Massachusetts Institute of Technology
Department of Economics
Working Paper Series

CONDITIONAL VALUE-AT-RISK: ASPECTS OF MODELING AND ESTIMATION

Victor Chernozhukov, MIT
Len Umantsev, Stanford

Working Paper 01-19
November 2000

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

CONDITIONAL VALUE-AT-RISK: ASPECTS OF MODELING AND ESTIMATION

Victor Chernozhukov, MIT
Len Umantsev, Stanford

Working Paper 01-19
November 2000

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Empirical Economics manuscript No.
(will be inserted by the editor)

Conditional Value-at-Risk: Aspects of Modeling and Estimation

Victor Chernozhukov[1], **Len Umantsev**[2*]

[1] Department of Economics
MIT
Cambridge, MA 02139
vchern@leland.stanford.edu

[2] Department of Management Science and Engineering
Stanford University
Stanford, CA 94305-4026
uman@stanford.edu

Received: September 30, 1999 / Revised version: November 20, 2000

Abstract This paper considers flexible conditional (regression) measures of market risk. Value-at-Risk modeling is cast in terms of the quantile regression function – the inverse of the conditional distribution function. A basic specification analysis relates its functional forms to the benchmark models of returns and asset pricing. We stress important aspects of measuring the extremal and intermediate conditional risk. An empirical application characterizes the key economic determinants of various levels of conditional risk.

* We thank Takeshi Amemiya, Herman Bierens, Emily Gallagher, Roger Koenker, Mary Ann Lawrence, Tom MaCurdy, Warren Huang, an anonymous referee, and participants of seminars at Stanford University, University of Mannheim, Midwest Finance Association Risk Session, CIRANO, International Conference on Economic Applications of Quantile Regression for many useful conversations and/or comments. Very special thanks to Bernd Fitzenberger who provided extremely useful comments and corrections as an editor.

1 Introduction and Conclusion

Value-at-Risk (hereafter, VaR) is a most widely used measure of market risk, employed in the financial industry for both the internal control and regulatory reporting. We explore various aspects of VaR modeling based on the median/ quantile regressions (cf. Hogg (1975), Bassett and Koenker (1978), Koenker and Bassett (1978)):

- Conditional VaR modeling is cast in terms of the regression quantile function – the inverse of the conditional distribution function. We relate the functional forms of conditional quantiles to the basic statistical models of returns and the models of asset pricing and arbitrage. The conditional quantile models are semi-parametric in nature and are considerably more flexible than most commonly used parametric methods (section 2).
- The key econometric aspects of conditional VaR measurement are displayed (section 3). In particular, we address estimation, inference, and specification analysis of high and intermediate conditional risk. A fundamental problem in measuring high conditional risk is the lack of data on high risk events, which requires the considerations of extreme value theory (see Chernozhukov (1999a) for a theoretical account).
- An extensive empirical section exposes the methods. We estimate and analyze the conditional market risk of an oil producer stock price as a function of the key economic variables. We find that these variables impact various quantiles of the return distribution in a very differential and nontrivial manner. The key determinants of the extremal and intermediate conditional risk are characterized. The market index (DJI) is found to be the only statistically significant determinant of the extremal risk. The other key variables may also exhibit very large effect. However, the direction of the effect can not be isolated due to the scarcity of data on high risk events.

We hope our views are useful to the reader. We also recommend other works that consider regression quantile modeling in value-at-risk and related problems in finance: Bassett and Chen (1999), Engle and Manganelli (1999), Heiler and Abberger (1999), Taylor (1999), among others.

2 Modeling Risk Conditionally

In this section we discuss modeling VaR and related Market Risk measures (MRMs) via conditional quantiles and other techniques.

2.1 Preliminaries

The setting we consider is as follows:

- Y_t is the price process of the security or portfolio of securities.
- X_t is taken to be a "state process" or "information" vector. In practical applications, X_t usually consists of prices (or returns) of securities, market indexes, interest rates, spreads, yields, and the like. We may allow X_t to grow with the length of the sampling period. Lagged values of Y_t and functions of such lagged values (e.g. exponentially-weighted sample volatility) may or may not be included in X_t.[1]

The *Return* of a portfolio with price process Y_t over $[t, t+h)$ is

$$y_t^h = \ln Y_{t+h} - \ln Y_t.$$

The *Conditional Value at Risk (VaR)*, $v_t^h(p)$, is defined as a level of return y_t^h over the period of $[t, t+h)$ that is exceeded with probability p $(p \in (0,1))$: [2]

$$v_t^h(p) \equiv \inf_v \left\{ v : \mathbb{P}_t(y_t^h \leq v) \geq 1-p \right\}.$$

Alternatively, this can be written in terms of the conditional distribution function of y_t^h:

$$v_t^h(p) = F_{y_t^h}^{-1}(1-p|X_t),$$

where $F_{y_t^h}^{-1}(\cdot|X_t)$ is an inverse of the cdf $F_{y_t^h}(\cdot|X_t)$ or the so-called *conditional quantile function*. Let us call p and $\tau = 1-p$ – the *confidence level* and the *index* of VaR, respectively. The *Conditional VaR curve* is the conditional VaR viewed as a function of the confidence level.

Similarly, the *Extreme* VaR may be defined as the *maximum possible* loss over a period of time. In this regard, the extreme VaR may be introduced as the limit form of the non-extreme VaR for confidence levels p approaching 1:

$$v_t^h(1) \equiv \lim_{p \nearrow 1} F_{y_t^h}^{-1}(1-p|X_t) = \inf \left\{ v : \mathbb{P}_t(y_t^h \leq v) > 0 \right\}.$$

Correspondingly, *extremal* VaR are VaR measures with p close to 1.

[1] Using the standard notation, time subscript under the expectation $\mathbb{E}[\cdot]$, probability $\mathbb{P}(\cdot)$, df $F(\cdot)$, or density $f(\cdot)$ denotes conditioning on the information set X_t.

[2] This definition allows to avoid ambiguity when the distribution of y_t^h is atomic. Otherwise, the definition is equivalent to $v_t^h(p) = \{v : \mathbb{P}_t(y_t^h \leq v) = 1-p\}$.

As noted, the superscript h specifies the length of the time interval. (We drop the superscript in the sequel to simplify notation.) In the practical applications, h is typically chosen to be 2 weeks for the regulatory reporting [3] (which typically translates to 10 business days for local applications or 12 business days for around-the-globe trading applications, such as Forex desks). 1-day intervals are used for the quality assessment of banks' VaR models by regulators. It is also commonly used for the internal risk management, although other values of h (up to one month) are not uncommon. There are two primary reasons why we are not interested in measuring market risk for time periods longer than one month: (1) market risk events typically happen during short intervals, and (2) losses due to market risk events can be relatively easily restored in short periods of time (reduction of balance sheet, refreshment of capital, etc.), especially if the market risk event is not accompanied by the liquidity or systemic risk events. This contrasts with the measures of the credit risk, which have to be estimated over the instruments' lifetime (which could be as long as 10 to 30 years).

2.2 Market Risk Measures and Conditioning

MRMs, as statistical estimators, can be parametric, semi-parametric, and non-parametric, depending on the strength of identifiability assumptions made by the methods.[4] The most commonly used parametric MRM is the variance-covariance method that assumes (conditionally) normal returns (implemented, for example, in J.P. Morgan's RiskMetrics). The conditional quantile methods with parametric and non-parametric functional forms fall into the semi-parametric and the non-parametric classes, respectively.

---

[3] See FedReg (1996) for the review of the 1996 Risk Amendment to the Basle Capital Accord of 1988.

[4] Another fundamentally different method is the stress-testing, in which one defines "basis of probable events" (such as various parallel shifts of the Treasury yield curve, changes of the yield curve slope, and others used in the DPG guidelines), "highly unlikely events" (such as a drop by 25% in the S&P500) and "structurally impossible events" (such as a drop by 25% in the S&P500 accompanied by a large increase in DJI) and examines the behavior of the portfolio under various such events or shocks. MRMs of the last type may offer valuable insight about the market-risk properties of a security or a portfolio. The methods are less statistical and more experimentation-based in their nature, yet they could be usefully combined with statistical methods, especially when data is not informative about the extremal events.

The type of conditioning is another vital dimension along which MRMs can be classified. The following two types are not mutually exclusive, yet the proposed distinction will be useful.

(a) Moving Window Sampling – Regime Conditioning

Many of the historical and parametric MRMs are based on the *samples* of certain (often moving) width (e.g. a window of the last 100 returns). The sample is then *weighted* or *not weighted* to compute the parameter estimates. For example, the "historical volatility" approaches[5] use geometric weighting. In this case, such measures may be seen as GARCH models that have recursive conditional variances.[6] Thus such methods could be interpreted as the regression measures described next.

Oftentimes, however, no weighting is done and no interpretation of the above kind is provided. What is achieved by such a form of conditioning? A primary goal is to robustify the MRM against the structural changes or regime switches. That is, the whole history may not be an appropriate sample for measurement/estimation due to structural changes that have occurred in the past. Indeed, the dynamics of oil prices in the 70s is probably very different from that now. Therefore, considering the equally weighted sample of moving width is, most of all, a method for disregarding un- or dis- informative data. Thus, such procedures could be seen primarily as methods of conditioning on the environment and as a way to guard against misspecification w.r.t. a given historical period.

(b) Regression – Conditioning on the information X_t

In addition to considering the informative data, a risk-modeler seeks to produce a measure of risk conditionally on the set of the key economic ("state") variables X_t. For example, one may wish to characterize volatility of the oil return as a function of the oil spot price, key exchange rates, etc. Such a form of conditioning is a *regression characterization* of risk. Regression seeks to describe the moments of return (generally, distribution or quantiles) as functions of X_t. The sample analogues of regression dependencies are regression estimators. Examples of MRM of this kind include frequently used parametric methods, such as the normal GARCH, and the class of quantile regression MRMs treated in this paper. Indeed, GARCH represents volatility as a function of X_t, the regression quantile

[5] Implemented in RiskMetrics.

[6] Standard GARCH models also assign geometric weights to the squares of past innovations, and the "moving window" results from truncating the sum once the weights are sufficiently small.

method describes the quantiles of return distribution of $y_t^h|X_t$. Note that most non-parametric or "historical" methods, such as unconditional quantile or moment estimation are not regression methods in that the regression dependence is simply not modeled. An example of a non-parametric method that gives a regression measure is a non-parametric estimator of the conditional density function (from which a conditional VaR estimate can be computed – see Ait-Sahalia and Lo (1998)).[7]

2.3 Quantile Regression

QR flexibly allows us to directly model the conditional VaR, utilizing only the pertinent information that determines quantiles of interest. This is in sharp contrast with the traditional methods that use information on the central moments of conditional distribution – mean, variance, kurtosis, etc. – to construct the VaR estimates. This primary feature of QR is especially important for modeling intermediate and extremal conditional VaR.

Envisioning the essence of the conditional quantile modeling is quite easy. Fix any p, a confidence level. Assume some functional dependence of VaR on the information variables X_t:

$$v_t^h(p) = F_{y_t^h}^{-1}(\tau|X_t) = m_p(X_t, \beta(p)).$$

For any p (or a set of p's) we could suitably pick one or the other form of m_p to match the observed historical data sufficiently well through a set of sample moment restrictions.[8] This matching is implemented through the means of quantile regression and related methods (see section 4). In this paper we exclusively confine our attention to the linear (polynomial) analysis.

3 A Basic Specification Analysis

Linear (Polynomial) Models

[7] To name a few disadvantages of the fully non-parametric approach: (1) computational, and (2) due to model complexity, it often entails a significant loss of predictive ability in moderate-sized samples and cases of many conditioning variables.

[8] In fact, the choice of m_p could be arbitrarily flexible - non-parametric. For example, m_p can be modeled through a composition of basis functions of a predetermined class.

A fundamental model is the linear model of conditional quantile function (VaR):

$$v_t(p) \equiv F_{y_t}^{-1}(\tau|X_t) = \widetilde{X}_t' \beta(p), \tag{3.1}$$

where $\widetilde{X}_t$ is a d-dimensional vector representing any desired transforms (powers, etc.) of X_t, including a constant. We next discuss both the plausibility and restrictiveness of such model.

(a) Location-Scale Models

Linear model (3.1) naturally arises, for example, from an AR formulation of the return dynamics. Indeed, consider a simple linear location-scale model:

$$\begin{gathered} y_t = \widetilde{X}_t'\alpha + \widetilde{X}_t'\lambda u_t, \\ u_t \text{ is independent of } X_t \ , \quad \mathbb{P}(u_t < 0) = 1/2. \end{gathered} \tag{3.2}$$

Model (3.2) is a location-scale model, where both location $\widetilde{X}_t'\alpha$ and scale $\widetilde{X}_t'\lambda > 0$ are parameterized as linear functions. Furthermore, location $\widetilde{X}_t'\alpha$ is the conditional median function. We should also impose scale restrictions to identify λ, e.g. $E|u_t| = 1$ or, alternatively, $\|\lambda\| = 1$. Since models like (3.2) are not used in sequel, we omit any further discussion of identification.

We can define the "shock" term u_t in a (perhaps) more familiar way:

$$\begin{gathered} y_t = \widetilde{X}_t'\alpha + \widetilde{X}_t'\lambda u_t, \\ u_t \text{ is independent of } X_t \ , \quad \mathbb{E}(u_t) = 0, \mathbb{E}\left(u_t^2\right) = 1. \end{gathered} \tag{3.3}$$

In this case, location $\widetilde{X}_t'\alpha$ is the conditional mean function, and $\widetilde{X}_t'\lambda > 0$ is the scale ($(\widetilde{X}_t'\lambda)^2$ is the conditional variance). Note that models such as (3.3) are directly justifiable from the standard APT and other factor models (see Campbell, MacKinlay, and Lo (1997)).

It is very plausible that either model generates a linear conditional quantile model. Denote by F_u the distribution function of u_t. Then clearly $v_t(p) = \widetilde{X}_t'\alpha + \widetilde{X}_t'\lambda F_u{}^{-1}(\tau) = \widetilde{X}_t'\beta(p)$.

(b) Non-Location-Scale Models

While a linear location-scale model implies a linear VaR/quantile function, the converse may not be true. Indeed, it is easy to see that the location-scale model necessarily involves monotone coefficients $\beta(p)$ in the quantile index p, whereas (3.1) imposes no such assumption. In either model (3.2) or (3.3), it is assumed that u_t is independent of X_t. In general u_t will be not independent of X_t, and all such cases form the class of non-location-scale models. It is not difficult to see how these cases can generate the linear/polynomial forms.

Polynomial Models as Approximations of Non-linear Models

More generally, the linear/polynomial specifications can be considered as approximations of nonlinear models of the form $y_t = \mu(X_t) + \sigma(X_t)u_t$. Since it is clear how this approximation works, we omit any further discussion.

Recursive Specifications

Models that we have considered so far represent VaR as a function of only a few key economic variables and their transforms, $\widetilde{X}_t$. It may be desirable to consider models that reflect the whole past information X_t. Nonlinear dynamic models allow for a wide variety of such specifications. For example, let $\dot{X}_t$ be the subset of information variables (or their transforms) that become available in the t-th period and X_{t-1} be variables $\{\dot{X}_j\}_{j=0}^{t-1}$, so that $X_t = (\dot{X}_t, X_{t-1})$. A general recursive specification can then be obtained as follows:

$$\begin{aligned} v_t(p) &= \dot{X}_t' a(p) + b(p) f_1(X_t, p) \\ f_1(X_t, p) &= c(p) f_1(X_{t-1}, p) + f_2(\dot{X}_t, d(p)) \end{aligned} \tag{3.4}$$

Linear forms in (3.4) can be replaced by nonlinear forms. Restrictions, guaranteeing stability of the model, must be imposed on a, b, c, f_1, f_2. Models of this sort were considered in Weiss (1991), Koenker and Zhao (1996), Engle and Manganelli (1999). Models like (3.4) are useful as parsimonious regressions that represent value-at-risk/quantiles as a function of all past information. In contrast, the simpler markovian specifications posit that most of relevant information is contained in a few past lags of the key variables.

The model (3.4) can typically be solved recursively to eliminate f_1, yielding nonlinear dynamic functional forms, which can be used in a usual nonlinear estimation. One example involves the model

$$f_1(X_t, p) = \sigma(Y_t - \mu_t | X_t), \tag{3.5}$$

where $\sigma^2(y_t - \mu_t|X_t)$ is the conditional variance of the de-meaned return. It can, for example, take the form of a GARCH model (see Koenker and Zhao (1996) for discussion). In practice, a simple strategy to accommodate a model like (3.4) -(3.5) into a linear framework is to first estimate $\sigma(y_t - \mu_t|X_t)$ via a GARCH model and use it as a regressor in the earlier linear model: $\widetilde{X}_t = (\dot{X}_t', \hat{\sigma}(y_t - \mu_t|X_t))'$. Of course, that the regressor is estimated can be accommodated in the inference analysis or, alternatively, can be ignored for all practical purposes. In the empirical section we will not make use of the recursivity, since we are more interested in the economic determinants of risk, but most of techniques discussed next apply both to the linear and nonlinear specifications.

In summary, it is important to stress that the conditional quantile model imposes *no* strong assumptions about the distribution function of the underlying error term. This underscores the *semi-parametric, flexible nature* of the conditional quantile models, which could be valuable to model the market risk.

4 Estimation/Inference

This section reviews some recent results on estimation and inference. Section 4.1 reviews regular asymptotic estimation and inference in the quantile regression literature. Section 4.2 reviews results of Chernozhukov (1999a) on estimating high and low (extremal) regression quantiles.

4.1 Estimation and Inference

The Conditional VaR function, $v_t(p)$, is parameterized as $m(X_t, \beta(\tau), \tau)$, which, in our empirical setting, takes the linear-quadratic form $\widetilde{X}_t'\beta(\tau)$. Hence we will discuss this case only (to keep notation simple). The nonlinear case is similar – just replace $\widetilde{X}_t$ by the derivative $\partial m(X_t, \beta, \tau)/\partial\beta|_{\beta(\tau)}$, and $\widetilde{X}_t'\beta$ by $m(X_t, \beta, \tau)$.

The sample analogue of moment conditions that define the quantile function, integrated with respect to $\beta(\tau)$ yields the quantile regression objective function Q_T (Koenker and Bassett, 1978). $\hat{\beta}(\tau)$ is defined as the argmin of Q_T:

$$\hat{\beta}(\tau) = \text{argmin}_{\,\beta}\left[Q_T(\beta, \tau) = \sum_t \rho_\tau(y_t - \widetilde{X}_t\beta)\right], \tag{4.6}$$

where $\rho_\tau(x) = \tau x^- + (1-\tau)x^+$. For a given τ, $\sqrt{T}(\hat{\beta}(\tau) - \beta(\tau))$ is asymptotically normal under general dependence and heterogeneity,[9] and, furthermore, the regression VaR coefficients converge to a Gaussian process $\mathbf{G}(\cdot)$, [10] as *functions* of τ[11]

$$\sqrt{T}\left(\hat{\beta}(\cdot) - \beta(\cdot)\right) \Rightarrow \mathbf{G}(\cdot). \tag{4.7}$$

[9] See e.g. Portnoy (1991), Fitzenberger (1998), Weiss (1991) (nonlinear cases), Koenker and Zhao (1996).

[10] See Portnoy (1991) and also the results in Chernozhukov (1999b) that allow for non-linear specifications and various forms of dependent data.

[11] Here $\Rightarrow$ denotes weak convergence in ℓ^∞ - see e.g. van der Vaart and Wellner (1996). $\mathbf{G}(\cdot) = J^{-1}(1-\cdot)\mathbb{G}_P f(W, \cdot)$, $f(W,\tau) = (1(Y \le \widetilde{X}'\beta(\tau)) - \tau)\,\widetilde{X}$, and $W = (X, Y)$. $\mathbb{G}_P(f(W,\tau))$ is zero-mean random Gaussian function of τ $(= 1-p)$,

Inference is facilitated by estimating appropriate variance-covariance matrices by a moving-block bootstrap (Politis and Romano (1994), Fitzenberger (1998)).

Test Processes and their Functionals

Testing is discussed in detail in e.g. Koenker and Portnoy (1999) and Weiss (1991). Koenker and Machado (1999) and also Chernozhukov (1999b) study several forms of tests and test processes (test statistics viewed as functions of τ). These test processes are variants of Wald, Score, quasi-LR, and specification test statistics, viewed as functions of τ in an interval $\mathcal{P}$ (e.g. $\mathcal{P} = [.1, .9]$). The quasi-LR, Wald, and rank-score test processes are introduced and studied in Koenker and Machado (1999) in the context of independent data and linear functional forms. These test processes, as well as some of their alternatives, are studied in Chernozhukov (1999b), under conditions of dependence and nonlinearities. Test processes are shown to be asymptotically distributed as quadratic forms of Gaussian processes (P-Bessel Processes), and each coordinate of the test statistics is asymptotically distributed as chi-squared. Bootstrap inference is also discussed there. The specification test process is introduced Chernozhukov (1999b). The coordinates $Sc(\tau)$ of the specification test process $Sc(\cdot)$ are defined as quadratic forms of the usual specification test statistic (such as gmm-like overidentification statistics or statistics like those in Bierens and Ginther (1999)). A simple, practical version defines $Sc(\tau)$ as a quadratic form of $\frac{1}{T}\sum_{t=1}^{T}\left[(1(y_t \le \widetilde{X}_t'\widehat{\beta}(\tau)) - \tau)Z_t\right]$, where Z_t's are functions of $\widetilde{X}_t$ or other information variables (other than $\widetilde{X}_t$). The specification test process converges weakly to a generalized P-Bessel process. By selecting various forms of Z_t one can check:

- *conditionality* – ability to incorporate all relevant past information (with Z_t equal to past information variables)
- *functional form validity* (with Z_t equal to various transforms of X_t).

---

exactly a P-Brownian Bridge –cf. van der Vaart and Wellner (1996), whose distribution is defined by the finite-dimensional normal distributions and the covariance kernel $\mathbb{CV}(f(W,\tau_i), f(W,\tau_j)) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(f(W_t,\tau_i)\,f(W_t,\tau_j)')$ $+ \sum_{k=1}^{T}\mathbb{E}\left[f(W_{t+k},\tau_i)f(W_t,\tau_j)' + f(W_t,\tau_i)f(W_{t+k},\tau_j)'\right]$; and finally, $J(\tau)$ is a fixed non-stochastic invertible matrix $J(\tau) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}$ $\left[f_{y_t}(F_{y_t}^{-1}(\tau|X_t)|X_t)\widetilde{X}_t\widetilde{X}_t'\right]$. Note that (4.7) is very succinct statement, conveniently characterizing the distribution. For example, $(\sqrt{T}(\hat{\beta}(\tau_i) - \beta(\tau_i), j = 1, ...l)$ converges in distribution to $N(0, A)$, where ij-th block $A_{ij} = A_{ji}'$ of matrix A is given by $J^{-1}(\tau_i)\mathbb{CV}(f(W,\tau_i), f(W,\tau_j))J^{-1}(\tau_j)'$.

Various functionals of these processes form test statistics that enable simultaneous tests on all $\tau \in \mathcal{P}$. For example, $\sup_{\tau \in \mathcal{P}} Sc(\tau)$ forms a global specification test of the Kolmogorov-Smirnov type. It examines validity of the functional form of the conditional quantile function for all τ in $\mathcal{P}$ simultaneously.[12]

Simpler versions of the (pointwise) conditionality tests (that disregard that $\hat{\beta}(\tau)$ is an estimated quantity) can be constructed by regressing $(1(y_t \leq \widetilde{X}_t'\widehat{\beta}(\tau)) - \tau)$ on the lagged values of itself and other information variables, and then checking if the regression coefficients are zero. These and other tests are suggested, to evaluate VaR models, by Lopez (1998), Christoffersen (1998), Crnkoviĉ and Drachman (1996), Diebold, Gunther, and Tay (1998), Engle and Manganelli (1999), among others. Such procedures offer a good simple way to quickly check if a given VaR model is more or less plausible. Lopez (1998) offers a valuable detailed discussion on the criteria with which VaR models can be evaluated.

Definitions of the Test Processes for the Empirical Section

In the empirical section we compute Wald, integrated Wald, and quasi-Score test processes at a sequence of values of p in order to test the hypothesis:

$$H_0 : \beta_1(\tau) = 0,$$

where $\beta = (\beta_0(\tau), \beta_1(\tau)')'$, and $\beta_0(\tau)$ is the intercept parameter. Hypothesis H_o states that the coefficients on all conditioning variables are zero. If H_o is true, it means that *conditioning is statistically irrelevant.*

To define these tests denote the constrained quantile regression coefficient as $\hat{\beta}^R(\tau) = \text{arginf}_\beta\ Q_T(\beta, \tau)$ s.t. $\beta_1(\tau) = 0$. Then define the Wald

[12] Given the stochastic equicontinuity of the test processes, their distribution can be approximated via finite subnets of the processes evaluated at the grid of equi-distant quantile indices $\{\tau_i, i \in 1, ...K\}$ for K sufficiently large. This means that only a finite number of evaluations should be considered to approximate well, e.g. $\sup_{\tau \in \mathcal{P}} Sc(\tau)$ via $\sup_{\tau_i} Sc(\tau_i)$. If the approximation error is to vanish, we need $K \to \infty$ as $T \to \infty$.

and Score test processes[13]:

$$W(\cdot) \equiv T(\hat{\beta}^R(\cdot) - \hat{\beta}(\cdot))'\Omega_{1T}(\cdot)(\hat{\beta}^R(\cdot) - \hat{\beta}(\cdot)),$$
$$S(\cdot) \equiv T\nabla_\beta Q_T(\hat{\beta}^R(\cdot), \cdot)'\Omega_{2T}(\cdot)\nabla_\beta Q_T(\hat{\beta}^R(\cdot), \cdot).$$

The specification test process, as a function of τ, is defined as:

$$Sc(\cdot) \equiv \mu_T(\cdot)'\Omega_{3T}(\cdot)\mu_T(\cdot), \text{ where } \mu_T(\tau) \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1(y_t \leq \widetilde{X}_t'\widehat{\beta}(\tau)) - \tau)Z_t$$

The specification test, as stated earlier, checks the *validity of the given functional form*, as well as the conditioning ability.[14]

4.2 Near-Extreme Regression Quantiles (VaR)

While the regular asymptotics is certainly useful in a large sample for *non-extremal* regression quantiles, a more cautious approach is needed to distinguish a separate kind of inference for near-extreme (extremal) regression quantiles, as developed in Chernozhukov (1999a). What follows summarizes some of these results. Intuitively, the sample quantile regression is a method of generalizing the notion of order statistics to the regression settings. Correspondingly, for any given sample size T, the τ-th regression quantile fit can be seen as the τT-th *conditional* order (or rank) statistic. Depending on this rank, asymptotic approximations reflect the *extremal* or *rare* data considerations. Indeed, consider a simple example with no covariates. A .01-th quantile estimator in the sample of 100 is the first order statistics, and

[13] (a) How to choose Ω_{1T}, Ω_{2T}, and Ω_{3T} ? Each of the tests processes is of the form $w_T(\tau)'\Omega_T(\tau)w_T(\tau)$. For any τ, $\Omega_T(\tau)$ is chosen to be an inverse of a consistent estimator of the asymptotic variance of $w_T(\tau)$. One can either exploit the analytical expressions for these variances (as discussed in Chernozhukov (1999b)) or, as employed in the empirical section, use a moving-block bootstrap to estimate them. The procedure is quite simple: using such form of bootstrap, compute statistics $w_{bT}(\tau)$ (for $b = 1, ...B$ denoting the bootstrap replications, setting B large). Then compute the variance matrix of the "sample" $\{w_{bT}(\tau), b = 1, ...B\}$. (This procedure is consistent under the local alternatives, since then only the asymptotic mean of $w_T(\tau)$ is shifted and variance is unaffected.) (b) $\nabla_\beta Q_T$ means $\frac{\partial}{\partial\beta}\sum_{t\in I}\rho_\tau(Y_t - \widetilde{X}_t'\hat{\beta}(\tau))$, $I = \{t : Y_t \neq \widetilde{X}_t'\hat{\beta}(\tau)\}$.

[14] In the empirical section we give the specification test process for the linear model with Z_t selected to be the polynomial forms of X_t. That is, we have devoted our attention to the first problem. However, one can check the conditionality by selecting the appropriate Z_t's, as discussed.

hence one would expect that regular asymptotic approximation does not apply here, but some other asymptotic theory does. Similar considerations apply to regression setting. Indeed, data scarcity is *amplified* by the presence of covariates. To that end, the concept of effective rank is useful. Effective rank, r, is the *ratio* of *rank* $k = \tau T$ (if $\tau < 1/2$, and $(1-\tau)T$ if $\tau \geq 1/2$) to the *number of regressors*, k/d. To motivate such a notion, consider the "regression" quantile problem in a sample of 1000 observations and 10 dummy regressors, in which the target is the $.01 - th$ conditional quantile function. The constructed estimates of slopes will be the 1-st lowest order statistics in each of 10 subsamples corresponding to the dummy variables. Hence an extremal situation still applies here.

Let us now distinguish two formal asymptotic considerations/statistical experiments that account for the data scarcity illustrated above. (In the sequel, we always assume $\tau < 1/2$. Replace τ by $(1-\tau)$ if $\tau \geq 1/2$):

(i) $\tau \to 0$, $\tau T \to \infty$ (intermediate rank behavior, or τ is relatively low as compared to the sample size),

(ii) $\tau \to 0$, $\tau T \to k$ (extreme rank behavior, or τ is very low as compared to the sample size).

Both (i) and (ii) intend to *asymptotically capture forms of data scarcity arising in the tails of distribution. Both can be applied for any given quantile index of interest in a fixed dataset*, and these notions give rise to alternative inference techniques that can be applied when dealing with the extremal regression quantiles (VaR). Intuitively, these alternative inference approaches should perform better for the near-extreme regression quantiles than for the central ones. In the empirical section, these alternative techniques are used to construct the confidence intervals of the regression quantile coefficients. A "rule-of-thumb" choice of the (more) appropriate approximation theory for inference purposes is as follows: if $r < 10$, one may select method (ii) (as illustrated in the example above), if $r \in (10, 25)$ – method (i), and if $r \geq 25$ – the regular or central asymptotic approximations (previous subsection). A brief description of the asymptotics under (i) and (ii) follows next.

Assume that the conditional distribution function of $y_t - \widetilde{X}_t'\beta(q)$ ($q = 1/2$ or 0) is tail-equivalent to some function $K(x)F_u(\cdot)$,[15] where F_u is a distribution function with the extremal tail types.

[15] We say that cdf F_1 is tail-equivalent to cdf F_2 at (say the lower) end-point x, if as $z \searrow x$, $F_1(z)/F_2(z) \to 1$.

In the case (i), the asymptotic distributions are normal, but the covariance matrix depends on the tail index: (for any $m > 0, m \neq 1$)

$$\frac{\sqrt{\tau T}}{\mu_X' \left(\beta(m\tau) - \beta(\tau)\right)} \left(\hat{\beta}(\tau) - \beta(\tau)\right) \xrightarrow{\mathcal{D}} N\left(0, Q_{\mathcal{H}}^{-1} Q_{xx} Q_{\mathcal{H}}^{-1} \frac{\xi^2}{(m^{-\xi} - 1)^2}\right),$$

where $Q_{xx} \equiv \lim_T \frac{1}{T}\sum_{t=1}^T E\widetilde{X}_t\widetilde{X}_t'$, $Q_{\mathcal{H}} = \lim_T \frac{1}{T}\sum_{t=1}^T E\widetilde{X}_t\widetilde{X}_t'/\mathcal{H}(\widetilde{X}_t)$, ξ is the tail index of F_u, $\mathcal{H}(x)$ is some function of x that depends on the tail type of F_u, and $\mu_X = \lim_T \frac{1}{T}\sum_{t=1}^T EX_t$. Furthermore, $\mu_X'(\beta(m\tau) - \beta(\tau))$ can be replaced by $\overline{\overline{X}}'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))$, without affecting the validity of the result. We do not state the explicit forms here since the result is used only indirectly to justify the resampling techniques (that will be used to construct the confidence intervals in the empirical section – see the next section).

In the case (**ii**), the asymptotic distributions are defined by a random variable that solves a stochastic optimization problem, where the objective function is an integral w.r.t. a Poisson point process:

$$a_T\left(\hat{\beta}(\tau) - \beta(\tau)\right) \xrightarrow{\mathcal{D}} \mathrm{c}(k) + \operatorname*{arginf}_z \left[-k\mu_X' z + \int (j - x'z)^- d\mathbf{N}(j,x)\right],$$

where $\mathbf{N}$ is a certain Poisson Point Process. The mean intensity function of $\mathbf{N}$, constants $\mathrm{c}(k)$, and scaling a_T depend on the underlying tail type of F_u and on the tail heterogeneity function $K(x)$. Again, we do not state the explicit forms here since the result is used only indirectly to justify the resampling techniques (that will be used to construct the confidence intervals in the empirical section – see the next section).

Furthermore, Chernozhukov (1999a) defines and studies inference processes analogous to those in the previous section[16] and shows how to conduct inference by asymptotic or resampling methods. In particular, estimates of tail index ξ, [17] tail heterogeneity function $K(x)$, and scaling constants a_T

[16] Construction of quantile and inference processes is done by introducing an index l in a set $[l_1, l_2]$, so that process $a_T\left(\hat{\beta}(\tau\cdot) - \beta(\tau\cdot)\right)$ is a function of l, etc.

[17] A simple rule-of-thumb estimator for the empirical section is deduced from the following relationship $\frac{\bar{X}(\hat{\beta}(m\tau)-\hat{\beta}(\tau))}{\bar{X}(\hat{\beta}(\tau)-\hat{\beta}(\tau m^{-1}))}/m^{-\xi} \xrightarrow{P} 1$, as $\tau T \to \infty, \tau \searrow 0$ (but more sophisticated estimators can be constructed – see (Chernozhukov, 1999a)). So that

$$\hat{\xi}(m,\tau) = -\ln \frac{\bar{X}(\hat{\beta}(m\tau) - \hat{\beta}(\tau))}{\bar{X}(\hat{\beta}(\tau) - \hat{\beta}(\tau m^{-1}))} / \ln m$$

In practice m should not be set too far away from 1. E.g. in the empirical section, we used $m = .75$, $m = 1.25$, and various values of τ s.t. r, the effective rank, is between 10 and 20. We then took the median of $\{\hat{\xi}(m_i, \tau_j)\}$ over all such values of m_i and τ_j to obtain the final estimate $\hat{\xi}$.

are offered, and the validity of subsampling is established. Alternative regression quantile estimators emerge from these results. For example, a regression quantile extrapolation estimator[18] is constructed as follows (for τ_e close to zero and τ not close), and any positive constant $m \neq 1$:

$$\widehat{F}_{y_t}^{-1}(\tau_e|x) = \frac{(\tau_e/\tau)^{-\hat{\xi}} - 1}{m^{-\hat{\xi}} - 1}\left[x'(\hat{\beta}(m\tau) - \hat{\beta}(\tau))\right] + x'\hat{\beta}(\tau)$$

This also implicitly defines the extrapolation quantile estimates for $\beta(\tau_e)$.

5 Empirical Analysis

This section considers estimating the VaR of the Occidental Petroleum (NYSE:OXY) security returns. The dataset consists of 2527 daily observations[19] on

- y_t, the one-day returns,
- X_t, a vector of returns (or prices, yields, etc.) of other securities that affect distribution of Y_t and/or lagged values of y_t itself: a constant, lagged one-day return of Dow Jones Industrials (DJI), the lagged return on the spot price of oil (NCL, front-month contract on crude oil on NYMEX), and the lagged return y_t.

Generally, to estimate the VaR of a stock return, X_t may contain such variables as a market index of corresponding capitalization and type (for instance, the $S\&P500$ Value for a large-cap value stock), the industry index, a price of commodity or some other traded risk that the firm is exposed to, and lagged values of its stock price. It is also conceivable to include some unobserved factors, such as Size, Value, Momentum, or Liquidity premiums, whose effect on stock returns and risk has been a subject of numerous studies. However, we chose not to include estimated variables in the information set for the sake of simplicity.

Functional Forms of Conditional Quantile Functions

Two functional forms of conditional VaR were estimated:

- Linear Model : $v_t^h(p) = X_t'\,\theta(p)$,
- Quadratic Model: $v_t^h(p) = X_t'\theta(p) + X_t\mathcal{B}(p)X_t'$.

---

[18] This is a direct regression analogue of the estimator of Dekkers and de Haan (1989) that was suggested for non-regression cases.

[19] From September 1986 to November 1998

Conditional Risk Surfaces

Figure 1 presents surfaces of the regression VaR functions plotted in the time-probability level coordinates, (t, p). Recall that p is called the *probability level* of VaR, and $\tau = 1 - p$ is the quantile index. We report VaR for *all* values of $p \in [.01, .99]$. The conventional VaR reporting typically involves the probability levels of $p = .99$ and $p = .95$. Clearly, the whole *VaR surface* formed by varying p in $[.01, .99]$ represents a more *complete depiction of conditional risk.* Note also that since one can be either long or short the security, estimation of VaR in both tails of the return distribution is of interest.

The dynamics depicted in figure 1 unambiguously indicate certain dates on which market risk tends to be much higher than its usual level. This by itself underscores the importance of conditional modeling. We also stress that the driving force behind the dynamics is the behavior of X_t.

Model Comparison

Figure 1 also compares the dynamic evolution of the linear and the quadratic VaR surfaces. Notably, the quadratic model predicts higher risk magnitudes than the linear model. Indeed, the fluctuations of the quadratic VaR surface are significantly larger. The linear model thus predicts a more "smoothed out" VaR surface.

Conditional Quantile and Quantile Coefficient Functions

The next series of figures presents the statistical aspects of the analysis. For brevity, we chose to present the results in a graphical form.[20]

Let us set the date at $t = 2500$ to analyze the VaR. Figure 2 depicts the estimated $VaR_{2500}(p)$ for values of p in the interval $[.01, .99]$.[21] This figure also shows the 95% confidence intervals (c.i.) obtained by the following procedures:[22] (1) regular inference, based on the asymptotic normal approximation (labeled as "asymptotic"), (2) resampling inference, by the stationary bootstrap, that is valid under regular and intermediate rank asymptotics, and (3) and (4): subsampling inference with different scaling schemes, denoted as "Subsampling I" and "Subsampling II," suited for dependent data, and valid under the extreme rank asymptotics. Method (1) is intended to

[20] We have not presented here the formal statistical analysis of the quadratic model for brevity. Umantsev and Chernozhukov (1999) offer a detailed analysis of the quadratic model.

[21] We computed VaR($\cdot$) and coefficients for values of p lying on a grid with cell size .01 and interpolated in between. This is a justifiable interpolation since VaR($\cdot$) and coefficient processes are stochastically equicontinuous.

[22] All methods are in a form that is suitable for dependent data.

give the confidence intervals that are best for the central values[23] $p \in [.1, .9]$, method (2) – for the intermediate (near-extreme) values, $p \in [.04, .96]$, and methods (3) and (4) – for the extreme values $p \in (0, .04]$ and $[.96, 1)$.[24]

As can be seen from figure 2, the c.i. by methods (2), (3), and (4) tend to be roughly 1.5, 2, and 2.5 times wider than the standard c.i., respectively. Hence *additional significant estimation uncertainty* is present in the tails, and it is important to *properly* account for it. Accounting for it means that, within the c.i. by methods (2)-(4), near-extreme VaR may actually be as much as two times higher than the point estimates suggest.

Figures 4-5 present the same analysis for the coefficient functions $\theta(\cdot)$ of the linear model. The methods and the results employed are like those we have just discussed. We will give an economic meaning to the coefficient shapes later.

Specification Analysis

Figure 6 presents the pointwise values of Wald and quasi-Score test statistics for testing the hypothesis:

– *Is the conventional (unconditional historical) VaR model statistically not different from the conditional VaR model?*

The answer is conclusive: the hypothesis is rejected pointwise. Note that the 'p-values'[25] for this case are all smaller than 0.01. That is, the *regression conditioning matters.* This can also be seen in figure 4, where the confidence intervals of slope coefficients are plotted throughout the interesting range of p – these confidence intervals exclude 0s.

Figure 6 (right) also depicts the specification test process (see earlier section). Results of the specification testing are clearly in *favor* of the *linear model*: the critical value (pointwise) for 10% level is 6.25, which is above

[23] Based on Monte-Carlo with the sample size of 1000 and the considerations of the previous section.

[24] In this application, for transparency and clarity, the subsampling methods were operationalized by assuming the tails are exactly algebraic, so that the rate of convergence or divergence is $a_T = T^{-\xi}$. ξ was estimated to be approximately .25 by the method described in the previous section. Hence $a_T = T^{-.25}$ defined a scaling for the subsampling procedure. As suggested in Chernozhukov (1999a), the centering constant was taken to be $\hat{\beta}_T(k/b)$. The subsample size b was set to be 1/10 of the whole sample T. The resulting confidence intervals are labeled "Subsampling II." For comparison, rate $a_T = T^{.01}$ was also used, and the resulting confidence intervals were labeled as "Subsampling I."

[25] Not to be confused with p in $VaR(p)$.

any of the values depicted. Obviously, since the critical value for the test statistic $\sup_{\tau\in[0.01,.99]} Sc(\tau)$ should be above 6.25, the linear model passes this stronger Kolmogorov-type test, too !

The Determinants of Risk

We now provide both a statistical and an economic interpretation of the coefficient functions $\theta_i(\cdot)$. Let us fix time period $t = 2500$ and suppose for a moment that $\theta_i(p) > 0$ for some $i > 0, p \in (0,1)$. As $VaR_t(p) = v_t(p) = \theta_0(p) + \sum_i \theta_i(p)X_{t,i}$, a positive coefficient in front of $X_{t,i}$ implies that higher values of $X_{t,i}$ correspond to higher values of $VaR_t(p)$, given that other elements of X_t are unchanged. Stated differently, if $\theta_i(p) > 0$, then increases (decreases) of $X_{t,i}$ are associated with upward (downward) shifts of $VaR_t(\cdot)$ at point p. Note that $VaR_t(\cdot)$ is the "reversed" inverse of the cdf of $y_t|X_t$, i.e. $F_{y_t}^{-1}(1-\cdot|X_t)$ [Take figure 3 (middle) and rotate it 90^o clockwise to get the conditional cdf.]. Thus positive shocks in $X_{t,i}$ shift the cdf $F_{y_t}(\cdot|X_t)$ to the *right.*

Similarly, if $\theta_i(p)$ is negative, effects of positive and negative shocks in the i^{th} information variable are reversed: positive shocks move $VaR_t(p)$ down and cdf of $y_t|X_t$ to the left and negative shocks move $VaR_t(p)$ up and cdf of $y_t|X_t$ to the right.

The effects described above are *local,* in the sense that they affect $VaR_t(\cdot)$ and $F_{y_t}(\cdot|X_t)$ only locally, around points p and $F_{y_t}^{-1}(1-p|X_t)$, respectively. Transformations of these functions at other points caused by such shocks depend on the sign and magnitude of $\theta_i(p)$ at other probability levels p.

Suppose next that θ is positive and decreasing in the right tail of distribution of $y_t|X_t$ (e.g. $\theta_1(\cdot)$ on $(0,.2)$, see figure 4). A positive shock in X_1 will now shift the entire right tail of cdf of $y_t|X_t$ to the right, and the effect will be greater for extreme points (those close to $p = 0$), at which $\theta_1(p)$ is higher. The effect on the density of $y_t|X_t$ is schematically depicted in figure 7. Thus, this particular shape of $\theta_1(\cdot)$ implies that positive shocks of the corresponding information variable result in the right tail of density of $y_t|X_t$ being stretched further to the right (more positive skewness in the right tail). A similar shape is observed for the coefficient function $\theta_2(p)$ for almost all values of p.

Thus, the shapes of $\theta_i(\cdot)$ such as $\theta_1(\cdot)$ or $\theta_2(\cdot)$ in figure 4 translate positive shocks of the corresponding information variables into the longer right tails (favorable for holding long positions in Y). On the other hand, shapes of $\theta_i(\cdot)$ similar to those of $\theta_3(\cdot)$ in figure 4 translate such shocks into shorter right-tails (averse effect for long holders of Y). (see figure 7).

Finally, we provide the economic interpretation of the slope coefficient functions $\theta_1(\cdot)$, $\theta_2(\cdot)$, $\theta_3(\cdot)$, corresponding to the lagged returns on oil spot price, X_1, equity index, X_2, and price of the security in question, X_3.

- $\theta_1(\cdot)$ is significantly positive and decreasing in the right tail of the distribution of $y_t|X_t$ (figures 4 and 5, $p < 1/4$). It is insignificantly positive in the middle part (for $p \in (.25, .95)$) and then it is increasing in the far left tail ($p > .96$), although values $\theta_1(p)$, $p > .96$, are not as high as those for $p < 1/4$. This suggests that the Spot price of Oil and the return on our stock are positively related, with the *right tail of equity return being much more sensitive to Oil price shocks, than the left tail.* This effect can be explained by, for example, real optionality intrinsic to the operation of the firm, or by a non-linear hedging policy (e.g., long positions in put options instead of swaps or futures, whose payoff is linear in the underlying price movements). The overall effect of a positive shock in the spot oil price X_1 is presented in figure 7.
- $\theta_2(\cdot)$, in contrast, is significantly positive for all values of p with the possible exception of the far right tail of $y_t|X_t$, ($p \in (0, 0.04)$). We also notice a moderate increase in the right tail and a sharp increase in the left tail (p close to 1). Thus, in addition to the *strong positive relation between the stock return on the individual stock and the market* return (DJI) (dictated by the fact that $\theta_2(\cdot) > 0$ on $(0, 1)$) there is also *additional sensitivity of the left tail of the security return to the market movements* (steep increase on $(.94, 1)$), which is *strongly consistent with the notion of highly correlated equity returns in market drops.* For high positive returns, in contrast, market return has a much weaker effect (low values on $(0, 0.04)$). The effect of positive shock in the market return (X_2) is depicted in figure 7.
- $\theta_3(\cdot)$, in contrast, is significantly negative, except for values of p close to 0. This may be clearly interpreted as a "*mean reversion*" effect in the central part of the distribution. However, X_3, the lagged return, does not appear to significantly shift the quantile function in the tails. Thus X_3 is more *important for the determination of intermediate risks* (values of p in $[.15, .85]$). The effect of a positive shock in X_3 is schematically portrayed in figure 7. Figures 4 and 7 also capture the *asymmetry of response to the negative and positive return shocks– a positive shock leads to mean reversion and intermediate risk contraction, whereas a negative shock leads to mean reversion and intermediate risk amplification.*

Note that the estimates of near-extreme VaR should be interpreted carefully, since the point estimates provided by regression quantiles are highly biased in the tails. Some correction can be achieved by using alternative estimators that use the regular variation properties of tails in order to construct regression-like estimates of the near-extreme VaR (see previous section). For example, as depicted in Figure 3, the *regression extrapolation estimator introduces a significant correction*, but within the confidence bands constructed by method (4). Further conclusions are stated in section 1.

Fig. 1 $VaR_t(p)$ for Linear (upper) and Qudratic (lower) Models ($VaR_t(p)$ is on the vertical axis, and (p,t) are on the horisontal axes.)

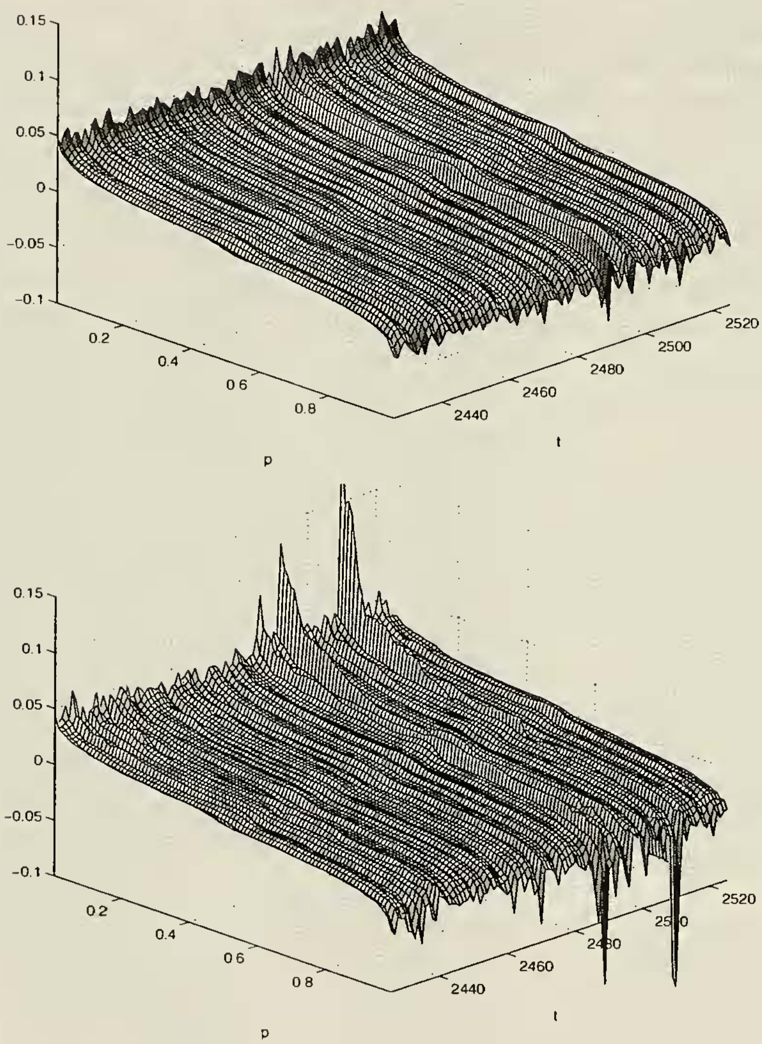


Fig. 2 Estimates and 95% (Pointwise) Confidence Intervals for $\widehat{VaR}^{1}_{2500}(\cdot)$

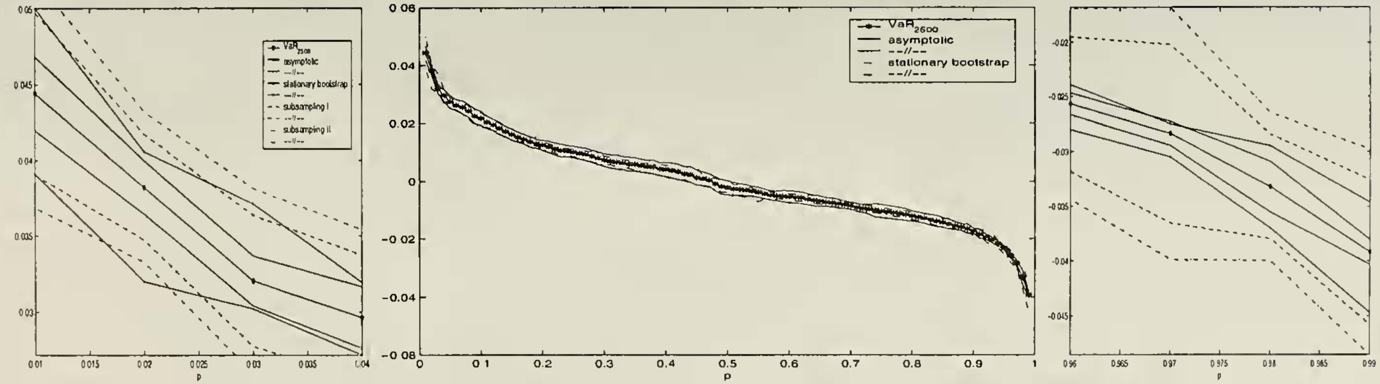


Fig. 3 Extrapolation Regression Quantiles (VaR) ($m = .75$ and $.25$) and Regression Quantiles in the Tails

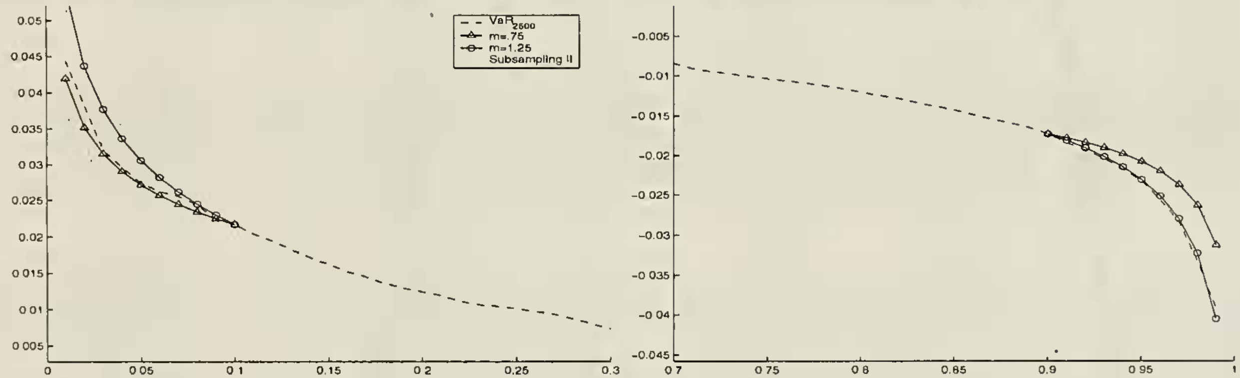


Fig. 4 Estimates and 95% (Pointwise) Confidence Intervals for $\theta_0(\cdot)$, $\theta_1(\cdot), \theta_2(\cdot), \theta_3(\cdot)$

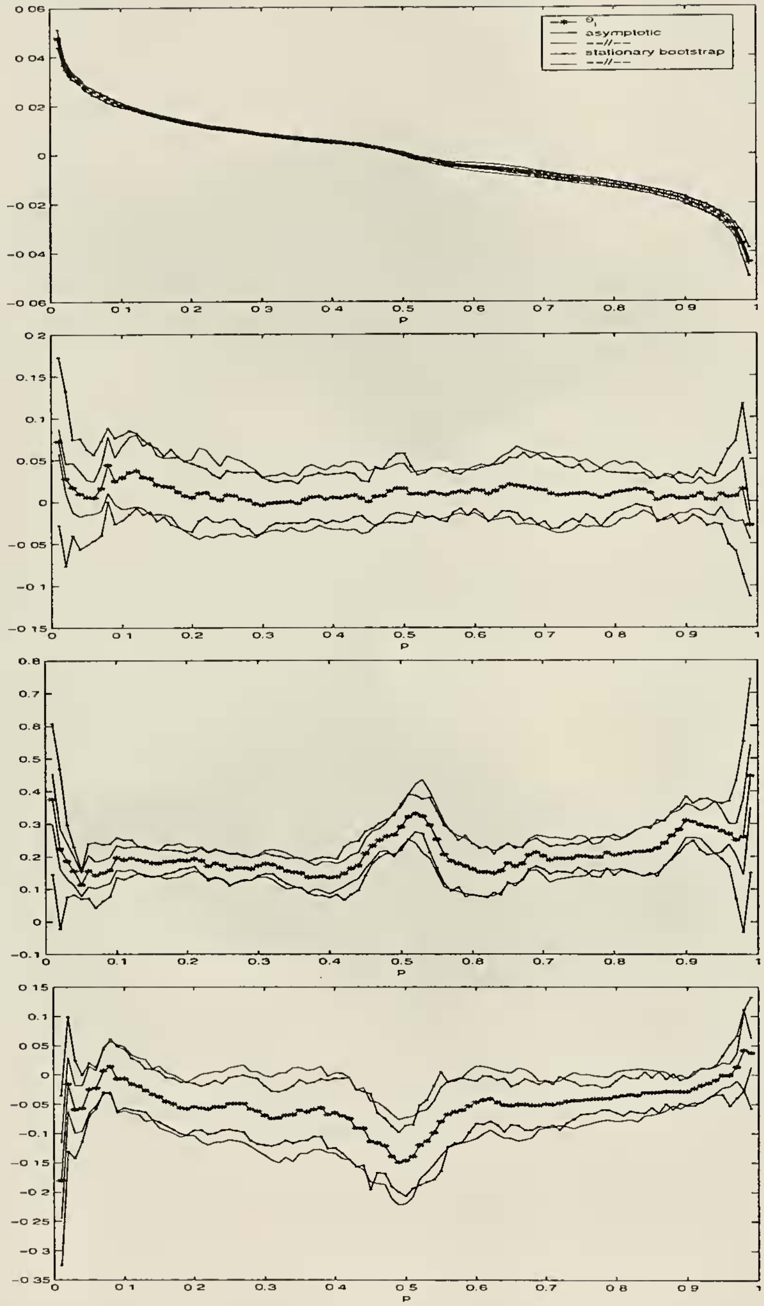


Fig. 5 Near-Extreme Quantile Coefficients: estimates and 95% Confidence Intervals for $\theta_0(\cdot), \theta_1(\cdot), \theta_2(\cdot), \theta_3(\cdot)$

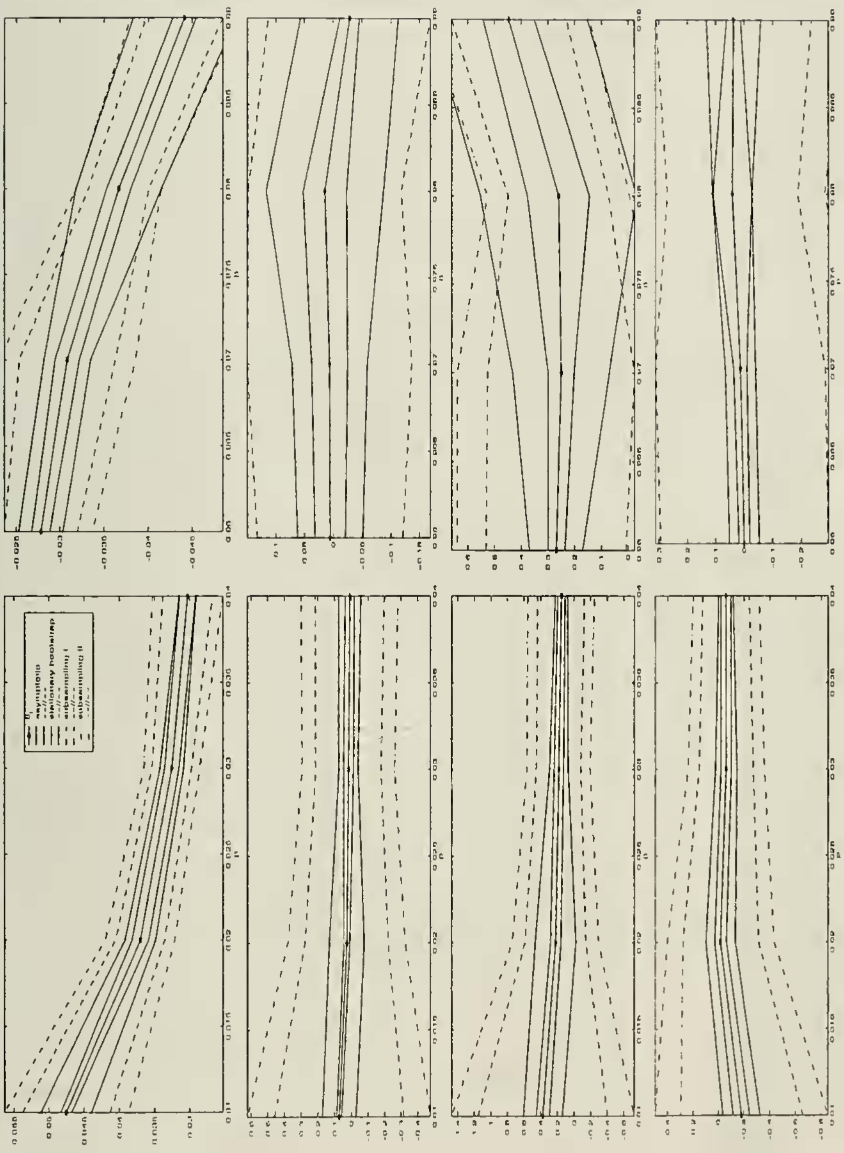


Fig. 6 Wald and Score Test Processes, and Specification Test Processes

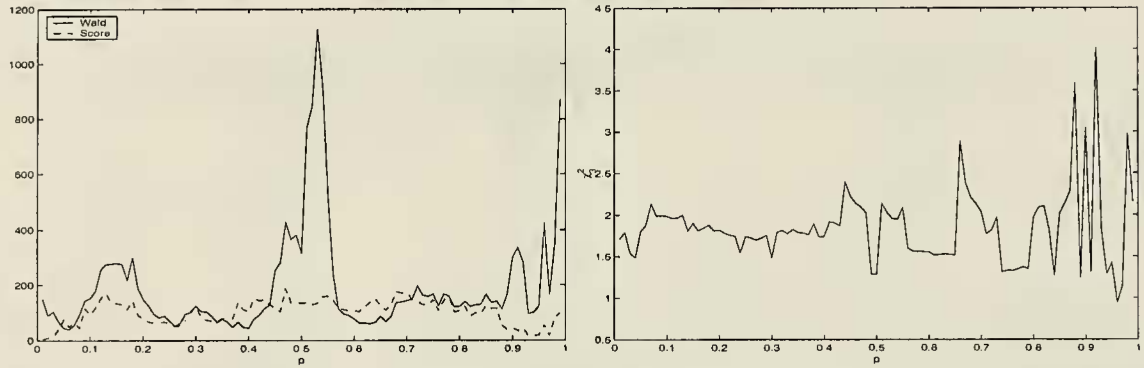
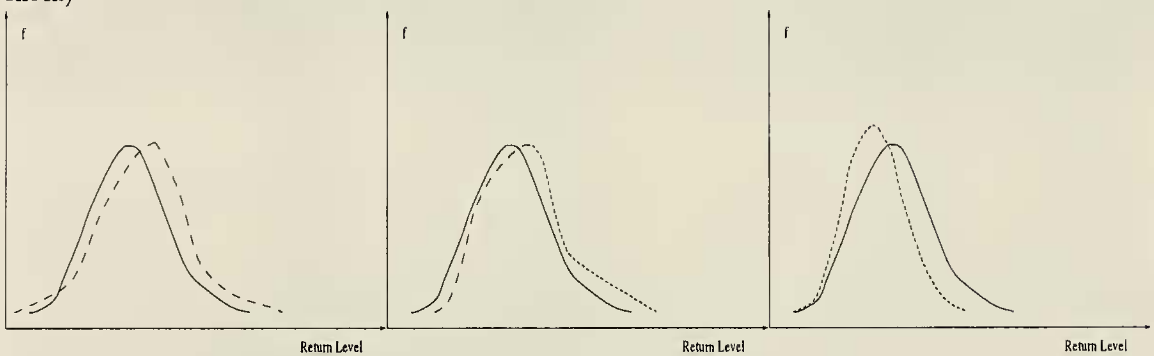


Fig. 7 Schematic Local Effects on the Density of Return by Positive Shocks in X_1, X_2, X_3 (solid line–before shock, dashed line – after shock)



References

Ait-Sahalia, Y., and A. W. Lo (1998): "Nonparametric Risk Management and Implied Risk Aversion," *preprint*.

Bassett, G., and H. Chen (1999): "Quantile Style: Quantiles to Assess Mutual Fund Investment Styles," a Working Paper, Presented at the International Conference on Economic Applications of Quantile Regression, Konstanz, 2000.

Bassett, Jr., G., and R. Koenker (1978): "Asymptotic theory of least absolute error regression," *J. Amer. Statist. Assoc.*, 73(363), 618–622.

Bierens, H., and D. K. Ginther (1999): "Integrated Conditional Moment Testing of Quantile Regression Models," a Working Paper, Presented at the International Conference on Economic Applications of Quantile Regression, Konstanz, 2000.

Campbell, J., C. MacKinlay, and A. Lo (1997): *Econometrics of Financial Markets*. MIT.

Chernozhukov, V. (1999a): "Conditional Extremes and Near-Extremes," November, Ph.D. dissertation draft, Stanford.

——— (1999b): "Specification and Other Test Processes for Quantile Regression," August, a Working Paper, Stanford.

Christoffersen, P. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39(4), 841–862.

Crnkovič, C., and J. Drachman (1996): "Quality Control," *Risk*, 9(9), 138–144.

Dekkers, A. L. M., and L. de Haan (1989): "On the estimation of the extreme-value index and large quantile estimation," *Ann. Statist.*, 17(4), 1795–1832.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39(4), 863–883.

Engle, R. F., and S. Manganelli (1999): "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," UCSD Economics Deaprtment Working Paper 99-20 , October.

FedReg (1996): *Risk-based capital standards: Market Risk*vol. 61. Federal Register.

Fitzenberger, B. (1998): "The moving blocks bootstrap and robust inference for linear least squares and quantile regressions," *J. Econometrics*, 82(2), 235–287.

Heiler, S., and K. Abberger (1999): "Applications of Nonparametric Quantile Regression to Financial Data," a Working Paper, Presented at the International Conference on Economic Applications of Quantile Regression, Konstanz, 2000.

Hogg, R. V. (1975): "Estimates of Percentile Regression Lines Using Salary Data," *J. Amer. Statist. Assoc.*, 70(349), 56–59.

Koenker, R., and G. S. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and J. Machado (1999): "Goodness of Fit and Related Inference Processes for Quantile Regression," preprint.

Koenker, R., and S. Portnoy (1997): "The Gaussian Hare and the Laplacian Tortoise," *Statistical Science*, 12, 279–300.

Koenker, R., and S. Portnoy (1999): "Regression Quantiles," preprint.

Koenker, R., and Q. Zhao (1996): "Conditional quantile estimation and inference for ARCH models," *Econometric Theory*, 12(5), 793–813.

Lopez, J. A. (1998): "Regulatory Evaluation of Value-at-Risk Models," *Journal of Risk*, 1(2), 37–63.

Politis, D. N., and J. P. Romano (1994): "The stationary bootstrap," *J. Amer. Statist. Assoc.*, 89(428), 1303–1313.

Portnoy, S. (1991): "Asymptotic behavior of regression quantiles in nonstationary, dependent cases," *J. Multivariate Anal.*, 38(1), 100–113.

Resnick, S. I. (1987): *Extreme values, regular variation, and point processes*. Springer-Verlag, New York-Berlin.

TAYLOR, J. (1999): "A Quantile Regression Approach to Estimating the Distribution of Multi-Period Returns," *J. Derivatives*, Fall, 64–78.
UMANTSEV, L., AND V. CHERNOZHUKOV (1999): "Polynomial Regression Quantile Modeling of Risk," September, a Working Paper, Stanford.
VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak convergence and empirical processes.* Springer-Verlag, New York.
WEISS, A. A. (1991): "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation," *Econometric Theory*, 47, 47–66.