FUNCTIONAL COEFFICIENT MOVING AVERAGE MODEL WITH APPLICATIONS TO FORECASTING CHINESE CPI

Song Xi Chen, Lihua Lei and Yundong Tu*

Peking University

Supplementary Material

Appendix: Lemmas and Proofs

Lemma 1 (Fan and Yao, 2006)**.** Suppose that

1. $\{X_t, Y_t\}$ are strictly stationary and α-mixing with $\sum_{l\geq 1} l^{\lambda}[\alpha(l)]^{1-\frac{2}{\delta}} \leq \infty$ and $E\{|Y_t|^{\delta}|X_t = x\} < \infty$ for some $\delta > 2$ and $\lambda > 1 - 2/\delta$.

2. The conditional density $f_{X_0,X_l|Y_0,Y_l}(x_0, x_l|y_0, y_l) \leq A < \infty$ for some $A > 0$ and all $l > 0$.

3. The conditional distribution of Y_t given $X_t = u$, denoted by $G(y|u)$ is continuous at the point $u = x$.

4. As $T \to \infty$, $h \to 0$ and there exists a sequence of positive integers $s_T \to \infty$ and $s_T = o((Th)^{1/2})$ such that $(T/h)^{1/2}\alpha(s_T) \to 0$ as $T \to \infty$.

5. $K(\cdot)$ is a symetric and bounded kernel with a bounded support $[-1, 1]$ such that $\int K(u)du = 1$.

6. $\sigma^2(\cdot) = Var(Y_t|X_t = \cdot)$ and the density function $f(\cdot)$ of X_t are continuous at the point x.

*Corresponding author. Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, China, 100871

Let $\hat{m}(x)$ be the local linear estimator of the conditional mean $m(x) = E(Y_t|X_t = x)$, then

$$\sqrt{Th}(\hat{m}(x) - m(x) - \frac{1}{2}\int u^2K(u)du\ m''(x)h^2) \xrightarrow{d} N(0, \frac{\sigma^2(x)}{f(x)}\int K^2(u)du)$$

Lemma 2. Suppose $x_t \sim FMA(1)$. For $j = 0, 1$, Let

$$(\hat{a}_j^*(z), \hat{b}_j^*(z)) = argmin_{(a,b)} \sum_{t=1}^{T}\{(x_t - \mu)(x_{t-j} - \mu) - a - b(z_t - z)\}^2 K(\frac{z_t - z}{h}),$$

then under the assumptions (A1)$\sim$(A6), it holds that

$$\sqrt{Th}\left(\begin{array}{c} \hat{a}_1^*(z) - (1+\theta^2(z))\sigma^2 - \frac{1}{2}\sigma_K^2\theta''(z)\sigma^2h^2 \\ \hat{a}_0^*(z) - \theta(z)\sigma^2 - \sigma_K^2(\theta(z)\theta''(z) + \theta'^2(z))\sigma^2h^2 \end{array}\right) \xrightarrow{d} N(0, \frac{\Gamma(z)}{p(z)}\sigma^4R(K)).$$

Proof. For any $v = (v_0, v_1)^T \in \mathbb{R}^2$, let $y_t(v) = v_0(x_t - \mu)^2 + v_1(x_t - \mu)(x_{t-1} - \mu)$. Denote $\hat{a}^*(z; v)$ by the local linear estimator of $E(y_t(v)|z_t = z) = v_0(1+\theta^2(z))\sigma^2 + v_1\theta(z)\sigma^2$, i.e.

$$(\hat{a}^*(z; v), \hat{b}^*(z; v)) = argmin_{a(z),b(z)} \sum_{t=2}^{T}(y_t(v) - a - b(z_t - z))^2K(\frac{z_t - z}{h})$$

Then it is easy to show that $\hat{a}^*(z; v) = v_0\hat{a}_0^*(z) + v_1\hat{a}_1^*(z)$. If we proved that

$$\begin{aligned} \sqrt{Th}(\hat{a}^*(z; v) - E(y_t(v)|z_t = z) - \frac{1}{2}\sigma^2h^2\sigma_K^2\ v^T\left(\begin{array}{c} \theta''(z) \\ 2(\theta(z)\theta''(z) + \theta'^2(z)) \end{array}\right)) \\ \xrightarrow{d} N(0, \frac{v^T\Gamma(z)v}{p(z)}\sigma^4R(K)). \end{aligned} \tag{1}$$

Then Lemma 2 will be proved by Cramér Device. Now we prove (1).

First, by Assumptions (A2) and (A3), $\{y_t(v), z_t\}$ is strictly stationary and α-mixing such that

$$E(|y_t(v)|^\delta|z_t = z) < C||v||^2E(|\epsilon_t|^{2\delta} + |\epsilon_{t-1}|^{2\delta} + |\epsilon_{t-2}|^{2\delta}|z_t = z) < \infty$$

and $\alpha(m) \le Am^{-\beta}$. Let $\lambda = \frac{\beta}{2} - \frac{1}{\delta}$, then $\lambda > 1$ since $\beta > (2\delta - 2)/(\delta - 2)$ and

$$\sum_{l \ge 1} l^{\lambda}(\alpha(l))^{1-\frac{2}{\delta}} \le A^{1-\frac{2}{\delta}} \sum_{l \ge 1} l^{-(1-\frac{2}{\delta})(\frac{\beta}{2}+\frac{1}{\delta-2})} < \infty.$$

Thus, the condition 1 of Lemma 1 is satisfied.

By Assumption (A1), it holds that $h = O(T^{-(1-\epsilon_0)})$. Let $s_T = [(Th)^{1/2}/\log T]$, then $s_T = o((Th)^{1/2})$ and

$$(T/h)^{1/2}\alpha(s_T) = O(T^{1-\frac{1+\beta}{2}\epsilon_0}(\log T)^{-\beta}) = o(1).$$

Thus, the condition 4 of Lemma 1 is satisfied.

Further, it follows Assumptions (A4), (A5) and (A6) that the conditions 2,3,5,6 of Lemma 1 hold. Therefore, (1) is proved by Lemma 1 and hence the lemma is proved by Cramér Device. $\square$

Lemma 3. Suppose that Assumptions (A1)∼(A6) holds. Then

$$|\hat{a}_j(z) - \hat{a}_j^*(z)| = O_p(\frac{1}{\sqrt{T}}) \tag{2}$$

Proof. First, we show that $\bar{x} = O_p(T^{-1/2})$.

$$TVar(\bar{x}) = \sum_{|j|<T}(1 - \frac{|j|}{T})\gamma(j) \le \sum_{-\infty}^{\infty}(1 - \frac{|j|}{T})\gamma(j) < \infty.$$

Thus $\lim_{T\to\infty} TVar(\bar{x}) = \sum_{-\infty}^{\infty}\gamma(h)$ and then $\bar{x} = O_p(T^{-1/2})$. Let $w_t(z) = K(\frac{z_t - z}{h})(s_{n,2} - (z_t - z)s_{n,1})$, where $s_{n,j} = \sum_{t=1}^{T} K(\frac{z_t - z}{h})(z_t - z)^j$, then

$$\hat{a}_j(z) = \frac{\sum_{t=j+1}^{T} w_t(z)(x_t - \bar{x})(x_{t-j} - \bar{x})}{\sum_{t=j+1}^{T} w_t(z)}, \quad \hat{a}_j^*(z) = \frac{\sum_{t=j+1}^{T} w_t(z)(x_t - \mu)(x_{t-j} - \mu)}{\sum_{t=j+1}^{T} w_t(z)}.$$

Notice that

$$|\hat{a}_j(z) - \hat{a}_j^*(z)| \le |\mu^2 - \bar{x}^2| + |\mu - \bar{x}|\left|\frac{\sum_{t=j+1}^{T} w_t(z)(x_t + x_{t-j})}{\sum_{t=j+1}^{T} w_t(z)}\right|.$$

On the one hand,

$$\mu^2 - \bar{x}^2 = (\mu - \bar{x})(\mu + \bar{x}) = O_p(\frac{1}{\sqrt{T}}).$$

On the other hand, $\frac{\sum_{t=j+1}^{T} w_t(z)(x_t + x_{t-j})}{\sum_{t=j+1}^{T} w_t(z)}$ is the local linear estimator of $E(x_t + x_{t-j}|z_t = z)$. Let Then by Lemma 1, it is easy to prove that

$$\left|\frac{\sum_{t=j+1}^{T} w_t(z)(x_t + x_{t-j})}{\sum_{t=j+1}^{T} w_t(z)}\right| = O_p(1)$$

and hence

$$|\hat{a}_j(z) - \hat{a}_j^*(z)| = O_p(\frac{1}{\sqrt{T}}).$$

□

Proof of Theorem 1.
Without loss of generality, we assume $\mu = 0$. Let $M_T = T^{-1}\sum_{t=1}^{T} K_h(z_t - z)$,

$$\hat{g}\{\theta(z)\} - g\{\theta(z)\} = \frac{\hat{a}_1(z)}{\hat{a}_0(z)} - \frac{\theta(z)}{1+\theta^2(z)} = \frac{\hat{a}_1^*(z) + O_p(T^{-\frac{1}{2}})}{\hat{a}_0^*(z) + O_p(T^{-\frac{1}{2}})} - \frac{\theta(z)}{1+\theta^2(z)}$$
$$= \frac{\theta(z)\sigma^2 + \frac{1}{2}\sigma_K^2\theta''(z)\sigma^2 h^2 + (Th)^{-\frac{1}{2}}A_1 + O_p(T^{-\frac{1}{2}})}{(1+\theta^2(z))\sigma^2 + \sigma_K^2(\theta(z)\theta''(z) + \theta'^2(z))\sigma^2 h^2 + (Th)^{-\frac{1}{2}}A_0 + O_p(T^{-\frac{1}{2}})} - \frac{\theta(z)}{1+\theta^2(z)}$$
$$= G(z)h^2 + (Th)^{-\frac{1}{2}}\frac{(1+\theta^2(z))A_1 - \theta(z)A_0 + O_p(1)}{(1+\theta^2(z))^2 + o_p(1)}$$

where the second equality follows from Lemma 3, the third quality follows from Lemma 2 and

$$\begin{pmatrix} A_0 \\ A_1 \end{pmatrix} \sim N(0, \frac{\Gamma(z)}{p(z)}\sigma^4 R(K))$$

Then it follows from Slusky Theorem that

$$\sqrt{Th}(\hat{g}\{\theta(z)\} - g\{\theta(z)\} - G(z)h^2) \xrightarrow{d} N(0, \nu(z)).$$

□

Proof of Theorem 2. For (i), by Theorem 1, it suffices to prove

$$\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - \hat{g}\{\theta(z)\}) \xrightarrow{d} 0.$$

For arbitrary $\epsilon > 0$,

$$\begin{aligned}
&P(\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - \hat{g}\{\theta(z)\}) > \epsilon) \le P(\tilde{g}\{\theta(z)\} \neq \hat{g}\{\theta(z)\}) \\
=&P(|\hat{g}\{\theta(z)\}| > \frac{1}{2}) \le P(|\hat{g}\{\theta(z)\} - g\{\theta(z)\}| > \frac{1}{2} - |g\{\theta(z)\}|) \to 0.
\end{aligned}$$

Thus, (i) is proved. Now turn to (ii). Notice that $G = 0$ when $g\{\theta(z)\} = \frac{1}{2}$, thus by Theorem 1, we know that

$$\sqrt{Th/\nu(z)}(\hat{g}\{\theta(z)\} - \frac{1}{2}) \xrightarrow{d} Z$$

where $Z \sim N(0,1)$. Let $f(x) = \min\{x, 0\}$, then

$$\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - \frac{1}{2}) = f[\sqrt{Th/\nu}(\hat{g}\{\theta(z)\} - \frac{1}{2})].$$

Since f is continuous, by continuous mapping theorem, we have

$$\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - \frac{1}{2}) \xrightarrow{d} f(Z),$$

where $f(Z) \sim \Phi^-$. Therefore, (ii) is proved and similarly (iii) is proved. □

Proof of Theorem 3. (i) is directly followed from Lemma 2 and Delta Method. Now we prove (ii) while (iii) can be dealt with in similar way. It follows Remark 1 that $G(z) = 0$ when $\theta(z) = 1$, by Theorem 1, we know that

$$\sqrt{Th/\nu(z)}(\hat{g}\{\theta(z)\} - \frac{1}{2}) \xrightarrow{d} N(0,1),$$

where $Z \sim N(0,1)$. For any positive d, we have

$$\begin{aligned}
&P\left\{\sqrt[4]{\frac{Th}{\nu(z)}}(\hat{\theta}(z) - 1) \le -r\right\} = P\left\{\hat{\theta}(z) \le 1 - \frac{r\sqrt[4]{\nu(z)}}{\sqrt[4]{Th}}\right\} \\
=&P\left\{g(\hat{\theta}(z)) \le g(1 - \frac{r\sqrt[4]{\nu(z)}}{\sqrt[4]{Th}})\right\}
\end{aligned}$$

$$
\begin{aligned}
=&P\left\{\sqrt{Th}[g(\hat{\theta}(z))-\frac{1}{2}]\le\sqrt{Th}[g(1-\frac{r\sqrt[4]{\nu(z)}}{\sqrt[4]{Th}})-g(1)]\right\}\\
=&P\left\{\sqrt{Th}[g(\hat{\theta}(z))-\frac{1}{2}]\le\sqrt{Th}[-\frac{r^2\sqrt{\nu(z)}}{4\sqrt{Th}}+o(\frac{1}{\sqrt{Th}})]\right\}\\
=&P\left\{\sqrt{Th/\nu(z)}(g(\hat{\theta}(z))-\frac{1}{2})\le-\frac{r^2}{4}+o(1)\right\}\to\Phi(-\frac{r^2}{4}).
\end{aligned}
$$

Also, since $\hat{\theta}(z)\le 1$, we have

$$\sqrt[4]{Th/\nu(z)}(\hat{\theta}(z)-\theta(z))\xrightarrow{d}H_\Phi^-.$$

$\square$

Proof of Theorem 4. Write $\hat{\theta}(z)=h[g\{\hat{a}_1(z)/\hat{a}_0(z)\}]=:q\{\hat{a}_1(z),\hat{a}_0(z)\}$, say. By Taylor expansion,

$$\hat{\theta}(z)\;=\;q\{a_1(z),a_0(z)\}+\sum_{j=0}^{1}\frac{\partial q\{a_1(z),a_0(z)\}}{\partial a_j(z)}\{\hat{a}_j(z)-a_j(z)\}+O_p((Th)^{-1/2}+h^2)$$

The test statistic can be expressed as

$$
\begin{aligned}
D_T\;=\;&Th^{1/2}\big[\int\{\hat{\theta}(z)-E\{\hat{\theta}(z)\}\}^2\pi(z)dz\\
&+2\int\{\hat{\theta}(z)-E\{\hat{\theta}(z)\}\}\{E\{\hat{\theta}(z)-\hat{\theta}\}\pi(z)dz\\
&+\int\{E\{\hat{\theta}(z)-\hat{\theta}\}^2\pi(z)dz\\
=\;&D_{T1}+D_{T2}+D_{T3},\quad\text{say.}
\end{aligned}\tag{3}
$$

Let us first study D_{T1}, which can be expressed as

$$D_{T1}\;=\;Th^{1/2}\int(\hat{a}(z)-E\{\hat{a}(z)\})^T\,\frac{\partial q(z)}{\vec{a}}\frac{\partial^T q(z)}{\vec{a}}\,(\hat{a}(z)-E\{\hat{a}(z))\,\pi(z)dz.$$

Take expectation, we have

$$E(D_{T1})\;=\;h^{-1/2}\int\{g^{'}\{\theta(z)\}^2\nu(z)\big]\pi(z)dz\{1+o(1)\}.\tag{4}$$

It can be shown (Chen, Gao and Tang, 2008; Chen and Gao, 2011) that

$$\begin{aligned} & Var(D_{T1}) \\ = \; & 2T^2 h \int\int Cov^2\{\hat{\theta}(z_1) - E\{\hat{\theta}(z_1)\}, \hat{\theta}(z_2) - E\{\hat{\theta}(z_2)\}\}\pi(z_1)\pi(z_2)dz_1dz_2 \\ = \; & 2T^2 h \int\int \Big[\sum_{i,j=0}^{1} \frac{\partial q(z_1)}{\partial a_i}\frac{\partial q(z_2)}{\partial a_j} Cov\{\hat{a}_i(z_1) - E\{\hat{a}_i(z_1)\}, \hat{a}_j(z_2) - E\{\hat{a}_j(z_2)\}\Big]^2 \\ & \times\pi(z_1)\pi(z_2)dz_1dz_2\{1+o(1)\}. \end{aligned}$$

Standard derivations in kernel regression estimation show that, for $i, j = 0$ or 1,

$$\begin{aligned} & Cov\{\hat{a}_i(z_1) - E\{\hat{a}_i(z_1)\}, \hat{a}_j(z_2) - E\{\hat{a}_j(z_2)\}\} \\ = \; & (Th)^{-1} f^{-1}(z_2) K^{(2)}\left(\frac{z_1 - z_2}{h}\right) m_{i,j}(z_1)\{1+o(1)\}. \end{aligned}$$

These lead to

$$\begin{aligned} & \sigma^2(D_{T1}) =: Var(D_{T1}) \\ = \; & 2h^{-1}\int\int \{K^{(2)}\left(\frac{z_1 - z_2}{h}\right)\}^2 f^{-2}(z_2)\Big[\sum_{i,j=0}^{1}\frac{\partial q(z_1)}{\partial a_i}\frac{\partial q(z_2)}{\partial a_j} m_{i,j}(z_1)\Big]^2 \\ & \times\pi(z_1)\pi(z_2)dz_1dz_2\{1+o(1)\} \\ = \; & 2K^{(4)}(0)\int f^{-2}(z)\Big[\sum_{i,j=0}^{1}\frac{\partial q(z)}{\partial a_i}\frac{\partial q(z)}{\partial a_j} m_{i,j}(z)\Big]^2 \pi^2(z)dz\{1+o(1)\} \end{aligned} \tag{5}$$

where $K^{(2)}(\cdot)$ is the convolution of K and $K^{(4)}(\cdot)$ is the convolution of $K^{(2)}$. Let $\sigma^2_{D_T}$ denote the leading order term of (5).

Employing the same technique used in Chen and Gao (2011) for goodness-of-test statistics for α-mixing sequences, it can be shown that the statistic D_{T1} is asymptotically normally distributed so that

$$\sigma^{-1}(D_{T1})\{D_{T1} - E(D_{T1})\} \xrightarrow{d} N(0,1) \quad \text{as} \quad T \to \infty. \tag{6}$$

To study the properties of D_{T2}, let θ^* be the limit of the estimator $\hat{\theta}$ such that $\hat{\theta} \xrightarrow{p} \theta^*$ at the rate of $\sqrt{T}$. Clearly, under H_0, $\theta^* = \theta(z)$ for any z. Under H_1, θ^* can be viewed as the one in Θ which minimizes a criterion function for the estimation, like a likelihood or a distance function. This is a common provision made in testing hypothesis (White, 1982; Chen and Gao, 2007). Let

$$D_{T21} = 2Th^{1/2} \int \{\hat{\theta}(z) - E\{\hat{\theta}(z_1)\}\}\{E\{\hat{\theta}(z_1)\} - \theta^*\}\pi(z)dz.$$

Under H_0, D_{T21} vanishes. Under H_1, $E(D_{T21}) = 0$ and $Var(D_{T21}) = O(Th)$. Hence, D_{T21} is at most $O_p\{(Th)^{1/2}\}$. As

$$D_{T2} = Th^{1/2} \int \{\hat{\theta}(z) - E\{\hat{\theta}(z_1)\}\}\{E\{\hat{\theta}(z_1)\} - \theta^* + \theta^* - \hat{\theta}\}\pi(z)dz,$$

$D_{T2} = O_p(h^{1/2})$ under H_0 and $O_p\{(Th)^{1/2}\}$ under H_1.

We can argue similarly that $D_{T3} = Th^{1/2} \int \{E\{\hat{\theta}(z)\} - \hat{\theta}\}^2\pi(z)dz$ is $o_p(1)$ under H_0, but diverges to positive infinity at the rate of $Th^{1/2}$ under H_1, which is faster than T_{T2}. In summary,

$$D_T - E(D_T) = D_{T1} - E(D_{T1}) + o_p(1) \quad \text{under} \quad H_0 \quad \text{and} \tag{7}$$

$$D_T - E(D_{T1}) = O_p(Th^{1/2}) \quad \text{under} \quad H_1. \tag{8}$$

These together with (6) lead to the conclusion of Theorem 4. □

References

Chen, S.X. and Gao, J. (2007). An Adaptive Empirical Likelihood Test For Time Series Models. *Journal of Econometrics* **141**, 950-972.

Chen, S.X. and Gao J. (2011). Simultaneous Specification Test for the Mean and Variance Structures for Nonlinear Time Series regression. *Econometric Theory* **27**, 792-843.

Chen, S.X., Gao J. and Tang C. Y. (2008). A Test for Model Specification of Diffusion Processes. *The Annals of Statistics* 36, 167-198.

Fan, J. and Gijbels I. (1996). Local Polynomial Modeling and Its Applications. Chapman and Hall.

Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society, Series B* **65**, 57-80.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica* **50**, 1-26.

Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, China, 100871
E-mail: csx@gsm.pku.edu.cn

School of Mathematics, Peking University, Beijing, China, 100871
E-mail: leilihuallh@126.com

Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, China, 100871
E-mail: yundong.tu@gsm.pku.edu.cn