ONLINE SUPPLEMENT TO "FORECASTING VALUE AT RISK VIA INTRA-DAY RETURN CURVES"

1. One-step VaR Forecasting based on Functional Linear Quantile Regression

In this supplementary material, we introduce a functional linear quantile regression (FLQR) model to forecast the VaR. This model incorporates the information in the OCIDR curves is to use them directly as covariates in order to model the conditional quantile within the framework of a quantile regression.

Let $Y_i \in \mathbb{R}$ denote the daily log return, which is defined by $Y_i = X_i(1) = \log P_i(1) - \log P_{i-1}(1)$, for $1 \leq i \leq N$. We posit that the VaR at level τ follows a quantile regression model of the form

$$\text{VaR}_i^\tau = \omega^\tau + \sum_{l=1}^{L} \beta_l^\tau \mathfrak{g}(Y_{i-l}) + \sum_{k=1}^{K} \int b_k^\tau(t) \cdot X_{i-k}(t)dt, \quad t \in [0,1] \tag{1.1}$$

where ω^τ, β_l^τ, $b_k^\tau(t)$ are parameters associated with the fixed τth quantile. We call this model a functional linear quantile regression model (FLQR(L,K)). The function $\mathfrak{g}(\cdot)$ is a non-linear function acting on the lagged values of Y_i. We set $\mathfrak{g}(Y_{i-l}) \equiv |Y_{i-l}|$, although one might consider alternative transformations in order to estimate conditional heteroscedasticity in the sequence Y_i.

This model specification is inspired by the CAViaR model (Engle and Manganelli, 2004), with the primary difference being that we allow the covariates to be functional data objects. Functional versions of linear quantile regression models are relatively new in the FDA field, and still under development. Kato (2012) considered a quantile regression with a scalar response and functional covariates, and established a consistent FPCA-based estimator with sharp convergence rates. However, these results are strictly speaking not suitable for applications in the functional time series framework (in particular with financial data) because they assume the responses and covariates are independent and identically distributed. An application to weakly dependent data was provided by Cabrera and Schulz (2017), who used a functional version of the linear quantile regression model to forecast electricity demand in the German electricity market, but their paper does not establish consistency of the estimators.

2. The FLQR(L,K) Model: Estimation, Model Selection and Simulations

In this section, we explain how to approximately estimate model (1.1). Similar to most of functional time series models, we first conduct a dimension reduction procedure using FPCA. Using the Karhunen-Loève (K-L) representation, we represent the demeaned process $X_i(t) = \sum_{m=1}^{\infty} \xi_{i,m}\psi_m(t)$, where the scores are given by $\xi_{i,m} = \langle X_i, \psi_m\rangle \in \mathbb{R}$. The basis functions ψ_m share the same properties with the orthonormal bases discussed in Section 2 in the main context. Under the assumption that the projections of X_i beyond M_3 principal curves capture negligible information regarding estimating the conditional quantile of Y_i, we make the approximation $X_i(t) \approx \sum_{m=1}^{M_3} \xi_{i,m}\psi_m(t)$. The model (1.1) then reduces to

$$\mathrm{VaR}_i^\tau \approx \omega^\tau + \sum_{l=1}^{L} \beta_l^\tau |Y_{i-l}| + \sum_{k=1}^{K}\sum_{m=1}^{M_3} b_{k,m}^\tau \xi_{i-k,m},$$

where $b_{k,m}^\tau = \langle b_k^\tau(t), \psi_m(t)\rangle$. Similar to the estimation of FGARCH(p,q) model, we can use TVE or cross-validation to determine the order of M_3. Another option is to use information criteria to select M_3, and we propose and study three such criteria below. Besides, as suggested by one of the reviewers, instead of using the empirical covariance operator in presenting a K-L representation, one can use the lagged covariance operator in order to overcome the microstructure errors (Bathia et al. 2010). We leave this as an option for future applications, and the results in this supplement still rely on the empirical covariance operator.

When the functions $\psi_m(t)$ are replaced with estimates of the principal components $\hat{\psi}_m(t)$, the model (1.1) may be re-expressed as

$$\mathrm{VaR}_i^\tau = \tilde{\mathbf{Z}}_i^\top \boldsymbol{\theta}_\tau = \omega^\tau + \sum_{l=1}^{L} \beta_l^\tau |Y_{i-l}| + \sum_{k=1}^{K}\sum_{m=1}^{M_3} \tilde{b}_{k,m}^\tau \tilde{\xi}_{i-k,m}, \tag{2.1}$$

where $\tilde{b}_{k,m}^\tau = \langle b_k^\tau(t), \hat{\psi}_m(t)\rangle$ and the score $\tilde{\xi}_{i,m} = \langle X_i(t), \hat{\psi}_m(t)\rangle$. Equation (2.1) is indeed an equality when $M_3 = N$, unlike the above equation. This is so because even when the K-L representation of X_i has infinitely many nonzero terms, it is always the case that its empirical counterpart (that is, the expansion of X_i obtained via the empirical covariance function) is a finite sum, and the latter representation yields exactly X_i (again, if we put $M_3 = N$).

We define the parameter set $\boldsymbol{\theta}_\tau = \{\omega^\tau, \beta_1^\tau, \ldots, \beta_L^\tau, \tilde{b}_{1,1}^\tau, \ldots, \tilde{b}_{1,M_3}^\tau, \ldots, \tilde{b}_{K,1}^\tau, \ldots, \tilde{b}_{K,M_3}^\tau\} \in \Theta_\tau \subset \mathbb{R}^{K\cdot M_3+L+1}$, and $\tilde{\boldsymbol{Z}}_{i-1} \equiv \{|Y_{i-1}|, \ldots, |Y_{i-L}|, \tilde{\xi}_{i-1,1}, \ldots, \tilde{\xi}_{i-1,M_3}, \ldots, \tilde{\xi}_{i-K,1}, \ldots, \tilde{\xi}_{i-K,M_3}\}$.

After these simplifications, estimating the FLQR(L,K) model turns out to be equivalent with estimating a scalar linear quantile regression model. Therefore, we are able to estimate $\boldsymbol{\theta}_\tau$ by solving the classic optimization problem presented in Koenker (2005),

$$\hat{\boldsymbol{\theta}}_\tau = \underset{\boldsymbol{\theta}_\tau \in \Theta_\tau}{\arg\min} (N-K)^{-1} \sum_{i=K+1}^{N} \rho_\tau(Y_i - \tilde{Y}_i), \tag{2.2}$$

where $\rho_\tau(x) = x \cdot (\tau - \mathbb{1}(x \leq 0))$ denotes the check function, and $\tilde{Y}_i = \mathrm{VaR}_i^\tau$ as shown in Equation (2.1).

Under natural ergodicity conditions on the returns Y_i and OCIDR curves $X_i(t)$, one can establish the consistency of $\hat{\boldsymbol{\theta}}_\tau$ along the lines of the consistency results in Engle and Manganelli (2004). Based on the estimator $\hat{\boldsymbol{\theta}}_\tau$, we can obtain the estimated kernel $b_k^\tau(t)$ in (1.1) as $\hat{b}^\tau(t) = \sum_{m=1}^{M_3} \hat{b}_m^\tau \hat{\psi}_m(t)$.

Following Kato (2012), we now introduce three types of information criteria (AIC, BIC, and HQ) to use in order to select the optimal order of L, K, and M_3 in (2.1). We treat $\hat{\boldsymbol{\theta}}_\tau$ specified in (2.2) as a conditional maximum likelihood estimator, and let the conditional distribution of Y_i based on $\mathcal{F}_{i-1}$ to follow an asymmetric Laplace density function with an unknown weighting parameter ς,

$$f(Y_i|\mathcal{F}_{i-1}, \tau, \varsigma) = \frac{\tau(1-\tau)}{\varsigma} exp[-\frac{1}{\varsigma}\rho_\tau(Y_i - \omega^\tau - \sum_{l=1}^{L} \beta_l^\tau |Y_{i-l}| - \sum_{k=1}^{K}\sum_{m=1}^{M_3} b_{k,m}^\tau \tilde{\xi}_{i-k,m})],$$

Since that coefficients ω^τ, β_k^τ and b_m^τ can be estimated through (2.2), we find a plug-in estimator of the unknown parameter ς given by,

$$\hat{\varsigma} = N^{-1} \sum_{i=K+1}^{N} [\rho_\tau(Y_i - \hat{\omega}^\tau - \sum_{l=1}^{L} \hat{\beta}_l^\tau |Y_{i-l}| - \sum_{k=1}^{K}\sum_{m=1}^{M_3} \hat{b}_{k,m}^\tau \tilde{\xi}_{i-k,m})],$$

and then substituting $\hat{\varsigma}$ produces the log-likelihood function,

$$\begin{aligned} \mathcal{L}_N =& N \cdot \log(\tau(1-\tau)) - N \cdot \log(\hat{\varsigma}) - \frac{1}{\hat{\varsigma}} \sum_{i=K+1}^{N} [\rho_\tau(Y_i - \hat{\omega}^\tau \\ & - \sum_{l=1}^{L} \beta_l^\tau |Y_{i-l}| - \sum_{k=1}^{K}\sum_{m=1}^{M_3} \hat{b}_{k,m}^\tau \tilde{\xi}_{i-k,m})]. \end{aligned} \tag{2.3}$$

We thereby can compute the AIC, BIC, and HQ criteria as,

$$
\begin{aligned}
AIC &= -2\mathcal{L}_N + 2(M_3 \times K + L + 1), \\
BIC &= -2\mathcal{L}_N + \log N \cdot (M_3 \times K + L + 1), \\
HQ &= -2\mathcal{L}_N + \log(\log N) \cdot (M_3 \times K + L + 1),
\end{aligned} \tag{2.4}
$$

respectively. The model candidate with the smallest information criteria is selected as the optimal model. The following provides a Monte Carlo simulation study to examine the finite sample performance of these information criteria in the FLQR framework, which suggests that they work well in general for selecting the correct model with sample sizes relevant for VaR forecasting.

We now conduct a Monte Carlo simulation study to assess the finite sample performance of the FLQR(L,K) model selection based on the proposed information criteria. Motivated by Kato (2012), the simulation considers two data generating processes (DGP) for Y_i,

(1) DGP 1:

$$
Y_i = 0.01 + \int b(t) X_{i-1}(t) dt + \varepsilon_i, \quad t \in [0,1] \tag{2.5}
$$

where ε_i is IID with the distribution $\mathcal{N}(0,1)$ and is independent of $X_{i-1}(t)$. We set the kernel function $b(t)$ as,

$$
b(t) = \sum_{m=1}^{M} b_j \phi_j(t)
$$

where $b_j = -(-1)^j \cdot (j+1)^{-1/2}$, and the orthonormal bases $\phi_j(t) = \sin(j \cdot \pi \cdot t)$. We generate $X_i(t)$ through a functional autoregression,

$$
X_i(t) = \sum_{j=1}^{M} \xi_{i,j} \phi_j(t)
$$

where each loadings series $\xi_{i,j}$ obeys an autoregressive process of order one $\xi_{i,j} = 0.5 \cdot \xi_{i-1,j} + v_{i,j}$ for all $j \in [1, M]$, and the collection $v_{i,j}$ is an IID sequence with $\mathcal{N}(0, \sigma_v^2)$.

(2) DGP 2:

$$Y_i = 0.01 + \int b(t)X_{i-1}(t)dt + \varepsilon_i, \quad t \in [0,1], \quad \varepsilon = \sigma_i \cdot z_i \tag{2.6}$$
$$\sigma_i^2 = 0.01 + 0.85 \cdot \sigma_{i-1}^2 + 0.05 \cdot \varepsilon_{i-1}^2$$

where the parameters and covariates in the mean equation are set to be as the same as DGP 1.

In the DGP1, the dynamics of Y_i only depends on the conditional mean equation; i.e., there is no GARCH effect on Y_i; thus L and K are set to be 0 and 1, respectively. Comparably, in the DGP 2, we set $L = 1$ and $K = 1$. For each DGP, the projection number M are set to be 5 or 10, and the sample sizes are chosen to be $N = 125$, 250 and 500. For the simulation, first 1,000 observations are burned, and the simulation is replicated 1,000 times.

We fit Y_i and $X_i(t)$ with a FLQR(L,K) model and concentrate on three quantiles $\tau = 0.025$, 0.01, 0.005. As the FLQR(L,K) model is estimated under finite dimensional projections (2.1), we work on the FLQR(L, K, M_3) model, and select the optimal specification on the permutation of $L = \{0, 1, 2\}$, $K = \{1, 2, \cdots, 5\}$ and $M_3 = \{1, 2, \cdots, 10\}$. The procedure is to compute AIC, BIC and HQ information criteria for each model candidate, and select the one with the smallest criteria. We assess the model performances by using the averaged root mean square error (ARMSE),

$$ARMSE = \frac{1}{1,000} \sum_{1}^{1,000} \{\frac{1}{N} \sum_{i=1}^{N} |\hat{Y}_i^\tau - Y_i^\tau|^2\}^{1/2}$$

where $\hat{Y}_i^\tau$ is the fitted quantile of Y_i and Y_i^τ is the true quantile. Table 2.1 reports the results. Overall, all of the three information criteria suggest precise orders of L, K, and M_3 for all DGPs, and the ARMSE is reduced along as the sample size N increases. Moreover, at a less extreme quantile (2.5%), the FLQR model suffers a larger error to tailed quantiles (0.5%), which is reasonable as the tailed behavior is usually more difficult to model and estimate.

Table 2.1. ARMSE of the Optimal Selected Models with the values in the parentheses representing for the optimal order selected by the information criteria, e.g., (1,1,5) means $L = 1$, $K = 1$, and $M_3 = 5$, respectively.

	N=500			N=750			N=1,000		
	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
DGP 1: $L = 0$, $K = 1$, $M_3 = 5$									
$\tau = 0.025$	0.3142, (0,1,5)	0.3089, (1,1,4)	0.3146, (0,1,5)	0.3008, (0,1,5)	0.3007, (0,1,5)	0.3008, (0,1,5)	0.2922, (0,1,5)	0.2923, (0,1,5)	0.2921, (0,1,5)
$\tau = 0.01$	0.3482, (0,1,5)	0.3451, (0,1,4)	0.3481, (0,1,5)	0.3245, (0,1,5)	0.3243, (0,1,5)	0.3249, (0,1,5)	0.3126, (0,1,5)	0.3129, (0,1,5)	0.3125, (0,1,5)
$\tau = 0.005$	0.3900, (1,1,5)	0.3838, (1,1,4)	0.3899, (0,1,5)	0.3593, (0,1,5)	0.3595, (0,1,5)	0.3596 , (0,1,5)	0.3310, (0,1,5)	0.3306, (0,1,5)	0.3312, (0,1,5)
$L = 0$, $K = 1$, $M_3 = 10$									
$\tau = 0.025$	0.3165, (0,1,9)	0.3112, (0,1,8)	0.3164, (0,1,10)	0.3051, (0,1,9)	0.3036, (0,1,8)	0.3049, (0,1,9)	0.2976, (0,1,9)	0.2976, (0,1,9)	0.2978, (0,1,10)
$\tau = 0.01$	0.3542, (1,1,10)	0.3513, (1,1,9)	0.3543, (0,1,10)	0.3283 , (0,1,10)	0.3251, (0,1,9)	0.3284, (0,1,10)	0.3195, (0,1,10)	0.3179, (0,1,10)	0.3196, (0,1,10)
$\tau = 0.005$	0.4400, (1,1,10)	0.4341, (1,1,9)	0.4399, (0,1,10)	0.3749, (0,1,10)	0.3706, (0,1,9)	0.3751, (0,1,10)	0.3445, (0,1,9)	0.3433, (0,1,9)	0.3433, (0,1,10)
DGP 2: $L = 1$, $K = 1$, $M_3 = 5$									
$\tau = 0.025$	0.3269, (1,1,5)	0.3253, (1,1,4)	0.3271, (1,1,5)	0.3144, (1,1,5)	0.3141, (1,1,5)	0.3144, (1,1,5)	0.3168, (1,1,5)	0.3167, (1,1,5)	0.3168, (1,1,5)
$\tau = 0.01$	0.3569, (1,1,5)	0.3552, (1,1,4)	0.3569, (1,1,5)	0.3400, (1,1,5)	0.3392, (1,1,5)	0.3401, (1,1,5)	0.3383, (1,1,5)	0.3382, (1,1,5)	0.3384, (1,1,5)
$\tau = 0.005$	0.4024, (1,1,5)	0.4000, (1,1,4)	0.4024, (1,1,5)	0.3723, (1,1,5)	0.3720, (1,1,5)	0.3723, (1,1,5)	0.3518, (1,1,5)	0.3509, (1,1,5)	0.3518, (1,1,5)
$L = 1$, $K = 1$, $M_3 = 10$									
$\tau = 0.025$	0.3232, (1,1,10)	0.3229, (1,1,8)	0.3234, (1,1,10)	0.3170, (1,1,10)	0.3162, (1,1,9)	0.3169, (1,1,10)	0.3170, (1,1,10)	0.3167, (1,1,10)	0.3168 , (1,1,10)
$\tau = 0.01$	0.3635, (1,1,10)	0.3613, (1,1,9)	0.3634, (1,1,10)	0.3518, (1,1,10)	0.3519, (1,1,9)	0.3517, (1,1,10)	0.3336, (1,1,10)	0.3338, (1,1,10)	0.3335, (1,1,10)
$\tau = 0.005$	0.4248, (1,1,10)	0.4202, (1,1,9)	0.4248, (1,1,10)	0.3709, (1,1,10)	0.3701, (1,1,9)	0.3709, (1,1,10)	0.3690, (1,1,10)	0.3692, (1,1,10)	0.3690, (1,1,10)

3. Application to equity and FOREX market returns

In this section, based on the empirical analysis in the main context, we document the empirical VaR forecasting results by using the FLQR model. Regarding the order of the FLQR(L,K) model, in each training sample, we select the optimal model specification (the orders to L, K, and M_3) from the pool of $L \in \{0, 1, 2, 3\}$, $K \in \{1, 2, \cdots, 5\}$ and $M_3 \in \{1, 2, \cdots, 10\}$ by using the AIC information criteria, given the fact that the AIC outperforms the other two information criteria in the simulation study. The optimal fitted models are used to predict 1-day-ahead VaR. Compared with other models reported in the main context, the FLQR model generally performed the worst. This is consistent with our expectations and the literature on quantile regression, as there appears to be a general consensus that it is difficult to accurately estimate extremal quantiles through quantile regression-type methods. The asymptotic arguments in Chernozhukov (2005) suggest that one can expect such methods to be feasible when the tail probability considered multiplied by the sample size is large. In our setting, using approximately three years of training data ($N \approx 750$) is still apparently prohibitive for estimating the 1% and lower tail quantiles. The FLQR model did not pass the backtests at any reasonable significance level. This result is in accordance with the simulation outcomes shown in Table A.1, which indicate that the estimation error of the FLQR model increases along with quantiles deviating from the center. This model is expected to work when a less tailed quantile is chosen, e.g. $\tau = 5\%$ - 10%.

TABLE 3.1. Violation rates and P-values of backtests for the VaR forecasts with the bold values indicating significance at the 5% level.

	S&P500			DAX30			CAC40			USD-Euro		
τ	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%
Panel A: VaR Violation Rates												
FLQR	0.040	0.035	0.022	0.055	0.037	0.028	0.049	0.040	0.037	0.083	0.070	0.047
Panel B: Unconditional Coverage Test												
FLQR	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.03**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Panel C: Conditional Coverage Test												
FLQR	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Panel D: Average Probability Scores valued with units 10^{-4}												
FLQR	8.42	6.50	5.02	10.41	6.12	4.67	10.40	8.14	6.23	6.02	4.93	2.62

REFERENCES

Bathia, N., Yao, Q., Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics 38*, 3352–3386.

Cabrera, B. L., Schulz, F. (2017). Forecasting generalized quantiles of electricity demand: A functional data approach. *Journal of the American Statistical Association 112*, 127–136.

Engle, R. F., Manganelli, S. (2004). CAViaR: conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economics Statistics 22*, 367–381.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics 40*, 3108–3136.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.