Efficient Estimation for Models With Nonlinear Heteroscedasticity

Zhanxiong Xu and Zhibiao Zhao

Department of Statistics, Penn State University, University Park, PA

ABSTRACT

We study efficient estimation for models with nonlinear heteroscedasticity. In two-step quantile regression for heteroscedastic models, motivated by several undesirable issues caused by the preliminary estimator, we propose an efficient estimator by constrainedly weighting information across quantiles. When the weights are optimally chosen under certain constraints, the new estimator can simultaneously eliminate the effect of preliminary estimator as well as achieve good estimation efficiency. When compared to the Cramér-Rao lower bound, the relative efficiency loss of the new estimator has a conservative upper bound, regardless of the model design structure. The upper bound is close to zero for practical situations. In particular, the new estimator can asymptotically achieve the optimal Cramér-Rao lower bound if the noise has either a symmetric density or the asymmetric Laplace density. Monte Carlo studies show that the proposed method has substantial efficiency gain over existing ones. In an empirical application to GDP and inflation rate modeling, the proposed method has better prediction performance than existing methods.

1. Introduction

Consider the heteroscedastic model

\[ Y_t = X_t^T \beta + \sigma_t \epsilon_t \quad \text{with} \quad \sigma_t = s(X_t, \alpha), \quad t = 1, \ldots, n, \quad (1) \]

where \( \beta \in \mathbb{R}^p \), \( s(X_t, \alpha) > 0 \) is some nonlinear function with unknown \( \alpha \in \mathbb{R}^k \), and the noise \( \epsilon_t \) is independent of covariates \( X_t \). On one hand, the linear location function framework is widely used to model many important financial and economic variables. For example, in finance, the celebrated capital asset pricing model (CAPM) asserts that the risk premium (relative to risk-free interest rates) of an individual security or portfolio equals the market risk premium multiplied by the beta; in economics, the well-known Fisher hypothesis states that nominal asset returns move one-to-one with inflation rate; and a third example is the predictability of inflation from interest rates (Fama 1975). On the other hand, numerous ARCH models impose different nonlinear structures on \( \sigma_t \), including square-root heteroscedasticity \( \sigma_t = (\alpha_0 + \alpha_1 X_{t1}^2 + \cdots + \alpha_p X_{tp}^2)^{1/2} \), linear heteroscedasticity \( \sigma_t = \alpha_0 + \alpha_1 X_{t1} + \cdots + \alpha_p X_{tp} \), and exponential-type heteroscedasticity \( \sigma_t = \exp(\alpha_0 + \alpha_1 X_{t1} + \cdots + \alpha_p X_{tp}) \) among others; see Bollerslev (2008) for more examples.

Since Koenker and Bassett (1978), quantile regression has become a powerful tool for studying heteroscedastic models. Let \( Q_{e}(\tau) \) be the \( \tau \)th quantile of \( \epsilon_t \). The conditional \( \tau \)th quantile of \( Y_t \) given \( X_t \) is the nonlinear function involving parameters \( (\beta, \alpha, Q_{e}(\tau)) \):

\[ Q_{Y_t|X_t}(\tau) = X_t^T \beta + s(X_t, \alpha)Q_{e}(\tau). \quad (2) \]

However, since nonlinear quantile regression usually requires stringent conditions and is technically and computationally challenging, the literature has almost exclusively focused on linear heteroscedasticity \( s(X_t, \alpha) = X_t^T \alpha \) so that \( Q_{Y_t|X_t}(\tau) = X_t^T \gamma(\tau) \), \( \gamma(\tau) = \beta + Q_{e}(\tau) \alpha \). This linear quantile representation plays an essential role in the quantile regression literature on heteroscedastic models (Koenker and Bassett 1982; Gutenbrunner and Jurečková 1992; Koenker and Zhao 1994, 1996; Zhou and Portnoy 1998; Xiao and Koenker 2009; Zhao and Xiao 2014; Zhu, Zheng, and Li 2018). The linearity enables linear quantile regression [throughout, \( \rho_t(z) = z(1 - I_{z<0}) \) is \( \tau \)-th quantile loss]:

\[ \hat{\gamma}(\tau) = \text{argmin}_b \sum_{t=1}^n \frac{\rho_t(Y_t - X_t^T b)}{X_t^T \hat{\alpha}^0}, \]

\[ \hat{\alpha}^0 \text{ is some preliminary estimator of } \alpha. \quad (3) \]

The denominator \( X_t^T \hat{\alpha}^0 \) accounts for heteroscedasticity. Under regularity conditions,

\[ \hat{\gamma}(\tau) = \tilde{\gamma}(\tau) + o_p(n^{-1/2}), \quad \text{where} \]

\[ \tilde{\gamma}(\tau) = \text{argmin}_b \sum_{t=1}^n \frac{\rho_t(Y_t - X_t^T b)}{X_t^T \alpha}. \quad (4) \]

That is, asymptotically speaking, the preliminary estimator \( \hat{\alpha}^0 \) has no effect on \( \tilde{\gamma}(\tau) \).

However, for nonlinear heteroscedasticity \( s(X_t, \alpha) \), we have a quite different phenomenon. With preliminary estimator \( \hat{\alpha}^0 \approx \alpha \), Equation (2) has the linear approximation \( Q_{Y_t|X_t}(\tau) \approx
\( X_i^T \beta + s(X_i, \hat{\alpha}^0)Q_e(\tau) \). Thus, similar to Equation (3), we can apply linear quantile regression

\[
(\hat{\beta}(\tau), \hat{Q}_e(\tau)) = \arg\min_{b,q} \sum_{i=1}^{n} \rho_{\tau}(Y_i - X_i^T b - \sigma_i q), \quad \text{where} \quad \sigma_i = s(X_i, \hat{\alpha}^0).
\]

As in Equation (4), denote by \( \tilde{\beta}(\tau) \) the estimate obtained by replacing \( \sigma_i \) with \( \sigma_i = s(X_i, \alpha) \) in Equation (5). Our first contribution is to establish a connection between \( \hat{\beta}(\tau) \) and \( \tilde{\beta}(\tau) \) in the form

\[
\sqrt{n}[\hat{\beta}(\tau) - \beta] = \sqrt{n}[\tilde{\beta}(\tau) - \beta] + \mathcal{O}(p(1)), \quad (6)
\]

where \( \mathcal{O}(\cdot) \) quantifies the effect of \( \hat{\alpha}^0 \) at quantile \( \tau \). Thus, unlike the linear heteroscedasticity case, \( \hat{\alpha}^0 \) has a non-negligible effect for nonlinear heteroscedasticity.

In Equation (6), the presence of the non-negligible term \( \mathcal{O}(\cdot) \) poses several undesirable issues in statistical inference. First, in Section 2.1 it is shown that \( \mathcal{O}(\cdot) \) deviates from zero at the rate of \( \sqrt{n}[\hat{\alpha}^0 - \alpha] \) and thus any large estimation error in \( \hat{\alpha}^0 \) will propagate to \( \beta(\tau) \). Second, it can be difficult to derive asymptotic normality for \( \hat{\beta}(\tau) \) and conduct-related statistical inference as we need to investigate the joint behavior of \( (\hat{\beta}(\tau), \hat{\alpha}^0) \), where the preliminary estimator \( \hat{\alpha}^0 \) usually requires a two-step estimation procedure and its distributional effect on \( \beta(\tau) \) can be nontrivial. Third, when a different method is used to obtain \( \hat{\alpha}^0 \), completely new analysis must be carried out. Consequently, researchers could make different conclusions just because of using a different preliminary estimator \( \hat{\alpha}^0 \). Fourth, if we are given \( \hat{\alpha}^0 \) from prior experience without knowing the exact method used to construct \( \hat{\alpha}^0 \), from Equation (6) it is infeasible to make statistical inference.

Motivated by the aforementioned issues, our main contribution is to construct efficient estimator of \( \beta \) that can simultaneously achieve two goals (i) it is immune to the effect of preliminary estimator \( \hat{\alpha}^0 \); and (ii) it can achieve good estimation efficiency or even asymptotically achieve the Cramér-Rao bound, at least in some situations. Our method is based on constrainedly optimally combining information across multiple quantiles. Given \( N \) quantiles \( 0 < \tau_1 < \cdots < \tau_N < 1 \), we consider a new estimator \( \tilde{\beta}(\omega) \) of the form

\[
\tilde{\beta}(\omega) = \sum_{i=1}^{N} \omega_i \hat{\beta}(\tau_i), \quad \text{subject to} \quad \sum_{i=1}^{N} \omega_i = 1 \quad \text{and} \quad \sum_{i=1}^{N} \omega_i \mathcal{O}(\cdot) = 0. \quad (7)
\]

By Equation (6), the constraint \( \sum_{i=1}^{N} \omega_i \mathcal{O}(\cdot) = 0 \) ensures that the undesirable term \( \mathcal{O}(\cdot) \) is eliminated, and the constraint \( \sum_{i=1}^{N} \omega_i = 1 \) ensures the consistency of \( \tilde{\beta}(\omega) \). We further propose the constrainedly optimally weighted quantile average estimator (COWQAE) by minimizing the limiting covariance matrix of \( \sqrt{n}[\tilde{\beta}(\omega) - \beta] \) subject to the constraints (7).

Although the idea of combining multiple single-quantile regression estimates has been studied in the literature, our approach makes novel theoretical and methodological contributions. Existing works focused on simple homoscedastic linear model (Koenker 1984; Koenker and Portnoy 1987; Portnoy and Koenker 1989; Zhao and Xiao 2014; Wang and Wang 2016), nonlinear homoscedastic model (Jurečková and Procházka 1994), linear model with linear heteroscedasticity (Gutenbrunner and Jurečková 1992; Koenker and Zhao 1994; Zhao and Xiao 2014; Zhu, Zheng, and Li 2018) and time-varying coefficient longitudinal model (Kim, Zhao, and Xiao 2018). It is non-trivial to generalize existing methods to nonlinear heteroscedasticity case, and the main issue is the non-negligible effect when we use estimated heteroscedasticity \( \hat{\sigma}_i \) in Equation (5). While this undesirable effect exists for each quantile, with multiple quantiles we can eliminate it via the constraint \( \sum_{i=1}^{N} \omega_i \mathcal{O}(\cdot) = 0 \) in Equation (7). To our knowledge, there has been no existing work in the literature that used weighted average of multiple quantile regression estimates with special constraints on the weights to suppress the effect of preliminary estimator.

We show that the proposed COWQAE has good estimation efficiency. When the number of quantiles goes to infinity, the relative efficiency loss (compared to the Cramér-Rao lower bound) can be factorized into two components, one depending only on the noise distribution and the other depending only on model design structure. We further derive a conservative upper bound for the relative efficiency loss, and it is illustrated that the upper bound is usually quite small for practical situations. In particular, the COWQAE can asymptotically achieve the Cramér-Rao lower bound under either case: (i) \( \epsilon \) has a symmetric density or (ii) \( \epsilon \) has the asymmetric Laplace density. These results are new to the literature. Both simulation studies and an empirical application to GDP and inflation rate modeling show that the proposed method outperforms existing ones.

While the proposed method is based on constrainedly weighted average of multiple single-quantile regression estimates, another different framework, the composite quantile regression (CQR) which combines loss function across quantiles, has also been extensively studied (Koenker 1984; Zou and Yuan 2008; Kai, Li, and Zou 2010, 2011; Bradic, Fan, and Wang 2011; Jiang, Jiang, and Song 2012; Xu, Wang, and Huang 2014), all for homoscedastic models. We envision that a weighted CQR approach can be developed for the nonlinear heteroscedastic model (1) by imposing appropriate constraints on the weights to eliminate the effect of preliminary estimate \( \hat{\alpha}^0 \), and we leave it as a future work.

We introduce some notation. The notation \( p \rightarrow \) means convergence in probability. For a matrix (or vector) \( A = \{a_{ij}\} \), write \( ||A|| = \sum_{i,j} |a_{ij}| \). For random variable \( Z \), write \( Z \in \mathcal{L}^{d} \) if \( \mathbb{E}[|Z|^d] < \infty \). For vectors \( U \in \mathbb{R}^p \) and \( V \in \mathbb{R}^q \), write \( \text{cov}(U,V) = \mathbb{E}(UV^T) - \mathbb{E}(U)\mathbb{E}(V^T) \in \mathbb{R}^{p \times q} \) as the cross-covariance matrix and \( \text{var}(U) = \text{cov}(U,U) \in \mathbb{R}^{p \times p} \). All technical proofs are gathered in the online Appendix.

2. Main Results

For \( f_{\epsilon}(u), F_{\epsilon}(u) \) and \( Q_{\epsilon}(\tau) \), respectively. For \( s(X_i, \alpha) \) as a function of \( \alpha \), denote its partial derivatives by \( \partial s(X_i, \alpha) / \partial \alpha \) and its Hessian matrix by \( \partial^2 s(X_i, \alpha) / \partial \alpha^2 \).
**Remark 1.** In Equation (1), we assume that $X_i^T \beta$ does not include a constant intercept. Otherwise, if there is an intercept $\beta_0$, when $\delta_t \approx \sigma$ is close to a constant (i.e., approximate homoscedasticity), the quantile regression (5) becomes approximately singular since $\beta_0 + \sigma q$ can be combined into one parameter $\beta_0 + \sigma q$ and we cannot identify them separately. In practice, we are more interested in slope parameters which measure the effect of covariates. See Section 4 for details on how to deal with models with an intercept.

### 2.2. Constrainedly Optimally Weighted Quantile Average Estimator

While the effect of $\hat{\alpha}^0$ exists for each quantile $\tau$ (as demonstrated in Section 2.1), we can combine information across quantiles to eliminate such effect. Throughout let $t_i = i/(N + 1), i = 1, \ldots, N$, be $N$ quantiles and consider a new estimator by linearly weighting $\hat{\beta}(t_1), \ldots, \hat{\beta}(t_N)$ with weight $\omega = (\omega_1, \ldots, \omega_N)^T$:

$$\tilde{\beta}(\omega) = \sum_{i=1}^{N} \omega_i \hat{\beta}(t_i),$$

$$\omega \in \Omega := \left\{ (\omega_1, \ldots, \omega_N)^T : \sum_{i=1}^{N} \omega_i = 1, \sum_{i=1}^{N} \omega_i Q_e(t_i) = 0 \right\}.$$  

Under the first constraint $\sum_{i=1}^{N} \omega_i = 1$, this type of estimator is studied for simple homoscedastic linear model (Koenker 1984; Koenker and Portney 1987; Portnoy and Koenker 1989; Zhao and Xiao 2014; Wang and Wang 2016) and linear model with linear heteroscedasticity (Gutenbrunner and Jurečková 1992; Koenker and Zhao 1994; Zhao and Xiao 2014; Zhu, Zheng, and Li 2018). For nonlinear heteroscedasticity model (1), we introduce the second constraint $\sum_{i=1}^{N} \omega_i Q_e(t_i) = 0$ to suppress the effect of $\hat{\alpha}^0$. By the two constraints in $\Omega$ and using the representation (9), we can obtain

$$\sqrt{n}[\tilde{\beta}(\omega) - \beta] = \sqrt{n} \sum_{i=1}^{N} \omega_i [\hat{\beta}(t_i) - \beta]$$

$$= \sum_{i=1}^{N} \omega_i \sqrt{n} [\hat{\alpha}(t_i) - \alpha] + o_p(1).$$
**Remark 2.** For the standard homoscedastic models, that is, $\sigma_i \equiv 1$ in (1), we replace $\hat{\sigma}_i$ in Equation (5) by constant 1 and modify (10) as $\hat{\beta}(\omega) = \sum_{i=1}^{N} \omega_i \hat{\beta}(\tau_i)$ for $\omega = (\omega_1, \ldots, \omega_N)^T$ satisfying $\sum_{i=1}^{N} \omega_i = 1$. By Zhao and Xiao (2014), in Theorem 3, the optimal weight becomes $\omega^* = H^{-1}u/(u^TH^{-1}u)$ and $\sqrt{n}\hat{\beta}(\omega^*) - \beta \Rightarrow N(0, \Sigma_1^{-1}(S(\omega)))$. Under regularity conditions, $\lim_{n \to \infty} u^TH^{-1}u = \int_0^1 [f^2(x)/f_n(x)]dx$, thus $\hat{\beta}(\omega^*)$ asymptotically achieve Cramér-Rao bound, regardless of whether the density is symmetric or not.

**Remark 3.** In Equation (9), consider $p = k = 1$ and write $U = \sqrt{n}\hat{\beta}(\tau) - \beta$, $V = \sqrt{n}(\hat{\alpha} - \alpha)$, $\epsilon = Q_e(\tau)\Sigma_1^{-1}\Sigma_2$. From $\text{var}(U - cV) = \text{var}(U) = \text{cov}(U, V)^2$, if $\text{cov}(U, V) = 0$ (e.g., if $\hat{\alpha}^0$ is obtained from an independent dataset so that $\text{cov}(U, V) = 0$), then $\text{var}(U - cV) = \text{var}(U) \geq \text{cov}(V)$ and a large value of $|c|$ leads to poor performance of $\hat{\beta}(\tau)$. As a special example, consider $\sigma = \sqrt{\sigma^2 + 0.2\lambda^2}$ and $X_1 \sim \text{Uniform}(0, 1)$, for $\alpha = 0.3, 0.1, 0.01$, $\Sigma_1^{-1}\Sigma_2 \approx -0.4, -2, -26$, respectively. Therefore, there could be a substantial loss of efficiency (proportional to $c^2$) to use $\hat{\beta}(\tau)$. On the other hand, if $\text{cov}(U, V) > \text{cov}(V)/2$, $\hat{\beta}(\tau)$ can perform even better than $\beta(\tau)$, though the potential efficiency gain $2\text{cov}(U, V) - \text{cov}(V)$ is unlikely very large due to the term $\text{cov}(V)$ growing at rate $c^2$. In practice, $c$ is determined by the underlying model and we have no control over it. Moreover, it may be infeasible to evaluate $\text{cov}(V)$ and $\text{cov}(U, V)$, especially when $\hat{\alpha}^0$ is given as an a priori knowledge without specifying how it was obtained. Therefore, it is desirable to eliminate such unknown effects via Equation (10).

**Remark 4.** By Equation (9), if $Q_e(\tau) = 0$, then $\hat{\beta}(\tau) = \hat{\beta}(\tau) + O_p(n^{-1/2})$. In particular, if $\epsilon_t$ has a symmetric density, then the median-quantile-regression estimator $\hat{\beta}(0.5)$ performs as well as the infeasible estimator $\hat{\beta}(0.5)$. In this case, while $\hat{\beta}(0.5)$ is not affected by the preliminary estimator $\hat{\alpha}^0$, in general it is not an efficient estimator of $\beta$ since it only uses the information from the single median quantile. For example, consider estimation of $\beta$ in the special case $Z_t = \beta + \epsilon_t$. If $\epsilon_t \sim N(0, 1)$, the sample mean is the maximum likelihood estimator (MLE) and much superior to the sample median. For a continuous random variable $Z$ with quantile function $Q_Z(\tau), \tau \in (0, 1)$, we have $E(Z) = \int_0^1 Q_Z(\tau)d\tau$. Thus, by using $\omega_i = 1/N$, $\hat{\beta}(\omega)$ in Equation (10) can approximate the sample mean, leading to better performance than $\hat{\beta}(0.5)$. On the other hand, if $\epsilon_t$ has a standard Laplace distribution, then the sample median is the MLE and much superior to the sample mean, and by assigning the unit weight on $\hat{\beta}(0.5)$ and zero weights on other quantiles, $\hat{\beta}(\omega)$ becomes exactly the sample median and is more efficient than sample mean. These phenomena are confirmed in our simulation studies in Section 3. In fact, by Theorem 5, if $\epsilon_t$ has a symmetric density, then $\hat{\beta}(\omega)$ can asymptotically achieve the Cramér-Rao bound.

As illustrated in Remarks 3 and 4, the new estimator (10) can achieve two important purposes: eliminating potentially unknown severe effects from preliminary estimator $\alpha^0$ as well as combining information across quantiles to improve estimation efficiency.

**Theorem 2.** Under the conditions in Theorem 1, $\sqrt{n}\hat{\beta}(\omega) - \beta \Rightarrow N(0, \Sigma_1^{-1}(S(\omega)))$, where

$$S(\omega) = \omega^TH\omega \quad \text{with} \quad H = \left[\min(t_i, \tau_i) - \tau_i t_i f_c(Q_e(\tau_i))/f_c(Q_e(\tau_i))\right]_{1 \leq i \leq N} \in R^{N \times N}.$$  \hspace{1cm} (11)

In Theorem 2, $\Sigma_1^{-1}$ depends on the model structure and the scalar value $S(\omega)$ depends on the weight $\omega \in \Omega$. Naturally, the optimal weight $\omega^*$ can be obtained as

$$\omega^* = \arg\min_{\omega \in \Omega} S(\omega), \quad \text{subject to} \ \omega \in \Omega.$$  \hspace{1cm} (12)

We call $\hat{\beta}(\omega^*)$ the constrainedly optimally weighted quantile average estimator (COWQAE). Zhao and Xiao (2014) used the term OWQAE without the constraint $\sum_{i=1}^{N} \omega_i Q_e(\tau_i) = 0$.

**Theorem 3.** Denote by $v = (Q_e(\tau_1), \ldots, Q_e(\tau_N))^T$ and $u = (1, \ldots, 1)^T$. For $\omega^*$ in (12),

$$\omega^* = \frac{H^{-1}}{\text{E}_N} \left[ u - \left( v^TH^{-1}u \right) v \right] \quad \text{with} \quad \text{E}_N = u^TH^{-1}u - \left( v^TH^{-1}u \right)^2.$$  \hspace{1cm} (13)

Using this optimal weight, we have

$$\sqrt{n}\hat{\beta}(\omega^*) - \beta \Rightarrow N(0, \Sigma_1^{-1}/\text{E}_N).$$  \hspace{1cm} (14)

**Remark 5.** In Equation (5), $\hat{\beta}(\tau)$ remains unchanged if we replace $\hat{\sigma}_i$ by $\hat{\sigma}_i/c$ for any $c > 0$ (this will only affect $\epsilon$ therein by the same scale). The optimal weights and asymptotic normality in Theorem 3 are invariant under the scale change: replace $(\epsilon_t, \sigma_t)$ in (1) by $(c\epsilon_t, \sigma_t/c)$.

**Assumption 6.** As a function of $\tau$, $f_c(Q_e(\tau)) \in \mathcal{G}(0, 1)$ and $Q_e(\tau)f_c(Q_e(\tau)) \in \mathcal{G}(0, 1)$, where $\mathcal{G}(0, 1)$ is the set of twice differentiable functions $g(\cdot)$ such that

$$\int_0^1 [g'(\tau)^2]d\tau < \infty, \quad \lim_{\tau \to 0} \frac{g^2(\tau) + g^2(1 - \tau)}{\tau} = 0,$$

$$\lim_{\tau \to 0} \int_0^1 [g''(\tau)^2]d\tau = 0.$$  \hspace{1cm} (15)

In Assumption 6, Equation (15) is adopted from Zhao and Xiao (2014). It can be verified that many commonly used distributions with support on $\mathbb{R}$ satisfy this condition, including, for example, all the distributions in Section 3.

**Theorem 4.** For $\text{E}_N$ in (14), under Assumption 6, $\lim_{n \to \infty} \text{E}_N = \vartheta_1 - \vartheta_3^2/\vartheta_3$, where

$$\vartheta_1 = \int_{\mathbb{R}} [f_c(x)]^2 f_c(x)dx, \quad \vartheta_2 = \int_{\mathbb{R}} [f_c(x)f_c(x) + xf_c(x)] f_c(x)dx,$$

$$\vartheta_3 = \int_{\mathbb{R}} [f_c(x) + xf_c(x)]^2 f_c(x)dx.$$  \hspace{1cm} (16)

From (14) and Theorem 4, when we optimally combine more and more quantiles, the limiting covariance matrix of the resulting estimator $\hat{\beta}(\omega^*)$ approaches
Remark 6. From Theorem 3, \( \sqrt{n}[\beta(\omega^*) - \beta] \Rightarrow N(0, \Sigma_{\beta}^{-1}/J_N) \) for fixed \( N \), and Theorem 4 establishes \( \lim_{N \to \infty} \Sigma_{\beta}^{-1}/J_N = \Sigma_{\beta} \) with \( \Sigma_{\beta} \) defined in (16). A natural question is whether we can strengthen Theorem 3 to establish \( \sqrt{n}[\beta(\omega^*) - \beta] \Rightarrow N(0, \Sigma_{\beta}) \) if we take \( N = N_n \to \infty \) as \( n \to \infty \). As in Portnoy and Koenker (1989) and Zhao and Xiao (2014), we do not pursue this direction and instead we examine finite sample performance through extensive simulation studies.

2.3. Comparison to Cramér-Rao Lower Bound

It is well known that the Cramér-Rao bound is the smallest possible variance for any unbiased estimator of the parameter. Write the likelihood for \((Y_t, X_t)\) as \( f(Y_t, X_t; \theta) = (\beta^T, \alpha^T)\).

By the independence between \( e_t \) and \( X_t \), the Fisher information matrix for \( \theta \) (assuming \( X_t/\sigma_t \) is centered so \( \mathbb{E}(X_t/\sigma_t) = 0 \)):

\[
I = \mathbb{E} \left[ \frac{\partial \log f(Y_t, X_t; \theta)}{\partial \theta} \frac{\partial \log f(Y_t, X_t; \theta)}{\partial \theta^T} \right]
\]

where \( \Sigma_1 \) and \( \Sigma_2 \) are defined in Theorem 1, \( \Sigma_3 \) is the standard normal density and distribution function, \( e_t \), \( e_2 \), and \( e_3 \) are defined in Theorem 4. By block matrix inverse, the Cramér-Rao bound for \( \beta \), denoted by \( CR_{\beta} \), is

\[
CR_{\beta} = p \times p \text{ submatrix of } I^{-1} = \left( e_1 \Sigma_1 - e_2 \Sigma_2 \Sigma_3^{-1} \Sigma_2^T \right)^{-1}
\]

To compare \( \Sigma_{\beta} \) [see (16)] and the Cramér-Rao bound \( CR_{\beta} \), we introduce some notations. For a symmetric matrix \( M \), write \( M \geq 0 \) if \( M \) is positive semi-definite; for positive semi-definite symmetric matrices \( M_1 \in \mathbb{R}^{p \times p} \) and \( M_2 \in \mathbb{R}^{p \times p} \), write \( M_1 \geq M_2 \) if \( M_1 - M_2 \geq 0 \).

Note that for \( M_1 \geq 0 \) and \( M_2 \geq 0 \), \( M_1 \geq M_2 \) if and only if \( M_1^{1/2} M_2^{-1/2} M_1^{1/2} \geq I_p \), where \( I_p \) is the \( p \times p \) identity matrix. Naturally, we can define the relative efficiency loss (REL) of \( \Sigma_{\beta} \), when compared to the Cramér-Rao bound \( CR_{\beta} \), as

\[
REL(\Sigma_{\beta}) = \Sigma_{\beta}^{1/2} CR_{\beta}^{-1} \Sigma_{\beta}^{1/2} - I_p.
\]

Then \( \Sigma_{\beta} \geq CR_{\beta} \) if and only if REL(\( \Sigma_{\beta} \)) \geq 0, and \( \Sigma_{\beta} \geq CR_{\beta} \) if and only if REL(\( \Sigma_{\beta} \)) = 0.

Theorem 5. Regarding REL(\( \Sigma_{\beta} \)) in Equation (18), the following results hold:

i. The factorization REL(\( \Sigma_{\beta} \)) = \( N_{\text{noise}} N_{\text{design}} \geq 0 \), where

\[
N_{\text{noise}} = e_2^2 / e_1 e_3 - e_2^2 \geq 0 \quad \text{and}
\]

\[
N_{\text{design}} = I_p - \Sigma_1^{-1/2} \Sigma_2 \Sigma_3^{-1} \Sigma_2 \Sigma_1^{-1/2} \geq 0.
\]

ii. We have \( \Sigma_{\beta} = CR_{\beta} \) (i.e., REL(\( \Sigma_{\beta} \)) = 0) if \( e_t \) has either a symmetric density or the asymmetric Laplace density \( f_{e}(x) = p(1-p) \exp(-x(p-1|x<0|), p \in (0,1)).

iii. Regardless of the model design structure, the upper bound always holds:

\[
REL(\Sigma_{\beta}) \leq N_{\text{noise}} I_p \quad \text{with}
\]

\[
N_{\text{noise}} = e_2^2 / e_1 e_3 - e_2^2.
\]

By Theorem 5(i), REL(\( \Sigma_{\beta} \)) can be factorized into two independent factors: \( N_{\text{noise}} \) depending only on the noise distribution and \( N_{\text{design}} \) depending only on the model design. We can analyze \( N_{\text{noise}} \) and \( N_{\text{design}} \) separately to evaluate the potential efficiency loss.

By Theorem 5(ii), under one of the two cases listed, \( \tilde{\beta}(\omega^*) \) can asymptotically achieve the Cramér-Rao bound. That is, by optimally weighting information across quantiles we can simultaneously eliminate the effect of \( \tilde{\sigma}^0 \) and achieve optimal estimation efficiency.

Remark 7. Beran (1974) and Bickel (1982) pioneered the work on adaptive estimators that can asymptotically achieve the information bound. In time series context, adaptive likelihood based methods are studied in, for example, Kreiss (1987) for ARMA models and Linton (1993) for ARCH models, where it is assumed that the noise has a symmetric density with finite fourth moment. The finite fourth moment condition excludes some heavy-tailed distributions such as Student-\( t_d \) distribution with degrees of freedom \( d \leq 4 \), and the symmetric density condition ensures that the effect from preliminary estimators vanishes. For asymmetric density case, these methods generally will not work for heteroscedastic models such as Equation (1), due to the nonvanishing effect from preliminary estimator.

While our method also cannot achieve the information bound for asymmetric densities, Theorem 5 establishes an upper bound for the potential efficiency loss which, as illustrated in Examples 1–2, is quite small for practical situations. Furthermore, our proposed method is computationally appealing as it involves simple linear quantile regressions.

By Theorem 5(iii), the potential efficiency loss has the upper bound \( N_{\text{noise}} I_p \) which depends only on the noise distribution. Since \( N_{\text{noise}} = 0 \) for any symmetric densities, below we calculate \( N_{\text{noise}} \) for some asymmetric densities only. For illustration, consider scalar-value parameter (\( p = 1 \)) so that REL(\( \Sigma_{\beta} \)) = \( \Sigma_{\beta} / CR_{\beta} - 1 \) can be interpreted as the percentage efficiency loss relative to the Cramér-Rao bound.

Example 1 (Mixture of normal). Consider the mixture of normal distribution \( p N(1 - p, 1) + (1 - p) N(-p, \sigma^2) \) for parameters \( p \in [0,1], \sigma^2 > 0 \). This distribution has zero mean. Figure 1 plots \( N_{\text{noise}} \) for two cases: fixed \( p = 0.8 \) and varying \( \sigma^2 \), and fixed \( \sigma^2 = 3 \) and varying \( p \). In both cases, \( N_{\text{noise}} \) is close to zero for the range of parameters considered.

Example 2 (Skew normal). Skew normal distributions are widely used to model data with asymmetric and/or nonnormal distribution. The density is \( f_{e}(x | \mu, \alpha) = 2 \phi(x - \mu) \Phi(\alpha(x - \mu)) \), where \( \mu \) is location parameter, \( \alpha \) is slant parameter, \( \phi(x) \) and \( \Phi(x) \) are the standard normal density and distribution function, respectively. By Andel, Netuka, and Svara (1984), \( f_{e}(x | \mu, \alpha) \) is
the stationary density of the process \(\{e_t\}\) defined by
\[
e_t - \mu = \rho |e_t - \mu| + \sqrt{1 - \rho^2} e_t, \quad \rho = \alpha/\sqrt{1 + \alpha^2},  \\
e_t: \text{iid } N(0,1). \tag{20}
\]
For \(\mu = -\alpha/\sqrt{1 + \alpha^2}/\sqrt{2\pi}\) so that the distribution has zero mean, we plot \(N_{\text{noise}}\) in Figure 2 as a function of the slant parameter \(\alpha\). For \(|\alpha| \in [-2, 2]\), the maximal efficiency loss is upper bounded by 10%. Note that \(|\alpha| = 2\) corresponds to \(|\rho| \approx 0.89\) in (20), which is quite extreme and close to the unit root case. For the even more extreme \(|\alpha| = 3\) (corresponding to \(|\rho| \approx 0.95\), which is unlikely in practice), the upper bound is about 28%.

**Example 3 (Asymmetric Laplace).** Asymmetric Laplace distribution has density \(f_\alpha(x; \mu, p) = p(1 - p) \exp(-(x - \mu)(p - 1)_{x<\mu < 0})\), where \(\mu\) is location parameter and \(p\) is skewness parameter [Koenker and Machado (1999)]. Calculations show \(N_{\text{noise}} = \mu^2 p(1 - p)\). Thus, if \(\mu = 0\), then \(N_{\text{noise}} = 0\) and we have asymptotic efficiency; see Theorem 5(ii).

As illustrated in Examples 1–3, the upper bound for efficiency loss is usually quite small for practical situations. We stress that the upper bound holds regardless of the model design structure and thus tends to be very conservative as it will be further multiplied by the factor \(N_{\text{design}} \leq 1\); see Theorem 5(i). For scalar-value parameter, \(N_{\text{design}} \in [0, 1]\).

### 2.4. Estimating the Optimal Weight

To estimate the unknown optimal weight \(\omega^*\) in Equation (13), we propose the procedure

i. Obtain preliminary estimate \((\hat{\alpha}^0, \hat{\beta}^0)\) of \((\alpha, \beta)\). We defer the discussion to Section 3.
ii. From (1), we estimate the noise \(e_t\) by
\[
\hat{e}_t = \frac{Y_t - X_t^T \hat{\beta}^0}{s(X_t, \hat{\alpha}^0)}, \quad t = 1, \ldots, n. \tag{21}
\]
iii. For any \(x\), we estimate \(f_\alpha(x)\) through the nonparametric kernel density estimator
\[
\hat{f}_\alpha(x) = \frac{1}{nb_n} \sum_{t=1}^n K\left(\frac{x - \hat{e}_t}{b_n}\right), \tag{22}
\]
for bandwidth \(b_n > 0\) and kernel function \(K(\cdot)\). In numerical analysis, we use standard normal density kernel \(K(\cdot)\) and Silverman’s (1986) rule-of-thumb bandwidth.
iv. Estimate \(Q_\alpha(t)\) by \(\hat{Q}_\alpha(t)\), the sample \(t\)-th quantile of \(\hat{e}_1, \ldots, \hat{e}_n\).

v. In (13), let \(\hat{v} = (\hat{Q}_\alpha(t_1), \ldots, \hat{Q}_\alpha(t_N))^T\) and \(\hat{H}\) be obtained by replacing \(f_\alpha(Q_\alpha(t))\) with \(\hat{f}_\alpha(Q_\alpha(t))\) in \(H\) [see (11)]. We propose the estimate
\[
\hat{\omega}^* = \frac{\hat{H}^{-1}}{u^T \hat{H}^{-1} u - \hat{\lambda}^2 v^T \hat{H}^{-1} v} (u - \hat{\lambda} \hat{v}) \quad \text{with} \quad \hat{\lambda} = \frac{v^T \hat{H}^{-1} u}{v^T \hat{H}^{-1} v}. \tag{23}
\]

**Assumption 7.** In (21), \(|\hat{\alpha}^0 - \alpha| + ||\hat{\beta}^0 - \beta|| = O_p(n^{-1/2})\). In (22), \(K(u)\) has bounded support and bounded derivative, and the bandwidth \(b_n\) satisfies \(b_n \to 0\), \(nb_n^2 \to \infty\) are very mild and satisfied for Silverman’s (1986) rule-of-thumb bandwidth \(b_n \propto n^{-1/5}\).
Theorem 6. Suppose Assumptions 1–3 and 7 hold. Then we have: (i) \(|\hat{\omega}^* - \omega^*| = o_p(1)\); and (ii) \(\hat{\beta}(\hat{\omega}^*)\) has the same asymptotic normality in Theorem 3.

2.5. Inferences

As discussed in Section 2.1, while it can be difficult to use \(\hat{\beta}(r)\) to conduct statistical inference, it is straightforward to use \(\hat{\beta}(\hat{\omega}^*)\) to address inference problems. By Theorem 6,
\[
\sqrt{n}[\hat{\beta}(\hat{\omega}^*) - \beta] \Rightarrow N(0, \frac{\Sigma^{-1}}{J_N}),
\]
where
\[
J_N = u^T H^{-1} u - \frac{(v^T H^{-1} u)^2}{v^T H^{-1} v}.
\]
(24)

Recall \(\hat{v}\) and \(\hat{H}\) in Section 2.4 and \(\hat{\sigma}_t = s(X_t, \hat{\omega}^*).\) We can estimate \(\Sigma_t\) and \(J_N\) by
\[
\hat{\Sigma}_1 = \frac{1}{n-1} \sum_{t=1}^n \left( \frac{X_t}{\hat{\sigma}_t} - \hat{\mathcal{M}} \right) \left( \frac{X_t}{\hat{\sigma}_t} - \hat{\mathcal{M}} \right)^T
\]
with
\[
\hat{\mathcal{M}} = \frac{1}{n} \sum_{t=1}^n \frac{X_t}{\hat{\sigma}_t},
\]
(25)

and
\[
\hat{J}_N = u^T \hat{H}^{-1} u - \frac{(v^T \hat{H}^{-1} u)^2}{v^T \hat{H}^{-1} v}.
\]
(26)

Proposition 1. Under Assumptions 1–3 and 7, \(\hat{\Sigma}_1 \overset{p}{\to} \Sigma_1\) and \(\hat{J}_N \overset{p}{\to} J_N.\)

We can use Equation (24) and Proposition 1 to construct confidence interval for each component of \(\beta.\) Also, if we are interested in the null hypothesis
\[
H_0 : A\hat{\beta} = B,
\]
for some matrix \(A \in \mathbb{R}^{r \times p}\) and vector \(B \in \mathbb{R}^r.\)

Assume \(A\) has full rank \(r \leq p.\) By Equation (24) and Proposition 1, we can form the \(\chi^2\)-type statistic
\[
T_n = n \hat{J}_N [A(\hat{\beta}(\hat{\omega}^*) - B)^T (A\hat{\Sigma}_1^{-1} A^T)^{-1} [A(\hat{\beta}(\hat{\omega}^*) - B].
\]
Under \(H_0, T_n\) converges in distribution to the \(\chi^2\) distribution with \(r\) degrees of freedom.

3. Monte Carlo Studies

Consider model of the form
\[
Y_t = \beta_1 X_{11} + \beta_2 X_{12} + s(X_{11}, X_{12}, \alpha)\epsilon_t, \quad t = 1, \ldots, n.
\]
(27)

We use iid standard normals for \(X_{11}\) and \(X_{12},\) and \((\beta_1, \beta_2, \alpha, \alpha_1, \alpha_2) = (0.5, 0.8, 0.5, 0.3, 0.6).\) We consider three widely used forms of heteroscedasticity:

- Heteroscedasticity 1: \(s(X_{11}, X_{12}, \alpha) = \sqrt{\alpha_0 + \alpha_1 X_{11}^2 + \alpha_2 X_{12}^2}\)
- Heteroscedasticity 2: \(s(X_{11}, X_{12}, \alpha) = \alpha_0 + \alpha_1 X_{11} + \alpha_2 X_{12}\)
- Heteroscedasticity 3: \(s(X_{11}, X_{12}, \alpha) = \exp[\alpha_0 + \alpha_1 X_{11} + \alpha_2 X_{12}]\)

Preliminary estimate: We need preliminary estimate \((\hat{\alpha}^*, \hat{\beta}^*)\) in (5) and step (i) of Section 2.4. For identifiability, assume \(\mathbb{E}(\epsilon_t) = 0.\) Then \(\mathbb{E}(Y_t | X_{11}, X_{12}) = \beta_1 X_{11} + \beta_2 X_{12}\) and we can use ordinary least-square method to obtain \(\hat{\beta}^0 = \arg\min_{\beta_1, \beta_2} \sum_{t=1}^n (Y_t - \beta_1 X_{11} - \beta_2 X_{12})^2.\) In Equation (27), write \(\epsilon_t = s(X_{11}, X_{12}, \alpha)\epsilon_t.\) With \(\hat{\beta}^0, \epsilon_t\) can be estimated by \(\hat{\epsilon}_t := Y_t - \hat{\beta}^0 X_{11} - \hat{\beta}^0 X_{12}.\) Next we obtain \(\hat{\alpha}^0.\) For Heteroscedasticity 1, by Remark 5, we assume without loss of generality (WLOG) that \(\mathbb{E}(\epsilon_t^2) = 1\) (otherwise, we can rescale \(\epsilon_t.\)) Then \(\mathbb{E}(\epsilon_t^2 | X_{11}, X_{12}) = \alpha_0 + \alpha_1 X_{11}^2 + \alpha_2 X_{12}^2\) and we can obtain \(\hat{\alpha}^0\) by regressing \(\hat{\epsilon}_t^2\) on \((X_{11}, X_{12}).\)

For Heteroscedasticity 2, we assume WLOG \(\mathbb{E}(\epsilon_t) = 1\) so \(\mathbb{E}(\epsilon_t^2 | X_{11}, X_{12}) = \alpha_0 + \alpha_1 X_{11} + \alpha_2 X_{12}.\) Then we can obtain \(\hat{\alpha}^0\) via regressing \(\hat{\epsilon}_t^2\) on \((X_{11}, X_{12}).\) For Heteroscedasticity 3, \(\mathbb{E}(\log(\epsilon_t^2 | X_{11}, X_{12}) = \alpha_0 + \alpha_1 X_{11} + \alpha_2 X_{12}, \alpha_0^* = \alpha_0 + \epsilon(\log(\epsilon_t)).\) Again, we assume WLOG that \(\mathbb{E}(\log(\epsilon_t)) = 0\) and obtain \(\hat{\alpha}^0\) by regressing log \(\hat{\epsilon}_t^2\) on \((X_{11}, X_{12}).\)

Remark 8. Under standard conditions in least-square estimation, it is not difficult to establish \(\sqrt{n}\)-consistency for \(\hat{\alpha}^0, \hat{\beta}^0.\) To keep length, we skip the details.

Remark 9. Under different identifiability conditions, other methods can be used to obtain \((\hat{\alpha}^0, \hat{\beta}^0).\) For example, if we assume in (27) the median of \(\epsilon_t\) is zero, then we use the least-absolute-deviation (LAD) or median quantile regression to obtain \(\hat{\beta}^0 = \arg\min_{\beta_1, \beta_2} \sum_{t=1}^n |Y_t - \beta_1 X_{11} - \beta_2 X_{12}|.\) For \(\hat{\alpha}^0,\) it can be obtained by the same method discussed above with the least-squares regression replaced by LAD regression. For example, for Heteroscedasticity 1, by Remark 5, we assume WLOG that \(\epsilon_t^2\) has median one, then we can obtain \(\hat{\alpha}^0\) via argmin \(\alpha_0, \alpha_1, \alpha_2; \sum_{t=1}^n |\epsilon_t^2 - \alpha_0 - \alpha_1 X_{11} - \alpha_2 X_{12}|.\)

For noise \(\epsilon_t,\) we consider ten distributions, all with zero mean for identifiability.

- Distribution 1: Standard normal \(N(0, 1).\)
- Distribution 2: Student \(t_3\) with 3 degrees of freedom.
- Distribution 3: Laplace distribution, \(f(x) = 0.5 \exp[-|x|].\)
- Distributions 4–6: mixture of normal (MN): MN1: 0.5 \(N(-2, 1) + 0.5 N(2, 1);\) MN2: 0.5\(N(0, 1) + 0.5 N(0, 0.5^2);\) MN3: 0.8\(N(0.2, 1) + 0.2 N(-0.8, 9).\)
- Distributions 7–8: skew normal (SN) with \(\alpha = 0.5\) and \(\alpha = 0.9,\) see Example 2.
- Distributions 9–10: asymmetric Laplace (AL), Example 3 with \(\mu = -(1 - 2p)/p(1 - p)\) so the distribution has zero mean, skewness parameter \(p = 0.3\) and \(p = 0.6.\)

We compare the following six methods to estimate \((\beta_1, \beta_2):\)

1. OLS: Ordinary least-square argmin \(\sum_{t=1}^n (Y_t - \beta_1 X_{11} - \beta_2 X_{12})^2.\)
2. WLS: Weighted least-square with estimated weights argmin \(\sum_{t=1}^n \frac{(Y_t - \beta_1 X_{11} - \beta_2 X_{12})^2}{s(X_{11}, X_{12})^2}\).
3. WLS-T: Weighted least-square with true weights argmin \(\sum_{t=1}^n \frac{(Y_t - \beta_1 X_{11} - \beta_2 X_{12})^2}{s(X_{11}, X_{12})^2}\).
4. WLS-T: Weighted LAD with true weights argmin \(\sum_{t=1}^n \frac{|Y_t - \beta_1 X_{11} - \beta_2 X_{12}|}{s(X_{11}, X_{12})^2}\). Since we include intercept and use true weight, we do not require \(\epsilon_t\) to have median zero.
5. COWQAE-T: The proposed COWQAE with true weight $\omega^*$ in (13). Throughout, we use nine quantiles $0.1, 0.2, \ldots, 0.9$ for all COWQAE.

6. COWQAE: The proposed COWQAE with estimated weight $\hat{\omega}^*$ in (23).

For estimate $\hat{\beta}_j$ of $\beta_j$, we calculate its mean squared errors (MSE) by averaging 1000 realizations of $(\hat{\beta}_j - \beta_j)^2$, and Tables 1–3 present the relative MSE or MSE(method)/MSE(COWQAE), for the three forms of heteroscedasticity in Equations (28)–(30), using sample size $n = 200$ and $n = 500$. Clearly, COWQAE with estimated optimal weight has almost indistinguishable performance from COWQAE-T with true optimal weight and both deliver substantially better performance than other methods. It is remarkable that COWQAE significantly outperforms WLS and WLAD even when the true weights are used in WLS and WLAD. To further appreciate the performance of COWQAE, we take the average of relative MSE over ten distributions and include these averages at the bottom of Tables 1–3. COWQAE outperforms other methods by a large margin and the result is consistent across the three forms of heteroscedasticity, the two sample sizes, and the two parameters.

4. Application: GDP and Inflation Rate Modeling

We consider two datasets, both available at Federal Reserve Bank of St. Louis web (https://fred.stlouisfed.org). The first dataset contains seasonally adjusted quarterly gross domestic product (GDP) growth rates (in annual percentage) from April 1947 to January 2019. The second dataset contains quarterly inflation rates, calculated as the percentage change of the inflation rates, calculated as the percentage change of the product (GDP) growth rates (in annual percentage) from April
Figure 3 suggest possible heteroscedasticity. Seasonally adjusted total Consumer Price Index, during the Figure 3.

Table 3. Relative MSE (COWQAE always has MSE one) for Heteroscedasticity 3 in (30). In each cell, the first number is the relative MSE for β1 and the bracketed number is the relative MSE for β2. Bottom two rows: average of relative MSE over ten distributions.

| distribution | OLS | WLS | WLS-T | WLAD-T | COWQAE-T |
|--------------|-----|-----|-------|--------|----------|
|              | n=200 |     |       |        |          |
| N(0,1)       |     |     |       |        |          |
|              | 6.78(21.29) | 0.96(1.21) | 0.79(0.75) | 1.26(1.48) | 0.92(1.08) |
|              |     |     |       |        |          |
| Student t3   |     |     |       |        |          |
|              | 10.66(33.73) | 2.05(3.32) | 1.40(1.45) | 1.10(1.23) | 0.93(1.00) |
|              |     |     |       |        |          |
| Laplace      |     |     |       |        |          |
|              | 14.70(41.74) | 3.40(3.10) | 1.53(1.54) | 1.13(1.22) | 0.94(1.01) |
|              |     |     |       |        |          |
| MN1          |     |     |       |        |          |
|              | 10.15(25.13) | 1.58(1.79) | 1.20(1.10) | 0.93(1.03) | 0.93(1.03) |
|              |     |     |       |        |          |
| MN2          |     |     |       |        |          |
|              | 11.68(34.62) | 1.52(1.52) | 1.33(1.28) | 0.91(0.97) | 0.91(0.97) |
|              |     |     |       |        |          |
| MN3          |     |     |       |        |          |
|              | 15.63(37.47) | 1.94(1.56) | 1.82(1.43) | 8.53(6.55) | 1.03(1.06) |
|              |     |     |       |        |          |
| SN, α = 0.5 |     |     |       |        |          |
|              | 5.78(19.49) | 1.05(1.21) | 0.85(0.78) | 1.39(1.59) | 0.99(1.08) |
|              |     |     |       |        |          |
| SN, α = 0.9 |     |     |       |        |          |
|              | 6.77(18.65) | 1.01(1.01) | 0.81(0.77) | 1.32(1.42) | 0.95(0.94) |
|              |     |     |       |        |          |
| AL, p = 0.3 |     |     |       |        |          |
|              | 7.90(21.62) | 1.31(1.70) | 0.97(0.81) | 1.10(1.56) | 0.91(1.14) |
|              |     |     |       |        |          |
| AL, p = 0.6 |     |     |       |        |          |
|              | 10.12(24.86) | 1.20(1.26) | 1.04(0.93) | 1.17(1.57) | 0.92(1.02) |
|              |     |     |       |        |          |
| Average      |     |     |       |        |          |
|              | 11.24(29.52) | 1.96(2.53) | 1.37(1.21) | 1.86(1.83) | 0.93(1.03) |
|              |     |     |       |        |          |
|              | 15.03(37.36) | 2.19(2.26) | 1.76(1.59) | 2.30(2.14) | 0.96(1.01) |

Figure 3. Quarterly GDP growth (left) and inflation rates (right) during April 1947–January 2019.

seasonally adjusted total Consumer Price Index, during the same period. There are \( n = 288 \) observations. The plots in Figure 3 suggest possible heteroscedasticity.

Model with intercept: In all subsequent models, we consider intercept. As discussed in Remark 1, the intercept may cause numerical issues. To address this, we use OLS to obtain the estimated intercept first, denoted by \( \hat{\beta}_0 \), subtract \( \hat{\beta}_0 \) from the response \( Y_t \), use the de-intercepted response in our COWQAE estimation procedure to obtain estimated slope parameters, and finally add \( \hat{\beta}_0 \) back to the model as the intercept.

We consider GDP forecasting. For model \( Y_t = \beta_0 + X_t^T \beta + \sigma(X_t, \alpha) \varepsilon_t \), we divide data into two parts: the first \( J \) observations \( (X_1, Y_1), \ldots, (X_J, Y_J) \) are used for model fitting (i.e., obtaining estimates \( \hat{\beta}_0, \hat{\beta} \)), and the remaining \( n - J \) observations \( (X_{J+1}, Y_{J+1}), \ldots, (X_n, Y_n) \) are used to evaluate the out of sample forecasting performance based on the prediction error \( \sum_{t=J+1}^{n} (Y_t - \hat{\beta}_0 - X_t^T \hat{\beta})^2 \). As in Section 3, we compare COWQAE to four other methods: OLS, WLS with estimated weights, LAD, and WLAD with estimated weights. For preliminary estimate, we follow the same procedure in Section 3. For ease of comparison, we consider relative prediction error defined as the prediction error of a method divided by the prediction error of COWQAE so that COWQAE always has relative prediction error one. We evaluate the relative prediction error using \( J = 100, 105, 110, \ldots, 200 \).

Three types of models are considered: (i) GDP on the past GDP; (ii) GDP on the past inflation rate; and (iii) GDP on the past GDP and the past inflation rate. Our goal is not to select the best model for GDP forecasting, but rather aim to compare the forecasting performance of our proposed estimation method to other estimation methods under various given models. For GDP on the past GDP, the partial autocorrelation function suggests that there is significant correlation at lag one, slightly significant correlation at lag two, and barely significant correlation at lag three, thus we consider up to lag two past GDPs. In all models, we use the most widely used square-root heteroscedasticity. We then have the following models:

Model 1: \( GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \sqrt{\alpha_0 + \alpha_1 GDP_{t-1}^2} \varepsilon_t \)

Model 2: \( GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 GDP_{t-2} + \sqrt{\alpha_0 + \alpha_1 GDP_{t-1}^2 + \alpha_2 GDP_{t-2}^2} \varepsilon_t \)
Figure 4. The relative prediction errors for Model 1 (top left), Model 2 (top right), Model 3 (bottom left), and Model 4 (bottom right). In each plot, the solid, dotted, dashed, and dot-dashed curves correspond to the relative prediction error for OLS, WLS, LAD, and WLAD, respectively.

Figure 5. Series of $\hat{\beta}_1$ in Model 1 when estimated using data in each 50-year moving time window based on five estimation methods: OLS (thin solid), WLS (thin dotted), LAD (thin dashed), WLAD (thin dot-dashed), and COWQAE (thick solid), along with the 95% confidence interval (thick dotted) for COWQAE.

Model 3: $GDP_t = \beta_0 + \beta_1\text{Inflation}_{t-1} \sqrt{\alpha_0 + \alpha_1\text{Inflation}_{t-1}^2} e_t$

Model 4: $GDP_t = \beta_0 + \beta_1GDP_{t-1} + \beta_2\text{Inflation}_{t-1} \sqrt{\alpha_0 + \alpha_1\text{GDP}_{t-1}^2 + \alpha_2\text{Inflation}_{t-1}^2} e_t$

The relative prediction errors for Models 1–4 are reported in Figure 4. The relative prediction errors almost always exceed one (except for very few exceptions where they lie between [0.95, 1]), suggesting almost uniformly better performance of COWQAE over other methods in all models and for all values of $J$ considered. Also, since most of the relative prediction errors lie between [1.05, 1.10], COWQAE has about 5%-10% efficiency gain in terms of prediction error. Note that we have the decomposition

$Y_t - \hat{\beta}_0 - X_t^T \hat{\beta} = [(\beta_0 - \hat{\beta}_0) + X_t^T (\beta - \hat{\beta})] + s(X_t, \alpha) e_t.$

Since $s(X_t, \alpha) e_t$ appears in the prediction error of all methods, it may shadow the better performance of $(\hat{\beta}_0, \hat{\beta})$, especially when $s(X_t, \alpha) e_t$ has a large magnitude. Thus, a 5%-10% efficiency gain could actually mean substantially better performance in $(\hat{\beta}_0, \hat{\beta})$.

Next we examine the estimated slope coefficients in details. For illustration, we consider $\beta_1$ in Model 1. For each $t = 5, 10, \ldots, 85$, we use data in the 50-year moving time window $[t, t + 200]$ to estimate $\beta_1$, and Figure 5 plots the series of $\hat{\beta}_1$ based on the five estimation methods: OLS (thin solid), WLS (thin dotted), LAD (thin dashed), WLAD (thin dot-dashed), and COWQAE (thick solid), along with the 95% confidence interval (thick dotted) for COWQAE at each $t$. The series $\hat{\beta}_1$ from the other four estimation methods are mostly contained within the confidence interval but there are occasional exceptions (suggesting significant difference). Since the confidence interval lies above zero for all $t$, we have strong evidence to reject the null hypothesis $H_0 : \beta_1 = 0$ for each
moving time window, and we conclude that GDP is positively correlated with past GDP.

Using \( J = 100 \) subsamples for model fitting, Figure 6 plots COWQAE predicted GDP vs. observed GDP. The points are reasonably scattered around the diagonal line.

Acknowledgments

We are grateful to an associate editor and a referee for their constructive comments. Part of this work is based on Zhanxiong Xu’s PhD dissertation at Penn State University.

References

Andel, J., Netuka, I., and Svara, K. (1984), “On Threshold Autoregressive Processes,” Kybernetika, 20, 89–106. [1502]
Beran, R. (1974), “Asymptotically Efficient Adaptive Rank Estimates in Location Models,” Annals of Statistics, 2, 248–266. [1502]
Bickel, P. (1982), “On Adaptive Estimation,” Annals of Statistics, 10, 647–671. [1502]
Bollerslev, T. (2008), “Glossary to ARCH (GARCH),” in Time Series Econometrics: Essays in Honor of Robert F. Engle, ed. T. Bollerslev, J. R. Russell, and M. Watson, Oxford: Oxford University Press.
Fama, E. (1975), “Short-Term Interest Rates as Predictors of Inflation,” American Economic Review, 65, 269–282. [1498]
Gutenbrunner, C., and Jurečková, J. (1992), “Regression Rank Scores and Regression Quantiles,” Annals of Statistics, 20, 305–330. [1498,1500]
Jiang, X., Jiang, J., and Song, X. (2012), “Oracle Model Selection for Nonlinear Models Based on Weighted Composite Quantile Regression,” Statistica Sinica, 22, 1479–1506. [1499]
Jurečková, J., and Procházka, B. (1994), “Regression Quantiles and Trimmed Least Squares Estimator in Nonlinear Regression Model,” Journal of Nonparametric Statistics, 3, 201–222. [1499]
Kai, B., Li, R., and Zou, H. (2010), “Local Composite Quantile Regression Smoothing: An Efficient and Safe Alternative to Local Polynomial Regression,” Journal of the Royal Statistical Society, Series B, 72, 49–69. [1499]
Koenker, R., and Portnoy, S. (1987), “L-Estimation for Linear Models,” Econometric Theory, 3, 248–266. [1502]
Koenker, R., and Zhao, Z. (2014), “Efficient Estimation for Time-Varying Coefficient Partial Linear Models,” Journal of Nonparametric Statistics, 30, 680–702. [1499]
Knight, K. (1998), “Limiting Distributions for \( L_1 \) Regression Estimators Under General Conditions,” Annals of Statistics, 26, 755–770.
Koenker, R., and Machado, J. (1999), “Goodness of Fit and Related Inference Processes for Quantile Regression,” Journal of the American Statistical Association, 94, 1296–1309. [1503]
Koenker, R., and Portnoy, S. (1987), “L-Estimation for Linear Models,” Journal of the American Statistical Association, 82, 851–857. [1499,1500]
Koenker, R., and Zhao, Q. (1994), “L-Estimation for Linear Heteroscedastic Models,” Journal of Nonparametric Statistics, 3, 223–235. [1498,1499,1500]
Koenker, R., and Bassett, G. (1978), “Regression Quantiles,” Econometrica, 46, 33–49. [1498]
Koenker, R., and Bassett, G. (1982), “Robust Tests for Heteroscedasticity Based on Regression Quantiles,” Econometrica, 46, 43–61. [1498]
Kreiss, J. P. (1987), “On Adaptive Estimation in Stationary ARMA Processes,” Annals of Statistics, 15, 112–133. [1502]
Linton, O. (1993), “Adaptive Estimation in ARCH Models,” Econometric Theory, 9, 539–569. [1502]
Pollard, D. (1991), “Asymptotics for Least Absolute Deviation Regression Estimators,” Econometric Theory, 7, 186–199.
Portnoy, S., and Koenker, R. (1989), “Adaptive L-Estimation of Linear Models,” Annals of Statistics, 17, 362–381. [1499,1500,1502]
Silverman, B.W. (1986), Density Estimation, London: Chapman and Hall. [1503]
Stout, W. F. (1974), Almost Sure Convergence, New York: Academic Press.
Wang, K., and Wang, H. (2016), “Optimally Combined Estimation for Tail Quantile Regression,” Statistica Sinica, 26, 295–311. [1499,1500]
Wu, W.B. (2005), Nonlinear System Theory: Another Look at Dependence Processes for Quantile Regression, Cambridge: Cambridge University Press. [1500]
Xia, Z., and Koenker, R. (2009), “Conditional Quantile Estimation and Inference for GARCH Models,” Journal of the American Statistical Association, 104, 1696–1712. [1498]
Xu, G., Wang, S., and Huang, J. (2014), “Focused Information Criterion and Model Averaging Based on Weighted Composite Quantile Regression,” Scandinavian Journal of Statistics, 41, 365–381. [1499]
Zhou, Q., Portnoy, S. (1998), “Statistical Inference on Heteroscedastic Models Based on Regression Quantiles,” Journal of Nonparametric Statistics, 9, 239–260. [1498]
Zhu, Q., Zheng, Y., and Li, G. (2018), “Linear Double Autoregression,” Journal of Econometrics, 207, 162–174. [1498,1499,1500]
Zou, H., and Yuan, M. (2008), “Composite Quantile Regression and The Oracle Model Selection Theory,” Annals of Statistics, 36, 1108–1126. [1499]