Online supplementary appendix to

New robust inference for predictive regressions

This supplement includes four appendices, Appendices A, B, C, and D. In Appendix A, we consider a class of nonlinear IV estimators, and discuss why the Cauchy estimator, which is a special case of nonlinear IV estimators, is useful in inference problems. Appendix B presents some useful lemmas, Lemmas S.1-S.8, and their proofs; Appendix C provides the proofs of the main results in the paper; and Appendix D provides some additional simulation results on finite sample performance of inference approaches dealt with.

Appendix A: Nonlinear IV Approaches

As mentioned in the main paper, the Cauchy estimator $\check{\beta}$ is the special case of a class of nonlinear IV estimators $\tilde{\beta}(\gamma)$ with $\gamma(\cdot) = sign(\cdot)$, where

$$\tilde{\beta}(\gamma) = \left(\sum_{t=1}^{T} \gamma(x_{t-1}) x_{t-1} \right)^{-1} \sum_{t=1}^{T} \gamma(x_{t-1}) y_t$$

for some function $\gamma : \mathbb{R} \to \mathbb{R}$. Clearly, $\tilde{\beta}(sign) = \check{\beta}$. Moreover, if we let ι be an identity function: $\iota(x) = x$, then $\tilde{\beta}(\iota)$ becomes the OLS estimator $\hat{\beta}$. In this section, we discuss why the choice of $\gamma(\cdot) = sign(\cdot)$, as in the Cauchy estimator, is important and useful when considering the class of nonlinear IV estimators in an inference problem. To explain the idea, we focus on the issues about the predictor x_t, and assume that $v_t = \sigma$ for all $t \geq 1$.

A.1 Nonstationary Predictor

The asymptotics for (near) unit root processes under various transformations are well known (see Park and Phillips (2001) and Park (2003)). For instance, let x_t be a unit root process and

$(x_t, \sum_{s=1}^t \varepsilon_s)$ satisfy the functional CLT with a limiting bivariate Brownian motion (X, W). It then follows from m Park and Phillips (2001) that for regularly integrable functions f and g^2

$$\begin{aligned} \frac{1}{T^{1/2}} \sum_{t=1}^T f(x_{t-1}) &\to_d L_X(1,0) \int_{-\infty}^{\infty} f(x)dx, \\ \frac{1}{T^{1/4}} \sum_{t=1}^T g(x_{t-1})\varepsilon_t &\to_d \left(L_X(1,0) \int_{-\infty}^{\infty} g^2(x)dx \right)^{1/2} Z(1), \end{aligned} \tag{S.1}$$

where $L_X(1,0)$ is the local time at the origin of X and Z is a Brownian motion independent of X. In particular, $\left(L_X(1,0) \int_{-\infty}^{\infty} g^2(x)dx\right)^{1/2} Z(1) =_d \mathbb{MN}\left(0, L_X(1,0) \int_{-\infty}^{\infty} g^2(x)dx\right)$, where $\mathbb{MN}$ is a mixed normal distribution.

On the other hand, if f and g^2 are asymptotically homogeneous functions, then

$$\begin{aligned} \frac{1}{T} f_\nu(T^{1/2}) \sum_{t=1}^T f(x_{t-1}) &\to_d \int_0^1 f_H(X(r))dr, \\ \frac{1}{T^{1/2} g_\nu(T^{1/2})} \sum_{t=1}^T g(x_{t-1})\varepsilon_t &\to_d \int_0^1 g_H(X(r))dW(r), \end{aligned} \tag{S.2}$$

where F_ν and F_H are, respectively, the asymptotic order and the limit homogeneous function of asymptotically homogeneous function $F = f, g$. The reader is referred to Section 3 of Park and Phillips (2001) for more detailed discussions about the asymptotics (S.1) and (S.2) as well as the precise definitions of the regularly integrable and asymptotically homogeneous functions.[1] Note also that for a near unit root process the asymptotics (S.1) and (S.2) remain valid if X is replaced by the limiting Ornstein-Uhlenbeck process of (x_t) (see, e.g., Section 3 of Park (2003)).

Importantly, the limit distribution of $\sum_{t=1}^T g(x_{t-1})\varepsilon_t$ in (S.2) is not Gaussian for an asymptotically homogeneous g^2 except in some special cases including $g_H(x) = sign(x)$. In particular, the sign function $sign(\cdot)$ is asymptotically homogeneous with $sign_\nu(\lambda) = 1$ for all λ and $sign_H(x) = sign(x)$. Since $\int_0^r sign(X(s))dW(s)$ is a Brownian motion by Lévy's characterization of Brownian motion, we have

$$\frac{1}{T^{1/2}} \sum_{t=1}^T sign(x_{t-1})\varepsilon_t \to_d W(1).$$

Using the asymptotics (S.1) and (S.2), one may construct a nonlinear IV estimator $\tilde{\beta}(\gamma)$ being asymptotically Gaussian for a proper choice of γ, i.e., for γ square integrable or the sign

[1] Similar asymptotic results for a diffusion process can be found in Kim and Park (2017).

function. For such γ, one may use the following test statistic

$$\tilde{\tau}(\gamma) = \frac{\sum_{t=1}^{T} \gamma(x_{t-1})x_{t-1}}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}} \times \tilde{\beta}(\gamma)$$

to test the null hypothesis of $\beta = 0$.[2]

Proposition S.1. Let Assumption 2.1 hold with $v_t = \sigma$ for all $t \geq 1$, and let (x_t) be a unit root process. Further assume that the convergences in (S.1) and (S.2) hold.

(a) Let $\gamma(\cdot) = sign(\cdot)$ or γ^2 be regularly integrable. Under $\beta = 0$,

$$\tilde{\tau}(\gamma) \to_d \mathbb{N}(0, \sigma^2)$$

(b) Let $\gamma(\cdot) = sign(\cdot)$. Under $\beta \neq 0$,

$$\frac{1}{T}\tilde{\tau}(\gamma) \to_d \beta \int_0^1 |B(r)|dr.$$

(c) Let $\gamma^2(x)$ be regularly integrable, and let $x\gamma(x)$ be either asymptotically homogeneous or regularly integrable such that $\int_{-\infty}^{\infty} x\gamma(x)dx \neq 0$. Under $\beta \neq 0$, we have $\tilde{\tau}(\gamma) \to_p \infty$ and $\tilde{\tau}(\gamma) = o_p(T)$.

Proof. Note that

$$\tilde{\tau}(\gamma) = \frac{\sum_{t=1}^{T} \gamma(x_{t-1})y_t}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}} = \beta \frac{\sum_{t=1}^{T} \gamma(x_{t-1})x_{t-1}}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}} + \frac{\sum_{t=1}^{T} \gamma(x_{t-1})u_t}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}}.$$

The stated results in the parts (a) and (b) then follow immediately from the convergences (S.1) and (S.2) since, in particular for $\gamma(\cdot) = sign(\cdot)$, we have

$$\sum_{t=1}^{T} sign^2(x_{t-1}) = T(1 + o_p(1)), \quad \frac{1}{T^{3/2}} \sum_{t=1}^{T} |x_{t-1}| \to_d \int_0^1 |B_r|dr.$$

We let $\iota(x) = x$. As for Part (c), we first let γ^2 and $\iota\gamma$ be regularly integrable. Then, under $\beta \neq 0$,

$$\tilde{\tau}(\gamma) = \beta \frac{\sum_{t=1}^{T} \gamma(x_{t-1})x_{t-1}}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}}(1 + o_p(1))$$

[2]When $\gamma(\cdot) = sign(\cdot)$, $\tilde{\tau}(\gamma) = T^{-1/2}\sum_{t=1}^{T} sign(x_{t-1})y_t$, which is a special case of $\tau(v)$ in (3) with $v_t = 1$.

and

$$\frac{1}{T^{1/4}} \frac{\sum_{t=1}^{T} \gamma(x_{t-1}) x_{t-1}}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}} \to_d \frac{L_X(1,0) \int_{-\infty}^{\infty} \gamma(x) x dx}{\left(L_X(1,0) \int_{-\infty}^{\infty} \gamma^2(x) dx\right)^{1/2}}$$

by (S.1), from which $\tilde{\tau}(\gamma) = o_p(T)$.

On the other hand, if $\iota\gamma$ is asymptotically homogeneous and γ^2 is regularly integrable, then

$$\frac{1}{T^{3/4}(\iota\gamma)_\nu(T^{1/2})} \frac{\sum_{t=1}^{T} \gamma(x_{t-1}) x_{t-1}}{\left(\sum_{t=1}^{T} \gamma^2(x_{t-1})\right)^{1/2}} \to_d \frac{\int_0^1 (\iota\gamma)_H(X_r) dr}{\left(L_X(1,0) \int_{-\infty}^{\infty} \gamma^2(x) dx\right)^{1/2}}$$

by (S.1) and (S.2). Since γ^2 is regularly integrable, $\gamma(\lambda)\lambda^{1/2} = o(1)$, and hence, $(\iota\gamma)(\lambda) = o(\lambda^{1/2})$. It then follows from the construction of the asymptotically homogeneous function that $(\iota\gamma)_\nu(T^{1/2}) = o(T^{1/4})$. Therefore, if γ^2 is regularly integrable and $\iota\gamma$ is either regularly integrable or asymptotically homogeneous, then $\tilde{\tau}(\gamma) = o_p(T)$ as required. □

According to Proposition S.1 (a) and (b), one may easily conduct a Gaussian inference using the nonlinear IV estimator with $\gamma(\cdot) = sign(\cdot)$. Moreover, Proposition S.1 (a) and (c) imply that a similar Gaussian inference can be conducted using a square integrable γ. However, Proposition S.1 (b) and (c) imply that under $\beta \neq 0$ the divergence rate of $\tilde{\tau}(sign)$ is faster than that of $\tilde{\tau}(\gamma)$ with a square integrable γ, which implies the test with $\tilde{\tau}(sign)$ tends to have a better power property than the test with $\tilde{\tau}(\gamma)$ in finite samples.

A.2 Stationary Predictor

For a stationary x_t, the asymptotic distribution of the nonlinear IV based test statistic $\tilde{\tau}(\gamma)$ can be obtained easily and is given by a Gaussian distribution under $\beta = 0$ when $\mathbb{E}|\gamma^2(x_{t-1})| < \infty$. However, if x_t has a heavy-tailed marginal distribution and $\mathbb{E}|\gamma^2(x_{t-1})|$ is unbounded for a given γ, then the limit distribution of $\tilde{\tau}(\gamma)$ is generally non-Gaussian. Therefore, the choice of γ, as in the unit root type predictor, is important in a Gaussian inference relying on the nonlinear IV. Importantly, the Cauchy based test statistic is always asymptotically Gaussian since $\mathbb{E}|\gamma^2(x_{t-1})| = 1$, and hence, the result of Proposition S.1 (a) remains valid. Moreover, it is shown in Theorem 3.3 that the test statistic based on the Cauchy estimator diverges under the alternative hypothesis and its divergence rate is no slower than the usual $\sqrt{T}$ rate for any nontrivial stationary process. As a conclusion of Section 2.3, the Cauchy estimator can be used to construct a robust inference having a Gaussian limit with no significant loss of testing power compared with other nonlinear IV based methods.

Appendix B: Useful Lemmas

Lemma S.1. Let Assumption 3.1 hold, and let $f_h(s) = f(s/h)$ for $f = K, K^2$. We have

$$\sup_{h \le r \le 1} \left| \frac{1}{hT} \sum_{t=1}^{T} f_h(r - t/T) - \int_0^1 f(s) ds \right| = O(1/(h^2 T)).$$

Proof. We only prove the result for the case $f = K$ since the argument in the case $f = K^2$ is similar. For the proof, we define a function $I_{r,h} : [0, 1] \to \{0, 1\}$ for $r \in [0, 1]$ and $h > 0$ as $I_{r,h}(s) = 1\{r - h \le s \le r\}$. We then write

$$\sup_{h \le r \le 1-h} \sum_{t=1}^{T} \int_{t-1}^{t} |K_h(r - t/T) - K_h(r - s/T)| ds = A_T(r) + B_T(r) + C_T(r),$$

where

$$\begin{aligned}
A_T(r) &= \sum_{t=1}^{T} \left(\int_{t-1}^{t} |K_h(r - t/T) - K_h(r - s/T)| ds \right) I_{r,h-1/T}(t/T), \\
B_T(r) &= \sum_{t=1}^{T} \left(\int_{t-1}^{t} |K_h(r - t/T) - K_h(r - s/T)| ds \right) (1 - I_{r,h+1/T}(t/T)), \\
C_T(r) &= \sum_{t=1}^{T} \left(\int_{t-1}^{t} |K_h(r - t/T) - K_h(r - s/T)| ds \right) (I_{r,h+1/T}(t/T) - I_{r,h-1/T}(t/T)).
\end{aligned}$$

By Assumption 3.1 (a) and (c), we have

$$\sup_{h \le r \le 1} A_T(r) \le C \sum_{t=1}^{T} \int_{t-1}^{t} \frac{|t - s|}{hT} ds \le C \frac{1}{h}.$$

Moreover, $\int_{t-1}^{t} |K_h(r - t/T) - K_h(r - s/T) = 0$ for all t satisfying $I_{r,h+1/T}(t/T) = 0$ by Assumption 3.1 (a), and hence, $\sup_{h \le r \le 1} B_T(r) = 0$. Moreover, it may be deduced from Assumption 3.1 (a) that

$$\sup_{h \le r \le 1} \sum_{t=1}^{T} |I_{r,h+1/T}(t/T) - I_{r,h-1/T}(t/T)| \le 2,$$

from which, jointly with Assumption 3.1 (b), we have $\sup_{h \le r \le 1} C_T(r) = O(1)$. The stated result for $f = K$ then follows immediately since

$$\frac{1}{hT} \int_0^T K_h(r - s/T) ds = \frac{1}{h} \int_0^1 K_h(r - s) ds = \int_0^1 K(s) ds \tag{S.3}$$

for $r \in [h, 1]$, by Assumption 3.1 (a) and the change of variable in integrals. □

Lemma S.2. Let Assumption 2.2 hold. As $h \to 0$,

$$\sup_{r \in \mathcal{C}_h, 0 < h' < h} |\sigma_T^2(r) - \sigma_T^2(r - h')| = o_{a.s.}(1).$$

Proof. We have

$$|\sigma_T^2(r) - \sigma_T^2(r - h')| \leq |\sigma_T^2(r) - \sigma^2(r)| + |\sigma_T^2(r - h') - \sigma^2(r - h')| + |\sigma^2(r) - \sigma^2(r - h')|.$$

Note that, under our conventions, $\sup_{0 \leq r \leq 1} |\sigma_T^2(r) - \sigma^2(r)| = o_{a.s.}(1)$ since $\sigma_T \to_d \sigma$ in Assumption 2.2 holds almost surely on $\mathbf{D}_{\mathbb{R}^+}[0, 1]$ endowed with the uniform topology. It follows that the first two terms are of $o_{a.s.}(1)$.

As for the last term, we note that a càdlàg function is uniformly right-continuous on finite closed intervals (see, e.g., Applebaum (2009), pp. 140). It follows that an associated càdlàg function $\sigma_-(r) = \sigma(r-)$ is uniformly left continuous, and hence,

$$\sup_{r \in [h,1], 0 < h' < h} |\sigma^2(r-) - \sigma^2(r - h')| = o_{a.s.}(1).$$

However, $\sigma^2(r-) = \sigma^2(r)$ for all $r \in \mathcal{C}_h$, which completes the proof. □

Lemma S.3. Let Assumptions 2.2 and 3.1 hold. If $h \to 0$ and $h^2 T \to \infty$, then for any fixed real number $c \geq 0$

$$\sup_{r \in \mathcal{C}_h} |\hat{\sigma}_1^2(r - c/T) - \sigma_T^2(r)| = o_p(1).$$

Proof. We have

$$\hat{\sigma}_1^2(r - c/T) - \sigma_T^2(r) = \frac{\sum_{t=1}^T (\sigma_T^2(t/T) - \sigma_T^2(r)) K_h(r - c/T - t/T)}{\sum_{t=1}^T K_h(r - c/T - t/T)}$$

and $\sup_{h \leq r \leq 1} hT / (\sum_{t=1}^T K_h(r - c/T - t/T)) = O_p(1)$ by Lemma S.1 and Assumption 3.1 (a).

To complete the proof, we write

$$\frac{1}{hT} \sum_{t=1}^T \left(\sigma_T^2(t/T) - \sigma_T^2(r)\right) K_h(r - c/T - t/T) = A_T(r) + B_T(r),$$

where

$$A_T(r) = \frac{1}{hT}\sum_{t=1}^{T}\left(\sigma_T^2(t/T)1\{t/T\in(r-h,r]\}-\sigma_T^2(r)\right)K_h(r-c/T-t/T)ds,$$
$$B_T(r) = \frac{1}{hT}\sum_{t=1}^{T}\sigma_T^2(t/T)\left(1\{(t+c)/T\in[r-h,r]\}-1\{t/T\in(r-h,r]\}\right)K_h(r-c/T-t/T)ds.$$

We can deduce from Lemmas S.1 and S.2 that

$$\sup_{r\in\mathcal{C}_h}|A_T(r)| = \sup_{r\in\mathcal{C}_h,0<h'<h}|\sigma_T^2(r)-\sigma_T^2(r-h')|\left(\frac{1}{hT}\sum_{t=1}^{T}K_h(r-t/T)\right) = o_p(1).$$

We note that σ_T and K are bounded due to Assumptions 2.1 (a) and 3.1 (b). Also, we have for a fixed $c\geq 0$ and large T

$$\begin{aligned}&\left|1\left\{\frac{t+c}{T}\in[r-h,r]\right\}-1\left\{\frac{t}{T}\in(r-h,r]\right\}\right|\\&\leq 1\left\{\frac{t}{T}\in\left[r-\frac{c}{T}-h,r-h\right]\right\}+1\left\{\frac{t}{T}\in\left[r-\frac{c}{T},r\right]\right\},\end{aligned}$$

and hence,

$$\sup_{h\leq r\leq 1}\sum_{t=1}^{T}|1\{(t+c)/T\in[r-h,r]\}-1\{t/T\in(r-h,r]\}|\leq 2c+2.$$

It follows that

$$\begin{aligned}\sup_{h\leq r\leq 1}|B_T(r)| &\leq \frac{1}{hT}\sup_{h\leq r\leq 1}|\sigma_T^2(r)K(r)|\sup_{h\leq r\leq 1}\sum_{t=1}^{T}|1\{(t+c)/T\in[r-h,r]\}-1\{t/T\in(r-h,r]\}|\\&= O_p(1/(hT)).\end{aligned}$$

This completes the proof. □

Lemma S.4. Let Assumptions 2.1, 3.1 and 3.3 hold. If $h\to 0$, $h^2T\to\infty$ and $h^\kappa T = O(1)$ for some $\kappa > 2$, then

$$\sup_{h\leq r\leq 1}|\hat{\sigma}_2^2(r)| = O_p\left((\log(hT)/(hT))^{1/2}T^{2q}\right).$$

Proof. Since $\sup_{h\leq r\leq 1} hT/(\sum_{t=1}^{T}K_h(r-t/T)) = O_p(1)$, it suffices to show that

$$\sup_{h\leq r\leq 1}\left|\sum_{t=1}^{T}v_t^2(\varepsilon_t^2-1)K_h(r-t/T)\right| = O_p((hT\log(hT))^{1/2}T^{2q}).$$

To complete the proof, we split the interval $[h,1]$ into $\bar{k}$ intervals of the form $I_k = 1\{r|r_k \leq r \leq r_{k+1}\}$, where $r_k = h + kh^{\bar{\kappa}}$ for $k = 0, \cdots, [(1-h)/h^{\bar{\kappa}}]$ for some $\bar{\kappa} \geq (\kappa+2)/2$. Then we can write

$$\sup_{h \leq r \leq 1} \left| \sum_{t=1}^{T} v_t^2(\varepsilon_t^2 - 1) K_h(r - t/T) \right| \leq \max_{0 \leq k \leq \bar{k}} |S_T(r_k)| + R_T,$$

where $S_t(r_k) = \sum_{s=1}^{t} v_s^2(\varepsilon_s^2 - 1) K_h(r_k - s/T)$ for $t = 1, \cdots, T$, and

$$R_T = \max_{0 \leq k \leq \bar{k}} \sup_{r \in I_k} \sum_{t=1}^{T} |v_t^2(\varepsilon_t^2 - 1)| \left| K_h(r - t/T) - K_h(r_k - t/T) \right|.$$

For R_T, we note that for each k, $r_{k+1} = r_k + h^{\bar{\kappa}}$ and

$$|K_h(r - t/T) - K_h(r_k - t/T)| = |K_h(r - t/T) - K_h(r_k - t/T)| \, 1\{r_k - h \leq t/T \leq r_k + h^{\bar{\kappa}}\}$$

for $r \in I_k$. It follows that

$$\begin{aligned}
R_T &= \max_{0 \leq k \leq \bar{k}} \sup_{r \in I_k} \sum_{t=1}^{T} |v_t^2(\varepsilon_t^2 - 1)| \left| K_h(r - t/T) - K_h(r_k - t/T) \right| 1\{r_k - h \leq t/T \leq r_k + h^{\bar{\kappa}}\} \\
&\leq C h^{\bar{\kappa}-1} \sum_{t=1}^{T} |v_t^2(\varepsilon_t^2 - 1)| 1\{r_k - h \leq t/T \leq r_k + h^{\bar{\kappa}}\} \\
&\leq C h^{\bar{\kappa}-1} \left(\max_{0 \leq k \leq \bar{k}} \sum_{t=1}^{T} v_t^4 1\{r_k - h \leq t/T \leq r_k + h^{\bar{\kappa}}\} \right)^{1/2} \left(\sum_{t=1}^{T} (\varepsilon_t^2 - 1)^2 \right)^{1/2} \\
&\leq O_p(h^{\bar{\kappa}-1} T^{1/2} (hT)^{1/2}),
\end{aligned}$$

where the second line follows from Assumption 3.1, the third line holds due to the Cauchy-Schwarz inequality, and the last line follows from the fact that for any $\bar{\kappa} \geq 1$

$$\max_{0 \leq k \leq \bar{k}} \sum_{t=1}^{T} v_t^4 1\{r_k - h \leq t/T \leq r_k + h^{\bar{\kappa}}\} \leq 2 \sup_{h \leq r \leq 1} \sum_{t=1}^{T} v_t^4 1\{r - h \leq t/T \leq r\} = O_p(hT).$$

Consequently, we have $R_T = O_p((hT)^{1/2})$ since $\bar{\kappa} \geq 2$ and $h^{2\bar{\kappa}-2} T \leq h^{\kappa} T = O(1)$.

As for $S_T(r_k)$, we note that for each $k = 1, \cdots, \bar{k}$, $(S_t(r_k), \mathcal{F}_t)$ is a square integrable martingale since $(\varepsilon_t^2 - 1, \mathcal{F}_t)$ is an MDS and v_t^2 is $\mathcal{F}_t$-adapted such that $\varepsilon_t^2 - 1$ has a finite second moment and $\sup_{r \in [h,1]} \sum_{t=1}^{T} v_t^2 1\{r - h \leq t/T \leq r\} = O_p(hT)$ by Assumptions 2.1. The predictive quadratic variation $\langle S(r_k) \rangle$ and the total quadratic variation $[S(r_k)]$ of $S(r_k)$ are respectively given by

$$\langle S(r_k)\rangle_T = \sum_{t=1}^{T} v_t^4 E[(\varepsilon_t^2-1)^2|\mathcal{F}_{t-1}]K_h^2(r_k-t/T),$$
$$[S(r_k)]_T = \sum_{t=1}^{T} v_t^4(\varepsilon_t^2-1)^2 K_h^2(r_k-t/T).$$

Clearly, $\max_{0\le k\le \bar{k}}\langle S(r_k)\rangle_T = O_p(hT)$ and $\max_{0\le k\le \bar{k}}[S(r_k)]_T = O_p(hTT^{4q})$. It follows that $\max_{0\le k\le \bar{k}}|\langle S(r_k)\rangle_T + [S(r_k)]_T| = O_p(hTT^{4q})$ which implies that for any δ, we can find a constant $M>0$ such that

$$P\left(\max_{0\le k\le \bar{k}}|\langle S(r_k)\rangle_T + [S(r_k)]_T| < M(hTT^{4q})\right)\le \delta.$$

Consequently, we have

$$\begin{aligned}
&P\left(\max_{0\le k\le \bar{k}}|S_T(r_k)| > M(hTT^{4q}\log(hT))^{1/2}\right)\\
&\le P\left(\max_{0\le k\le \bar{k}}|S_T(r_k)| > M(hTT^{4q}\log(hT))^{1/2}, \max_{0\le k\le \bar{k}}|\langle S(r_k)\rangle_T + [S(r_k)]_T| < M(hTT^{4q})\right)+\delta\\
&\le \sum_{k=0}^{\bar{k}} P\left(|S_T(r_k)| > M(hTT^{4q}\log(hT))^{1/2}, \langle S(r_k)\rangle_T + [S(r_k)]_T < M(hTT^{4q})\right)+\delta\\
&\le \frac{2}{h^{\bar{\kappa}}}\exp\left(-M\log(hT)/2\right)+\delta,
\end{aligned}$$

where the last line follows from the two-sided martingale exponential inequality (see, e.g., Theorem 2.1 of Bercu and Touati (2008)). Moreover, $h^{-\bar{\kappa}}\exp\left(-M\log(hT)/2\right) = (h^{\bar{\kappa}}(hT)^{M/2})^{-1}\to 0$ for all $M\ge 2\bar{\kappa}$ as $h^2T\to\infty$. This completes the proof. □

Lemma S.5. Let Assumptions 2.1, 3.1 and 3.2 hold. If $h\to 0$ and $h^2T\to\infty$, then

$$\sup_{h\le r\le 1}|\hat{\sigma}_3^2(r)| = O_p\left(T^{2p}/(hT)\right).$$

Proof. By Lemma 3.1, we have $(\hat{\beta}-\beta)^2 = O_p\left(T^{2p}\left(\sum_{t=1}^T x_{t-1}^2\right)^{-1}\right)$. Moreover, we may deduce from Lemma S.1 with Assumptions 3.1 that for some $0<M<\infty$

$$\sup_{h\le r\le 1}\left|\frac{\sum_{t=1}^T x_{t-1}^2 K_h(r-t/T)}{\sum_{t=1}^T K_h(r-t/T)}\right| \le \frac{M}{hT}\left(\sum_{t=1}^{T} x_{t-1}^2\right)(1+o_p(1)),$$

from which we have the stated result. □

Lemma S.6. Let Assumptions 2.1, 3.1 and 3.2 hold. If $h \to 0$ and $h^2 T \to \infty$, then

$$\sup_{h \le r \le 1} |\hat{\sigma}_4^2(r)| = O_p\left(T^{2p}/(hT)\right).$$

Proof. As in the proof of Lemma S.5, we have $\sup_{h \le r \le 1} |\sum_{t=1}^T K_h(r - t/T)|^{-1} = O_p((hT)^{-1})$ and $\hat{\beta} - \beta = O_p\left(T^p \left(\sum_{t=1}^T x_{t-1}^2\right)^{-1/2}\right)$. Moreover, we have

$$\sup_{h \le r \le 1} \left| \sum_{t=1}^T x_{t-1} u_t K_h(r - t/T) \right| = O_p\left(T^p \left(\sum_{t=1}^T x_{t-1}^2 \right)^{1/2} \right)$$

by Assumption 3.2. This completes the proof.

□

Lemma S.7. Let Assumptions 2.1, 3.1 and 3.2 hold. If $h \to 0$, $h^2 T \to \infty$ and $h^\kappa T = O(1)$ for some $\kappa > 2$, then

$$\sup_{h \le r \le 1} \frac{1}{\hat{\sigma}^2(r)}, \quad \sup_{h \le r \le 1} \frac{1}{\hat{\sigma}_1^2(r) + \hat{\sigma}_2^2(r)} = O_p(1).$$

Proof. It follows from Lemmas S.4-S.6 that

$$\frac{1}{\hat{\sigma}(r)} = \frac{\sum_{t=1}^T K_h(r - t/T)}{\sum_{t=1}^T \sigma_T^2(t/T) K_h(r - t/T)} + o_p(1)$$

uniformly in $r \in [h, 1]$. However, $\sigma_T^2(r) \ge \underline{v} > 0$ by Assumption 2.1, and hence, $\sup_{h \le r \le 1} 1/\hat{\sigma}^2(r) \ge \underline{v} + o_p(1) = O_p(1)$. Similarly, we can show that $\sup_{h \le r \le 1} 1/(\hat{\sigma}_1^2(r) + \hat{\sigma}_2^2(r)) = O_p(1)$. □

Lemma S.8. Under Assumptions 2.1 and 2.2,

$$\left(W_T, \sigma_T, \int \sigma_T(r) dW_T(r) \right) \to_d \left(W, \sigma, \int \sigma(r) dW(r) \right)$$

in $\mathbf{D}_{\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}}[0, 1]$.

Proof. The lemma follows from Theorem 4.6 of Kurtz and Protter (1991) (see also Theorem 2.1 of Hansen (1992). □

Appendix C: Proofs of the Main Results

Proof of Lemmas 2.1. The stated result follows immediately from Lemma S.8. □

Proof of Lemma 3.1. The stated result follows immediately from Assumptions 2.1 and 3.2. □

Proof of Proposition 3.2. Assumption 3.4 implies that $h^2 T \to \infty$ and $h^\kappa T = O(1)$ for some $\kappa > 2$. The stated results then follow immediately from Lemmas S.3-S.6. $\square$

Proof of Theorem 3.3. We define $\tilde{\sigma}$ by $\tilde{\sigma}(r) = \hat{\sigma}_1(r - 1/T) + \hat{\sigma}_2(r - 1/T)$ for $r \in [h, 1]$ and $\tilde{\sigma}(r) = \sigma_T(r)$ for $r \in [0, h)$. We also define $\bar{\sigma}$ as

$$\bar{\sigma}(r) = \tilde{\sigma}(r)1\{\sigma(s) = \sigma(s-), s \in (r - h, r]\} + \sigma_T(r)1\{\sigma(s) \neq \sigma(s-), s \in (r - h, r]\}$$

for $r \in [h, 1]$, and $\bar{\sigma}(r) = \sigma_T(r)$ for $r \in [0, h)$.

Clearly, $\bar{\sigma}(r)$ is a $\mathcal{F}_{Tr}$-adapted càdlàg process such that $\bar{\sigma} \to_d \sigma$ since

$$\sup_{0 \le r \le 1} |\bar{\sigma}^2(r) - \sigma_T^2(r)| \le \sup_{r \in \mathcal{C}_h} |\bar{\sigma}^2(r) - \sigma_T^2(r)| \le \sup_{r \in \mathcal{C}_h} |\hat{\sigma}_1^2(r - 1/T) - \sigma_T^2(r)| + \sup_{h \le r \le 1} |\hat{\sigma}_2(r)|$$

which is of $o_p(1)$ by Lemmas S.3 and S.4. It then follows from Lemma S.8 that under $\beta = 0$

$$\tau(\bar{\sigma}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{sign(x_{t-1}) y_t}{\bar{\sigma}(t/T)} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} sign(x_{t-1}) \varepsilon_t \frac{\sigma_T(t/T)}{\bar{\sigma}(t/T)} \to_d \mathbb{N}(0, 1).$$

To complete the proof, we show that

$$\tau(\hat{\sigma}) - \tau(\tilde{\sigma}),\ \tau(\tilde{\sigma}) - \tau(\bar{\sigma}) = o_p(1).$$

We write

$$\tau(\hat{\sigma}) - \tau(\tilde{\sigma}) = A_T - B_T + C_T,$$

where

$$A_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{hT} sign(x_{t-1}) \varepsilon_t \frac{\sigma_T(t/T)}{\hat{\sigma}((t-1)/T)}, \quad B_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{hT} sign(x_{t-1}) \varepsilon_t \frac{\sigma_T(t/T)}{\tilde{\sigma}(t/T)},$$
$$C_T = \frac{1}{\sqrt{T}} \sum_{t=hT+1}^{T} sign(x_{t-1}) \varepsilon_t \left(\frac{\sigma_T(t/T)}{\hat{\sigma}((t-1)/T)} - \frac{\sigma_T(t/T)}{\tilde{\sigma}(t/T)} \right).$$

To show $A_T, B_T = o_p(1)$, we note that

$$\frac{1}{\sqrt{hT}} \sum_{t=1}^{hT} sign(x_{t-1}) \varepsilon_t \sigma_T(t/T), \quad \frac{1}{\sqrt{hT}} \sum_{t=1}^{hT} sign(x_{t-1}) \varepsilon_t = O_p(1)$$

by Lemma S.8. It follows that

$$A_T = \sqrt{h}\frac{1}{\hat{\sigma}(h)}\left(\frac{1}{\sqrt{hT}}\sum_{t=1}^{hT} sign(x_{t-1})\varepsilon_t\sigma_T(t/T)\right) = O_p(\sqrt{h})$$

since $1/\hat{\sigma}(h) = O_p(1)$ by Lemma S.7, and

$$B_T = \sqrt{h}\left(\frac{1}{\sqrt{hT}}\sum_{t=1}^{hT} sign(x_{t-1})\varepsilon_t\right) = O_p(\sqrt{h}).$$

For C_T, we have

$$\begin{aligned}|C_T| &\leq \frac{1}{\sqrt{T}}\sum_{t=hT+1}^{T}|\sigma_T(t/T)\varepsilon_t|\left|\frac{\tilde{\sigma}^2(t/T)-\hat{\sigma}^2((t-1)/T)}{\hat{\sigma}((t-1)/T)\tilde{\sigma}(t/T)(\tilde{\sigma}(t/T)+\sigma_T(t/T))}\right| \\ &\leq O_p(\sqrt{T})\times\sup_{h\leq r\leq 1}|\tilde{\sigma}^2(r)-\hat{\sigma}^2(r-1/T)|\end{aligned}$$

since $\sum_{t=hT+1}^{T}|\sigma_T(t/T)\varepsilon_t| = O_p(T)$ and $1/\hat{\sigma}(r), 1/\tilde{\sigma}(r) = O_p(1)$ uniformly in $h \leq r \leq 1$ by Lemma S.7. However, it follows from Lemmas S.5 and S.6 that

$$\sup_{h\leq r\leq 1}|\tilde{\sigma}^2(r)-\hat{\sigma}^2(r-1/T)| \leq \sup_{h\leq r\leq 1}|\hat{\sigma}_3^2(r)| + \sup_{h\leq r\leq 1}|\hat{\sigma}_4^2(r)| = O_p(T^{2p}/(hT)),$$

from which we have

$$C_T = O_p(T^{2p}/(hT^{1/2})) = o_p(1)$$

due to Assumption 3.4 (a). Thus, $\tau(\hat{\sigma})-\tau(\tilde{\sigma}) = o_p(1)$.

We write $\tau(\tilde{\sigma})-\tau(\bar{\sigma}) = T^{-1/2}D_T$, where

$$\begin{aligned}D_s &= \sum_{t=1}^{s}\varepsilon_t\left(sign(x_{t-1})\left(\frac{\sigma_T(t/T)}{\tilde{\sigma}(t/T)}-\frac{\sigma_T(t/T)}{\bar{\sigma}(t/T)}\right)1\{\sigma(r)\neq\sigma(r-), r\in(t/T-h, t/T]\}\right) \\ &\equiv \sum_{t=1}^{s}\varepsilon_t z_t.\end{aligned}$$

We note that $(D_t, \mathcal{F}_t)$ is a square integrable martingale due, in particular, to Assumption 2.1 and our constructions of $\tilde{\sigma}$ and $\bar{\sigma}$. Moreover, the predictive quadratic variation $\langle D\rangle$ and the total quadratic variation $[D]$ of D satisfy

$$\langle D\rangle_T = \sum_{t=1}^{T} z_t^2 = O_p(hT), \quad [D]_T = \sum_{t=1}^{T} z_t^2\varepsilon_t^2 = O_p(hTT^{2q})$$

by Lemma S.7 and Assumption 3.3. It then follows from the two-sided martingale exponential inequality as in the proof of Lemma S.4 that $D_T = O_p(T^q\sqrt{hT})$, and hence, $\tau(\tilde{\sigma}) - \tau(\bar{\sigma}) = O_p(T^q\sqrt{h}) = o_p(1)$ by Assumption 3.4 (b).

Now we let $\beta \neq 0$. We have

$$\begin{aligned}
|\tau(\hat{\sigma})| &= \left| \frac{\beta}{\sqrt{T}} \sum_{t=1}^{T-1} \frac{|x_t|}{\hat{\sigma}((t-1)/T)} + \frac{\beta}{\sqrt{T}} \sum_{t=1}^{T-1} sign(x_{t-1}) \frac{\sigma_T(t)}{\hat{\sigma}((t-1)/T)} \right| \\
&\geq \left| \frac{\beta}{\sqrt{T}} \sum_{t=1}^{T-1} \frac{|x_t|}{\hat{\sigma}((t-1)/T)} \right| - \left| \frac{\beta}{\sqrt{T}} \sum_{t=1}^{T-1} sign(x_{t-1}) \frac{\sigma_T(t)}{\hat{\sigma}((t-1)/T)} \right| \\
&= \frac{|\beta|}{\sqrt{T}} \sum_{t=1}^{T-1} \frac{|x_t|}{\hat{\sigma}((t-1)/T)} + O_p(1) \\
&\geq \frac{1}{\sqrt{T}} \frac{|\beta|}{\underline{v}} \sum_{t=1}^{T-1} |x_t| + O_p(1),
\end{aligned}$$

where the second line follows from the reverse triangle inequality, the third line holds due to Theorem 3.3 (a), and the last line can be deduced from the proof of Lemma S.7. The stated result then follows immediately from the condition $\sum_{t=1}^{T-1} |x_t|/\sqrt{T} \to \infty$ in the theorem. □

Proof of Corollary 4.1. We note that Lemma 3.1 holds under Assumption 4.1 as long as Assumption 3.2 holds. Also, we can show the stated results in Proposition 3.2 (b)-(d) under Assumption 4.1 similar to the proofs of Proposition 3.2 (b)-(d).

To show that Proposition 3.2 (a) holds under Assumption 4.1, we write for any fixed $c \geq 0$

$$\hat{\sigma}_1^2(r - c/T) - \sigma_T^2(r) = \frac{\sum_{t=1}^T (\sigma_T(t/T)w_t^2 - \sigma_T(r))K_h(r - c/T - t/T)}{\sum_{t=1}^T K_h(r - c/T - t/T)} = A_T(r) + B_T(r),$$

where

$$\begin{aligned}
A_T(r) &= \frac{\sum_{t=1}^T (\sigma_T(t/T) - \sigma_T(r))w_t^2 K_h(r - c/T - t/T)}{\sum_{t=1}^T K_h(r - c/T - t/T)}, \\
B_T(r) &= \sigma_T(r) \frac{\sum_{t=1}^T (w_t^2 - 1)K_h(r - c/T - t/T)}{\sum_{t=1}^T K_h(r - c/T - t/T)}.
\end{aligned}$$

Similar to the proof of Lemma S.3, we can show that $\sup_{r \in \mathcal{C}_h} |A_T(r)| = o_p(1)$. Moreover, by applying an exponential inequality for a strong mixing process (see, e.g., Vogt (2012), Theorem 4.1 with $d = 0$), we may show under Assumption 4.1 that

$$\sup_{h \leq r \leq 1} |B_T(r)| = O_p\left((\log T/(hT))^{1/2}\right) = o_p(1).$$

This shows that Proposition **??** (a) holds under Assumption **??**.

As for the validity of Theorem 3.3 under Assumption 4.1, we consider $\bar{\sigma}$ as defined in the proof of Theorem 3.3. It is easy to see that $\bar{\sigma}(r)$ is a $\mathcal{F}_{Tr}$-adapted càdlàg process such that $\bar{\sigma} \to_d \sigma$. Therefore, the validity of Theorem 3.3 under Assumption 4.1 can be shown similarly to the proof of Theorem 3.3. □

Appendix D: Additional Figures

In this section, we present the finite sample power properties for Models CNST (continuous time) and Models CNST and ARCH (discrete time). The simulation settings are the same as those in Section 5.2 of the main paper.

Figure S.1 presents the results on finite sample power properties of the tests for the constant volatility case in continuous time. One can see that the power curves of the Cauchy RT test and our test are very close to each other. The other tests have higher size-adjusted power for the local-to-unit root regressor cases and comparable size-adjusted power for purely non-stationary regressors.

Figures S.2-S.4 present the numerical results on power properties under discrete time settings for all the tests considered except Cauchy RT which is inapplicable in discrete time settings. Results in the figures are provided for the cases of constant volatility (Figure S.2); the ARCH cases with $\alpha = 0.5773$ (Figure S.3) and $\alpha = 0.7325$ (Figure S.4).

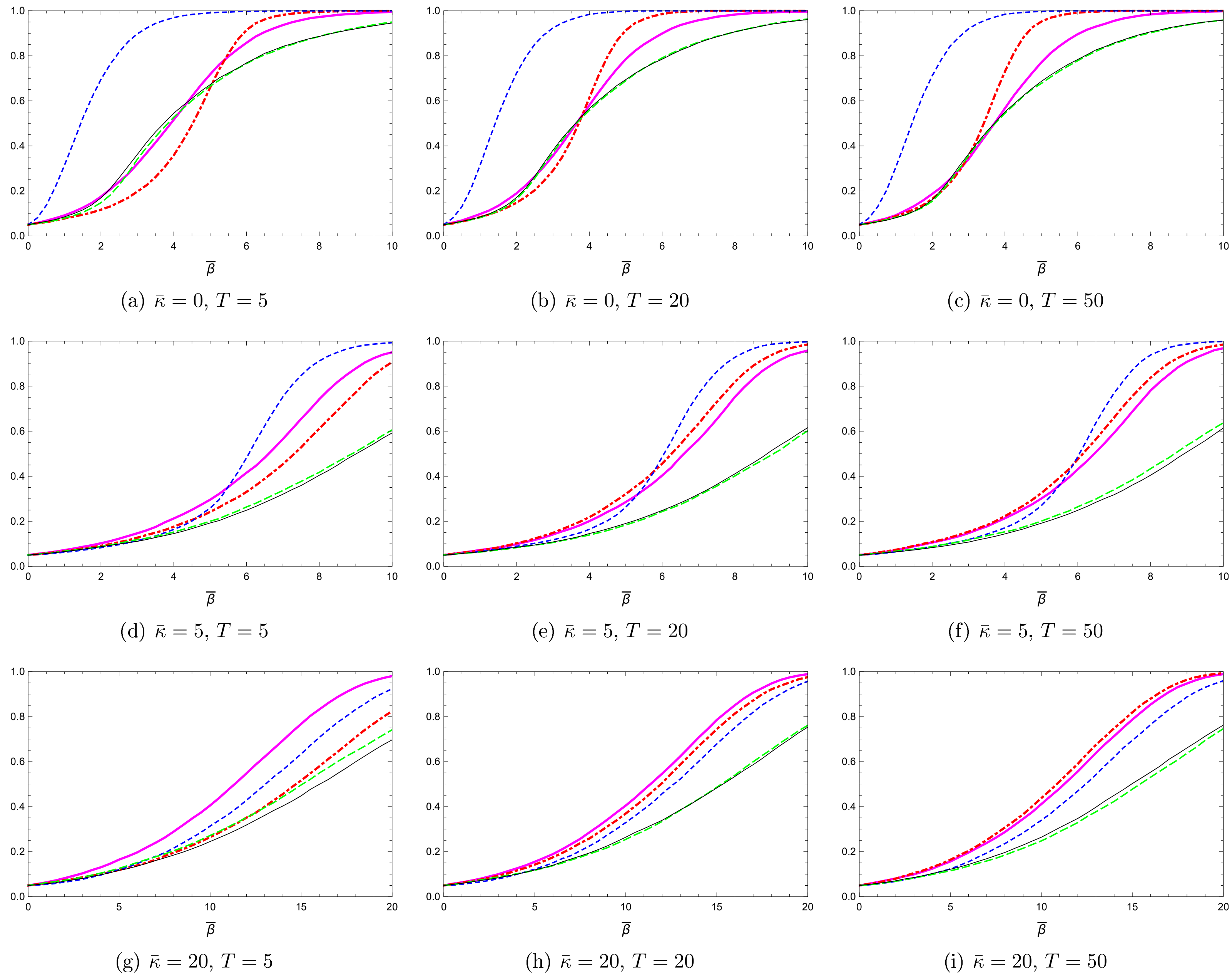


Figure S.1: Power for CNST (continuous time)
OLS: , Bonf. Q: RLRT: , Cauchy RT: , NP:

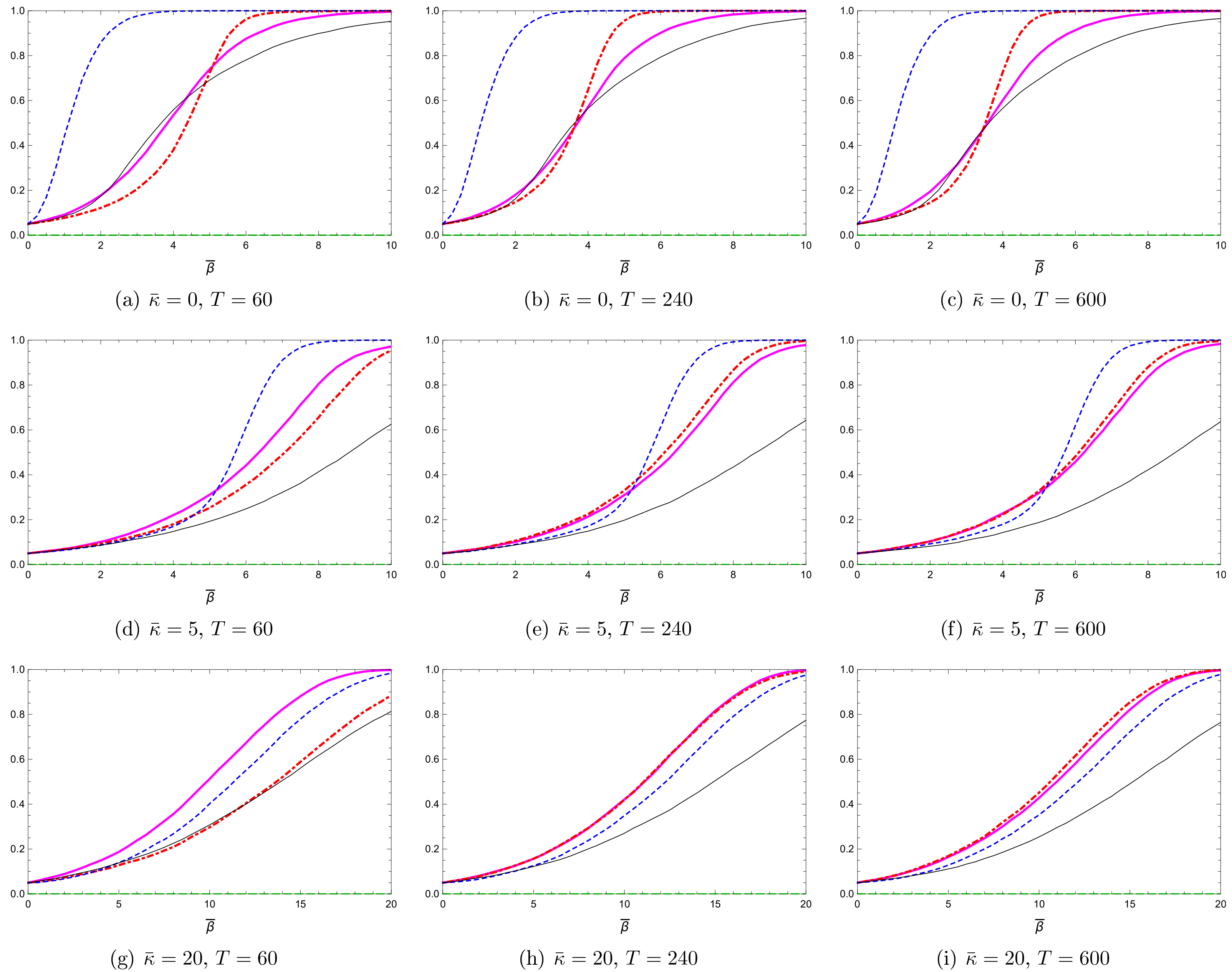
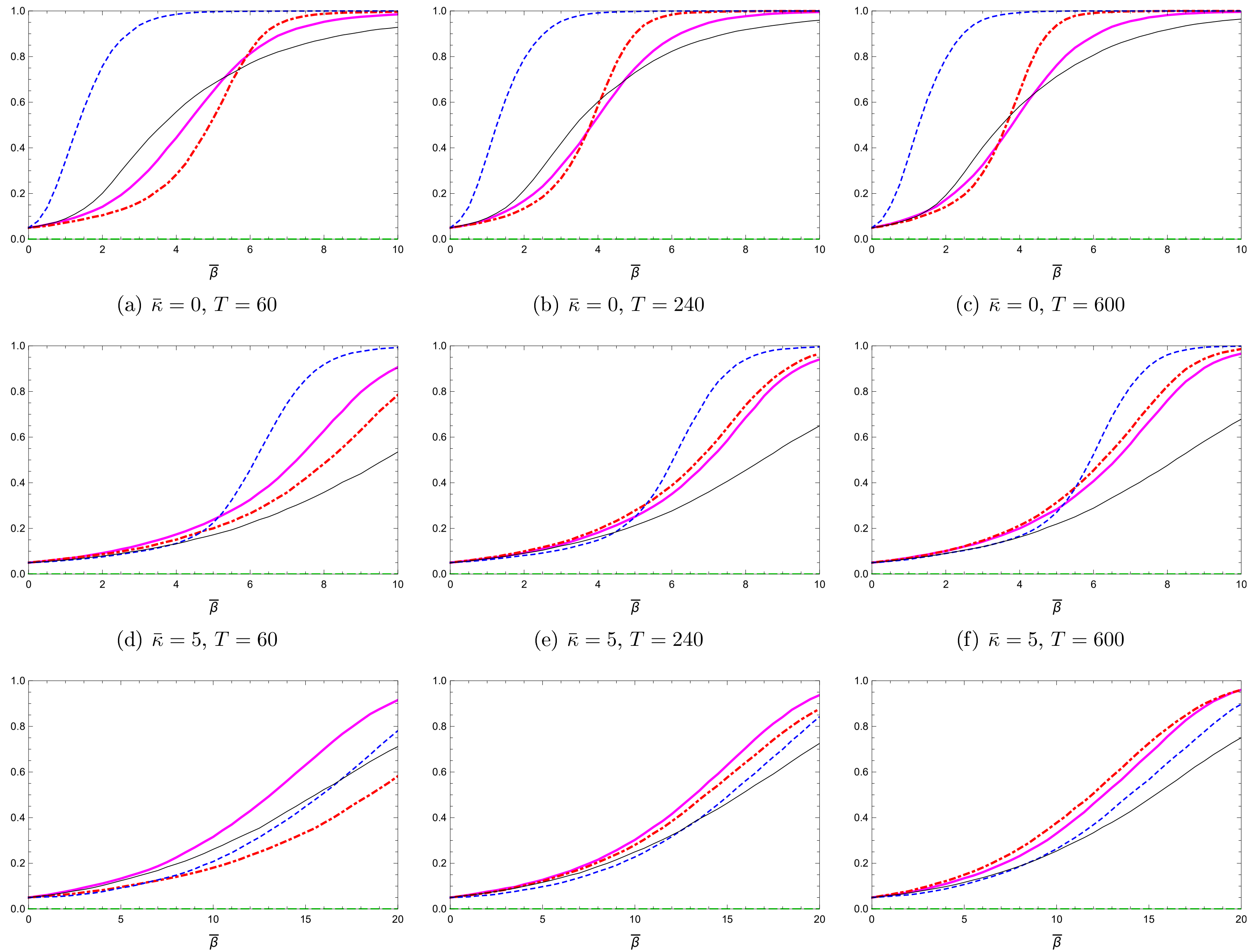


Figure S.2: Power for CNST (discrete time)
OLS: , Bonf. Q: RLRT: , NP:

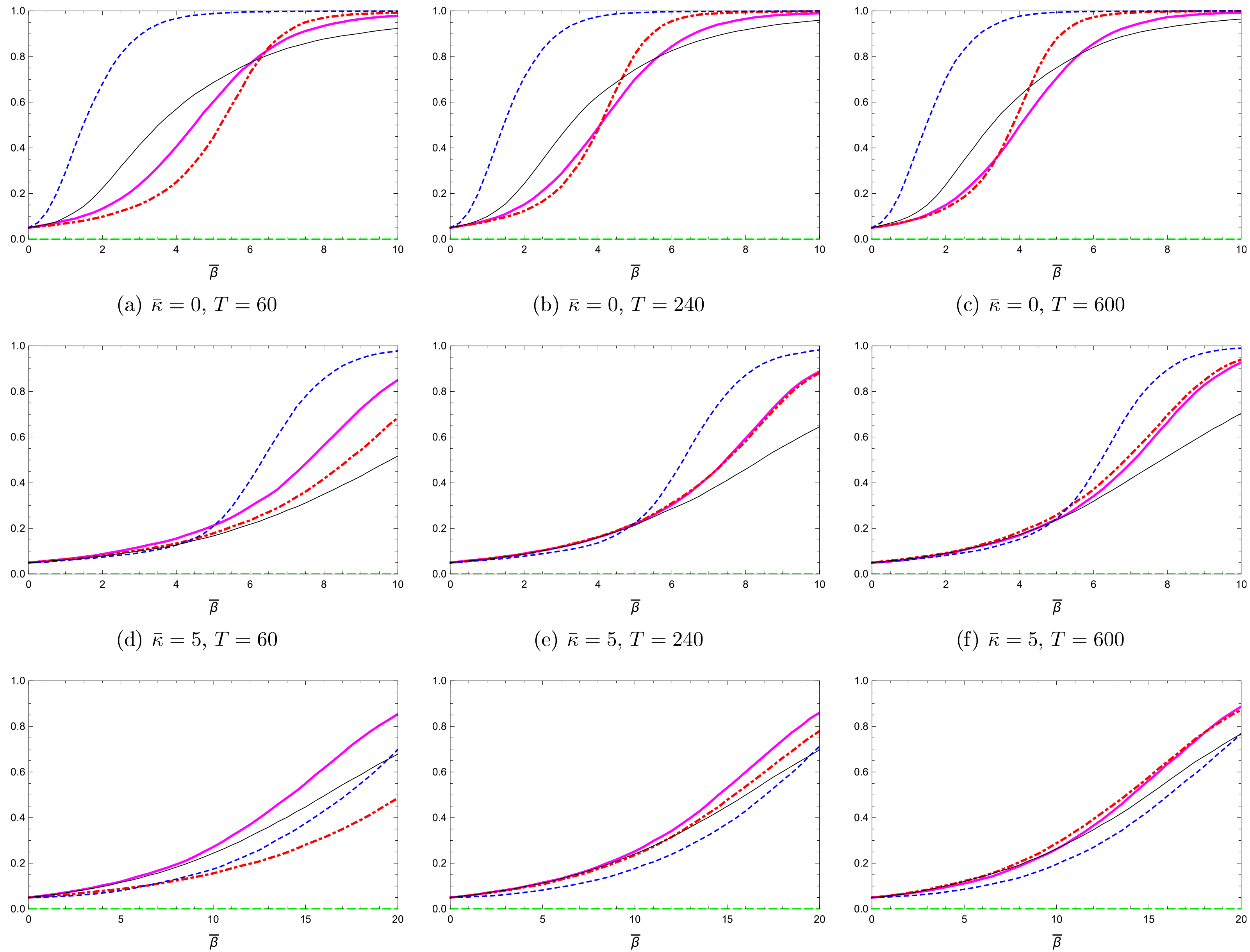


(a) $\bar{\kappa} = 0$, $T = 60$ (b) $\bar{\kappa} = 0$, $T = 240$ (c) $\bar{\kappa} = 0$, $T = 600$

(d) $\bar{\kappa} = 5$, $T = 60$ (e) $\bar{\kappa} = 5$, $T = 240$ (f) $\bar{\kappa} = 5$, $T = 600$

(g) $\bar{\kappa} = 20$, $T = 60$ (h) $\bar{\kappa} = 20$, $T = 240$ (i) $\bar{\kappa} = 20$, $T = 600$

Figure S.3: Power for ARCH with $\alpha = 0.5773$ (discrete time)
OLS:, Bonf. Q: RLRT:, NP:



(a) $\bar{\kappa} = 0$, $T = 60$ (b) $\bar{\kappa} = 0$, $T = 240$ (c) $\bar{\kappa} = 0$, $T = 600$

(d) $\bar{\kappa} = 5$, $T = 60$ (e) $\bar{\kappa} = 5$, $T = 240$ (f) $\bar{\kappa} = 5$, $T = 600$

(g) $\bar{\kappa} = 20$, $T = 60$ (h) $\bar{\kappa} = 20$, $T = 240$ (i) $\bar{\kappa} = 20$, $T = 600$

Figure S.4: Power for ARCH with $\alpha = 0.7325$ (discrete time)
OLS:, Bonf. Q: RLRT:, NP:

References

Applebaum, D. (2009), 'Lévy Processes and Stochastic Calculus'. Cambridge University Press.

Bercu, B., and Touati, A. (2008), 'Exponential inequalities for self-normalized martingales with applications', *Annals of Applied Probability* **18**, 1848–1869.

Hansen, B. E. (1992), 'Convergence to stochastic integrals for dependent heterogeneous processes', *Econometric Theory* **8**, 489–500.

Kurtz, T. G. and Protter, P. (1991), 'Weak limit theorems for stochastic integrals and stochastic differential equations', *Annals of Probability* **19**, 1035–1070.

Park, J. Y. (2003), 'Weak unit roots', Unpublished Manuscript. URL: http://www.ruf.rice.edu/econ/papers/2003papers/17park.pdf

Park, J. Y. and Phillips, P. C. B. (2001), 'Nonlinear regressions with integrated time series', *Econometrica* **69**, 117–161.

Vogt, M. (2012), 'Nonparametric regression for locally stationary time series', *Annals of Statistics* **40**, 2601–2633.