CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles

Robert F. ENGLE
Stern School of Business, New York University, New York, NY 10012-1126 (*rengle@stern.nyu.edu*)

Simone MANGANELLI
DG-Research, European Central Bank, 60311 Frankfurt am Main, Germany (*simone.manganelli@ecb.int*)

Value at risk (VaR) is the standard measure of market risk used by financial institutions. Interpreting the VaR as the quantile of future portfolio values conditional on current information, the conditional autoregressive value at risk (CAViaR) model specifies the evolution of the quantile over time using an autoregressive process and estimates the parameters with regression quantiles. Utilizing the criterion that each period the probability of exceeding the VaR must be independent of all the past information, we introduce a new test of model adequacy, the dynamic quantile test. Applications to real data provide empirical support to this methodology.

KEY WORDS: Nonlinear regression quantile; Risk management; Specification testing.

1. INTRODUCTION

The importance of effective risk management has never been greater. Recent financial disasters have emphasized the need for accurate risk measures for financial institutions. As the nature of the risks has changed over time, methods of measuring these risks must adapt to recent experience. The use of quantitative risk measures has become an essential management tool to be placed in parallel with models of returns. These measures are used for investment decisions, supervisory decisions, risk capital allocation, and external regulation. In the fast-paced financial world, effective risk measures must be as responsive to news as are other forecasts and must be easy to grasp in even complex situations.

Value at risk (VaR) has become the standard measure of market risk used by financial institutions and their regulators. VaR is a measure of how much a certain portfolio can lose within a given time period, for a given confidence level. The great popularity that this instrument has achieved among financial practitioners is essentially due to its conceptual simplicity; VaR reduces the (market) risk associated with any portfolio to just one monetary amount. The summary of many complex bad outcomes in a single number naturally represents a compromise between the needs of different users. This compromise has received the blessing of a wide range of users and regulators.

Despite VaR's conceptual simplicity, its measurement is a very challenging statistical problem, and none of the methodologies developed so far gives satisfactory solutions. Because VaR is simply a particular quantile of future portfolio values, conditional on current information, and because the distribution of portfolio returns typically changes over time, the challenge is to find a suitable model for time-varying conditional quantiles. The problem is to forecast a value each period that will be exceeded with probability $(1-\theta)$ by the current portfolio, where $\theta \in (0, 1)$ represents the confidence level associated with the VaR. Let $\{y_t\}_{t=1}^{T}$ denote the time series of portfolio returns and T denote the sample size. We want to find VaR_t such that $\Pr[y_t < -VaR_t | \Omega_t] = \theta$, where Ω_t denotes the information set available at time t. Any reasonable methodology should address the following three issues: (1) provide a formula for calculating VaR_t as a function of variables known at time $t-1$ and a set of parameters that need to be estimated; (2) provide a procedure (namely, a loss function and a suitable optimization algorithm) to estimate the set of unknown parameters; and (3) provide a test to establish the quality of the estimate.

In this article we address each of these issues. We propose a conditional autoregressive specification for VaR_t, which we call *conditional autoregressive value at risk* (CAViaR). The unknown parameters are estimated using Koenker and Bassett's (1978) regression quantile framework. (See also Chernozhukov and Umantsev 2001 for an application of linear regression quantile to VaR estimation.) Consistency and asymptotic results build on existing contributions of the regression quantile literature. We propose a new test, the dynamic quantile (DQ) test, which can be interpreted as an overall goodness-of-fit test for the estimated CAViaR processes. This test, which has been independently derived by Chernozhukov (1999), is new in the literature on regression quantiles.

The article is structured as follows. Section 2 reviews the current approaches to VaR estimation, and Section 3 introduces the CAViaR models. Sections 4 reviews the literature on regression quantiles and establishes consistency and asymptotic normality of the estimator. Section 5 introduces the DQ test, and Section 6 presents an empirical application to real data. Section 7 concludes the article.

2. VALUE AT RISK MODELS

VaR was developed in the early 1990s in the financial industry to provide senior management with a single number that could quickly and easily incorporate information about the risk of a portfolio. Today VaR is part of every risk manager's toolbox. Indeed, VaR can help management estimate the cost of positions in terms of risk, allowing them to allocate risk in a more efficient way. Also, the Basel Committee on Banking Supervision (1996) at the Bank for International Settlements

Journal of Business & Economic Statistics
October 2004, Vol. 22, No. 4
DOI 10.1198/073500104000000370

uses VaR to require financial institutions, such as banks and investment firms, to meet capital requirements to cover the market risks that they incur as a result of their normal operations. However, if the underlying risk is not properly estimated, these requirements may lead financial institutions to overestimate (or underestimate) their market risks and consequently to maintain excessively high (low) capital levels. The result is an inefficient allocation of financial resources that ultimately could induce firms to move their activities into jurisdictions with less-restrictive financial regulations.

The existing models for calculating VaR differ in many aspects, but all follow a common structure, which can be summarized as follows: (1) The portfolio is marked-to-market daily, (2) the distribution of the portfolio returns is estimated, and (3) the VaR of the portfolio is computed. The main differences among VaR models are related to the second aspect. VaR methodologies can be classified initially into two broad categories: factor models, such as RiskMetrics (1996), and portfolio models, such as historical quantiles. In the first case, the universe of assets is projected onto a limited number of factors whose volatilites and correlations have been forecast. Thus time variation in the risk of a portfolio is associated with time variation in the volatility or correlation of the factors. The VaR is assumed to be proportional to the computed standard deviation of the portfolio, often assuming normality. The portfolio models construct historical returns that mimic the past performance of the current portfolio. From these historical returns, the current VaR is constructed based on a statistical model. Thus changes in the risk of a particular portfolio are associated with the historical experience of this portfolio. Although there may be issues in the construction of the historical returns, the interesting modeling question is how to forecast the quantiles. Several different approaches have been used. Some first estimate the volatility of the portfolio, perhaps by a generalized autoregressive conditional heteroscedasticity (GARCH) or exponential smoothing, and then compute VaR from this, often assuming normality. Others use rolling historical quantiles under the assumption that any return in a particular period is equally likely. A third approach appeals to extreme value theory.

It is easy to criticize each of these approaches. The volatility approach assumes that the negative extremes follow the same process as the rest of the returns and that the distribution of the returns divided by standard deviations will be iid, if not normal. The rolling historical quantile method assumes that for a certain window, such as a year, any return is equally likely, but a return more than a year old has zero probability of occurring. It is easy to see that the VaR of a portfolio will drop dramatically just 1 year after a very bad day. Implicit in this methodology is the assumption that the distribution of returns does not vary over time, at least within a year. An interesting variation of the historical simulation method is the hybrid approach proposed by Boudoukh, Richardson, and Whitelaw (1998), which combines volatility and historical simulation methodologies by applying exponentially declining weights to past returns of the portfolio. However, both the choice of the parameters of interest and the procedure behind the computation of the VaR seem to be ad hoc and based on empirical justifications rather than on sound statistical theory.

Applications of extreme quantile estimation methods to VaR have recently been proposed (see, e.g., Danielsson and de Vries 2000). The intuition here is to exploit results from statistical extreme value theory and to concentrate the attention on the asymptotic form of the tail, rather than modeling the whole distribution. There are two problems with this approach. First, it works only for very low probability quantiles. As shown by Danielsson and de Vries (2000), the approximation may be very poor at very common probability levels (such as 5%), because they are not "extreme" enough. Second, and most important, these models are nested in a framework of iid variables, which is not consistent with the characteristics of most financial datasets, and, consequently, the risk of a portfolio may not vary with the conditioning information set. Recently, McNeil and Frey (2000) suggested fitting a GARCH model to the time series of returns and then applying the extreme value theory to the standardized residuals, which are assumed to be iid. Although it is an improvement over existing applications, this approach still suffers from the same problems as the volatility models. Chernozhukov (2000) and Manganelli and Engle (2004) have shown how extreme value theory can be incorporated into the regression quantile framework.

3. CAVIAR

We propose a different approach to quantile estimation. Instead of modeling the whole distribution, we model the quantile directly. The empirical fact that volatilities of stock market returns cluster over time may be translated in statistical words by saying that their distribution is autocorrelated. Consequently, the VaR, which is tightly linked to the standard deviation of the distribution, must exhibit similar behavior. A natural way to formalize this characteristic is to use some type of autoregressive specification. We propose a conditional autoregressive quantile specification, which we call CAViaR.

Suppose that we observe a vector of portfolio returns, $\{y_t\}_{t=1}^T$. Let θ be the probability associated with VaR, let $\mathbf{x}_t$ be a vector of time t observable variables, and let $\boldsymbol{\beta}_\theta$ be a p-vector of unknown parameters. Finally, let $f_t(\boldsymbol{\beta}) \equiv f_t(\mathbf{x}_{t-1}, \boldsymbol{\beta}_\theta)$ denote the time t θ-quantile of the distribution of portfolio returns formed at time $t-1$, where we suppress the θ subscript from $\boldsymbol{\beta}_\theta$ for notational convenience. A generic CAViaR specification might be the following:

$$f_t(\boldsymbol{\beta}) = \beta_0 + \sum_{i=1}^{q} \beta_i f_{t-i}(\boldsymbol{\beta}) + \sum_{j=1}^{r} \beta_j l(\mathbf{x}_{t-j}), \tag{1}$$

where $p = q + r + 1$ is the dimension of $\boldsymbol{\beta}$ and l is a function of a finite number of lagged values of observables. The autoregressive terms $\beta_i f_{t-i}(\boldsymbol{\beta}), i = 1, \ldots, q$, ensure that the quantile changes "smoothly" over time. The role of $l(\mathbf{x}_{t-j})$ is to link $f_t(\boldsymbol{\beta})$ to observable variables that belong to the information set. This term thus has much the same role as the news impact curve for GARCH models introduced by Engle and Ng (1993). A natural choice for $\mathbf{x}_{t-1}$ is lagged returns. Indeed, we would expect the VaR to increase as y_{t-1} becomes very negative, because one bad day makes the probability of the next somewhat greater. It might be that very good days also increase VaR, as would be the case for volatility models. Hence, VaR could depend symmetrically on $|y_{t-1}|$.

Next we discuss some examples of CAViaR processes that we estimate. Throughout, we use the notation $(x)^+ = \max(x, 0)$, $(x)^- = -\min(x, 0)$.

Adaptive:

$$f_t(\beta_1) = f_{t-1}(\beta_1) + \beta_1\{[1 + \exp(G[y_{t-1} - f_{t-1}(\beta_1)])]^{-1} - \theta\},$$

where G is some positive finite number. Note that as $G \to \infty$, the last term converges almost surely to $\beta_1[I(y_{t-1} \le f_{t-1}(\beta_1)) - \theta]$, where $I(\cdot)$ represents the indicator function; for finite G, this model is a smoothed version of a step function. The adaptive model incorporates the following rule: Whenever you exceed your VaR, you should immediately increase it, but when you do not exceed it, you should decrease it very slightly. This strategy obviously will reduce the probability of sequences of hits and will also make it unlikely that there will never be hits. But it learns little from returns that are close to the VaR or are extremely positive, when G is large. It increases the VaR by the same amount regardless of whether the returns exceeded the VaR by a small margin or a large margin. This model has a unit coefficient on the lagged VaR. Other alternatives are as follows:

Symmetric absolute value:

$$f_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 f_{t-1}(\boldsymbol{\beta}) + \beta_3|y_{t-1}|.$$

Asymmetric slope:

$$f_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 f_{t-1}(\boldsymbol{\beta}) + \beta_3(y_{t-1})^+ + \beta_4(y_{t-1})^-.$$

Indirect GARCH(1, 1):

$$f_t(\boldsymbol{\beta}) = (\beta_1 + \beta_2 f_{t-1}^2(\boldsymbol{\beta}) + \beta_3 y_{t-1}^2)^{1/2}.$$

The first and third of these respond symmetrically to past returns, whereas the second allows the response to positive and negative returns to be different. All three are mean-reverting in the sense that the coefficient on the lagged VaR is not constrained to be 1.

The indirect GARCH model would be correctly specified if the underlying data were truly a GARCH(1, 1) with an iid error distribution. The symmetric absolute value and asymmetric slope quantile specifications would be correctly specified by a GARCH process in which the standard deviation, rather than the variance, is modeled either symmetrically or asymmetrically with iid errors. This model was introduced and estimated by Taylor (1986) and Schwert (1988) and analyzed by Engle (2002). But the CAViaR specifications are more general than these GARCH models. Various forms of non-iid error distributions can be modeled in this way. In fact, these models can be used for situations with constant volatilities but changing error distributions, or situations in which both error densities and volatilities are changing.

4. REGRESSION QUANTILES

The parameters of CAViaR models are estimated by regression quantiles, as introduced by Koenker and Bassett (1978). Koenker and Bassett showed how to extend the notion of a sample quantile to a linear regression model. Consider a sample of observations $y_1, \ldots, y_T$ generated by the model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}^0 + \varepsilon_{\theta t}, \qquad Quant_\theta(\varepsilon_{\theta t}|\mathbf{x}_t) = 0, \tag{2}$$

where $\mathbf{x}_t$ is a p-vector of regressors and $Quant_\theta(\varepsilon_{\theta t}|\mathbf{x}_t)$ is the θ-quantile of $\varepsilon_{\theta t}$ conditional on $\mathbf{x}_t$. Let $f_t(\boldsymbol{\beta}) \equiv \mathbf{x}_t\boldsymbol{\beta}$. Then the θth regression quantile is defined as any $\hat{\boldsymbol{\beta}}$ that solves

$$\min_{\boldsymbol{\beta}} \frac{1}{T}\sum_{t=1}^{T}[\theta - I(y_t < f_t(\boldsymbol{\beta}))][y_t - f_t(\boldsymbol{\beta})]. \tag{3}$$

Regression quantiles include as a special case the least absolute deviation (LAD) model. It is well known that LAD is more robust than ordinary least squares (OLS) estimators whenever the errors have a fat-tailed distribution. Koenker and Bassett (1978), for example, ran a simple Monte Carlo experiment and showed how the empirical variance of the median, compared with the variance of the mean, is slightly higher under the normal distribution but much lower under all of the other distributions considered.

Analysis of linear regression quantile models has been extended to cases with heteroscedastic (Koenker and Bassett 1982) and nonstationary dependent errors (Portnoy 1991), time series models (Bloomfield and Steiger 1983), simultaneous equations models (Amemiya 1982; Powell 1983), and censored regression models (Powell 1986; Buchinsky and Hahn 1998). Extensions to the autoregressive quantiles have been proposed by Koenker and Zhao (1996) and Koul and Saleh (1995). These approaches differ from the one proposed in this article in that all of the variables are observable and the models are linear in the parameters. In the nonlinear case, asymptotic theory for models with serially independent (but not identically distributed) errors have been proposed by, among others, Oberhofer (1982), Dupacova (1987), Powell (1991), and Jureckova and Prochazka (1993). There is relatively little literature that considers nonlinear quantile regressions in the context of time series. The most important contributions are those by White (1994, cor. 5.12), who proved the consistency of the nonlinear regression quantile, both in the iid and stationary dependent cases, and by Weiss (1991), who showed consistency, asymptotic normality and asymptotic equivalence of Lagrange multiplier (LM) and Wald tests for LAD estimators for nonlinear dynamic models. Finally, Mukherjee (1999) extended the concept of regression and autoregression quantiles to nonlinear time series models with iid error terms.

Consider the model

$$y_t = f(y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1; \boldsymbol{\beta}^0) + \varepsilon_{t\theta} \quad [Quant_\theta(\varepsilon_{t\theta}|\Omega_t) = 0]$$
$$\equiv f_t(\boldsymbol{\beta}^0) + \varepsilon_{t\theta}, \qquad t = 1, \ldots, T, \tag{4}$$

where $f_1(\boldsymbol{\beta}^0)$ is some given initial condition, $\mathbf{x}_t$ is a vector of exogenous or predetermined variables, $\boldsymbol{\beta}^0 \in \Re^p$ is the vector of true unknown parameters that need to be estimated, and $\Omega_t = [y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1, f_1(\boldsymbol{\beta}^0)]$ is the information set available at time t. Let $\hat{\boldsymbol{\beta}}$ be the parameter vector that minimizes (3).

Theorems 1 and 2 show that the nonlinear regression quantile estimator $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal. Theorem 3 provides a consistent estimator of the variance–covariance matrix. In Appendix A we give sufficient conditions on f in (4), together with technical assumptions, for these results to hold. The proofs are extensions of work of Weiss (1991) and Powell (1984, 1986, 1991) and are in Appendix B. We denote the conditional density of $\varepsilon_{t\theta}$ evaluated at 0 by $h_t(0|\Omega_t)$, denote the $1 \times p$ gradient of $f_t(\boldsymbol{\beta})$ by $\nabla f_t(\boldsymbol{\beta})$, and define $\nabla f(\boldsymbol{\beta})$ to be a $T \times p$ matrix with typical row $\nabla f_t(\boldsymbol{\beta})$.

Theorem 1 (Consistency). In model (4), under assumptions C0–C7 (see App. A), $\hat{\boldsymbol\beta} \xrightarrow{p} \boldsymbol\beta^0$, where $\hat{\boldsymbol\beta}$ is the solution to

$$\min_{\boldsymbol\beta} T^{-1} \sum_{t=1}^{T} \{[\theta - I(y_t < f_t(\boldsymbol\beta))] \cdot [y_t - f_t(\boldsymbol\beta)]\}.$$

Proof. See Appendix B.

Theorem 2 (Asymptotic normality). In model (4), under assumptions AN1–AN4 and the conditions of Theorem 1,

$$\sqrt{T}\mathbf{A}_T^{-1/2}\mathbf{D}_T(\hat{\boldsymbol\beta} - \boldsymbol\beta^0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{I}),$$

where

$$\mathbf{A}_T \equiv E\left[T^{-1}\theta(1-\theta)\sum_{t=1}^{T} \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^0)\right],$$

$$\mathbf{D}_T \equiv E\left[T^{-1}\sum_{t=1}^{T} h_t(0|\Omega_t)\nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^0)\right],$$

and $\hat{\boldsymbol\beta}$ is computed as in Theorem 1.

Proof. See Appendix B.

Theorem 3 (Variance–covariance matrix estimation). Under assumptions VC1–VC3 and the conditions of Theorems 1 and 2, $\hat{\mathbf{A}}_T - \mathbf{A}_T \xrightarrow{p} \mathbf{0}$ and $\hat{\mathbf{D}}_T - \mathbf{D}_T \xrightarrow{p} \mathbf{0}$, where

$$\hat{\mathbf{A}}_T = T^{-1}\theta(1-\theta)\nabla' f(\hat{\boldsymbol\beta})\nabla f(\hat{\boldsymbol\beta}),$$

$$\hat{\mathbf{D}}_T = (2T\hat{c}_T)^{-1}\sum_{t=1}^{T} I(|y_t - f_t(\hat{\boldsymbol\beta})| < \hat{c}_T)\nabla' f_t(\hat{\boldsymbol\beta})\nabla f_t(\hat{\boldsymbol\beta}),$$

$\mathbf{A}_T$ and $\mathbf{D}_T$ have been defined in Theorem 2, and $\hat{c}_T$ is a bandwidth defined in assumption VC1.

Proof. See Appendix B.

In the proof of Theorem 1, we apply corollary 5.12 of White (1994), which establishes consistency results for nonlinear models of regression quantiles in a dynamic context. Assumption C1, requiring continuity in the vector of parameters $\boldsymbol\beta$ of the quantile specification, is clearly satisfied by all of the CAViaR models considered in this article. Assumptions C3 and C7 are identification conditions that are common in the regression quantile literature. Assumptions C4 and C5 are dominance conditions that rule out explosive behaviors (e.g., the CAViaR equivalent of an indirect integrated GARCH (IGARCH) process would not be covered by these conditions).

Derivation of the asymptotic distribution builds on the approximation of the discontinuous gradient of the objective function with a smooth differentiable function, so that the usual Taylor expansion can be performed. Assumptions AN1 and AN2 impose sufficient conditions on the function $f_t(\boldsymbol\beta)$ and on the conditional density function of the error terms to ensure that this smooth approximation will be sufficiently well behaved. The device for obtaining such an approximation is provided by an extension of theorem 3 of Huber (1967). This technique is standard in the regression quantile and LAD literature (Powell 1984, 1991; Weiss 1991). Alternative strategies for deriving the asymptotic distribution are the approach suggested by Amemiya (1982), based on the approximation of the regression quantile objective function by a continuously differentiable function, and the approach based on empirical processes as suggested by, for example, van de Geer (2000).

Regarding the variance–covariance matrix, note that $\hat{\mathbf{A}}_T$ is simply the outer product of the gradient. Estimation of the $\mathbf{D}_T$ matrix is less straightforward, because it involves the $h_t(0|\Omega_t)$ term. Following Powell (1984, 1986, 1991), we propose an estimator that combines kernel density estimation with the heteroscedasticity-consistent covariance matrix estimator of White (1980). Our Theorem 3 is a generalization of Powell's (1991) theorem 3 that accommodates the nonlinear dependent case. Buchinsky (1995) reported a Monte Carlo study on the estimation of the variance–covariance matrices in quantile regression models.

Note that all of the models considered in Section 3 satisfy the continuity and differentiability assumptions C1 and AN1 of Appendix A. The others are technical assumptions that are impossible to verify in finite samples.

5. TESTING QUANTILE MODELS

If model (4) is the true data generating process (DGP), then $\Pr[y_t < f_t(\boldsymbol\beta^0)] = \theta \;\forall t$. This is equivalent to requiring that the sequence of indicator functions $\{I(y_t < f_t(\boldsymbol\beta^0))\}_{t=1}^T$ be iid. Hence a property that any VaR estimate should satisfy is that of providing a filter to transform a (possibly) serially correlated and heteroscedastic time series into a serially independent sequence of indicator functions. A natural way to test the validity of the forecast model is to check whether the sequence $\{I(y_t < f_t(\boldsymbol\beta^0))\}_{t=1}^T \equiv \{I_t\}_{t=1}^T$ is iid, as was done by, for example, Granger, White, and Kamstra (1989) and Christoffersen (1998). Although these tests can detect the presence of serial correlation in the sequence of indicator functions $\{I_t\}_{t=1}^T$, this is only a necessary but not sufficient condition to assess the performance of a quantile model. Indeed, it is not difficult to generate a sequence of independent $\{I_t\}_{t=1}^T$ from a given sequence of $\{y_t\}_{t=1}^T$. It suffices to define a sequence of independent random variables $\{z_t\}_{t=1}^T$, such that

$$z_t = \begin{cases} 1 & \text{with probability } \theta \\ -1 & \text{with probability } (1-\theta). \end{cases} \quad (5)$$

Then setting $f_t(\boldsymbol\beta^0) = Kz_t$, for K large, will do the job. Notice, however, that once z_t is observed, the probability of exceeding the quantile is known to be almost 0 or 1. Thus the unconditional probabilities are correct and serially uncorrelated, but the conditional probabilities given the quantile are not. This example is an extreme case of quantile measurement error. Any noise introduced into the quantile estimate will change the conditional probability of a hit given the estimate itself.

Therefore, none of these tests has power against this form of misspecification and none can be simply extended to examine other explanatory variables. We propose a new test that can be easily extended to incorporate a variety of alternatives. Define

$$Hit_t(\boldsymbol\beta^0) \equiv I(y_t < f_t(\boldsymbol\beta^0)) - \theta. \quad (6)$$

The $Hit_t(\boldsymbol{\beta}^0)$ function assumes value $(1-\theta)$ every time y_t is less than the quantile and $-\theta$ otherwise. Clearly, the expected value of $Hit_t(\boldsymbol{\beta}^0)$ is 0. Furthermore, from the definition of the quantile function, the conditional expectation of $Hit_t(\boldsymbol{\beta}^0)$ given any information known at $t-1$ must also be 0. In particular, $Hit_t(\boldsymbol{\beta}^0)$ must be uncorrelated with its own lagged values and with $f_t(\boldsymbol{\beta}^0)$, and must have expected value equal to 0. If $Hit_t(\boldsymbol{\beta}^0)$ satisfies these moment conditions, then there will be no autocorrelation in the hits, no measurement error as in (5), and the correct fraction of exceptions. Whether there is the right proportion of hits in each calendar year can be determined by checking the correlation of $Hit_t(\boldsymbol{\beta}^0)$ with annual dummy variables. If other functions of the past information set, such as rolling standard deviations or a GARCH volatility estimate, are suspected of being informative, then these can be incorporated.

A natural way to set up a test is to check whether the test statistic $T^{-1/2}\mathbf{X}'(\hat{\boldsymbol{\beta}})\mathbf{Hit}(\hat{\boldsymbol{\beta}})$ is significantly different from 0, where $\mathbf{X}_t(\hat{\boldsymbol{\beta}})$, $t=1,\dots,T$, the typical row of $\mathbf{X}(\hat{\boldsymbol{\beta}})$ (possibly depending on $\hat{\boldsymbol{\beta}}$), is a q-vector measurable Ω_t and $\mathbf{Hit}(\hat{\boldsymbol{\beta}}) \equiv [Hit_1(\hat{\boldsymbol{\beta}}),\dots,Hit_T(\hat{\boldsymbol{\beta}})]'$.

Let $\mathbf{M}_T \equiv (\mathbf{X}'(\boldsymbol{\beta}^0) - E[T^{-1}\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0)]\mathbf{D}_T^{-1} \times \nabla' f(\boldsymbol{\beta}^0))$, where $\mathbf{H}$ is a diagonal matrix with typical entry $h_t(0|\Omega_t)$. Theorem 4 derives the in-sample distribution of the DQ test. The out-of-sample case is considered in Theorem 5.

Theorem 4 (In-sample dynamic quantile test). Under the assumptions of Theorems 1 and 2 and assumptions DQ1–DQ6,

$$\left[\theta(1-\theta)E(T^{-1}\mathbf{M}_T\mathbf{M}_T')\right]^{-1/2}T^{-1/2}\mathbf{X}'(\hat{\boldsymbol{\beta}})\mathbf{Hit}(\hat{\boldsymbol{\beta}}) \overset{d}{\sim} \mathrm{N}(\mathbf{0},\mathbf{I}).$$

If assumption DQ7 and the conditions of Theorem 3 also hold, then

$$DQ_{IS} \equiv \frac{\mathbf{Hit}'(\hat{\boldsymbol{\beta}})\mathbf{X}(\hat{\boldsymbol{\beta}})(\hat{\mathbf{M}}_T\hat{\mathbf{M}}_T')^{-1}\mathbf{X}'(\hat{\boldsymbol{\beta}})\mathbf{Hit}'(\hat{\boldsymbol{\beta}})}{\theta(1-\theta)} \overset{d}{\sim} \chi_q^2 \quad \text{as } T\to\infty$$

where

$$\hat{\mathbf{M}}_T \equiv \mathbf{X}'(\hat{\boldsymbol{\beta}}) - \left\{(2T\hat{c}_T)^{-1}\sum_{t=1}^{T} I\big(|y_t - f_t(\hat{\boldsymbol{\beta}})| < \hat{c}_T\big) \times \mathbf{X}_t'(\hat{\boldsymbol{\beta}})\nabla f_t(\hat{\boldsymbol{\beta}})\right\}\hat{\mathbf{D}}_T^{-1}\nabla' f(\hat{\boldsymbol{\beta}}).$$

Proof. See Appendix B.

If $\mathbf{X}(\hat{\boldsymbol{\beta}})$ contains $m<q$ lagged $Hit_{t-i}(\hat{\boldsymbol{\beta}})$ $(i=1,\dots,m)$, then $\mathbf{X}(\hat{\boldsymbol{\beta}}), \mathbf{Hit}(\hat{\boldsymbol{\beta}})$, and $\nabla f(\hat{\boldsymbol{\beta}})$ are not conformable, because $\mathbf{X}(\hat{\boldsymbol{\beta}})$ contains only $(T-m)$ elements. Here we implicitly assume, without loss of generality, that the matrices are made conformable by deleting the first m rows of $\nabla f(\hat{\boldsymbol{\beta}})$ and $\mathbf{X}(\hat{\boldsymbol{\beta}})$. Note that if we choose $\mathbf{X}(\hat{\boldsymbol{\beta}}) = \nabla f(\hat{\boldsymbol{\beta}})$, then $\mathbf{M}=\mathbf{0}$, where $\mathbf{0}$ is a (p,p) matrix of 0's. This is consistent with the fact that $T^{-1/2}\nabla' f(\hat{\boldsymbol{\beta}})\mathbf{Hit}(\hat{\boldsymbol{\beta}}) = o_p(1)$, by the first-order conditions of the regression quantile framework.

To derive the out-of-sample DQ test, let T_R denote the number of in-sample observations and let N_R denote the number of out-of-sample observations (with the dependence of T_R and N_R on R as specified in assumption DQ8). Make explicit the dependence of the relevant variables on the number of observations, using appropriate subscripts. Define the q-vector measurable Ω_n $\mathbf{X}_n(\hat{\boldsymbol{\beta}}_{T_R})$, $n=T_R+1,\dots,T_R+N_R$, as the typical row of $\mathbf{X}(\hat{\boldsymbol{\beta}}_{T_R})$, possibly depending on $\hat{\boldsymbol{\beta}}_{T_R}$, and $\mathbf{Hit}(\hat{\boldsymbol{\beta}}_{T_R}) \equiv [Hit_{T_R+1}(\hat{\boldsymbol{\beta}}_{T_R}),\dots,Hit_{T_R+N_R}(\hat{\boldsymbol{\beta}}_{T_R})]'$.

Theorem 5 (Out-of-sample dynamic quantile test). Under the assumptions of Theorems 1 and 2 and assumptions DQ1–DQ3, DQ8, and DQ9,

$$DQ_{OOS} \equiv N_R^{-1}\mathbf{Hit}'\big(\hat{\boldsymbol{\beta}}_{T_R}\big)\mathbf{X}\big(\hat{\boldsymbol{\beta}}_{T_R}\big)\big[\mathbf{X}'\big(\hat{\boldsymbol{\beta}}_{T_R}\big)\cdot\mathbf{X}\big(\hat{\boldsymbol{\beta}}_{T_R}\big)\big]^{-1} \times \mathbf{X}'\big(\hat{\boldsymbol{\beta}}_{T_R}\big)\mathbf{Hit}'\big(\hat{\boldsymbol{\beta}}_{T_R}\big)/\big(\theta(1-\theta)\big) \overset{d}{\sim} \chi_q^2 \quad \text{as } R\to\infty.$$

Proof. See Appendix B.

The in-sample DQ test is a specification test for the particular CAViaR process under study and it can be very useful for model selection purposes. The simpler version of the out-of-sample DQ test, instead, can be used by regulators to check whether the VaR estimates submitted by a financial institution satisfy some basic requirements that every good quantile estimate must satisfy, such as unbiasedeness, independent hits, and independence of the quantile estimates. The nicest features of the out-of-sample DQ test are its simplicity and the fact that it does not depend on the estimation procedure: to implement it, the evaluator (either the regulator or the risk manager) just needs a sequence of VaRs and the corresponding values of the portfolio.

6. EMPIRICAL RESULTS

To implement our methodology on real data, a researcher needs to construct the historical series of portfolio returns and to choose a specification of the functional form of the quantile. We took a sample of 3,392 daily prices from Datastream for General Motors (GM), IBM, and the S&P 500, and computed the daily returns as 100 times the difference of the log of the prices. The samples range from April 7, 1986, to April 7, 1999. We used the first 2,892 observations to estimate the model and the last 500 for out-of-sample testing. We estimated 1% and 5% 1-day VaRs, using the four CAViaR specifications described in Section 3. For the adaptive model, we set $G=10$, where G entered the definition of the adaptive model in Section 3. In principle, the parameter G itself could be estimated; however, this would go against the spirit of this model, which is simplicity. The 5% VaR estimates for GM are plotted in Figure 1, and all of the results are reported in Table 1.

The table presents the value of the estimated parameters, the corresponding standard errors and (one-sided) p values, the value of the regression quantile objective function [eq. (3)], the percentage of times the VaR is exceeded, and the p value of the DQ test, both in-sample and out-of-sample. To compute the VaR series with the CAViaR models, we initialize $f_1(\boldsymbol{\beta})$ to the empirical θ-quantile of the first 300 observations. The instruments used in the out-of-sample DQ test were a constant, the VaR forecast and the first four lagged hits. For the in-sample DQ test, we did not include the constant and the VaR forecast, because for some models there was collinearity

Table 1. *Estimates and Relevant Statistics for the Four CAViaR Specification*

	Symmetric absolute value			*Asymmetric slope*			*Indirect GARCH*			*Adaptive*		
	GM	*IBM*	*S&P 500*	*GM*	*IBM*	*S&P 500*	*GM*	*IBM*	*S&P 500*	*GM*	*IBM*	*S&P 500*
1% VaR												
Beta1	**.4511**	.1261	**.2039**	.3734	.0558	**.1476**	1.4959	1.3289	**.2328**	**.2968**	**.1626**	**.5562**
Standard errors	.2028	.0929	.0604	.2418	.0540	.0456	.9252	1.9488	.1191	.1109	.0736	.1150
p values	.0131	.0872	.0004	.0613	.1509	.0006	.0530	.2477	.0253	.0037	.0136	0
Beta2	**.8263**	**.9476**	**.8732**	**.7995**	**.9423**	**.8729**	**.7804**	**.8740**	**.8350**			
Standard errors	.0826	.0501	.0507	.0869	.0247	.0302	.0590	.1133	.0225			
p values	0	0	0	0	0	0	0	0	0			
Beta3	**.3305**	.1134	.3819	**.2779**	.0499	−.0139	.9356	**.3374**	1.0582			
Standard errors	.1685	.1185	.2772	.1398	.0563	.1148	1.2619	.0953	1.0983			
p values	.0249	.1692	.0842	.0235	.1876	.4519	.2292	.0002	.1676			
Beta4				**.4569**	**.2512**	**.4969**						
Standard errors				.1787	.0848	.1342						
p values				.0053	.0015	.0001						
RQ	172.04	182.32	109.68	169.22	179.40	105.82	170.99	183.43	108.34	179.61	192.20	117.42
Hits in-sample (%)	1.0028	.9682	1.0028	.9682	1.0373	.9682	1.0028	1.0028	1.0028	.9682	1.2448	.9336
Hits out-of-sample (%)	1.4000	1.6000	1.8000	1.4000	1.6000	1.6000	1.2000	1.6000	1.8000	1.8000	1.6000	1.2000
DQ in-sample (*p* values)	.6349	.5375	.3208	.5958	.7707	.5450	.5937	.5798	.7486	.0117	0*	.1697
DQ out-of-sample (*p* values)	.8965	.0326	.0191	.9432	.0431	.0476	.9305	.0350	.0309	.0017*	.0009*	.0035*
5% VaR												
Beta1	**.1812**	.1191	**.0511**	**.0760**	**.0953**	**.0378**	**.3336**	**.5387**	**.0262**	**.2871**	**.3969**	**.3700**
Standard errors	.0833	.0839	.0083	.0249	.0532	.0135	.1039	.1569	.0100	.0506	.0812	.0767
p values	.0148	.0778	0	.0011	.0366	.0026	.0007	.0003	.0043	0	0	0
Beta2	**.8953**	**.9053**	**.9369**	**.9326**	**.8892**	**.9025**	**.9042**	**.8259**	**.9287**			
Standard errors	.0361	.0500	.0224	.0194	.0385	.0144	.0134	.0294	.0061			
p values	0	0	0	0	0	0	0	0	0			
Beta3	**.1133**	**.1481**	**.1341**	.0398	**.0617**	**.0377**	.1220	.1591	.1407			
Standard errors	.0122	.0348	.0517	.0322	.0272	.0224	.1149	.1152	.6198			
p values	0	0	.0047	.1088	.0117	.0457	.1441	.0836	.4102			
Beta4				**.1218**	**.2187**	**.2871**						
Standard errors				.0405	.0465	.0258						
p values				.0013	0	0						
RQ	550.83	522.43	306.68	548.31	515.58	300.82	552.12	524.79	305.93	553.79	527.72	312.06
Hits in-sample (%)	4.9793	5.0138	5.0484	4.9101	4.9793	5.0138	4.9793	5.0484	5.0138	4.9101	4.8409	4.7372
Hits out-of-sample (%)	4.8000	6.0000	5.6000	5.0000	7.4000	6.4000	4.6000	7.4000	5.8000	6.0000	5.0000	4.6000
DQ in-sample (*p* values)	.3609	.0824	.3685	.9132	.6149	.9540	.1037	.1727	.2661	.0543	.0032*	.0380
DQ out-of-sample (*p* values)	.9855	.0884	.0005*	.9235	.0071*	.0007*	.8770	.1208	.0001*	.3681	.5021	.0240

NOTE: Significant coefficients at 5% formatted in bold; "*" denotes rejection from the DQ test at 1% significance level.

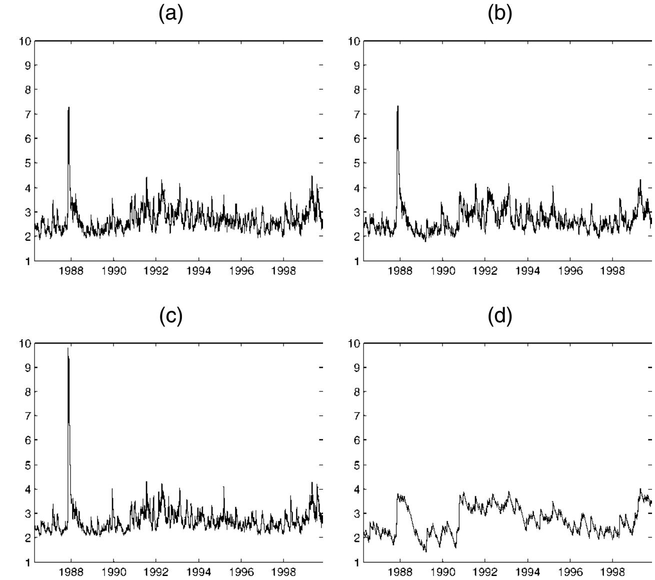

Figure 1. 5% Estimated CAViaR Plots for GM: (a) Symmetric Absolute Value; (b) Asymmetric Slope; (c) GARCH; (d) Adaptive. Because VaR is usually reported as a positive number, we set $\hat{VaR}_{t-1} = -f_{t-1}(\hat{\beta})$. The sample ranges from April 7, 1986, to April 7, 1999. The spike at the beginning of the sample is the 1987 crash. The increase in the quantile estimates toward the end of the sample reflects the increase in overall volatility following the Russian and Asian crises.

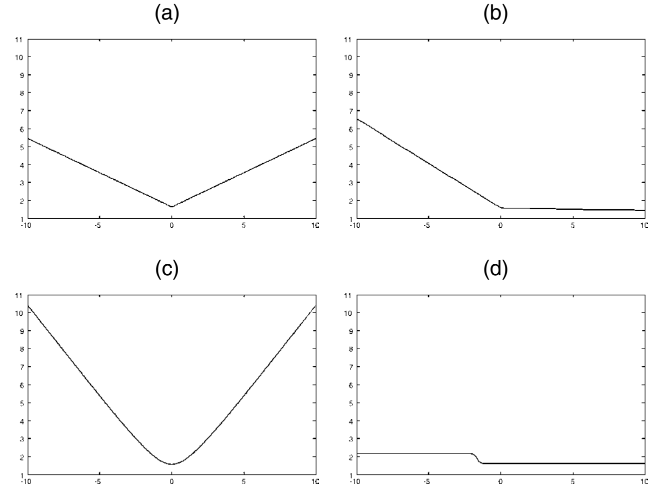

Figure 2. 1% CAViaR News Impact Curve for S&P 500 for (a) Symmetric Absolute Value, (b) Asymmetric Slope, (c) Indirect GARCH, and (d) Adaptive. For given estimated parameter vector $\hat{\beta}$ and setting (arbitrarily) $\hat{VaR}_{t-1} = -1.645$, the CAViaR news impact curve shows how $\hat{VaR}_t$ changes as lagged portfolio returns y_{t-1} vary. The strong asymmetry of the asymmetric slope news impact curve suggests that negative returns might have a much stronger effect on the VaR estimate than positive returns.

with the matrix of derivatives. We computed the standard errors and the variance–covariance matrix of the in-sample DQ test as described in Theorems 3 and 4. The formulas to compute $\hat{\mathbf{D}}_T$ and $\hat{\mathbf{M}}_T$ were implemented using k-nearest neighbor estimators, with $k = 40$ for 1% VaR and $k = 60$ for 5% VaR.

As optimization routines, we used the Nelder–Mead simplex algorithm and a quasi-Newton method. All of the computations were done in MATLAB 6.1, using the functions *fminsearch* and *fminunc* as optimization algorithms. The loops to compute the recursive quantile functions were coded in C.

We optimized the models using the following procedure. We generated n vectors using a uniform random number generator between 0 and 1. We computed the regression quantile (RQ) function described in equation (3) for each of these vectors and selected the m vectors that produced the lowest RQ criterion as initial values for the optimization routine. We set $n = [10^4, 10^5, 10^4, 10^4]$ and $m = [10, 15, 10, 5]$ for the symmetric absolute value, asymmetric slope, Indirect GARCH, and adaptive models. For each of these initial values, we ran first the simplex algorithm. We then fed the optimal parameters to the quasi-Newton algorithm and chose the new optimal parameters as the new initial conditions for the simplex. We repeated this procedure until the convergence criterion was satisfied. Tolerance levels for the function and the parameters values were set to 10^{-10}. Finally, we selected the vector that produced the lowest RQ criterion. An alternative optimization routine is the interior point algorithm for nonlinear regression quantiles suggested by Koenker and Park (1996).

Figure 2 plots the CAViaR news impact curve for the 1% VaR estimates of the S&P 500. Notice how the adaptive and the asymmetric slope news impact curves differ from the others. For both indirect GARCH and symmetric absolute value models, past returns (either positive or negative) have a symmetric impact on VaR. In contrast, for the adaptive model, the most important news is whether or not past returns exceeded the previous VaR estimate. Finally, the sharp difference between the impact of positive and negative returns in the asymmetric slope model suggests that there might be relevant asymmetries in the behavior of the 1% quantile of this portfolio.

Turning our attention to Table 1, the first striking result is that the coefficient of the autoregressive term (β_2) is always very significant. This confirms that the phenomenon of clustering of volatilities is relevant also in the tails. A second interesting point is the precision of all the models, as measured by the percentage of in-sample hits. This is not surprising, because the objective function of RQ models is designed exactly to achieve this kind of result. The results for the 1% VaR show that the symmetric absolute value, asymmetric slope, and indirect GARCH models do a good job describing the evolution of the left tail for the three assets under study. The results are particularly good for GM, producing a rather accurate percentage of hits out-of-sample (1.4% for the symmetric absolute value and the asymmetric slope and 1.2% for the indirect GARCH). The performance of the adaptive model is inferior both in-sample and out-of-sample, even though the percentage of hits seems reasonably close to 1. This shows that looking only at the number of exceptions, as suggested by the Basle Committee on Banking Supervision (1996), may be a very unsatisfactory way of evaluating the performance of a VaR model. But 5% results present a different picture. All the models perform well with GM. Note the remarkable precision of the percentage of out-of-sample hits generated by the asymmetric slope model (5.0%). Also notice that this time also the adaptive model is not rejected by the DQ tests. For IBM, the asymmetric slope, of which the symmetric absolute value is a special case, tends to overfit in-sample, providing a very poor performance out-of-sample. Finally, for the S&P 500 5% VaR, only the adaptive model survives the DQ test at the 1% confidence level, producing a rather accurate number of out-of-sample hits (4.6%). The poor out-of-sample performance of the other models can probably be explained by the fact that the last part of the sample of the S&P 500 is characterized by a sudden spur of volatility and roughly coincides with our out-of-sample period. Finally, it is interesting to note in the asymmetric slope model how the coefficients of the negative part of lagged returns are always strongly significant, whereas those associated to positive returns are sometimes not significantly different from 0. This indicates the presence of strong asymmetric impacts on VaR of lagged returns.

The fact that the DQ tests select different models for different confidence levels suggests that the process governing the tail behavior might change as we move further out in the tail. In particular, this contradicts the assumption behind GARCH and RiskMetrics, because these approaches implicitly assume that the tails follow the same process as the rest of the returns. Although GARCH might be a useful model for describing the evolution of volatility, the results in this article show that it might provide an unsatisfactory approximation when applied to tail estimation.

7. CONCLUSION

We have proposed a new approach to VaR estimation. Most existing methods estimate the distribution of the returns and then recover its quantile in an indirect way. In contrast, we directly model the quantile. To do this, we introduce a new class of models, the CAViaR models, which specify the evolution of the quantile over time using a special type of autoregressive process. We estimate the unknown parameters by minimizing the RQ loss function. We also introduced the DQ test, a new test to evaluate the performance of quantile models. Applications to real data illustrate the ability of CAViaR models to adapt to new risk environments. Moreover, our findings suggest that the process governing the behavior of the tails might be different from that of the rest of the distribution.

ACKNOWLEDGMENTS

The authors thank Hal White, Moshe Buchinsky, Wouter Den Haan, Joel Horowitz, Jim Powell, and the participants of the UCSD, Yale, MIT-Harvard, Wharton, Iowa, and NYU econometric seminars for their valuable comments. The views expressed in this article are those of the authors and do not necessarily reflect those of the European Central Bank.

APPENDIX A: ASSUMPTIONS

Consistency Assumptions

C0. (Ω, F, P) is a complete probability space, and $\{\varepsilon_{t\theta}, \mathbf{x}_t\}$, $t = 1, 2, \ldots$, are random vectors on this space.

C1. The function $f_t(\boldsymbol{\beta}) : \Re^{k_t} \times B \rightarrow \Re$ is such that for each $\boldsymbol{\beta} \in B$, a compact subset of $\Re^p$, $f_t(\boldsymbol{\beta})$ is measurable with respect to the information set Ω_t and $f_t(\cdot)$ is continuous in B, $t = 1, 2, \ldots$, for a given choice of explanatory variables $\{y_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1\}$.

C2. Conditional on all of the past information Ω_t, the error terms $\varepsilon_{t\theta}$ form a stationary process, with continuous conditional density $h_t(\varepsilon|\Omega_t)$.

C3. There exists $h > 0$ such that for all t, $h_t(0|\Omega_t) \geq h$.

C4. $|f_t(\boldsymbol{\beta})| < K(\Omega_t)$ for each $\boldsymbol{\beta} \in B$ and for all t, where $K(\Omega_t)$ is some (possibly) stochastic function of variables that belong to the information set, such that $E(|K(\Omega_t)|) \leq K_0 < \infty$, for some constant K_0.

C5. $E[|\varepsilon_{t\theta}|] < \infty$ for all t.

C6. $\{[\theta - I(y_t < f_t(\boldsymbol{\beta}))][y_t - f_t(\boldsymbol{\beta})]\}$ obeys the uniform law of large numbers.

C7. For every $\xi > 0$, there exists a $\tau > 0$ such that if $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \geq \xi$, then $\liminf_{T\rightarrow\infty} T^{-1} \sum P[|f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0)| > \tau] > 0$.

Asymptotic Normality Assumptions

AN1. $f_t(\boldsymbol{\beta})$ is differentiable in B and for all $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in a neighborhood υ_0 of $\boldsymbol{\beta}^0$, such that $\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| \leq d$ for d sufficiently small and for all t:
  (a) $\|\nabla f_t(\boldsymbol{\beta})\| \leq F(\Omega_t)$, where $F(\Omega_t)$ is some (possibly) stochastic function of variables that belong to the information set and $E(F(\Omega_t)^3) \leq F_0 < \infty$, for some constant F_0.
  (b) $\|\nabla f_t(\boldsymbol{\beta}) - \nabla f_t(\boldsymbol{\gamma})\| \leq M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = O(\|\boldsymbol{\beta} - \boldsymbol{\gamma}\|)$, where $M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is some function such that $E[M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})]^2 \leq M_0\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$ and $E[M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})F(\Omega_t)] \leq M_1\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$ for some constants M_0 and M_1.

AN2. (a) $h_t(\varepsilon|\Omega_t) \leq N < \infty \; \forall\, t$, for some constant N.
  (b) $h_t(\varepsilon|\Omega_t)$ satisfies the Lipschitz condition $|h_t(\lambda_1|\Omega_t) - h_t(\lambda_2|\Omega_t)| \leq L|\lambda_1 - \lambda_2|$ for some constant $L < \infty \; \forall\, t$.

AN3. The matrices $\mathbf{A}_T \equiv E[T^{-1}\theta(1-\theta)\sum_{t=1}^T \nabla' f_t(\boldsymbol{\beta}^0) \times \nabla f_t(\boldsymbol{\beta}^0)]$ and $\mathbf{D}_T \equiv E[T^{-1}\sum_{t=1}^T h_t(0|\Omega_t)\nabla' f_t(\boldsymbol{\beta}^0) \times \nabla f_t(\boldsymbol{\beta}^0)]$ have the smallest eigenvalues bounded below by a positive constant for T sufficiently large.

AN4. The sequence $\{T^{-1/2}\sum_{t=1}^T [\theta - I(y_t < f_t(\boldsymbol{\beta}^0))] \cdot \nabla' f_t(\boldsymbol{\beta}^0)\}$ obeys the central limit theorem.

Variance–Covariance Matrix Estimation Assumptions

VC1. $\hat{c}_T/c_T \xrightarrow{p} 1$, where the nonstochastic positive sequence c_T satisfies $c_T = o(1)$ and $c_T^{-1} = o(T^{1/2})$.

VC2. $E(|F(\Omega_t)|^4) \leq F_1 < \infty$ for all t and for some constant F_1, where $F(\Omega_t)$ has been defined in assumption AN1(a).

VC3. $T^{-1}\theta(1-\theta)\sum_{t=1}^T \nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) - \mathbf{A}_T \xrightarrow{p} \mathbf{0}$ and $T^{-1}\sum_{t=1}^T h_t(0|\Omega_t)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) - \mathbf{D}_T \xrightarrow{p} \mathbf{0}$.

In-Sample Dynamic Quantile Test Assumption

DQ1. $\mathbf{X}_t(\boldsymbol{\beta})$ is different element wise from $\nabla f_t(\boldsymbol{\beta})$, is measurable Ω_t, $\|\mathbf{X}_t(\boldsymbol{\beta})\| \leq W(\Omega_t)$, where $W(\Omega_t)$ is some (possibly) stochastic function of variables that belong to the information set, such that $E[W(\Omega_t) \times M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})] \leq W_0\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$ and $E[[W(\Omega_t) \cdot F(\Omega_t)]^2] < W_1 < \infty$ for some finite constants W_0 and W_1, and $F(\Omega_t)$ and $M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})$ have been defined in AN1.

DQ2. $\|\mathbf{X}_t(\boldsymbol{\beta}) - \mathbf{X}_t(\boldsymbol{\gamma})\| \le S(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})$, where $E[S(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})] \le S_0\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$, $E[W(\Omega_t)S(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})] \le S_1\|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$, and for some constant S_0.

DQ3. Let $\{\varepsilon_t^1, \dots, \varepsilon_t^{J_i}\}$ the set of values for which $\mathbf{X}_t(\boldsymbol{\beta})$ is not differentiable. Then $\Pr(\varepsilon_{t\theta} = \varepsilon_t^j) = 0$ for $j = 1, \dots, J_i$. Whenever the derivative exists, $\|\nabla \mathbf{X}_t(\boldsymbol{\beta})\| \le Z(\Omega_t)$, where $Z(\Omega_t)$ is some (possibly) stochastic function of variables that belong to the information set, such that $E[Z(\Omega_t)^r] < Z_0 < \infty$, $r = 1, 2$, for some constant Z_0.

DQ4. $T^{-1}\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0) - E[T^{-1}\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0)] \xrightarrow{p} \mathbf{0}$.

DQ5. $T^{-1}\mathbf{M}_T\mathbf{M}_T' - T^{-1}E(\mathbf{M}_T\mathbf{M}_T') \xrightarrow{p} \mathbf{0}$, where $\mathbf{M}_T \equiv \mathbf{X}'(\boldsymbol{\beta}^0) - E[T^{-1}\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0)] \cdot \mathbf{D}_T^{-1} \cdot \nabla' f(\boldsymbol{\beta}^0)$.

DQ6. The sequence $\{T^{-1/2}\mathbf{M}_T\mathbf{Hit}(\boldsymbol{\beta}^0)\}$ obeys the central limit theorem.

DQ7. $T^{-1}E(\mathbf{M}_T\mathbf{M}_T')$ is a nonsingular matrix.

Out-of-Sample Dynamic Quantile Test Assumptions

DQ8. $\lim_{R\to\infty} T_R = \infty$, $\lim_{R\to\infty} N_R = \infty$, and $\lim_{R\to\infty} N_R/T_R = 0$.

DQ9. The sequence $\{N_R^{-1/2}\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{Hit}(\boldsymbol{\beta}^0)\}$ obeys the central limit theorem.

APPENDIX B: PROOFS

Proof of Theorem 1

We verify that the conditions of corollary 5.12 of White (1994, p. 75) are satisfied. We check only assumptions 3.1 and 3.2 of White's corollary, because the others are obviously satisfied.

Let $Q_T(\boldsymbol{\beta}) \equiv T^{-1}\sum_{t=1}^T q_t(\boldsymbol{\beta})$, where $q_t(\boldsymbol{\beta}) \equiv [\theta - I(y_t < f_t(\boldsymbol{\beta}))][y_t - f_t(\boldsymbol{\beta})]$. First, we need to show that $E[q_t(\boldsymbol{\beta})]$ exists and is finite for every $\boldsymbol{\beta}$. This can be easily checked as follows: $E[q_t(\boldsymbol{\beta})] < E|y_t - f_t(\boldsymbol{\beta})| \le E|\varepsilon_{t\theta}| + E|f_t(\boldsymbol{\beta})| + E|f_t(\boldsymbol{\beta}^0)| < \infty$, by assumptions C4 and C5. Moreover, because f is continuous in $\boldsymbol{\beta}$ by assumption C1, $q_t(\boldsymbol{\beta})$ is continuous [because the regression quantile objective function is continuous in $f(\boldsymbol{\beta})$], and hence its expected value (which we just showed to be finite) will be also continuous. It remains to show that $E[V_T(\boldsymbol{\beta})] = E[Q_T(\boldsymbol{\beta}) - Q_T(\boldsymbol{\beta}^0)]$ is uniquely minimized at $\boldsymbol{\beta}^0$ for T sufficiently large. Let $v_t(\boldsymbol{\beta}) \equiv q_t(\boldsymbol{\beta}) - q_t(\boldsymbol{\beta}^0)$. Note that $q_t(\boldsymbol{\beta}) = [\theta - I(\varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}))][\varepsilon_{t\theta} - \delta_t(\boldsymbol{\beta})]$, where $\delta_t(\boldsymbol{\beta}) \equiv f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0)$. Then

$$v_t(\boldsymbol{\beta}) = \begin{cases} (1-\theta)\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} < 0 \\ (1-\theta)\delta_t(\boldsymbol{\beta}) - \varepsilon_{t\theta} & \text{if } \varepsilon_{t\theta} < \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} > 0 \\ \varepsilon_{t\theta} - \theta\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} \ge \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} < 0 \\ -\theta\delta_t(\boldsymbol{\beta}) & \text{if } \varepsilon_{t\theta} \ge \delta_t(\boldsymbol{\beta}) \text{ and } \varepsilon_{t\theta} > 0. \end{cases}$$

After some algebra, it can be shown that

$$E[v_t(\boldsymbol{\beta})|\Omega_t] = I\big(\delta_t(\boldsymbol{\beta}) < 0\big)\int_{-|\delta_t(\boldsymbol{\beta})|}^0 \big(\lambda + |\delta_t(\boldsymbol{\beta})|\big)h_t(\lambda|\Omega_t)\,d\lambda + I\big(\delta_t(\boldsymbol{\beta}) > 0\big)\int_0^{|\delta_t(\boldsymbol{\beta})|}\big(|\delta_t(\boldsymbol{\beta})| - \lambda\big)h_t(\lambda|\Omega_t)\,d\lambda.$$

Reasoning following Powell (1984), the continuity of $h_t(\cdot|\Omega_t)$ (assumption C2) and assumption C3 imply that there exist $h_1 > 0$ such that $h_t(\lambda|\Omega_t) > h_1$ whenever $|\lambda| < h_1$. Hence, for any $0 < \tau < h_1$,

$$\begin{aligned} E[v_t(\boldsymbol{\beta})|\Omega_t] &\ge I\big(\delta_t(\boldsymbol{\beta}) < -\tau\big)\int_{-\tau}^0 [\lambda + \tau]h_1\,d\lambda \\ &\quad + I\big(\delta_t(\boldsymbol{\beta}) > \tau\big)\int_0^\tau [\tau - \lambda]h_1\,d\lambda \\ &= \tfrac{1}{2}\tau^2 h_1 I\big(|\delta_t(\boldsymbol{\beta})| > \tau\big). \end{aligned}$$

Therefore, taking the unconditional expectation,

$$\begin{aligned} E[V_T(\boldsymbol{\beta})] &\equiv E\left[T^{-1}\sum_{t=1}^T v_t(\boldsymbol{\beta})\right] \\ &\ge \tfrac{1}{2}\tau^2 h_1 T^{-1}\sum_{t=1}^T \Pr\big[|f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0)| > \tau\big], \end{aligned}$$

which is greater than 0 by assumption C7 if $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \ge \xi$.

Proof of Theorem 2

The proof builds on Huber's (1967) theorem 3. Weiss (1991) showed that Huber's conclusion also holds in the case of non-iid dependent random variables. Define $Hit_t(\boldsymbol{\beta}) \equiv I(y_t < f_t(\boldsymbol{\beta})) - \theta$ and $g_t(\boldsymbol{\beta}) \equiv Hit_t(\boldsymbol{\beta})\nabla' f_t(\boldsymbol{\beta})$. The strategy of the proof follows three steps: (1) Show that Huber's theorem holds; (2) apply Huber's theorem; and (3) apply the central limit theorem to $T^{-1/2}\sum_{t=1}^T g_t(\boldsymbol{\beta}^0)$.

To verify Huber's conditions, define $\lambda_T(\boldsymbol{\beta}) \equiv T^{-1} \times \sum_{t=1}^T E[g_t(\boldsymbol{\beta})]$ and $\mu_t(\boldsymbol{\beta}, d) \equiv \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \le d}\|g_t(\boldsymbol{\tau}) - g_t(\boldsymbol{\beta})\|$. Here we only show that $T^{-1/2}\sum_{t=1}^T g_t(\hat{\boldsymbol{\beta}}) = o_p(1)$ and that assumptions (N2) and (N3) in the proof of Theorem 3 of Weiss (1991) are satisfied, because the other conditions are easily checked. We follow Ruppert and Carroll's (1980) strategy (see Lemmas A1 and A2). Let $\{\mathbf{e}_j\}_{j=1}^p$ be the standard basis of $\Re^p$ and define $Q_j(a) \equiv -T^{-1/2}\sum_{t=1}^T q_t(\hat{\boldsymbol{\beta}} + a\mathbf{e}_j)$, where a is a scalar. Let $G_j(a)$ be the (finite) one-sided derivative of $Q_j(a)$, that is,

$$G_j(a) \equiv -T^{-1/2}\sum_{t=1}^T \nabla_j f_t(\hat{\boldsymbol{\beta}} + a\mathbf{e}_j)Hit_t(\hat{\boldsymbol{\beta}} + a\mathbf{e}_j).$$

Because $Q_j(a)$ is continuous in a and achieves a maximum at 0, it must be that in a neighborhood of 0, for some $\xi > 0$,

$$\begin{aligned} |G_j(0)| &\le G_j(\xi) - G_j(-\xi) \\ &= T^{-1/2}\sum_{t=1}^T\big[-\nabla_j f_t(\hat{\boldsymbol{\beta}} + \xi\mathbf{e}_j)Hit_t(\hat{\boldsymbol{\beta}} + \xi\mathbf{e}_j) \\ &\qquad + \nabla_j f_t(\hat{\boldsymbol{\beta}} - \xi\mathbf{e}_j)Hit_t(\hat{\boldsymbol{\beta}} - \xi\mathbf{e}_j)\big]. \end{aligned}$$

Now note that by taking the limit for $\xi \to 0$,

$$\begin{aligned} |G_j(0)| &\le T^{-1/2}\sum_{t=1}^T |\nabla_j f_t(\hat{\boldsymbol{\beta}})| I\big(y_t = f_t(\hat{\boldsymbol{\beta}})\big) \\ &\le T^{-1/2}\Big[\max_{1\le t\le T} F(\Omega_t)\Big] \cdot \sum_{t=1}^T I\big(y_t = f_t(\hat{\boldsymbol{\beta}})\big). \end{aligned}$$

But $T^{-1/2}[\max_{1\le t\le T} F(\Omega_t)] = o_p(1)$ by assumption AN1(a) and $\sum_{t=1}^{T} I(y_t = f_t(\hat{\boldsymbol\beta})) = O_{\text{a.s.}}(1)$ by assumption C2. Therefore, $G_j(0) \xrightarrow{p} 0$. Since this holds for every j, this implies that $T^{-1/2}\sum_{t=1}^{T} g_t(\hat{\boldsymbol\beta}) = o_p(1)$.

For (N2) of Weiss's (1991) proof of Theorem 3, note that $\lambda_T(\boldsymbol\beta^0)$ is well defined, because $\boldsymbol\beta^0$ is an interior point of B by assumption AN1. Then $E[g_t(\boldsymbol\beta^0)] = E[E(Hit_t(\boldsymbol\beta^0)|\Omega_t) \times \nabla' f_t(\boldsymbol\beta^0)] = 0$, by the assumption that model (4) is specified correctly.

For N3(i), using the mean value theorem to expand $\lambda_T(\boldsymbol\beta)$ around $\boldsymbol\beta^0$ (applying Leibnitz's rule for differentiating under the integral sign), we get

$$\begin{aligned}\lambda_T(\boldsymbol\beta) &= T^{-1}\sum_{t=1}^{T} E\biggl\{\nabla' f_t(\boldsymbol\beta^*)\nabla f_t(\boldsymbol\beta^*)\\ &\qquad\times\biggl[I\bigl(\delta_t(\boldsymbol\beta^*)>0\bigr)\int_0^{\delta_t(\boldsymbol\beta^*)} h_t(\lambda|\Omega_t)\,d\lambda\\ &\qquad - I\bigl(\delta_t(\boldsymbol\beta^*)<0\bigr)\int_{\delta_t(\boldsymbol\beta^*)}^{0} h_t(\lambda|\Omega_t)\,d\lambda\biggr]\biggr\}\\ &\quad + T^{-1}\sum_{t=1}^{T} E\bigl\{\nabla' f_t(\boldsymbol\beta^*)\cdot\nabla f_t(\boldsymbol\beta^*) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\bigr\}\\ &\quad\times(\boldsymbol\beta-\boldsymbol\beta^0)\\ &\equiv \boldsymbol\Lambda_T(\boldsymbol\beta^*)(\boldsymbol\beta-\boldsymbol\beta^0),\end{aligned}$$

where $\boldsymbol\beta^*$ lies between $\boldsymbol\beta$ and $\boldsymbol\beta^0$. We now show that $\boldsymbol\Lambda_T(\boldsymbol\beta^*) = \mathbf{D}_T + O(\|\boldsymbol\beta-\boldsymbol\beta^0\|)$, where $\mathbf{D}_T$ is defined in assumption AN3.

$$\begin{aligned}&\|\boldsymbol\Lambda_T(\boldsymbol\beta^*) - \mathbf{D}_T\|\\ &\quad = \biggl\| T^{-1}\sum_{t=1}^{T} E\biggl\{\nabla' f_t(\boldsymbol\beta^*)\nabla f_t(\boldsymbol\beta^*)\\ &\qquad\times\biggl[I\bigl(\delta_t(\boldsymbol\beta^*)>0\bigr)\int_0^{\delta_t(\boldsymbol\beta^*)} h_t(\lambda|\Omega_t)\,d\lambda\\ &\qquad - I\bigl(\delta_t(\boldsymbol\beta^*)<0\bigr)\int_{\delta_t(\boldsymbol\beta^*)}^{0} h_t(\lambda|\Omega_t)\,d\lambda\biggr]\biggr\}\\ &\qquad + T^{-1}\sum_{t=1}^{T} E\bigl[\nabla' f_t(\boldsymbol\beta^*)\nabla f_t(\boldsymbol\beta^*) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\\ &\qquad - \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^0) h_t(0|\Omega_t)\bigr]\biggr\|.\end{aligned}$$

The first line of the foregoing expression can be shown to be $O(\|\boldsymbol\beta^*-\boldsymbol\beta^0\|)$; therefore, $O(\|\boldsymbol\beta-\boldsymbol\beta^0\|)$, by a mean value expansion of the integrals around $\boldsymbol\beta^0$. For the second line, invoking assumptions AN1(a), AN1(b), AN2(a), and AN2(b), note that it can be rewritten as

$$\begin{aligned}&\biggl\| T^{-1}\sum_{t=1}^{T} E\Bigl[\nabla' f_t(\boldsymbol\beta^*)\nabla f_t(\boldsymbol\beta^*) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\\ &\qquad - \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^*) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\\ &\qquad + \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^*) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\\ &\qquad - \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^0) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\\ &\qquad + \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^0) h_t\bigl(\delta_t(\boldsymbol\beta^*)|\Omega_t\bigr)\\ &\qquad - \nabla' f_t(\boldsymbol\beta^0)\nabla f_t(\boldsymbol\beta^0) h_t(0|\Omega_t)\Bigr]\biggr\|\\ &\quad\le T^{-1}\sum_{t=1}^{T} E\bigl[M(\Omega_t,\boldsymbol\beta^*,\boldsymbol\beta^0)\cdot F(\Omega_t)\cdot N\\ &\qquad + N\cdot F(\Omega_t)\cdot M(\Omega_t,\boldsymbol\beta^*,\boldsymbol\beta^0)\\ &\qquad + F(\Omega_t)^3\cdot L\|\boldsymbol\beta-\boldsymbol\beta^0\|\bigr]\\ &\quad\le T^{-1}\sum_{t=1}^{T}\bigl(2N\cdot M_1\cdot\|\boldsymbol\beta-\boldsymbol\beta^0\| + F_0 L\|\boldsymbol\beta-\boldsymbol\beta^0\|\bigr)\\ &\quad\le (2N\cdot M_1 + F_0 L)\|\boldsymbol\beta-\boldsymbol\beta^0\|\\ &\quad = O(\|\boldsymbol\beta-\boldsymbol\beta^0\|).\end{aligned}$$

Therefore,

$$\lambda_T(\boldsymbol\beta) = \mathbf{D}_T(\boldsymbol\beta-\boldsymbol\beta^0) + O(\|\boldsymbol\beta-\boldsymbol\beta^0\|^2). \tag{B.1}$$

But because $\mathbf{D}_T$ is positive definite for T sufficiently large by assumption AN3, the result follows.

For N3(ii), noting that $|Hit_t(\boldsymbol\tau) - Hit_t(\boldsymbol\beta)| = I(|y_t - f_t(\boldsymbol\beta)| < |f_t(\boldsymbol\tau) - f_t(\boldsymbol\beta)|)$, we have

$$\begin{aligned}&\mu_t(\boldsymbol\beta,\delta)\\ &\quad\le \sup_{\|\boldsymbol\tau-\boldsymbol\beta\|\le d}\|\nabla' f_t(\boldsymbol\tau) - \nabla' f_t(\boldsymbol\beta)\|\\ &\qquad + \sup_{\|\boldsymbol\tau-\boldsymbol\beta\|\le d}\|\nabla' f_t(\boldsymbol\beta)\|\cdot I\bigl(|y_t - f_t(\boldsymbol\beta)| < |f_t(\boldsymbol\tau) - f_t(\boldsymbol\beta)|\bigr).\end{aligned}$$

Thus,

$$\begin{aligned}&\mu_t(\boldsymbol\beta,\delta)\\ &\quad\le M(\Omega_t,\boldsymbol\beta,\boldsymbol\tau) + F(\Omega_t)\cdot I\bigl(|y_t - f_t(\boldsymbol\beta)| < |f_t(\boldsymbol\tau) - f_t(\boldsymbol\beta)|\bigr)\end{aligned}$$

and

$$\begin{aligned}E[\mu_t(\boldsymbol\beta,\delta)] &\le M_0 d + E\bigl[F(\Omega_t)\cdot 2|\nabla f_t(\boldsymbol\tau^*)\cdot(\boldsymbol\tau-\boldsymbol\beta)|\cdot N\bigr]\\ &\le M_0 d + 2NF_0 d\\ &= O(d),\end{aligned}$$

where d is defined in assumption AN1.

Finally, for N3(iii), we have

$$\begin{aligned}E[\mu_t(\boldsymbol\beta,\delta)^2] &\le M_0 d + E\bigl[F(\Omega_t)^2\cdot 2F(\Omega_t)\cdot\|\boldsymbol\tau-\boldsymbol\beta\|\cdot N\\ &\qquad + 2M(\Omega_t,\boldsymbol\beta,\boldsymbol\tau)\cdot F(\Omega_t)\bigr]\\ &\le M_0 d + 2F_0 N d + 2M_1 d\\ &= O(d).\end{aligned}$$

We can therefore apply Huber's theorem,

$$T^{1/2}\lambda_T(\hat{\boldsymbol\beta}) = -T^{-1/2}\sum_{t=1}^{T} g_t(\boldsymbol\beta^0) + o_p(1).$$

Consistency of $\hat{\boldsymbol\beta}$ and application of Slutsky's theorem to (B.1) give

$$T^{1/2}\lambda_T(\hat{\boldsymbol\beta}) = \mathbf{D}_T\cdot T^{1/2}(\hat{\boldsymbol\beta}-\boldsymbol\beta^0) + o_p(1).$$

This, together with Huber's result, yields

$$\mathbf{D}_T \cdot T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = -T^{-1/2}\sum_{t=1}^{T} g_t(\boldsymbol{\beta}^0) + o_p(1). \qquad \text{(B.2)}$$

Application of the central limit theorem (assumption AN4) completes the proof.

Proof of Theorem 3

The proof that $\hat{\mathbf{A}}_T - \mathbf{A}_T \xrightarrow{p} \mathbf{0}$ is standard and is omitted. Following Powell (1991), define

$$\tilde{\mathbf{D}}_T \equiv (2Tc_T)^{-1}\sum_{t=1}^{T} I(|\varepsilon_{t\theta}| < c_T)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0).$$

We first show that $\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T = o_p(1)$ and then that $\tilde{\mathbf{D}}_T - \mathbf{D}_T = o_p(1)$.

Define $\hat{\varepsilon}_t \equiv y_t - f_t(\hat{\boldsymbol{\beta}})$. Then

$$\begin{aligned}
&\|\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T\| \\
&\quad = \frac{c_T}{\hat{c}_T}\Bigg\|(2Tc_T)^{-1} \\
&\qquad \times \sum_{t=1}^{T}\Bigg\{\big[I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T)\big]\nabla' f_t(\hat{\boldsymbol{\beta}})\nabla f_t(\hat{\boldsymbol{\beta}}) \\
&\qquad\qquad + I(|\varepsilon_{t\theta}| < c_T)[\nabla' f_t(\hat{\boldsymbol{\beta}}) - \nabla' f_t(\boldsymbol{\beta}^0)]\nabla f_t(\hat{\boldsymbol{\beta}}) \\
&\qquad\qquad + I(|\varepsilon_{t\theta}| < c_T)\nabla' f_t(\boldsymbol{\beta}^0)[\nabla f_t(\hat{\boldsymbol{\beta}}) - \nabla f_t(\boldsymbol{\beta}^0)] \\
&\qquad\qquad + \frac{c_T - \hat{c}_T}{c_T} I(|\varepsilon_{t\theta}| < c_T)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\Bigg\}\Bigg\|.
\end{aligned}$$

The indicator functions in the first line satisfy the inequality

$$\begin{aligned}
&|I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T)| \\
&\quad \le I\big(|\varepsilon_{t\theta} - c_T| < |\delta_t(\hat{\boldsymbol{\beta}})| + |\hat{c}_T - c_T|\big) \\
&\qquad + I\big(|\varepsilon_{t\theta} + c_T| < |\delta_t(\hat{\boldsymbol{\beta}})| + |\hat{c}_T - c_T|\big).
\end{aligned}$$

Thus

$$\begin{aligned}
&\|\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T\| \\
&\quad \le \frac{c_T}{\hat{c}_T}(2Tc_T)^{-1} \\
&\qquad \times \sum_{t=1}^{T}\big\{I\big(|\varepsilon_t - c_T| < |\delta_t(\hat{\boldsymbol{\beta}})| + |\hat{c}_T - c_T|\big) \\
&\qquad\qquad + I\big(|\varepsilon_t + c_T| < |\delta_t(\hat{\boldsymbol{\beta}})| + |\hat{c}_T - c_T|\big)\big\} \cdot F(\Omega_t)^2 \\
&\qquad + I(|\varepsilon_t| < c_T) \cdot M(\Omega_t, \hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0) \cdot F(\Omega_t) \\
&\qquad + I(|\varepsilon_t| < c_T) \cdot F(\Omega_t) \cdot M(\Omega_t, \hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0) \\
&\qquad + \frac{c_T - \hat{c}_T}{c_T} I(|\varepsilon_{t\theta}| < c_T) \cdot F(\Omega_t)^2 \\
&\quad \equiv \frac{c_T}{\hat{c}_T}(A_1 + 2A_2 + A_3).
\end{aligned}$$

Now suppose that for given $d > 0$ (which can be chosen arbitrarily small), T is sufficiently large that $|\frac{c_T - \hat{c}_T}{c_T}| < d$ and $c_T^{-1}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\| < d$. It is possible to show that $E(A_i) = O(d)$, for $i = 1, 2, 3$. This implies that we found a bounding function for $\|\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T\|$, which can be made arbitrarily small in probability, by choosing d sufficiently small. Here we show only that $E(A_1) = O(d)$, because the others are easily derived:

$$\begin{aligned}
E(A_1) &\le E\Bigg\{(2Tc_T)^{-1} \\
&\qquad \times \sum_{t=1}^{T}\big\{I\big(|\varepsilon_t - c_T| < \|\nabla f_t(\boldsymbol{\beta}^*)\| \cdot \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\| \\
&\qquad\qquad + |\hat{c}_T - c_T|\big) \\
&\qquad + I\big(|\varepsilon_t + c_T| < \|\nabla f_t(\boldsymbol{\beta}^*)\| \cdot \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\| \\
&\qquad\qquad + |\hat{c}_T - c_T|\big)\big\} \cdot F(\Omega_t)^2\Bigg\} \\
&\le E\Bigg\{(2Tc_T)^{-1}\sum_{t=1}^{T}\big\{I\big(|\varepsilon_t - c_T| < dc_T[F(\Omega_t) + 1]\big) \\
&\qquad + I\big(|\varepsilon_t + c_T| < dc_T[F(\Omega_t) + 1]\big)\big\} \\
&\qquad \times F(\Omega_t)^2\Bigg\} \\
&\le E\Bigg\{(2Tc_T)^{-1}\sum_{t=1}^{T} 4dc_T[F(\Omega_t) + 1]HF(\Omega_t)^2\Bigg\} \\
&\le T^{-1}\sum_{t=1}^{T} 4HF_0 \cdot d = 4HF_0 \cdot d = O(d).
\end{aligned}$$

To show that $\tilde{\mathbf{D}}_T - \mathbf{D}_T = o_p(1)$, rewrite this difference as

$$\begin{aligned}
&\tilde{\mathbf{D}}_T - \mathbf{D}_T \\
&\quad = (2Tc_T)^{-1}\sum_{t=1}^{T}\big\{I(|\varepsilon_{t\theta}| < c_T)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) \\
&\qquad\qquad - E\big[I(|\varepsilon_{t\theta}| < c_T)|\Omega_t\big]\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\big\} \\
&\qquad + T^{-1}\sum_{t=1}^{T}\big\{(2c_T)^{-1}E\big[I(|\varepsilon_{t\theta}| < c_T)|\Omega_t\big]\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) \\
&\qquad\qquad - E\big[h_t(0|\Omega_t)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\big]\big\}.
\end{aligned}$$

For the first term, note that it has expectation 0 and variance equal to

$$\begin{aligned}
&E\Bigg\{(2Tc_T)^{-1}\sum_{t=1}^{T} I(|\varepsilon_{t\theta}| < c_T)\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) \\
&\qquad - E\big[I(\varepsilon_{t\theta}| < c_T)|\Omega_t\big]\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\Bigg\}^2 \\
&\quad = (2Tc_T)^{-2}E\Bigg\{\sum_{t=1}^{T}\big\{I(|\varepsilon_{t\theta}| < c_T) - E\big[I(\varepsilon_{t\theta}| < c_T)|\Omega_t\big]\big\}^2 \\
&\qquad\qquad \times [\nabla' f_t(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)]^2\Bigg\}
\end{aligned}$$

$$\le (2Tc_T)^{-2}\sum_{t=1}^{T}E[F(\Omega_t)F(\Omega_t)]^2$$
$$= (4Tc_T^2)^{-1}F_1 = o(1),$$

where the first equality holds because all of the cross-products are 0 by the law of iterated expectations, and the inequality exploits assumption VC2. Hence the first term converges to 0 in mean square, and therefore converges to 0 in probability. For the second term, note that

$$\begin{aligned}&\left|(2c_T)^{-1}E\big[I(|\varepsilon_{t\theta}|<c_T)|\Omega_t\big]-h_t(0|\Omega_t)\right|\\&=\left|(2c_T)^{-1}\int_{-c_T}^{c_T}h_t(\lambda|\Omega_t)\,d\lambda-h_t(0|\Omega_t)\right|\\&\le\left|(2c_T)^{-1}2c_Th_t(c^*|\Omega_t)-h_t(0|\Omega_t)\right|\\&\le L|c_T|=o_p(1),\end{aligned}$$

where $h_t(c^*|\Omega_t)\equiv\max_{\lambda\in[-c_T,c_T]}h_t(\lambda|\Omega_t)$ and the second inequality exploits AN2(b). Substituting and using assumption VC3, the second term also converges to 0 in probability. Therefore, $\tilde{\mathbf{D}}_T-\mathbf{D}_T\xrightarrow{p}\mathbf{0}$, and the result follows.

Proof of Theorem 4

We first approximate the discontinuous function $Hit_t(\hat{\boldsymbol\beta})$ with a continuously differentiable function. We then apply the mean value theorem around $\boldsymbol\beta^0$ and show that the approximated test statistic converges in distribution to the normal distribution stated in Theorem 4. Finally, we prove that this approximation converges in probability to the test statistic defined in Theorem 4.

Define

$$\begin{aligned}Hit_t^{\oplus}(\hat{\boldsymbol\beta})&\equiv\big[1+\exp\{c_T^{-1}\hat\varepsilon_t\}\big]^{-1}-\theta\\&\equiv I^*(\hat\varepsilon_t)-\theta,\end{aligned}$$

where $\hat\varepsilon_t\equiv y_t-f_t(\hat{\boldsymbol\beta})$ and c_T is a nonstochastic sequence such that $\lim_{T\to\infty}c_T=0$. Then

$$\begin{aligned}\nabla_{\boldsymbol\beta}Hit_t^{\oplus}(\hat{\boldsymbol\beta})&=c_T^{-1}\exp\{c_T^{-1}\hat\varepsilon_t\}\big[1+\exp\{c_T^{-1}\hat\varepsilon_t\}\big]^{-2}\nabla f_t(\hat{\boldsymbol\beta})\\&\equiv k_{c_T}(\hat\varepsilon_t)\cdot\nabla f_t(\hat{\boldsymbol\beta}).\end{aligned}$$

Note that $k_{c_T}(\hat\varepsilon_t)$ is the pdf of a logistic with mean 0 and parameter c_T. In matrix form, we write $\nabla_{\boldsymbol\beta}\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta})=\mathbf{K}(\hat\varepsilon_t)\nabla f(\hat{\boldsymbol\beta})$, where $\mathbf{K}(\hat\varepsilon_t)$ is a diagonal matrix with typical entry $k_{c_T}(\hat\varepsilon_t)$. Now, because $\mathbf{X}_t(\hat{\boldsymbol\beta})$ is bounded in probability and $Hit_t^{\oplus}(\hat{\boldsymbol\beta})$ is bounded between $-\theta$ and $1-\theta$, note that

$$\begin{aligned}&T^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta})\\&=T^{-1/2}\sum_{t=1}^{T}\left[\mathbf{X}_t'(\hat{\boldsymbol\beta})Hit_t^{\oplus}(\hat{\boldsymbol\beta})\cdot\left(1-\sum_{j=1}^{J_i}I(\varepsilon_{t\theta}=\varepsilon_t^j)\right)\right]\\&\quad+o_p(1),\end{aligned}$$

because the points over which $\mathbf{X}_t(\hat{\boldsymbol\beta})$ is nondifferentiable form a set of measure 0 by assumption DQ3. In the following, for simplicity of notation, we assume that $\mathbf{X}_t(\hat{\boldsymbol\beta})$ is differentiable everywhere. The case of nondifferentiability would be covered by working with $T^{-1/2}\sum_{t=1}^{T}[\mathbf{X}_t'(\hat{\boldsymbol\beta})Hit_t^{\oplus}(\hat{\boldsymbol\beta})\cdot(1-\sum_{j=1}^{J_i}I(\varepsilon_{t\theta}=\varepsilon_t^j))]$ rather than with $T^{-1/2}\mathbf{X}_t'(\hat{\boldsymbol\beta})Hit_t^{\oplus}(\hat{\boldsymbol\beta})$. An application of the mean value theorem gives

$$\begin{aligned}&T^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta})\\&\quad=T^{-1/2}\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^0)\\&\qquad+T^{-1/2}\big[\nabla\mathbf{X}(\boldsymbol\beta^*)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^*)+\mathbf{X}(\boldsymbol\beta^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol\beta^*)\big]\\&\qquad\times(\hat{\boldsymbol\beta}-\boldsymbol\beta^0),\end{aligned}$$

where $\boldsymbol\beta^*$ lies between $\hat{\boldsymbol\beta}$ and $\boldsymbol\beta^0$, $\varepsilon_t^*\equiv y_t-f_t(\boldsymbol\beta^*)$, $\mathbf{X}'(\boldsymbol\beta)\equiv[\mathbf{X}_1'(\boldsymbol\beta),\dots,\mathbf{X}_T'(\boldsymbol\beta)]$, $\nabla\mathbf{X}(\boldsymbol\beta)\equiv[\nabla\mathbf{X}_1(\boldsymbol\beta),\dots,\nabla\mathbf{X}_T(\boldsymbol\beta)]$, $\mathbf{Hit}^{\oplus}(\boldsymbol\beta)\equiv[Hit_1^{\oplus}(\boldsymbol\beta),\dots,Hit_T^{\oplus}(\boldsymbol\beta)]'$, $\mathbf{K}(\varepsilon_t^*)\equiv\mathrm{diag}([k_{c_T}(\varepsilon_1^*),\dots,k_{c_T}(\varepsilon_T^*)])$, and $\nabla f(\boldsymbol\beta)\equiv[\nabla'f_1(\boldsymbol\beta),\dots,\nabla'f_T(\boldsymbol\beta)]'$. By adding and subtracting appropriate terms, we can rewrite the foregoing expression as

$$\begin{aligned}&T^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta})\\&\quad=T^{-1/2}\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0)\\&\qquad-E[T^{-1}\mathbf{X}(\boldsymbol\beta^0)\mathbf{H}\nabla f(\boldsymbol\beta^0)]\cdot\mathbf{D}_T^{-1}\cdot T^{-1/2}\nabla'f(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0)\\&\qquad+E[T^{-1}\mathbf{X}(\boldsymbol\beta^0)\mathbf{H}\nabla f(\boldsymbol\beta^0)]\mathbf{D}_T^{-1}T^{-1/2}\nabla'f(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0)\\&\qquad-T^{-1}[\mathbf{X}(\boldsymbol\beta^0)\mathbf{H}\nabla f(\boldsymbol\beta^0)]\mathbf{D}_T^{-1}T^{-1/2}\nabla'f(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0)\\&\qquad+T^{-1/2}\mathbf{X}(\boldsymbol\beta^0)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^0)-T^{-1/2}\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0)\\&\qquad+T^{-1/2}\big[\nabla\mathbf{X}(\boldsymbol\beta^*)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^*)+\mathbf{X}(\boldsymbol\beta^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol\beta^*)\big]\\&\qquad\times(\hat{\boldsymbol\beta}-\boldsymbol\beta^0)\\&\qquad+T^{-1}[\mathbf{X}(\boldsymbol\beta^0)\mathbf{H}\nabla f(\boldsymbol\beta^0)]\\&\qquad\times\mathbf{D}_T^{-1}\cdot T^{-1/2}\nabla'f(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0).\end{aligned}\tag{B.3}$$

We first need to show that the terms in the last five lines are $o_p(1)$. The term in the fourth and fifth lines is $o_p(1)$ by assumption DQ4. For the term in the sixth line, noting that $I^*(|\varepsilon_{t\theta}|)=1-I^*(-|\varepsilon_{t\theta}|)$, we have, for each t,

$$\begin{aligned}&\left|Hit_t^{\oplus}(\boldsymbol\beta^0)-Hit_t(\boldsymbol\beta^0)\right|\\&\quad\le I^*(|\varepsilon_{t\theta}|)\big[I(|\varepsilon_{t\theta}|\ge T^{-d})+I(|\varepsilon_{t\theta}|<T^{-d})\big]\\&\quad\equiv C_t+D_t,\end{aligned}$$

where d is a positive number greater than $1/2$, such that $\lim_{T\to\infty}c_TT^d=0$. Therefore,

$$\begin{aligned}&T^{-1/2}\sum_{t=1}^{T}\left\|\mathbf{X}_t(\boldsymbol\beta^0)[Hit_t^{\oplus}(\boldsymbol\beta^0)-Hit_t(\boldsymbol\beta^0)]\right\|\\&\quad\le T^{-1/2}\sum_{t=1}^{T}\|\mathbf{X}_t(\boldsymbol\beta^0)\|\cdot|Hit_t^{\oplus}(\boldsymbol\beta^0)-Hit_t(\boldsymbol\beta^0)|\\&\quad\le T^{-1/2}\sum_{t=1}^{T}W(\Omega_t)\cdot(C_t+D_t),\end{aligned}$$

where $C_t\equiv I^*(|\varepsilon_{t\theta}|)\cdot I(|\varepsilon_{t\theta}|\ge T^{-d})$ and $D_t\equiv I^*(|\varepsilon_{t\theta}|)\cdot I(|\varepsilon_{t\theta}|<T^{-d})$.

Noting that $I^*(|\varepsilon_{t\theta}|)$ is decreasing in $|\varepsilon_{t\theta}|$, we have $C_t \le I^*(T^{-d})$. Therefore,

$$\begin{aligned}T^{-1/2}&\sum_{t=1}^{T} E[W(\Omega_t)C_t]\\ &\le E[W(\Omega_t)]\big[1+\exp(c_T^{-1}T^{-d})\big]^{-1}\\ &= T^{-1/2}\sum_{t=1}^{T} W_0\big[1+\exp(c_T^{-1}T^{-d})\big]^{-1}\\ &= T^{1/2}W_0\big[1+\exp(c_T^{-1}T^{-d})\big]^{-1} \overset{T\to\infty}{\to} 0.\end{aligned}$$

For D_t, note that $D_t \le I(|\varepsilon_{t\theta}| < T^{-d})$, because $I^*(|\varepsilon_{t\theta}|)$ is bounded between 0 and 1. Therefore, for any $\xi > 0$,

$$\begin{aligned}T^{-1/2}&\sum_{t=1}^{T}\Pr\big(W(\Omega_t)D_t > \xi\big)\\ &\le T^{-1/2}\xi^{-1}\sum_{t=1}^{T} E\bigg[W(\Omega_t)\int_{-T^{-d}}^{T^{-d}} h_t(\lambda|\Omega_t)\,d\lambda\bigg]\\ &\le T^{-1/2}\xi^{-1}\sum_{t=1}^{T} W_0\cdot 2T^{-d}N\\ &= 2\xi^{-1}W_0NT^{-d+1/2} \overset{T\to\infty}{\to} 0.\end{aligned}$$

Rewrite the term in the seventh line of (B.3) as

$$T^{-1/2}\big[\nabla\mathbf{X}(\boldsymbol{\beta}^*)\mathbf{Hit}^{\oplus}(\boldsymbol{\beta}^*) + \mathbf{X}(\boldsymbol{\beta}^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol{\beta}^*)\big] \times T^{-1/2}\mathbf{D}_T^{-1}\mathbf{D}_T T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0).$$

Then the last two lines of (B.3) will cancel each other asymptotically if

$$\mathbf{D}_T T^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = -T^{-1/2}\nabla' f(\boldsymbol{\beta}^0)\mathbf{Hit}(\boldsymbol{\beta}^0) + o_p(1) \qquad \text{(B.4)}$$

and

$$T^{-1}\big[\nabla\mathbf{X}(\boldsymbol{\beta}^*)\mathbf{Hit}^{\oplus}(\boldsymbol{\beta}^*) + \mathbf{X}(\boldsymbol{\beta}^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol{\beta}^*)\big]\mathbf{D}_T^{-1} = T^{-1}[\mathbf{X}(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0)]\cdot\mathbf{D}_T^{-1} + o_p(1). \qquad \text{(B.5)}$$

(B.4) is equivalent to (B.2). For (B.5) note that, by the consistency of $\hat{\boldsymbol{\beta}}$,

$$T^{-1}\big[\nabla\mathbf{X}(\boldsymbol{\beta}^*)\mathbf{Hit}^{\oplus}(\boldsymbol{\beta}^*) + \mathbf{X}(\boldsymbol{\beta}^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol{\beta}^*)\big] - T^{-1}\big[\nabla\mathbf{X}(\boldsymbol{\beta}^0)\mathbf{Hit}^{\oplus}(\boldsymbol{\beta}^0) + \mathbf{X}(\boldsymbol{\beta}^0)\mathbf{K}(\varepsilon_{t\theta})\nabla f(\boldsymbol{\beta}^0)\big] \overset{p}{\to} \mathbf{0}.$$

For $T^{-1}\nabla\mathbf{X}(\boldsymbol{\beta}^0)\mathbf{Hit}^{\oplus}(\boldsymbol{\beta}^0)$, we have already shown that $T^{-1}\times\nabla\mathbf{X}(\boldsymbol{\beta}^0)\mathbf{Hit}^{\oplus}(\boldsymbol{\beta}^0) - T^{-1}\nabla\mathbf{X}(\boldsymbol{\beta}^0)\mathbf{Hit}(\boldsymbol{\beta}^0) \overset{p}{\to} \mathbf{0}$. Moreover,

$$\begin{aligned}E&\bigg[T^{-1}\sum_{t=1}^{T}\nabla\mathbf{X}_t(\boldsymbol{\beta}^0)Hit_t(\boldsymbol{\beta}^0)\bigg]\\ &= E\bigg[T^{-1}\sum_{t=1}^{T}\nabla\mathbf{X}_t(\boldsymbol{\beta}^0)E\big[Hit_t(\boldsymbol{\beta}^0)|\Omega_t\big]\bigg]\\ &= 0\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}&\bigg[T^{-1}\sum_{t=1}^{T}\nabla\mathbf{X}_t(\boldsymbol{\beta}^0)Hit_t(\boldsymbol{\beta}^0)\bigg]\\ &\le T^{-2}\sum_{t=1}^{T} E\big[\nabla\mathbf{X}_t(\boldsymbol{\beta}^0)\nabla_j'\mathbf{X}_t(\boldsymbol{\beta}^0)\big]\\ &\le T^{-2}\sum_{t=1}^{T} E[Z(\Omega_t)Z(\Omega_t)]\\ &= T^{-1}Z_0 \overset{T\to\infty}{\to} 0.\end{aligned}$$

It remains to show that $T^{-1}[\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{K}(\varepsilon_{t\theta})\nabla f(\boldsymbol{\beta}^0)] - T^{-1}[\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0)] = o_p(1)$. Rewrite this term as

$$\begin{aligned}T^{-1}&\big[\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{K}(\varepsilon_{t\theta})\nabla f(\boldsymbol{\beta}^0)\big] - T^{-1}\big[\mathbf{X}'(\boldsymbol{\beta}^0)\mathbf{H}\nabla f(\boldsymbol{\beta}^0)\big]\\ &= T^{-1}\sum_{t=1}^{T}\big[k_{c_T}(\varepsilon_{t\theta}) - E\big(k_{c_T}(\varepsilon_{t\theta})\big|\Omega_t\big)\big]\mathbf{X}_t'(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\\ &\quad + T^{-1}\sum_{t=1}^{T}\big[E\big(k_{c_T}(\varepsilon_{t\theta})\big|\Omega_t\big) - h_t(0|\Omega_t)\big]\mathbf{X}_t'(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0).\end{aligned}$$

First, we show that the expected value of $k_{c_T}(\varepsilon_{t\theta})$, given Ω_t, converges to $h_t(0|\Omega_t)$. Let $k(u) \equiv e^u[1+e^u]^{-2}$. Then

$$\begin{aligned}E&\big[k_{c_T}(\varepsilon_{t\theta})|\Omega_t\big]\\ &= \int_{-\infty}^{\infty} k(u)h_t(uc_T|\Omega_t)\,du\\ &= \int_{-\infty}^{\infty} k(u)\big[h_t(0|\Omega_t) + h_t'(0|\Omega_t)uc_T + o(c_T)\big]\,du\\ &= h_t(0|\Omega_t) + o(c_T),\end{aligned}$$

where in the first equality we performed a change of variables, in the second we applied the Taylor expansion to $h_t(uc_T|\Omega_t)$ around 0, and the last equality comes from the fact that $k(u)$ is a density function with first moment equal to 0.

Next, we need to show that $T^{-1}\sum_{t=1}^{T}[k_{c_T}(\varepsilon_{t\theta}) - E(k_{c_T}(\varepsilon_{t\theta})|\Omega_t)]\mathbf{X}_t'(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0) = o_p(1)$. It obviously has 0 expectation. If, in addition, its variance converges to 0, then the result follows from application of the Chebyshev inequality,

$$\begin{aligned}&\bigg\|E\bigg[T^{-1}\sum_{t=1}^{T}\big[k_{c_T}(\varepsilon_{t\theta}) - E\big(k_{c_T}(\varepsilon_{t\theta})\big|\Omega_t\big)\big]\mathbf{X}_t'(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)\bigg]^2\bigg\|\\ &= \bigg\|E\bigg[T^{-2}\sum_{t=1}^{T}\big[k_{c_T}(\varepsilon_{t\theta}) - E\big(k_{c_T}(\varepsilon_{t\theta})\big|\Omega_t\big)\big]^2 \times [\mathbf{X}_t'(\boldsymbol{\beta}^0)\nabla f_t(\boldsymbol{\beta}^0)]^2\bigg]\bigg\|\\ &\le T^{-2}\sum_{t=1}^{T} E\big[O(c_T^{-1})[W(\Omega_t)F(\Omega_t)]^2\big]\\ &\le T^{-2}\sum_{t=1}^{T} W_1 O(c_T^{-1})\\ &= T^{-1}W_1 O(c_T^{-1}) \overset{T\to\infty}{\to} 0,\end{aligned}$$

where the first equality holds because all of the cross-products are 0 by the law of iterated expectations, and the two inequalities follow from assumptions DQ1, AN1(a) and

$$\begin{aligned}
&E\{[k_{c_T}(\varepsilon_{t\theta}) - E(k_{c_T}(\varepsilon_{t\theta})|\Omega_t)]^2|\Omega_t\} \\
&\quad = E[k_{c_T}(\varepsilon_{t\theta})^2|\Omega_t] - E[k_{c_T}(\varepsilon_{t\theta})|\Omega_t]^2 \\
&\quad = \int_{-\infty}^{\infty} k_{c_T}(\lambda)^2 h(\lambda|\Omega_t)\,d\lambda - h(0|\Omega_t)^2 + o(c_T) \\
&\quad = c_T^{-1}\int_{-\infty}^{\infty} k(u)^2 h(uc_T|\Omega_t)\,du - h(0|\Omega_t)^2 + o(c_T) \\
&\quad \le 1/4c_T^{-1}\int_{-\infty}^{\infty} k(u)[h(0|\Omega_t) + h'(0|\Omega_t)uc_T + o(c_T)]\,du \\
&\qquad - h(0|\Omega_t)^2 + o(c_T) \\
&\quad \le 1/4c_T^{-1}[h(0|\Omega_t) + o(c_T)] - h(0|\Omega_t)^2 + o(c_T) \\
&\quad = O(c_T^{-1}).
\end{aligned}$$

Therefore, (B.3) can be rewritten as

$$\begin{aligned}
&T^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta}) \\
&\quad = T^{-1/2}[\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0) \\
&\qquad - E[T^{-1}\mathbf{X}(\boldsymbol\beta^0)\mathbf{H}\nabla f(\boldsymbol\beta^0)]\cdot\mathbf{D}_T^{-1}\cdot\nabla' f(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0)] \\
&\qquad + o_p(1) \\
&\quad \equiv T^{-1/2}\mathbf{M}_T\mathbf{Hit}(\boldsymbol\beta^0) + o_p(1).
\end{aligned}$$

Finally, analogously to how we showed that $T^{-1/2}\mathbf{X}(\boldsymbol\beta^0)\times\mathbf{Hit}^{\oplus}(\boldsymbol\beta^0) - T^{-1/2}\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}(\boldsymbol\beta^0) = o_p(1)$, it is also possible to show that $T^{-1/2}\mathbf{X}(\hat{\boldsymbol\beta})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta}) - T^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta})\mathbf{Hit}(\hat{\boldsymbol\beta}) = o_p(1)$. Therefore,

$$T^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta})\mathbf{Hit}(\hat{\boldsymbol\beta}) = T^{-1/2}\mathbf{M}_T\mathbf{Hit}(\boldsymbol\beta^0) + o_p(1).$$

We can now apply the central limit theorem to $T^{-1/2}\mathbf{M}_T \times \mathbf{Hit}(\boldsymbol\beta^0)$, by assumption DQ6. The law of iterated expectations can be used to show that this term has expectation equal to 0 and variance equal to $\theta(1-\theta)E\{T^{-1}\mathbf{M}_T\mathbf{M}_T'\}$. The result follows.

To conclude the proof of Theorem 4, it remains to show that $T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}_T' - E(T^{-1}\mathbf{M}_T\mathbf{M}_T') \xrightarrow{p} \mathbf{0}$, where

$$\hat{\mathbf{M}}_T \equiv \mathbf{X}'(\hat{\boldsymbol\beta}) - \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla' f(\hat{\boldsymbol\beta})$$

and

$$\hat{\mathbf{G}}_T \equiv (2T\hat{c}_T)^{-1}\sum_{t=1}^{T} I(|y_t - f_t(\hat{\boldsymbol\beta})| < \hat{c}_T)\mathbf{X}_t'(\hat{\boldsymbol\beta})\nabla f_t(\hat{\boldsymbol\beta}).$$

Expanding the product, we get

$$\begin{aligned}
T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}_T' = T^{-1}[&\mathbf{X}'(\hat{\boldsymbol\beta})\cdot\mathbf{X}(\hat{\boldsymbol\beta}) - 2\mathbf{X}'(\hat{\boldsymbol\beta})\nabla f(\hat{\boldsymbol\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}_T \\
&+ \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla' f(\hat{\boldsymbol\beta})\nabla f(\hat{\boldsymbol\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}_T].
\end{aligned}$$

Adopting the same strategy used in the proof of Theorem 3, each of these terms can be shown to converge in probability to the analogous terms of $E(T^{-1}\mathbf{M}_T\mathbf{M}_T')$.

Proof of Theorem 5

As in the proof of Theorem 4, we first approximate the discontinuous function $Hit_t(\hat{\boldsymbol\beta}_{T_R})$ with a continuously differentiable function and then apply the mean value expansion to this approximation,

$$\begin{aligned}
&N_R^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta}_{T_R})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta}_{T_R}) \\
&\quad = N_R^{-1/2}\{\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^0) \\
&\qquad + [\nabla\mathbf{X}(\boldsymbol\beta^*)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^*) + \mathbf{X}(\boldsymbol\beta^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol\beta^*)] \\
&\qquad \times (\hat{\boldsymbol\beta}_{T_R} - \boldsymbol\beta^0)\},
\end{aligned}$$

where $\boldsymbol\beta^*$ lies between $\hat{\boldsymbol\beta}$ and $\boldsymbol\beta^0$ and the variables are defined in the proof of Theorem 4. Assumption DQ8, consistency of $\hat{\boldsymbol\beta}_{T_R}$, and Slutsky's theorem yield

$$\begin{aligned}
&\lim_{R\to\infty} N_R^{-1/2}\mathbf{X}'(\hat{\boldsymbol\beta}_{T_R})\mathbf{Hit}^{\oplus}(\hat{\boldsymbol\beta}_{T_R}) \\
&\quad = \lim_{R\to\infty}\Bigg\{N_R^{-1/2}\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^0) \\
&\qquad + \left(\frac{N_R}{T_R}\right)^{1/2} \\
&\qquad \times \frac{1}{N_R}[\nabla\mathbf{X}(\boldsymbol\beta^*)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^*) \\
&\qquad + \mathbf{X}(\boldsymbol\beta^*)\mathbf{K}(\varepsilon_t^*)\nabla f(\boldsymbol\beta^*)]T_R^{1/2}(\hat{\boldsymbol\beta}_{T_R} - \boldsymbol\beta^0)\Bigg\} \\
&\quad = \lim_{R\to\infty} N_R^{-1/2}\mathbf{X}'(\boldsymbol\beta^0)\mathbf{Hit}^{\oplus}(\boldsymbol\beta^0).
\end{aligned}$$

The rest of the proof follows the analogous parts in Theorem 4.

[Received July 2002. Revised January 2004.]

REFERENCES

Amemiya, T. (1982), "Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

Basle Committee on Banking Supervision (1996), "Amendment to the Capital Accord to Incorporate Market Risks," Basel Committee Publications, www.bis.org/publ/bcbs24a.htm.

Bloomfield, P., and Steiger, W. L. (1983), *Least Absolute Deviations: Theory, Applications and Algorithms*, Boston: Birkhauser.

Boudoukh, J., Richardson, M., and Whitelaw, R. F. (1998), "The Best of Both Worlds," *Risk*, 11, 64–67.

Buchinsky, M. (1995), "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models. A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338.

Buchinsky, M., and Hahn, J. (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671.

Chernozhukov, V. (1999), "Specification and Other Test Processes for Quantile Regression," mimeo, Stanford University.

——— (2000), "Conditional Extremes and Near-Extremes," Working Paper 01-21, MIT, Dept. of Economics.

Chernozhukov, V., and Umantsev, L. (2001), "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics*, 26, 271–293.

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

Danielsson, J., and de Vries, C. G. (2000), "Value-at-Risk and Extreme Returns," *Annales d'Economie et de Statistique*, 60, 239–270.

Dupacova, J. (1987), "Asymptotic Properties of Restricted L_1-Estimates of Regression," in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, ed. Y. Dodge, Amsterdam: North-Holland.

Engle, R. F. (2002), "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425–446.

Engle, R. F., and Ng, V. (1993), "Measuring and Testing the Impact of News On Volatility," *Journal of Finance*, 48, 1749–1778.

Granger, C. W. J., White, H., and Kamstra, M. (1989), "Interval Forecasting: An Analysis Based Upon ARCH-Quantile Estimators," *Journal of Econometrics*, 40, 87–96.

Huber, P. J. (1967), "The Behaviour of Maximum Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium*, 4, 221–233.

Jureckova, J., and Prochazka, B. (1993), "Regression Quantiles and Trimmed Least Squares Estimators in Nonlinear Regression Models," *Journal of Nonparametric Statistics*, 3, 202–222.

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

——— (1982), "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43–61.

Koenker, R., and Park, B. J. (1996), "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265–283.

Koenker, R., and Zhao, Q. (1996), "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793–813.

Kould, H. L., and Saleh, A. K. E. (1995), "Autoregression Quantiles and Related Rank-Scores Processes," *The Annals of Statistics*, 23, 670–689.

Manganelli, S., and Engle, R. F. (2004), "A Comparison of Value at Risk Models in Finance," in *Risk Measures for the 21st Century*, ed. G. Szegö, Chichester, U.K.: Wiley.

McNeil, A. J., and Frey, R. (2000), "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach," *Journal of Empirical Finance*, 7, 271–300.

Mukherjee, K. (1999), "Asymptotics of Quantiles and Rank Scores in Nonlinear Time Series," *Journal of Time Series Analysis*, 20, 173–192.

Oberhofer, W. (1982), "The Consistency of Nonlinear Regression Minimizing the L_1 Norm," *The Annals of Statistics*, 10, 316–319.

Portnoy, S. (1991), "Asymptotic Behavior of Regression Quantiles in Nonstationary, Dependent Cases," *Journal of Multivariate Analysis*, 38, 100–113.

Powell, J. (1983), "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569–1575.

——— (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

——— (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

——— (1991), "Estimation of Monotonic Regression Models Under Quantile Restriction," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, eds. W. A. Barnett, J. Powell, and G. E. Tauchen, Cambridge University Press.

RiskMetrics (1996), *Technical Document*, New York: Morgan Guarantee Trust Company of New York.

Ruppert, D., and Carroll, R. J. (1980), "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828–838.

Schwert, G. W. (1988), "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance*, 44, 1115–1153.

Taylor, S. J. (1986), *Modelling Financial Time Series*, Chichester, U.K.: Wiley.

van de Geer, S. (2000), *Empirical Processes in M-Estimation*, Cambridge, U.K.: Cambridge University Press.

Weiss, A. (1991), "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation," *Econometric Theory*, 7, 46–68.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

——— (1994), *Estimation, Inference and Specification Analysis*, Cambridge, U.K.: Cambridge University Press.