asymptotic distribution of

$$
\sqrt{Th^{d+2}}\left(\frac{1}{h^2}[\tilde{m}_T(c) - c] - \eta^2 \frac{\partial \log f_0(c)}{\partial y} \right) \tag{5.5}
$$

is equal to the asymptotic distribution of

$$
\sqrt{Th^{d+2}}\,\frac{1}{f_0(c)}\left(\frac{1}{Th^{d+2}} \sum_{t=1}^{T} K\left(\frac{y_t - c}{h}\right)(y_t - c) - \eta^2 \frac{\partial f_0(c)}{\partial y} \right), \tag{5.6}
$$

which is $N[0, (1/f_0(c)) \int uu' K^2(u)\,du]$.

By the δ-method we find the asymptotic distribution of the local pseudomaximum likelihood estimator and the asymptotic distribution of the log-derivative of the true p.d.f..

Proposition 5.9. *Under Assumptions 5.1–5.8 one has the following.*

(i)

$$
\sqrt{Th^{d+2}}\left(\frac{\partial \log f\left(c; \tilde{\tilde{\theta}}_T(c)\right)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y} \right) \xrightarrow{d} N\left[0, \frac{(\eta^2)^{-1}}{f_0(c)} \int uu' K^2(u)\,du \left[\eta^2\right]^{-1} \right]. \tag{5.7}
$$

(ii)

$$\sqrt{Th^{d+2}}\left(\tilde{\tilde{\theta}}_T(c) - \tilde{\theta}(c; f_0)\right) \xrightarrow{d} N\left[0, \left[\frac{\partial^2 \log f\left[c; \tilde{\theta}(c; f_o)\right]}{\partial\theta\partial y'}\right]^{-1} \frac{[\eta^2]^{-1}}{f_0(c)}\right.$$

$$\left. \times \int uu' K^2(u) du \left[\eta^2\right]^{-1} \left[\frac{\partial^2 \log f\left[c; \tilde{\theta}(c; f_0)\right]}{\partial y\partial\theta'}\right]^{-1}\right]$$

(5.8)

The first-order asymptotic properties of the estimator of the log-derivative of the density function do not depend on the pseudofamily, whereas the value of the estimator does. (It is beyond the scope of this paper to discuss the effect of the pseudofamily when dimension p is strictly larger than d. Nevertheless, by analogy to the literature on local estimation of nonparametric regression and density functions (see, e.g., the discussion in [7]), we expect that the finite sample bias in the associated estimator of the density will diminish when the pseudofamily is enlarged, that is, when the dimension of the pseudoparameter vector increases.) For a univariate proces (y_t), the functional estimator of the log-derivative $\partial \log f_0(c)/\partial y$ may be compared to the standard estimator

$$\frac{\partial \log \widehat{f}_0(c)}{\partial y} = \frac{\sum_{t=1}^{T}(1/h)\dot{K}((y_t - c)/h)}{\sum_{t=1}^{T}(1/h)K((y_t - c)/h)},$$

(5.9)

where $\dot{K}$ is the derivative of the kernel of the standard estimator. The standard estimator has a rate of convergence equal to that of the estimator introduced in this paper and the following asymptotic distribution:

$$\sqrt{Th^3}\left(\frac{\partial \log \widehat{f}_0(c)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y}\right) \xrightarrow{d} N\left[0, \frac{1}{\eta^4 f_0(c)}\int \dot{K}(u)^2 du\right].$$

(5.10)

The asymptotic distributions of the two estimators of the log-derivative of the density function are in general different, except when $|dK(u)/du| = |uK(u)|$, which, in particular, arises when the kernel is Gaussian. In such a case the asymptotic distributions of the estimators are identical.

5.2. Asymptotic versus Finite Sample Properties

In kernel-based estimation methods, the asymptotic distributions of estimators do not depend on serial dependence and are computed as if the data were i.i.d. However, serial dependence affects the finite sample properties of estimators and the accuracy of the theoretical approximation. Pritsker [16] (see also work by Conley et al. in [17]) illustrates this point by considering the finite sample properties of Ait-Sahalia's test of continuous time model of the short-term interest rate [18] in an application to data generated by the Vasicek model.

The impact of serial correlation depends on the parameter of interest, in particular on whether this parameter characterizes the marginal or the conditional density. This problem is not specific to the kernel-based approaches, but arises also in other methods such as the OLS. To see that, consider a simple autoregressive model $y_t = \rho y_{t-1} + \epsilon_t$, where ϵ_t is $\text{IIN}(0, \sigma^2)$. The expected value of y_t is commonly estimated from the empirical mean $\widehat{m} = \overline{y}_T$ that has asymptotic variance $V\widehat{m} \approx (\eta^2/T)(1/(1-\rho)-1)$, where $\eta^2 = Vy_t = \sigma^2/(1-\rho^2)$. In contrast, the autoregressive coefficient is estimated by $\widehat{\rho} = \sum_t y_t y_{t-1} / \sum_t y_{t-1}^2$ and has asymptotic variance $V\widehat{\rho} \approx (1 - \rho^2)/T$.

If serial dependence is disregarded, both estimators $\widehat{m}$ and $\widehat{\rho}$ have similar asymptotic efficiencies that are η^2/T and $1/T$, respectively. However, when ρ tends to one while η^2 remains fixed, the variance of $\widehat{m}$ tends to infinity whereas the variance of $\widehat{\rho}$ tends to zero. This simple example shows that omission of serial dependence does not have the same effect on the marginal parameters as opposed to the conditional ones. Problems considered by Conley et al. [17] or Pritsker [16] concern the marginal (long run) distribution of y_t, while our application is focused on a conditional parameter, which is the conditional VaR. This parameter is derived from the analysis of the joint pdf $f(y_t, y_{t-1})$ as in the previous example $\widehat{\rho}$ was derived from the bivariate vector $((1/T) \sum_t y_t y_{t-1}, (1/T) \sum_t y_{t-1}^2)$. Due to cointegration between y_t and y_{t-1} in the case of extreme persistence, we can reasonably expect that the estimator of the conditional VaR has good finite sample properties, even when the point estimators $\widehat{f}(y_t, y_{t-1})$ do not. The example shows that in finite sample the properties of the estimator of a conditional parameter can be even better than those derived under the i.i.d. assumption.

6. Conclusions

This paper introduces a local likelihood method of VaR computation for univariate or multivariate data on portfolio returns. Our approach relies on a local approximation of the unknown density of returns by means of a misspecified model. The method allows us to estimate locally the conditional density of returns, and to find the local conditional moments, such as a tail mean and tail variance. For a Gaussian pseudofamily, these tail moments can replace the global moments in the standard Gaussian formula used for computing the VaR's. Therefore, our method based on the Gaussian pseudofamily is convenient for practitioners, as it justifies computing the VaR from the standard Gaussian formula, although with a different input, which accommodates both the thick tails and path dependence of financial returns. The Monte-Carlo experiments indicate that tail-adjusted VaRs are more accurate than other VaR approximations used in the industry.

Appendices

A. Proof of Proposition 2.2

Let us derive the expansion of the objective function

$$
\begin{aligned}
A_h(\theta) = {} & E_0\left[\frac{1}{h^d}K\left(\frac{Y-c}{h}\right)\log f(Y;\theta)\right] \\
& - E_0\left[\frac{1}{h^d}K\left(\frac{Y-c}{h}\right)\right]\log\int\frac{1}{h^d}K\left(\frac{y-c}{h}\right)f(y;\theta)dy,
\end{aligned}
\tag{A.1}
$$

when h approaches zero. By using the equivalence (see Assumption (A.1))

$$\int \frac{1}{h^d} K\left(\frac{y-c}{h}\right) g(y) dy = g(c) + \frac{h^2}{2} \operatorname{Tr}\left[\frac{\partial^2 g(c)}{\partial y \partial y'} \eta^2\right] + o\left(h^2\right), \tag{A.2}$$

where Tr is the trace operator, we find that

$$A_h(\theta) = f_0(c) \log f(c; \theta) + \frac{h^2}{2} \operatorname{Tr}\left\{\frac{\partial^2}{\partial y \partial y'}[f_0(c) \log f(c; \theta)] \eta^2\right\}$$

$$- \left[f_0(c) + \frac{h^2}{2} \operatorname{Tr}\left(\eta^2 \frac{\partial^2 f_0(c)}{\partial y \partial y'}\right)\right] \log\left\{f(c; \theta) + \frac{h^2}{2} \operatorname{Tr}\left[\eta^2 \frac{\partial^2 f(c; \theta)}{\partial y \partial y'}\right]\right\} + o\left(h^2\right)$$

$$= -\frac{h^2}{2} \frac{\partial \log f_0(c)}{\partial y'} \eta^2 \frac{\partial \log f_0(c)}{\partial y}$$

$$+ \frac{h^2}{2}\left[\frac{\partial \log f(c; \theta)}{\partial y'} - \frac{\partial \log f_0(c)}{\partial y'}\right] \eta^2 \left[\frac{\partial \log f(c; \theta)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y}\right] + o\left(h^2\right). \tag{A.3}$$

The result follows.

The expansion above provides a local interpretation of the asymptotic objective function at order h^2 as a distance between the first-order derivatives of the logarithms of the pseudo and true pdf's. In this respect the asymptotic objective function clearly differs from the objective function proposed by Hjort and Jones [7], whose expansion defines an l^2-distance between the true and pseudo pdfs.

B. Proof of Proposition 2.4

For a Gaussian kernel $K(\cdot) = \phi(\cdot)$ of dimension d, we get

$$\int \frac{1}{h^d} \phi\left(\frac{y-c}{h}\right) f(y; \theta) dy = \int \frac{1}{h^d} \phi\left(\frac{c-y}{h}\right) \phi(y - \theta) dy$$

$$= \frac{1}{(1+h^2)^{d/2}} \phi\left(\frac{c-\theta}{\sqrt{1+h^2}}\right). \tag{B.1}$$

We have

$$
\tilde{\theta}_T(c) = \underset{\theta}{\text{Argmax}} \left\{ \sum_{t=1}^{T} \frac{1}{h^d} \phi\left(\frac{y_t - c}{h}\right) \left[-\frac{d}{2} \log 2\pi - \frac{\|y_t - \theta\|^2}{2} \right] \right.
$$

$$
\left. - \sum_{t=1}^{T} \frac{1}{h^d} \phi\left(\frac{y_t - c}{h}\right) \left[-\frac{d}{2} \log 2\pi - \frac{d}{2} \log\left(1 + h^2\right) - \frac{\|c - \theta\|^2}{2(1 + h^2)} \right] \right\}
$$

(B.2)

$$
= \frac{1 + h^2}{h^2} \frac{\sum_{t=1}^{T} (1/h^d) \phi((y_t - c)/h) y_t}{\sum_{t=1}^{T} (1/h^d) \phi((y_t - c)/h)} - \frac{1}{h^2} c
$$

$$
= \frac{1 + h^2}{h^2} \tilde{m}_T(c) - \frac{1}{h^2} c.
$$

Moreover we have

$$
\tilde{\theta}_T(c) - c = \frac{1 + h^2}{h^2} \left(\tilde{m}_T(c) - c\right)
$$

$$
= \frac{1 + h^2}{h^2} \frac{\sum_{t=1}^{T} (1/h^d)(y_t - c)\ \phi((y_t - c)/h)}{\sum_{t=1}^{T} (1/h^d)\ \phi((y_t - c)/h)}
$$

(B.3)

$$
= \left(1 + h^2\right) \frac{(\partial/\partial c)\left[\sum_{t=1}^{T} (1/h^d)\ \phi((y_t - c)/h)\right]}{\left[\sum_{t=1}^{T} (1/h^d)\ \phi((y_t - c)/h)\right]}
$$

$$
= \left(1 + h^2\right) \frac{\partial \log}{\partial c} \tilde{f}_T(c).
$$

C. Consistency

Let us consider the normalized objective function

$$
\tilde{A}_{T,h}(\theta) = \frac{1}{Th^2} \left[\sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \log f(y_t; \theta) \right.
$$

$$
\left. - \sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \log \int \frac{1}{h^d} K\left(\frac{y - c}{h}\right) f(y; \theta) dy \right].
$$

(C.1)

It can be written as

$$
\tilde{A}_{T,h}(\theta) = \frac{1}{Th^2} \sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right)
$$

$$
\times \left[\log f(c;\theta) + \frac{\partial \log f(c;\ \theta)}{\partial y'}(y_t - c) + \frac{1}{2}\left(\frac{y'_t - c}{h}\right)\frac{\partial^2 \log f(c;\ \theta)}{\partial y \partial y'}\left(\frac{y_t - c}{h}\right) \right.
$$

$$
\left. + R_1(y_t;\theta)\|y_t - c\|^3 \right] - \frac{1}{Th^2}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t - c}{h}\right)
$$

$$
\times \left[\log f(c;\theta) + \frac{h^2}{2}\frac{1}{f(c;\theta)}\operatorname{Tr}\left[\eta^2 \frac{\partial^2 f(c;\theta)}{\partial y \partial y'}\right] + h^3 R_2(\theta, h) \right]
$$

(C.2)

where $R_1(y_t;\ \theta)$, $R_2(\theta; h)$ are the residual terms in the expansion. We deduce:

$$
\tilde{A}_{T,h}(\theta) = \frac{1}{Th^2}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t - c}{h}\right)\frac{\partial \log f(c;\theta)}{\partial y'}(y_t - c)
$$

$$
+ \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t - c}{h}\right)\frac{1}{2}\left(\frac{y_t - c}{h}\right)'\frac{\partial^2 \log f(c;\theta)}{\partial y \partial y'}\left(\frac{y_t - c}{h}\right)
$$

$$
- \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t - c}{h}\right)\frac{1}{f(c;\theta)}\operatorname{Tr}\left[\eta^2 \frac{\partial^2 f(c;\theta)}{\partial y \partial y'}\right] + \text{residual terms.}
$$

(C.3)

Under the assumptions of Proposition 5.7, the residual terms tend almost surely to zero, uniformly on Θ, while the main terms tend almost surely uniformly on Θ to

$$
A_\infty = -\frac{\partial \log f_0(c)}{\partial y'}\eta^2 \frac{\partial \log f_0(c)}{\partial y}
$$

$$
+ \frac{1}{2}\left[\frac{\partial \log f(c;\theta)}{\partial y'} - \frac{\partial \log f_0(c)}{\partial y'}\right]\eta^2\left[\frac{\partial \log f(c;\theta)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y}\right],
$$

(C.4)

which is identical to $\lim_{h\to o}(A_h(\theta)/h^2)$ (see Appendix A).

Then, by Jennrich theorem [19] and the identifiability condition, we conclude that the estimator $\tilde{\theta}_T(c)$ exists and is strongly consistent of $\theta(c; f_0)$.

D. Asymptotic Equivalence

The main part of the objective function may also be written as

$$
\tilde{A}_{T,h}(\theta) \approx \frac{1}{Th^2} \sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \frac{\partial \log f(c;\theta)}{\partial y'} (y_t - c)
$$
$$
- \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \frac{\partial \log f(c;\theta)}{\partial y'} \eta^2 \frac{\partial \log f(c;\theta)}{\partial y}.
$$

$$(D.1)$$

We deduce that the local parameter function can be asymptotically replaced by the solution $\tilde{\tilde{\theta}}_T(c)$ of

$$
\frac{\partial \log f\left(c;\tilde{\tilde{\theta}}_T(c)\right)}{\partial y} = \left(\eta^2\right)^{-1} \frac{\left(1/Th^2\right) \sum_{t=1}^{T} (1/h^d) K((y_t - c)/h)(y_t - c)}{(1/T) \sum_{t=1}^{T} (1/h^d) K((y_t - c)/h)}.
$$

$$(D.2)$$

E. The First- and Second-Order Asymptotic Moments

Let us restrict the analysis to the numerator term $(1/Th^{d+2}) \sum_{t=1}^{T} K((Y_t - c)/h)(Y_t - c)$, which implies the nonstandard rate of convergence.

(1) First-Order Moment

We get

$$
E\left[\frac{1}{Th^{d+2}} \sum_{t=1}^{T} K\left(\frac{Y_t - c}{h}\right)(Y_t - c)\right] = \frac{1}{h^{d+2}} E\left[K\left(\frac{Y_t - c}{h}\right)(Y_t - c)\right]
$$
$$
= \frac{1}{h^{d+2}} \int K\left(\frac{y - c}{h}\right)(y - c) f_0(y) dy
$$
$$
= \frac{1}{h} \int K(u) \, u \, f_0(c + uh) du
$$
$$
= \frac{1}{h} \int K(u) u \left[f_0(c) + h \frac{\partial f_0(c)}{\partial y'} u + \frac{h^2}{2} u' \frac{\partial^2 f_0(c)}{\partial y \partial y'} u + o\left(h^2\right)\right] du
$$
$$
= \eta^2 \frac{\partial f_0(c)}{\partial y} + \frac{h}{2} \int K(u) u u' \frac{\partial^2 f_0(c)}{\partial y \partial y'} u du + o(h).
$$

$$(E.1)$$

(2) Asymptotic Variance

We have

$$
V\left[\frac{1}{Th^{d+2}}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right] = \frac{1}{Th^{2d+4}}V\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]
$$

$$
= \frac{1}{Th^{2d+4}}\left\{E\left[K^2\left(\frac{Y_t-c}{h}\right)(Y_t-c)(Y_t-c)'\right]\right.
$$

$$
\left. - E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]'\right]
$$

$$
= \frac{1}{Th^{d+2}}\left[\int K^2(u)uu'f_0(c+uh)du - h^{d+2}\eta^2\frac{\partial f_0(c)}{\partial y}\frac{\partial f_0(c)}{\partial y'}\eta^2\right]
$$

$$
= \frac{1}{Th^{d+2}}f_0(c)\int uu'K^2(u)du + o\left(\frac{1}{Th^{d+2}}\right),
$$

$$(E.2)$$

which provides the rate of convergence $(Th^{d+2})^{-1/2}$ of the standard error. Moreover the second term of the bias will be negligible if $h(Th^{d+2})^{1/2} \to 0$ or $Th^{d+4} \to 0$.

(3) Asymptotic Covariance

Finally we have also to consider:

$$
\mathrm{Cov}\left[\frac{1}{Th^{d+2}}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)(Y_t-c), \frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)\right]
$$

$$
= \frac{1}{Th^{d+3}}\mathrm{Cov}\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c), K\left(\frac{Y_t-c}{h}\right)\right]
$$

$$
= \frac{1}{Th^{d+3}}\left\{E\left[K^2\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right] - E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]EK\left(\frac{Y_t-c}{h}\right)\right\} \quad (E.3)
$$

$$
= \frac{1}{Th^{d+3}}\left\{h^2\int K^2(u)uf_0(c+uh)du + O\left(h^4\right)\right\}
$$

$$
= \frac{1}{Th^{d+1}}f_0(c)\int uK^2(u)du + o\left(\frac{1}{Th^{d+1}}\right).
$$

Acknowledgment

References

[1] Basle Committee on Banking Supervision, *An Internal Model-Based Approach to Market Risk Capital Requirements*, Basle, Switzerland, 1995.

[2] Basle Committee on Banking Supervision, *Overview of the Amendment to the Capital Accord to Incorporate Market Risk*, Basle, Switzerland, 1996.

[3] C. Gourieroux and J. Jasiak, "Value-at-risk," in *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. P. Hansen, Eds., pp. 553–609, Elsevier, Amsterdam, The Netherlands, 2009.

[4] J. P. Morgan, *RiskMetrics Technical Manual*, J.P. Morgan Bank, New York, NY, USA, 1995.

[5] R. F. Engle and S. Manganelli, "CAViaR: conditional autoregressive value at risk by regression quantiles," *Journal of Business & Economic Statistics*, vol. 22, no. 4, pp. 367–381, 2004.

[6] C. Gourieroux and J. Jasiak, "Dynamic quantile models," *Journal of Econometrics*, vol. 147, no. 1, pp. 198–205, 2008.

[7] N. L. Hjort and M. C. Jones, "Locally parametric nonparametric density estimation," *The Annals of Statistics*, vol. 24, no. 4, pp. 1619–1647, 1996.

[8] C. R. Loader, "Local likelihood density estimation," *The Annals of Statistics*, vol. 24, no. 4, pp. 1602–1618, 1996.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, UK, 1986.

[10] E. Nadaraya, "On estimating regression," *Theory of Probability and Its Applications*, vol. 10, pp. 186–190, 1964.

[11] G. S. Watson, "Smooth regression analysis," *Sankhyā. Series A*, vol. 26, pp. 359–372, 1964.

[12] J. Hull and A. White, "Incorporating volatility updating into the historical simulation for VaR," *The Journal of Risk*, vol. 1, pp. 5–19, 1998.

[13] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, vol. 110 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1996.

[14] G. G. Roussas, "Nonparametric estimation in mixing sequences of random variables," *Journal of Statistical Planning and Inference*, vol. 18, no. 2, pp. 135–149, 1988.

[15] G. G. Roussas, "Asymptotic normality of the kernel estimate under dependence conditions: application to hazard rate," *Journal of Statistical Planning and Inference*, vol. 25, no. 1, pp. 81–104, 1990.

[16] M. Pritsker, "Nonparametric density estimation and tests of continuous time interest rate models," *Review of Financial Studies*, vol. 11, no. 3, pp. 449–487, 1998.

[17] T. G. Conley, L. P. Hansen, and W.-F. Liu, "Bootstrapping the long run," *Macroeconomic Dynamics*, vol. 1, no. 2, pp. 279–311, 1997.

[18] Y. Aït-Sahalia, "Testing continuous-time models of the spot interest rate," *Review of Financial Studies*, vol. 9, no. 2, pp. 385–426, 1996.

[19] R. I. Jennrich, "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, vol. 40, pp. 633–643, 1969.