Simulation Study

Simulated Models

Consider the nonparametric model

$$X_{t,T} = m\Big(\frac{t}{T}, X_{t-1,T}\Big) + \varepsilon_t \tag{1}$$

and the additive framework

$$X_{t,T} = m\Big(\frac{t}{T}, X_{t-1,T}\Big) + g\Big(\frac{t}{T}, X_{t-2,T}\Big) + \varepsilon_t \tag{2}$$

with

$$\begin{aligned} m(u,x) &= 1.5(1+u)\cos(0.25\pi x) \\ g(u,x) &= 0.5ux \end{aligned}$$

and residuals ε_t that are i.i.d. standard normally distributed. With the help of a Monte Carlo experiment, I now tackle the question whether the asymptotic normality result

$$Q_T(u,x) := \sqrt{Th_u h_x}\big(\hat{m}(u,x) - m(u,x)\big) \xrightarrow{d} N(B_{u,x}, V_{u,x}) \tag{3}$$

gives a good approximation to the distribution of $Q_T(u,x)$ in small samples. Here, $\hat{m}(u,x)$ is either the NW estimate of $m(u,x)$ in (1) or the smooth backfitting estimate of $m(u,x)$ in (2), depending on which model we are talking about. Moreover, $B_{u,x}$ and $V_{u,x}$ are the corresponding asymptotic bias and variance expressions. Finally, h_u denotes the bandwidth in time direction and h_x the bandwidth in the direction of the regressor $X_{t-1,T}$.

Remark 1. Model (1) belongs to the tvNAR framework introduced in Section 3. Note however that the regression function m does not fulfill condition (M3). In particular, there is no real constant K_1 with $|\partial m(u,x)| \le \delta < 1$ for all $u \in [0,1]$ and $|x| > K_1$, where ∂m is the derivative of m with respect to x. To fix this problem, m may be replaced by the function

$$m_\tau(u,x) = \begin{cases} 1.5(1+u)\cos(0.25\pi x) & \text{for } u \in [0,1] \text{ and } |x| \le \tau := 8 \cdot 10^n \\ 1.5(1+u) & \text{for } u \in [0,1] \text{ and } |x| > \tau := 8 \cdot 10^n, \end{cases}$$

which fulfills (M1)–(M3). If we take n very large (say 100), then the simulated process (1) will virtually never wander into the region where m_τ differs from m (at least not for any practically relevant sample size). Thus, the difference between m and m_τ can be completely ignored in the simulation. This shows that from a practical point of view, (M3) is not a very harsh assumption and can usually be assumed to hold. (A similar point can be made with respect to the boundedness assumption (M1).)

Remark 2. Model (2) also belongs to the tvNAR framework. Similarly as above, m and g may be replaced by a pair of functions that fulfill the main conditions of the theory.

Remark 3. The constants which show up in the function m (and in g) influence the signal-to-noise ratio in the models. For example, if the constant 1.5 in the function m of model (1) is replaced by a smaller/larger constant, then the variation of the signal $m(\frac{t}{T}, X_{t-1,T})$ gets smaller/larger compared to the variation of the error term, and thus the signal-to-noise ratio decreases/increases. I have tried to choose the constants in a way that yields a reasonable signal-to-noise ratio. In particular, the standard deviation of the signal (which is $m(\frac{t}{T}, X_{t-1,T})$ in model (1) and $m(\frac{t}{T}, X_{t-1,T}) + g(\frac{t}{T}, X_{t-2,T})$ in model (2)) amounts to roughly 1.5 locally around the time point $u = 0.5$ where estimation is performed in what follows. This compares to a standard deviation of the error term which equals 1.

Simulation Setup

For each of the two models, I simulate $N = 1000$ samples of size T, considering four different lengths $T = 500, 1000, 1500, 2000$. To exclude boundary effects, I examine the normality result (3) at a point (u, x) which lies more or less in the center of the data support both in time direction and in the direction of $X_{t-1,T}$. In particular, I pick $u = 0.5$ and $x = 0$. Given this, I proceed as follows:

(i) For each sample, I calculate $\hat{m}(u, x)$ and use this estimate to compute the term of interest $Q_T(u, x)$. This provides me with N sample values of $Q_T(u, x)$. I then calculate the empirical distribution from these N values. This yields a simulated version of the distribution of $Q_T(u, x)$.

(ii) To compare the (simulated) distribution of $Q_T(u, x)$ with its normal limit, I compute the bias $B_{u,x}$ and the variance $V_{u,x}$ of the latter.
To calculate $V_{u,x}$, I need to estimate the density $f(u, x)$ of the approximating variables $X_t(u)$. All other terms occurring in $V_{u,x}$ are known. To get a precise estimate of $f(u, x)$, I simulate a long sample of the process $\{X_t(u)\}$ (of size 50000) and calculate the kernel density from it.
To calculate the asymptotic bias of the NW estimate, I neglect the terms that depend on the first derivatives of $f(u, x)$ (see the formula in Theorem 4.3). As the remaining terms are all known, the bias can then be computed without performing any estimation. The same term is used as a proxy for the asymptotic bias in the smooth backfitting case. In my simulation setup, this should be a fair choice for the following reason: The main ingredient of the asymptotic bias of the backfitting estimates is an additive projection of the function β specified in Lemma C4. Neglecting the components of β that depend on the first derivatives

of the density $p(u, x)$, this projection is identical to the bias approximation used in the NW case.

There are different reasons why the normal limit in (3) may be a poor approximation of the (simulated) distribution of $Q_T(u, x)$:

(a) The distribution of $Q_T(u, x)$ is highly non-normal.

(b) The distribution of $Q_T(u, x)$ is approximately normal but $\mathbb{E}[Q_T(u, x)]$ strongly differs from $B_{u,x}$.

(c) The distribution of $Q_T(u, x)$ is approximately normal but $\mathrm{Var}(Q_T(u, x))$ strongly differs from $V_{u,x}$.

To discern between these different reasons, I examine the following three issues:

(a) I check whether the standardized version of $Q_T(u, x)$ is approximately standard normal, i.e. I check whether the result

$$\frac{Q_T(u, x) - \mathbb{E}[Q_T(u, x)]}{\sqrt{\mathrm{Var}(Q_T(u, x))}} \xrightarrow{d} N(0, 1)$$

is approximately true in small samples.

(b) I check whether $B_{u,x}$ is a reasonable approximation of $\mathbb{E}[Q_T(u, x)]$.

(c) I check whether $V_{u,x}$ is a reasonable approximation of $\mathrm{Var}(Q_T(u, x))$.

Results for Model (1)

The results were produced using an Epanechnikov kernel and bandwidth vectors of the form

$$(h_u, h_x) = (h_u(\lambda), h_x(\lambda))$$

with

$$h_u(\lambda) = \lambda T^{-1/6} \qquad \text{and} \qquad h_x(\lambda) = s_X \lambda T^{-1/6}.$$

Here, λ takes the values 0.3, 0.5, 0.7 and 0.9. Moreover, s_X measures how large the support of the covariate observations is. In particular, I let $s_X = (q_{99} - q_1)$, where q_α is the average of the α%-quantiles of the simulated samples. (For the sample size $T = 1000$, for instance, these definitions yield the bandwidths $h_u \approx 0.09, 0.15, 0.22, 0.28$ and $h_x = s_X h_u$ with $s_X \approx 8$.) Note that the bandwidths shrink by a factor of $T^{-1/6}$ as the sample size increases. This reflects the fact that they are theoretically of the order $O(T^{-1/6})$.

I first report the results on issue (a). Figure 1 compares the (simulated) quantiles of the standardized version of $Q_T(u, x)$ with the theoretical quantiles of a standard normal law.

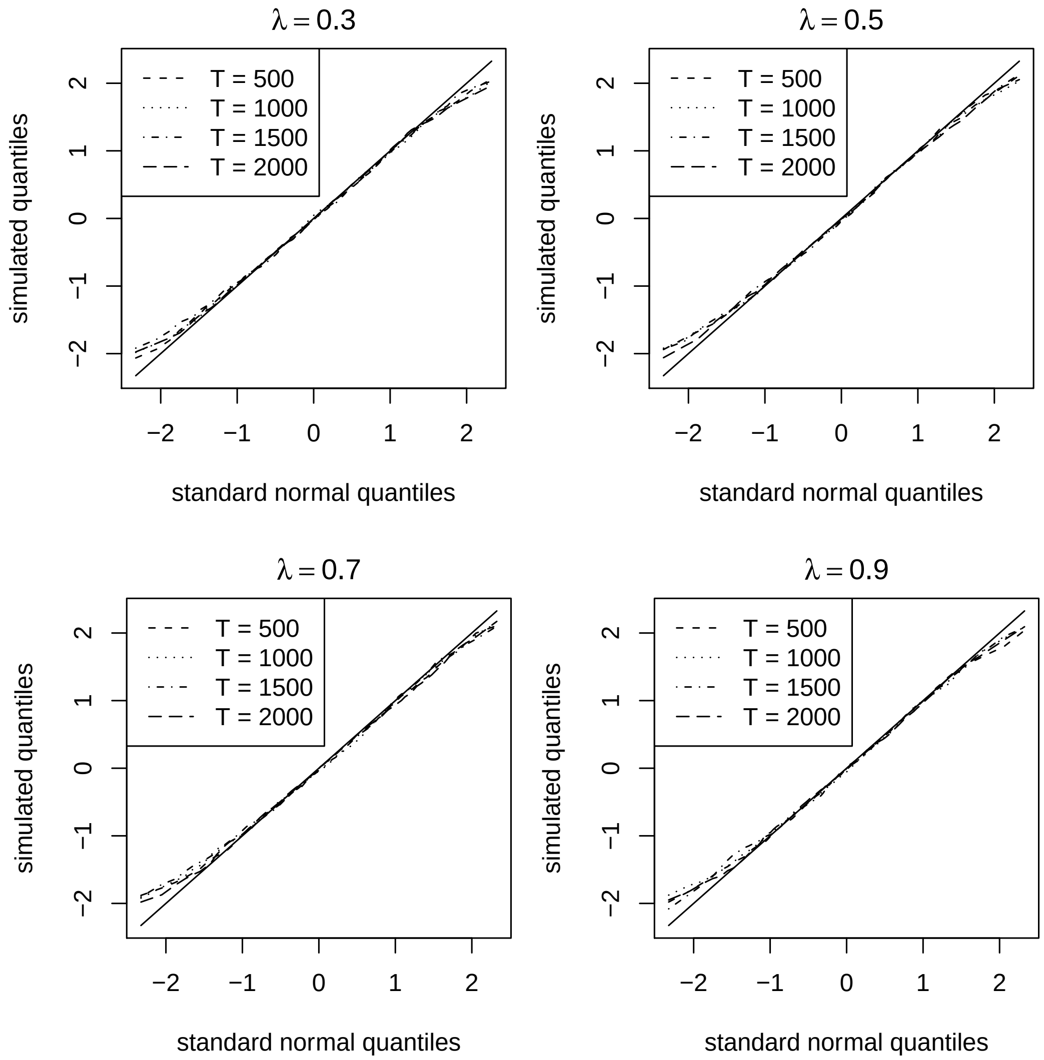

Figure 1: Quantile-quantile plots for different bandwidths and sample sizes. The plots compare the (simulated) quantiles of the standardized version of $Q_T(u, x)$ (y-axis) with the theoretical quantiles of a standard normal distribution (x-axis).

As can be seen, the quantiles of a standard normal distribution match the quantiles of the standardized version of $Q_T(u, x)$ pretty well even for small sample sizes. The only noticeable deviations occur for very extreme quantiles. Thus, a standard normal distribution appears to be a reasonable approximation.

I next turn to the issues (b) and (c), i.e. I compare the mean $B_T = \mathbb{E}[Q_T(u, x)]$ and the variance $V_T = \text{Var}(Q_T(u, x))$ with their asymptotic counterparts $B_\infty = B_{u,x}$ and $V_\infty = V_{u,x}$. Note that the asymptotic bias B_∞ in Table 1 varies for different bandwidths and sample sizes. The reason is that it depends on the constant c_h which is defined as the limit of Th^6 in Theorem 4.3. In the simulations, c_h is approximated by the value of Th^6.

$\lambda = 0.3$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-0.513	-0.587	0.075	1.994	2.026	-0.032
$T = 1000$	-0.605	-0.607	0.002	1.928	2.026	-0.098
$T = 1500$	-0.671	-0.610	-0.061	1.969	2.026	-0.057
$T = 2000$	-0.611	-0.613	0.003	1.941	2.026	-0.085

$\lambda = 0.5$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-2.626	-2.719	0.093	1.979	2.026	-0.047
$T = 1000$	-2.745	-2.809	0.064	1.907	2.026	-0.119
$T = 1500$	-2.867	-2.825	-0.042	1.992	2.026	-0.034
$T = 2000$	-2.766	-2.838	0.073	2.034	2.026	0.009

$\lambda = 0.7$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-6.841	-7.461	0.620	2.146	2.026	0.121
$T = 1000$	-7.233	-7.708	0.476	2.066	2.026	0.041
$T = 1500$	-7.516	-7.752	0.236	2.049	2.026	0.024
$T = 2000$	-7.456	-7.789	0.333	2.047	2.026	0.022

$\lambda = 0.9$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-13.36	-15.86	2.493	2.100	2.026	0.075
$T = 1000$	-14.45	-16.38	1.938	1.997	2.026	-0.026
$T = 1500$	-14.98	-16.48	1.495	2.035	2.026	0.010
$T = 2000$	-15.10	-16.55	1.452	2.040	2.026	0.015

Table 1: Comparison of the mean B_T and the variance V_T with their asymptotic counterparts B_∞ and V_∞ for different bandwidths and sample sizes.

The tables show that the variance V_T is pretty well approximated by the limit expression V_∞ for most bandwidths. Only if the estimate $\hat{m}(u,x)$ is strongly oversmoothed (i.e. in the case with $\lambda = 0.9$), the approximation gets considerably worse. Moreover, the values of V_T can overall be seen to get closer to the limit point V_∞ as the sample size increases. A similar picture arises for the mean B_T. The latter is reasonably close to B_∞ as long as $\hat{m}(u,x)$ is not strongly oversmoothed. In addition, the values of B_T can again be seen to tend to the limit B_∞ as T gets larger.

Results for Model (2)

The results for model (2) are presented in exactly the same way as those for model (1). For the estimation, I again use an Epanechnikov kernel and the bandwidth vectors from the NW case. To keep the results easily presentable, the bandwidth in the direction of the second covariate $X_{t-2,T}$ is fixed throughout. In particular, I choose it as $0.9 s_X T^{-1/6}$ which gives a good fit of the function g.

I start with the discussion of issue (a). Figure 2 shows that similarly to the findings for model (1), the distribution of the normalized variables $Q_T(u,x)$ is reasonably well approximated by a standard normal law.

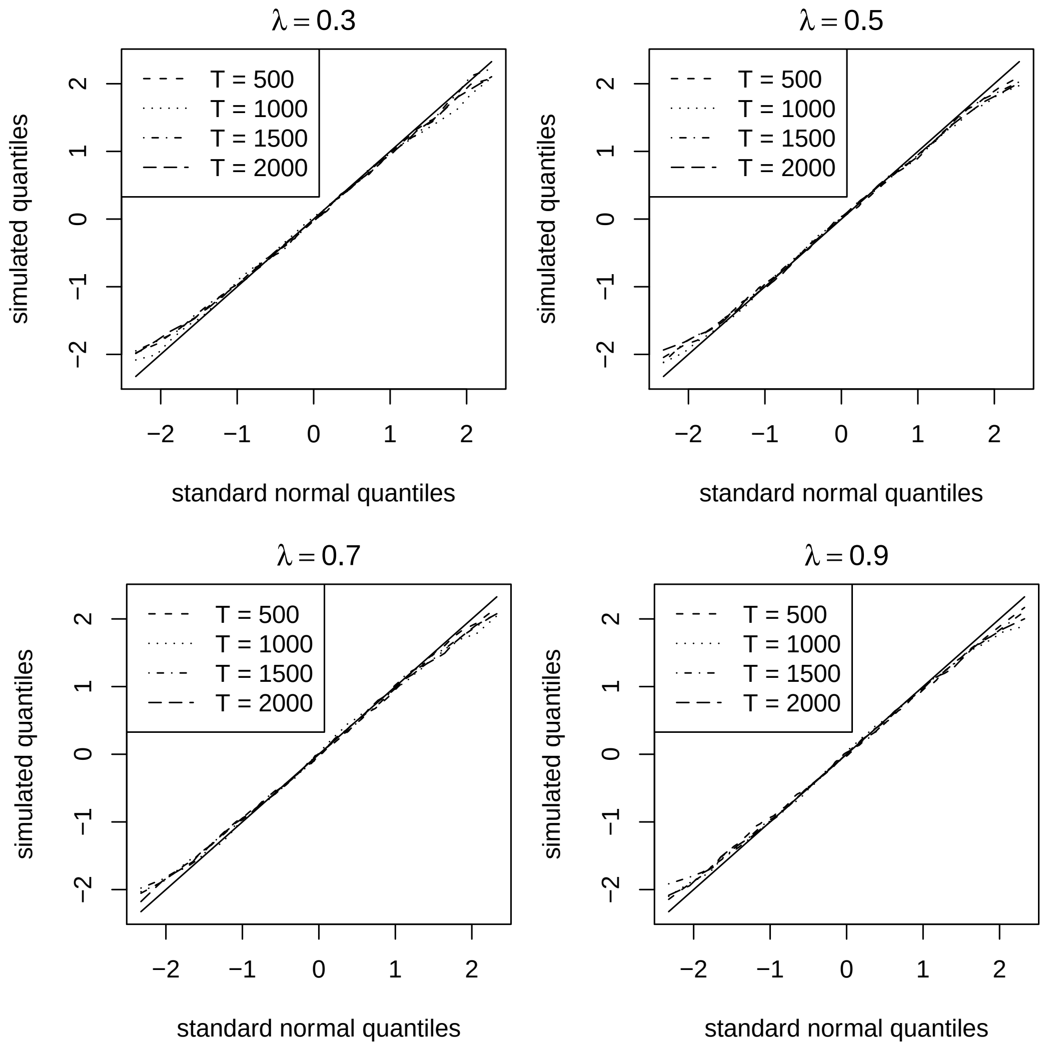

Figure 2: Quantile-quantile plots for different bandwidths and sample sizes. The values on the y-axis give the (simulated) quantiles of the standardized version of $Q_T(u,x)$.

I next compare the bias B_T and the variance V_T to their limits B_∞ and V_∞. The findings are again similar to those for model (1): On the whole, B_T and V_T are reasonably close to their limits as long as the bandwidths are not chosen too large.

$\lambda = 0.3$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-0.853	-0.758	-0.095	2.066	2.148	-0.082
$T = 1000$	-0.843	-0.793	-0.050	2.060	2.148	-0.088
$T = 1500$	-0.929	-0.805	-0.124	2.082	2.148	-0.066
$T = 2000$	-0.972	-0.806	-0.166	2.048	2.148	-0.010

$\lambda = 0.5$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-3.580	-3.561	-0.019	2.151	2.148	0.004
$T = 1000$	-3.797	-3.659	-0.138	1.983	2.148	-0.165
$T = 1500$	-4.017	-3.729	-0.289	2.226	2.148	0.079
$T = 2000$	-4.103	-3.741	-0.362	2.173	2.148	0.026

$\lambda = 0.7$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-9.079	-9.764	0.686	2.172	2.148	0.025
$T = 1000$	-9.768	-10.07	0.306	2.119	2.148	-0.029
$T = 1500$	-10.23	-10.24	0.010	2.283	2.148	0.136
$T = 2000$	-10.50	-10.28	-0.216	2.217	2.148	0.070

$\lambda = 0.9$

	B_T	B_∞	$B_T - B_\infty$	V_T	V_∞	$V_T - V_\infty$
$T = 500$	-17.53	-20.81	3.281	2.139	2.148	-0.009
$T = 1000$	-19.05	-21.46	2.417	2.161	2.148	0.014
$T = 1500$	-19.98	-21.82	1.842	2.258	2.148	0.111
$T = 2000$	-20.54	-21.84	1.292	2.133	2.148	-0.015

Table 2: Comparison of the mean B_T and the variance V_T with their asymptotic counterparts B_∞ and V_∞ for different bandwidths and sample sizes.

To sum up, the simulations show that both in the NW and the smooth backfitting case, the asymptotic normality result (3) gives a reasonable approximation in small samples (at least for the models considered). However, one can also see that the results are sensitive to the choice of bandwidths to a certain extent. This indicates that an appropriate choice is crucial when applying the normality result (3). Some ideas how to select the bandwidths in my model framework can be found in my replies to the referee reports and in Section 6 of the paper.