Functional Coefficient Moving Average Model with Applications to forecasting Chinese CPI

Song Xi Chen and Lihua Lei and Yundong Tu

Peking University, University of California, Berkeley, Peking University

2014

FUNCTIONAL COEFFICIENT MOVING AVERAGE MODEL WITH APPLICATIONS TO FORECASTING CHINESE CPI

Song Xi Chen, Lihua Lei and Yundong Tu*

Peking University

Abstract: This article establishes the functional coefficient moving average model (FMA), which allows the coefficient of the classical moving average model to adapt with a covariate. The functional coefficient is identified as a ratio of two conditional moments. Local linear estimation technique is used for estimation and asymptotic properties of the resulting estimator are investigated. Its convergence rate depends on whether the underlying function reaches its boundary or not, and asymptotic distribution could be nonstandard. A model specification test in the spirit of Härdle-Mammen (1993) is developed to check the stability of the functional coefficient. Intensive simulations have been conducted to study the finite sample performance of our proposed estimator, and the size and the power of the test. The real data example on CPI data from China Mainland shows the efficacy of FMA. It gains more than 20% improvement in terms of relative mean squared prediction error compared to moving average model.

Key words and phrases: Moving Average model, functional coefficient model, forecasting, Consumer Price Index.

1 Introduction

Autoregressive Integrated Moving Average (ARIMA) models have been popular in time series analysis due to the simplicity and adaptability. An ARIMA(p, d, q) model has the following expression:

$$(1-B)^d(1-\phi_1 B-\cdots-\phi_p B^p)x_t=\mu+(1+\theta_1 B+\cdots+\theta_q B^q)\epsilon_t$$

where B is the lagged operator and $\{\epsilon_t\}$ is a white noise series with zero mean and finite variance. On the one hand, it describes a special dependence structure

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

of data while on the other hand it can be regarded as an approximation to all stationary process according to Wold Decomposition Theorem. In the past decades, numerous works in statistics and econometrics have been devoted into studying and extending the ARIMA model and its applications (to cite a few, Box and Jenkins, 1970; Box and Tiao, 1975; Dahlhaus, 1989; Cleveland and Tiao, 1976 ; Granger and Joyeux, 1980; Hannan and Deistler, 1988; Engle and Granger, 1987).

One important application of ARIMA model is to forecast Consumer Price Index (CPI). The growth rate of CPI can be regarded as a proxy of inflation rate, which is a chief target of **macro-economic management by various governments and is an** important economic indicator for investors. One popular model for the CPI is ARIMA(0,1,1) (Nelson and Schwert, 1977; Schwert 1987; Barsky 1987) :

$$(1-B)x_t = \mu + (1-\theta B)\epsilon_t$$

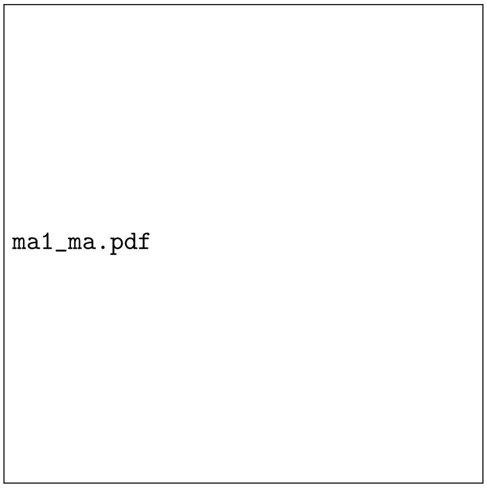
where $\{x_t\}$ represents logarithm of CPI. Although the model is easy to implement, it puts rather stringent restrictions to the inflation dynamics that the autocovariance should be constant over time. However, it is observed for the US data that the estimates of θ are not stable over time and are fairly volatile. Stock and Watson (2006) interpreted this instability as the variation of variance, which changes inversely with the magnitude of MA coefficient estimates. The parameter instability is also observed **in our analysis** when analyzing monthly CPI data of China Mainland from January 1990 to March 2014. We build the ARIMA(0,1,1) model on the year-on-year CPI monthly growth data, and estimate the **MA** coefficient θ on an expanding window basis and a rolling window basis with a 60-month **window**-width. These estimates are plotted in Figure 1. It can be seen that the estimates of θ are quite variable.

Based on the above observations, we consider an extension of ARIMA(0,1,1) model in which the MA coefficient is a smooth function of a state **covariate** z_t such that

$$(1-B)x_t = \mu + (1-\theta(z_t)B)\epsilon_t. \tag{1}$$

This model is called Functional Moving Average (FMA) model of order 1, or

Shall we call it a co-state variable or covariate ?



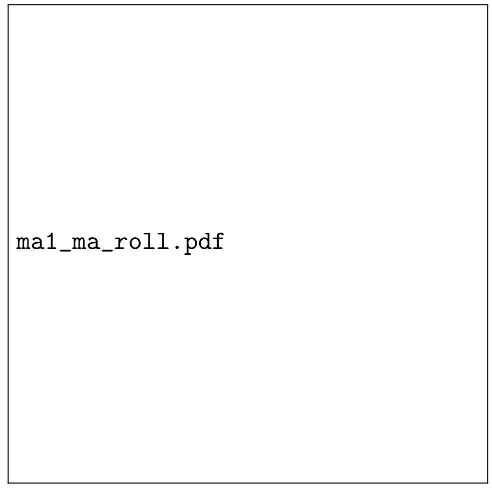


Figure 1: Estimates of θ on an expanding window basis (panel (a)) and a rolling window basis (panel (b))

FMA(1). **The co-state variable z_t contains information that affects the dynamics of x_t, and does not has to be exogeneous. The dynamic admissible to z_t is very general as indicated by the Assumption A3-A5 given in Section 2.2. We provides a testing procedure that determines whether a given variable is qualified as co-state variable which can be used to improve the inference and prediction of x_t in Section 2.4.** In this paper, we focus on inference on FMA(1). The extension to higher order FMA **will be discussed later in the conclusion section**. The choice of z_t can be made based on, for example, related economic theory, or through a data driven procedure. In this paper, we develope a test procedure to check if a variable z_t is adequate to function as a co-state variable. **We note that our FMA model is related to the state-dependent models of Priestley (1980) and the autoregressive functional moving average (ARFMA) model of Wang (2008), where the latter is a specific form of the former. However, the ARFMA model has the functional coefficient of the MA parts being functions of the legged values of the state variabe x_t itself. Our FMA framework has the functional coefficient depends on a co-variable z_t, which is not necessarily legged value of the state variable. Of course, ARFMA and FMA can be united under a more generakl framework**

with multivariate state variables. In any case, the asymptotic properties of the estimators for the state-dependent and the ARFMA models have yet to be made. And we provide such results in this paper for the FMA(1) model.

In econometrics and time series literatures (Hamilton, 1994), the MA coefficients are often explained as the Impulse Response (IR). To be precise, for any series x_t that can be written in a MA(∞) form:

$$x_t = \mu + \sum_{j \geq 0} \theta_j \epsilon_{t-j},$$

the j-th order IR is $\frac{\partial x_t}{\partial \epsilon_{t-j}} = \theta_j$ for any $j \geq 0$. It measures the effect of a shock on the response after j periods. For FMA(1) model, the 1-st order IR is $\theta(z_t)$, which is a function of the state variable rather than a constant as in the MA(1) model. This flexibility brings closer linkage to the real world as the effect of a shock is often affected by the state of the world.

Our work is closely related to a large body of literature on varying coefficient models. They have been well developed in nonparametric statistics and time series analysis, including ARCH/GARCH (Engle, 1982; Bollerslev, 1986), TAR (Tong, 1983; Chan and Tong, 1986; Tong, 1990; Tiao and Tsay, 1994; Caner and Hansen, 2001), EXPAR (Haggan and Ozaki, 1981; Ozaki, 1982) and FAR (Chen and Tsay, 1993; Cai, Fan and Li, 2000; Fan, Yao and Cai, 2003). This literature focuses mainly on extending the AR component of the ARIMA model, while the current work aims to relax the flexibility of the MA component. See also Priestley (1980) and Wang (2008).

The unique feature **in the inference for the** FMA(1) model **is the** estimation technique. Unlike the FAR(1) model which has a regression form, local polynomial regression cannot be directly applied to FMA (1). Nevertheless, we find that the functional coefficient is identified via the conditional autocovariance function. As a result, the functional coefficient can be consistently estimated by first estimating the autocovariance function. To this end, local linear least square is used to obtain estimates of conditional moments.

We note to the readers that this paper could be extended in several directions. First, an AR component could be incorporated to allow for more general dependence structure. Second, the FMA(1) model could also be generalized to allow for multiple state variables Z_t. To avoid the curse of dimensionality, a single index structure for $\theta(\cdot)$, such as $\theta(Z_t^\top \gamma)$, could be imposed and estimation procedure adapted from Ichimura (1993) can be used. Nevertheless, the identification and estimation technique proposed in this paper would not simply apply in either case. We leave these complicated extensions for future research.

The rest of paper is structured as follows. The next section introduces the details for identification and estimation of the FMA model. The asymptotic distribution of the proposed estimator is established and a model specification test is also developed. Section 3 presents simulation results that evaluate the finite sample performance of our estimator, and the size and power of the model specification test. Section 4 shows the efficacy of FMA model by forecasting Chinese CPI data and compare it to MA(1) models. Section 5 concludes with remarks on future work. All technical lemmas and proofs are left in the appendix.

2 Theoretical Property

2.1 Identification and Estimation

For MA(1) model

$$x_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

where $\{\epsilon_t\}$ is a white noise process with variance σ^2, the variance and the first autocovariance of x_t is

$$E((x_t - \mu)^2) = (1 + \theta^2)\sigma^2,$$

$$E((x_t - \mu)(x_{t-1} - \mu)) = \theta\sigma^2.$$

Higher order autocovariances are all 0's. Then θ could be estimated via the ratio of two moments after certain transformation.

Now suppose that x_t follows a FMA model with the state variable z_t, i.e.

$$x_t = \mu + \epsilon_t + \theta(z_t)\epsilon_{t-1}.$$

where $\{\epsilon_t\}$ is a white noise with variance σ^2, $\theta(z_t)$ is a smooth function with

$|\theta(z_t)| \leq 1$. Conditional on z_t, its autocovariance functions follows a similar structure as those of MA(1). To see this, it follows from the definition that

$$E((x_t-\mu)^2|z_t=z) = E(\epsilon_t^2|z_t=z) + 2\theta(z)E(\epsilon_t\epsilon_{t-1}|z_t=z) + \theta^2(z)E(\epsilon_{t-1}^2|z_t=z)$$

$$E((x_t-\mu)(x_{t-1}-\mu)|z_t=z) = E(\epsilon_t\epsilon_{t-1}|z_t=z) + E(\theta(z_{t-1})\epsilon_t\epsilon_{t-2}|z_t=z)$$
$$+\theta(z)\{E(\epsilon_{t-1}^2|z_t=z) + E(\theta(z_{t-1})\epsilon_{t-1}\epsilon_{t-2}|z_t=z)\}.$$

If for $j,k=0,1$,

$$E(\epsilon_{t-k}\epsilon_{t-j}|z_t) = E(\epsilon_{t-k}\epsilon_{t-j}) = \sigma^2 I(j=k) \text{ and} \tag{2}$$

$$E(\epsilon_{t-j}\epsilon_{t-2}|z_t, z_{t-1}) = E(\epsilon_{t-j}\epsilon_{t-2}) = 0, \tag{3}$$

then

$$E((x_t-\mu)^2|z_t=z) = (1+\theta^2(z))\sigma^2 \text{ and} \tag{4}$$

$$E((x_t-\mu)(x_{t-1}-\mu)|z_t=z) = \theta(z)\sigma^2. \tag{5}$$

Now the two conditional moments have the same form with those of the MA(1) model. The condition (2) and (3) are satisfied if (z_t, z_{t-1}) is independent of $(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2})$ for all t. In practice, z_t is often taken as lagged variables (e.g., x_{t-d}, for some $d > 2$) that contain the state information, as in FAR (Cai, Fan and Yao, 2000). This requirement is not as stringent as it appears, since it is often reasonable in application to assume the independence between the future innovations and the past variables. This condition is precisely described in Assumption (A5) below.

Nonparametric method of moments can be used to estimate $\theta(z)$. To do so, we need to estimate two conditional moments (4) and (5). Many nonparametric estimators could be used such as the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964) and the local polynomial estimator (Fan and Gijbel, 1996). We prefer to using the local linear estimators due to its attractive statistical properties including the minimax efficiency, automatic boundary correction and a simpler form of the asymptotic bias. Denote the local linear estimator of the

---

(2) and (3) give the impression that the z_t are exogeneous.

variance and the autocovariance by $\hat{a}_0(z)$ and $\hat{a}_1(z)$, i.e.

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

for $j = 0, 1$, where $\bar{x} = T^{-1} \sum_{t=1}^{T} x_t$ is a consistent estimator for μ, $k(\cdot)$ is a kernel function and h is the smoothing parameter.

Denote $g(w) = w/(1 + w^2)$, which is monotone in $w \in [-1, 1]$. A natural estimator for $g\{\theta(z)\}$ is

$$\hat{g}\{\theta(z)\} = \frac{\hat{a}_1(z)}{\hat{a}_0(z)}. \tag{6}$$

Note that $|g(w)| \leq 1/2$ for all $w \in [-1, 1]$. To incorporate this restriction, we consider the constrained estimator

$$\tilde{g}\{\theta(z)\} = \hat{g}\{\theta(z)\} I(|\hat{g}\{\theta(z)\}| \leq \tfrac{1}{2}) + \tfrac{1}{2} I(\hat{g}\{\theta(z)\} > \tfrac{1}{2}) - \tfrac{1}{2} I(\hat{g}\{\theta(z)\} < -\tfrac{1}{2}). \tag{7}$$

Then, $\theta(z)$ can be estimated by

$$\hat{\theta}(z) = h(\tilde{g}\{\theta(z)\}),$$

where $h : [-1/2, 1/2] \to [-1, 1]$, and

$$h(x) = g^{-1}(x) = \begin{cases} \frac{1-\sqrt{1-4x^2}}{2x} & (\text{ if } x \neq 0); \\ 0 & (\text{ if } x = 0). \end{cases}$$

It is noted that our estimation for $g(\theta(z))$ is based on a ratio estimator and may not be efficient. Therefore, efficient estimator for $\theta(z)$ may be constructed, which is left for further investigation.

2.2 Large Sample Theory

To maximize the clarity of presentation, we only consider the case where z_t is a scalar. The extension to allow for multi-dimensional state variables follows in a similar fashion and is further remarked in the conclusion. The following regularity conditions are assumed to obtain the large sample properties.

(A1) $h = O(T^{\epsilon_0 - 1})$ as $T \to \infty$ for some $\epsilon_0 \in (0, 1)$.

(A2) The kernel function $K(\cdot)$ is symmetric and Liptchitz on its support $S_K = [-1, 1]$, in that there exists a $M > 0$ such that $|K(x) - K(y)| \leq M|x - y|$ for all $x, y \in S_K$.

(A3) (i) $\{\epsilon_t\}$ is a white noise sequence with $E\epsilon_t^2 = \sigma^2 < \infty$, $E|\epsilon_t|^{2\delta} < \infty$ for some $\delta > 2$; (ii) $\{\epsilon_t, z_t\}$ is a strictly stationary α-mixing process with the mixing coefficients satisfying the condition $\alpha(k) < ck^{-\beta}$ for some $\beta > \max\{\frac{2\delta-2}{\delta-2}, \frac{2-\epsilon_0}{\epsilon_0}\}$ and constant $c > 0$.

(A4) (i) The density function $p(z)$ of z_t has a bounded second derivative; (ii) the conditional density function of (z_1, z_m) given $(x_1, \cdots, x_m)$ is bounded by a $C_0 > 0$ uniformly with $m \geq 0$; (iii) the conditional density of x_t given z_t is continuous.

(A5) (i) For each t and $j, k = 0, 1$, $E(\epsilon_{t-k}\epsilon_{t-j}|z_t) = \sigma^2 I(j = k)$ and $E(\epsilon_{t-j}\epsilon_{t-2}|z_t, z_{t-1}) = 0$; (ii) $E(|\epsilon_{t-j}|^{2\delta}|z_t = z) \leq M < \infty$ for some M and $j = 0, 1, 2$, and the same δ in (A2).

(A6) The coefficient function $\theta(z)$ has continuous second derivative and $|\theta(z)| \leq 1$ for any $z \in \mathbb{R}$.

Conditions (A1) and (A2) are standard assumptions in the kernel smoothing literature. For instance, the second-order Epanechnikov kernel satisfies this requirement and is used throughout the paper. Conditions (A3) and (A4) are used by Masry and Fan (1997) for α-mixing processes. The condition imposed on β in (A3) is a technical requirement. If ϵ_t satisfies the Cramér Condition, i.e. $Ee^{\lambda|\epsilon_t|^\alpha} < \infty$ for some $\lambda, \alpha > 0$, then δ can be arbitrarily large and hence (A3) can be reduced to $\beta > 2$ if $\epsilon_0 > \frac{2}{3}$. Condition (A5.i) is needed for identification of the model, which has been discussed in the last subsection. (A5.ii) is a technical condition in order to apply the result of Masry and Fan (1997). It holds under (A3) if z_t is independent of $(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2})$. (A6) places smoothness condition on the functional coefficient. **We note that In particular, z_t does not have to be exogeneous. The dynamic admissible to z_t is very general as indicated by the Assumption A3-A5, which are largely of the mixing condition, the conditional moment conditions and conditions regarding the conditional densities.**

We begin with the asymptotic normality of $\hat{g}\{\theta(z)\}$. The following quantities are needed to present the asymptotic distribution of $\hat{g}\{\theta(z)\}$. Let

$$G(z) = \frac{u(z)^\top A u''(z)}{2[1+\theta^2(z)]^2}\sigma_K^2, \nu(z) = \frac{u(z)^\top \Gamma(z) u(z)}{[1+\theta^2(z)]^4 p(z)} R(K), A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \text{ and}$$

$$\Gamma(z) = Cov\left[(x_t-\mu)(x_{t-1}-\mu), (x_t-\mu)^2 | z_t = z\right]/\sigma^4,$$

where $\sigma_K^2 = \int u^2 K(u)du$, $R(K) = \int K^2(u)du$, $u(z) = (1+\theta^2(z), -\theta(z))^\top$, and $\theta'(z)$ and $\theta''(z)$ are the first and second derivatives of $\theta(z)$. Let $\mathbb{S} = \{z | p(z) > 0\}$.

Theorem 1. Under Assumptions (A1)$\sim$(A6), it holds for $z \in \mathbb{S}$ that as $T \to \infty$,

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

Remark 1. Let $\mathbb{M} = \{z : \theta(z) = \pm 1\}$. Since $|\theta(z)| \leq 1$ for all $z \in \mathbb{R}$, the points in $\mathbb{M}$ are local extremas of $\theta(z)$. Thus, $\theta'(z) = 0$ for all $z \in \mathbb{M}$. It is easily shown that $G(z) = 0$, for $z \in \mathbb{M}$.

The next theorem establishes the asymptotic property of $\tilde{g}\{\theta(z)\}$, the constrained estimator of $g\{\theta(z)\}$.

Theorem 2. Under Assumptions (A1)$\sim$(A6), it holds for $z \in \mathbb{S}$ that

(i) If $|g\{\theta(z)\}| < \frac{1}{2}$,

$$\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - g\{\theta(z)\} - G(z)h^2) \xrightarrow{d} \Phi;$$

(ii) If $g\{\theta(z)\} = \frac{1}{2}$,

$$\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - g\{\theta(z)\}) \xrightarrow{d} \Phi^-;$$

(iii) If $g\{\theta(z)\} = -\frac{1}{2}$,

$$\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - g\{\theta(z)\}) \xrightarrow{d} \Phi^+,$$

where Φ is the standard normal distribution function, and

$$\Phi^-(x) = \Phi(x)I(x<0) + I(x \geq 0), \Phi^+(x) = \Phi(x)I(x \geq 0).$$

The above theorem reveals the distribution discontinuity at the boundaries of $g\{\theta(z)\}$. Intuitively, when $|g\{\theta(z)\}| < 1/2$, the unconstraint estimator $\hat{g}\{\theta(z)\}$ will be the same as $\tilde{g}\{\theta(z)\}$ for sample size large enough. Therefore, the unconstraint estimator and the constraint estimator are asymptotically equivalent. However, when $|g\{\theta(z)\}| = 1/2$, the constraint becomes binding, i.e. $\hat{g}\{\theta(z)\} \neq \tilde{g}\{\theta(z)\}$, with positive probability. In this case, the asymptotic distribution of the constrained estimator will be different from that of the unconstrained one.

Now we are in a position to state the asymptotic property of $\hat{\theta}(z) = h(\tilde{g}\{\theta(z)\})$. Note that $h(x)$ is differentiable when $|x| < 1/2$. The delta-method can be applied to Theorem 2 to obtain the asymptotic distribution of $\hat{\theta}(z)$. At $|x| = 1/2$, the asymptotic distribution can be derived directly. See the appendix for details.

Theorem 3. Under Assumptions (A1)$\sim$(A6), it holds for $z \in \mathbb{S}$ that

(i) If $|\theta(z)| < 1$,

$$\sqrt{Th/\nu(z)}g'\{\theta(z)\}(\hat{\theta}(z) - \theta(z) - g'\{\theta(z)\}^{-1}G(z)h^2) \xrightarrow{d} \Phi;$$

(ii) If $\theta(z) = 1$,

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

(iii) If $\theta(z) = -1$,

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

where $H_\Phi^-(x) = \Phi(-x^2/4)I(x < 0) + I(x \geq 0)$ and $H_\Phi^+(x) = \Phi(x^2/4)I(x \geq 0)$.

It is seen that the convergence rate of $\hat{\theta}(z)$ depends on whether $|\theta(z)| < 1$. When $\theta(z) = \pm 1$, it converges at a slower rate and its asymptotic distribution is nonstandard.

We note that asymptotic variance of the above estimators rely on the unknown parameter σ^2. It could be consistently estimated by the sample average of the squared innovation residuals $\hat{\epsilon}_t^2$, for $t = 1, \cdots, T$ under Assumption (A3), where $\hat{\epsilon}_t$ could be obtained in a similar iterative procedure like that in the moving average models, with $\hat{\theta}$ replaced by the estimated function $\hat{\theta}(z_t)$.

2.3 Bandwidth Selection

The theoretical optimal bandwidth for estimating $\theta(z)$ minimizing the asymptotic mean squared error of $\hat{\theta}(z)$ can be shown as

$$\hat{h}^{opt} = (\frac{\nu(z)g'(\theta(z))^2}{4G(z)^2 T})^{\frac{1}{5}} = \left(c_K \frac{u(z)^\top \Gamma(z) u(z) g'(\theta(z))^2}{u(z)^\top \Lambda(z) u(z) p(z)}\right)^{\frac{1}{5}} T^{-\frac{1}{5}} \tag{8}$$

where $c_K = R(K)/\sigma_K^4$ and $\Lambda(z) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} u''(z)u''^\top(z) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. This theoretical optimal bandwidth depends on the unknown elements $\theta(z)$, $\Gamma(z)$, $\Lambda(z)$ and $p(z)$. In practice, these terms can be estimated consistently with a prior bandwidth.

A practical way of bandwidth selection is to adopt the Residual Squares Criterion (RSC) proposed by Fan and Gijbel (1995), which avoids the above complication. Let

$$\hat{\Gamma}(z,h) = \frac{1}{\Delta}\sum_{t=2}^{T}(Y_t - \hat{Y}_t)(Y_t - \hat{Y}_t)^\top K(\frac{z_t - z}{h})$$

where $\Delta = tr(W - WZ(Z'WZ)^{-1}Z'W)$, $Z = [(1, z_2 - z)^\top, \ldots, (1, z_T - z)^\top]^\top$, $W = diag\{K(\frac{z_2 - z}{h}), \cdots, K(\frac{z_T - z}{h})\}$, $Y_t = ((x_t - \mu)^2, (x_t - \mu)(x_{t-1} - \mu))^\top$ and $\hat{Y}_t = (\hat{a}_0^*(z_t), \hat{a}_1^*(z_t))^\top$, where

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

With the similar arguments of Fan and Gijbel (1995), it can be shown that

$$E(\hat{\Gamma}(z,h)|z_2, \cdots, z_T) = \Gamma(z) + d_K \Lambda(z) h^4 + o_p(h^4) \tag{9}$$

where $d_K = \int u^4 K(u) du - \sigma_K^4$.

As a result, our criterion for bandwidth choice is defined as

$$R(z,h) = u(z)^\top \hat{\Gamma}(z,h) u(z)(1 + g'(\theta(z))^2 V), \tag{10}$$

where V is the first diagonal element of $(Z'WZ)^{-1}(Z'W^2Z)(Z'WZ)^{-1}$. Denote the minimizer of $R(z,h)$ as $\bar{h}$. Following Fan and Gijbel (1995), one can show that $adj_K\bar{h}$ offers a reasonable approximation for $\hat{h}^{opt}$ in practice, where

$$adj_K = \left(\frac{4c_K d_K}{R(K)}\right)^{\frac{1}{5}} = 4^{\frac{1}{5}}\left(\frac{\int u^4K(u)du}{\left(\int u^2K(u)du\right)^2} - 1\right)^{\frac{1}{5}}.$$

To see this, by Fan and Gijbel (1995), we have

$$V = \frac{R(K)}{Thp(z)}(1 + o_p(1)). \tag{11}$$

In addition, it follows from (10) and (11) that

$$E(R(z,h)|z_2,\cdots,z_T) = u(z)^\top\Gamma(z)u(z) + d_K u(z)^\top\Lambda(z)u(z)h^4$$
$$+R(K)\frac{u(z)^T\Gamma(z)u(z)g'(\theta(z))^2}{Thp(z)} + o_p(h^4 + \frac{1}{Th}).$$

It can be shown that the minimizer of the leading term of the above expression is

$$\hat{h}_o = \hat{h}^{opt}/adj_K.$$

Note that $R(z,h)$ depends on the unknown $\theta(z)$, we can use $\hat{\theta}(z)$ with a prior bandwidth h to replace $\theta(z)$. Furthermore, the constant adj_K is determined by the chosen kernel function. For example, $adj_K = (92/7)^{\frac{1}{5}}$ for the Epanechnikov kernel.

To obtain a globally optimal bandwidth, one can minimize

$$IR(h) = \int R(z,h)dz$$

and use $adj_K \cdot argmin_h IR(h)$ as the bandwidth. For implementation, the integral can be approximated by a discrete summation over the observed data. Finally, we note that undersmoothing is often desired as one would like to avoid the bias estimation in practice.

2.4 Model Specification Test

When the coefficient function $\theta(z)$ is a constant, the FMA(1) model is an MA(1) model, using an FMA model can result in a loss in estimation efficiency.

On the other hand, when the underlying model is not an MA(1) model but an FMA model, using a misspecified an MA(1) model can produce erroneous inference. Therefore, a model specification test is needed to check if the specification of FMA model is adequate.

Various approaches can be taken to construct such a specification test, for example, following Fan and Li (1996) or Chen and Gao (2007) among others. We are to adopt the L2 norm based test for regression functions (degenerated to a parameter in our case) proposed by Härdle and Mammen (1993) for testing the constancy of $\theta(z)$, due to its simple nature in implementation. The null hypothesis is

$$H_0 : P(\theta(z) \equiv \theta \text{ for some } \theta \in \mathbb{R}) = 1,$$

while the alternative is

$$H_1 : P(\theta(z) \equiv \theta \text{ for some } \theta \in \mathbb{R}) < 1.$$

Similar to Härdle and Mammen's approach, we consider the following statistic:

$$D_T = Th^{1/2} \int_R (\hat{\theta}(z) - \hat{\theta})^2 \pi(z) dz$$

where $\hat{\theta}$ is maximum likelihood estimator under H_0. Note that our test statistic does not have a smoothing operator on the parametric part, contrasted to the original Härdle and Mammen's (1993) test, as the parametric part is a degenerated function (i.e., a constant). To approximate the finite sample distribution of D under H_0, we use the following parametric bootstrap method in the spirit of Chen and Gao (2007):

Step 1 Apply the MA(1) model to x_t and obtain the estimator of the mean $\hat{\mu}$, the coefficient $\hat{\theta}$ and the variance $\hat{\sigma}^2$.

Step 2 Generate a bootstrap re-sample according to $x_t^* = \hat{\mu} + \epsilon_t^* + \hat{\theta}\epsilon_{t-1}^*$ for $t = 1, 2, \cdots, T$, where $\{\epsilon_t^*\}_{1 \le t \le T}$ are independent $N(0, \hat{\sigma}^2)$ variables and obtain an estimate $\hat{\theta}(z)$ based on the resample.

Step 3 Repeat Step 2 B times for a large integer B and obtain $\{\hat{\theta}^{(i)}(z)\}_{i=1}^B$.

Step 4 Calculate

$$D_T^{(i)} = Th^{1/2} \int_R (\hat{\theta}^{(i)}(z) - \hat{\theta})^2 \pi(z) dz, \; i = 1, 2, \cdots, B$$

and calculate the $(1-\alpha)$-th quantile of $\{D_T^{(i)}\}_{1 \le i \le B}$ as the critical value of the test.

For simplicity, one can set $\pi(z) = 1$ and use the discrete sum to approximate D_T. In the next section, we will use numerical simulations to study the size and the power of this proposed test.

Prof. Chen would add something related to the theoretical properties of the above test.

3 Finite Sample Investigation

In this section, we generate the state variable z_t from ARIMA(1,0,1) process:

$$(1 - 0.5B)z_t = (1 + 0.5B)u_t,$$

where $\{u_t\}$ is a Gaussian white noise. The response x_t is generated according to

$$x_t = \epsilon_t + s\theta(z_t)\epsilon_{t-1}$$

for some $s \in [0, 1]$, where $\{\epsilon_t\}$ is an Gaussian white noise that is independent of $\{u_t\}$. Three functions chosen for $\theta(\cdot)$ are

(1) $\theta_1(z) = 2e^{-z^2} - 1$;

(2) $\theta_2(z) = sin(3z)$;

(3) $\theta_3(z) = (e^{2z} - 1)/(e^{2z} + 1)$.

These functions are selected to describe three common features: humped, oscillated and monotone functional forms. We use these different functions to check on the sensitivity of our procedures to the pattern of the coefficient functions.

3.1 Performance of Estimation

It is known from Theorem 3 that our estimator has slower convergence rate when $\theta(z) = \pm 1$. Therefore, we only consider the case when $\theta(z) < 1$ for the finite sample study. To do so, we shrink the chosen functions by setting $s = 0.8$. For each choice of $\theta(z)$ and each $T \in \{100, 200, 500, 1000\}$, we generate $\{(x_t, z_t)\}_{t \leq T}$ for 1000 times and obtain 1000 estimates of $\theta(z)$, the mean value of which is plotted in solid line in Figure 2 to Figure 4. The dashed line in each figure represents the true function $0.8 \cdot \theta(z)$ and the dotted lines are the mean value plus and minus the standard deviation. It is seen that the proposed estimator provides accurate estimation in all the three specifications.

func1_estim.pdf

Figure 2: Plot of the true function $0.8 \cdot \theta_1(z)$ (dashed lines), averaged estimates (solid lines) and the associated one standard deviation confidence bands (dotted lines)

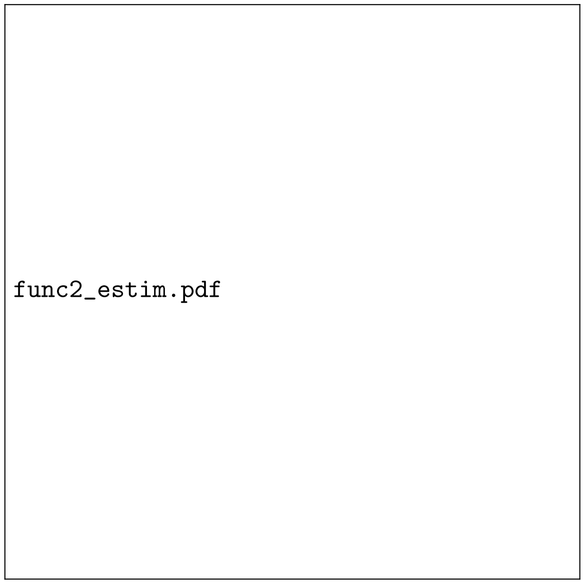


Figure 3: Plot of the true function $0.8 \cdot \theta_1(z)$ (dashed lines), averaged estimates (solid lines) and the associated one standard deviation confidence bands (dotted lines)

3.2 Finite Sample Distribution

Next we approximate the distribution of $\hat{\theta}(z)$ by simulations. Theorem 3 indicates that the asymptotic distribution of $\hat{\theta}(z)$ is determined by whether the true value lies on the boundary or not. We treat these two cases separately. We set $T = 100, 200, 500, 1000$. With each T, we generate a sample $\{(x_t, z_t)\}_{t \le T}$ for 1000 times, and obtain 1000 estimates of $\theta(z)$, denoted by $\hat{\theta}^{(1)}(z), \hat{\theta}^{(2)}(z), \cdots, \hat{\theta}^{(1000)}(z)$. Their kernel density is calculated and compared to the asymptotic distribution of $\hat{\theta}(z)$.

Note that when $\theta(z) = \pm 1$, the asymptotic distribution function of $\hat{\theta}(z)$ is discrete at ± 1, with the size of the atom being $1/2$ respectively at the origin. Even if $|\theta(z)| < 1$, there are still some estimates concentrating on ± 1 when the sample size is not large enough. Thus, if we use kernel density as the empirical density, there might be two peaks at -1 and 1, which is not desirable for comparison. To circumvent this annoying feature, we turn to the asymptotic conditional distribution of $\sqrt{Th/\nu(z)}(\hat{\theta}(z) - \theta(z) - G(z)h^2)$ given $|\hat{\theta}(z)| < 1$. When $|\theta(z)| < 1$, this distribution function will still be $\Phi(z)$ since $P(|\hat{\theta}(z)| < 1) \to 1$; when $\theta(z) = 1$, func3_estim.pdf

Figure 4: Plot of the true function $0.8 \cdot \theta_1(z)$ (dashed lines), averaged estimates (solid lines) and the associated one standard deviation confidence bands (dotted lines)

the conditional distribution will be $2\Phi(-z^2/4)$ for $z \in (-\infty, 0)$; when $\theta(z) = -1$, the conditional distribution will be $2\Phi(-z^2/4)$ for $z \in (0, \infty)$. We compare the kernel density of $\{\hat{\theta}^{(i)}(z) : |\hat{\theta}^{(i)}(z)| < 1\}$, to the corresponding asymptotic distribution. In addition, we also compute the fraction that $|\hat{\theta}^{(i)}(z)| = 1$, as denoted by $P(A)$. It should be close to 0 when $|\theta(z)| < 1$ and 0.5 when $\theta(z) = \pm 1$ for large enough T.

To save space, we set $s = 1$ and $\theta(z) = \theta_1(z)$ to illustrate the findings. First, we consider the estimation of $\theta(z)$ at $z_0 = \sqrt{\log 2}$. It is noted that $\theta(z_0) = 0 \in (-1, 1)$. The empirical conditional density of the standardized data are plotted in Figure 5 and the probability $P(A)$ is reported at the bottom of each subfigure. The bandwidth of kernel density is selected by cross validation. The red dashed line is the standard normal density and the black solid line is the kernel density. Note that two lines are close to each other even for moderate T and $P(A)$ decreases to 0 when the sample size becomes larger.

To study the boundary issue, we estimate $\theta(z)$ at $z_0 = 0$ ($\theta(z_0) = 1$). The conditional kernel density of the standardized data are plotted in Figure 6. Note that two lines are close to each other even for moderate T and $P(A)$ increases to

0.5 when the sample size increases.

Figure 5: The finite sample distribution of $\hat{\theta}_1(z)$ at $z_0 = \sqrt{\log 2}$ (solid lines) and the theoretical asymptotic distribution (dashed lines) together with the probability of $A = \{\hat{\theta}_1(z_0) = \pm 1\}$

3.3 Size and Power of the Test

In this subsection, we study the size and the power of the model specification test via simulation. The size is estimated by the proportion of rejection under the null hypothesis while the power is estimated by that under the alternative. As for the size, we consider the following DGP:

$$x_t = \epsilon_t + \theta \epsilon_{t-1}$$

The coefficient θ is set to be $0.2, 0.4, 0.6, 0.8, 1.0$, respectively. For each θ and each sample size $T \in \{100, 200\}$, we generate 500 sets of data and calculate the proportion of rejection when the significance level α is 0.05. The results are reported in table 1. It can be seen that the test has proper size.

As for the power, we consider the following DGPs.

$$x_t = \epsilon_t + s \cdot \theta_j(z_t)\epsilon_{t-1}$$

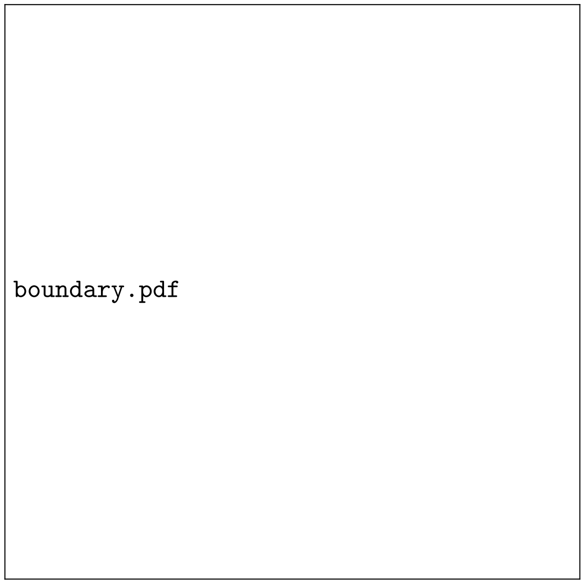


Figure 6: The finite sample conditional distribution of $\hat{\theta}_1(z)$ at $z_0 = 0$ given $|\hat{\theta}_1(z_0)| < 1$ (solid lines) and the theoretical asymptotic distribution (dashed lines) together with the probability of $A = \{\hat{\theta}_1(z_0) = \pm 1\}$

Table 1: Rejection Rate Under H_0 ($\alpha = 0.05$)

T	s=0.2	0.4	0.6	0.8	1.0
100	0.052	0.040	0.042	0.048	0.050
200	0.048	0.044	0.050	0.050	0.048

where $j \in \{1, 2, 3\}$ and $s \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$. For each design and sample size $T \in \{100, 200\}$, we generate 100 sets of data and calculate the proportion of rejection when the significance level α is 0.05. The results are reported in table 2 and it is seen that the rejection rate gets larger when s increases. For moderate value of s, the power is desirable.

4 Application to Chinese CPI

In this section, we apply a FMA model to Chinese CPI data and compare its forecast performance to that of MA model. The year-on-year CPI monthly growth data ranging from Jan. 1990 to Mar. 2014 is downloaded from Wind

Table 2: Rejection Rate Under H_1 ($\alpha = 0.05$)

$\theta(z)$	T	$s = 0.2$	0.4	0.6	0.8	1.0
$\theta_1(z)$	100	0.02	0.06	0.36	0.54	0.66
	200	0.10	0.32	0.64	0.84	0.86
$\theta_2(z)$	100	0.10	0.02	0.24	0.38	0.48
	200	0.06	0.14	0.26	0.60	0.74
$\theta_3(z)$	100	0.16	0.42	0.70	0.72	0.78
	200	0.26	0.38	0.70	0.92	0.98

database (www.wind.com.cn). The raw data is plotted in panel (a) of Figure 7. It is clear that the data is nonstationary (the p value of ADF test is less than 0.01). The first order difference of the data is plotted in panel (b) of Figure 7.

rawdata.pdf

Figure 7: The CPI monthly growth rate (panel (a)) and its first order difference (panel (b))

Our target is to forecast the data ranging from Jan. 2011 - Mar. 2014. If we use the MA(1) model for the first-order differenced log CPI (or equivalently, ARIMA(0,1,1) for CPI), the root mean squared forecast error (RMSE) of MA(1) model is 0.589.

Now we turn to the forecast using FMA(1). The first set of state variables we consider are various measures of money supply, including M0, M1, M2, as the neutrality of money implies that increase in money supply will eventually convert to the increase in price level. Other economic variables which may affect the level of price includes export (Ex), import (Im), retail sales (RS) and PPI are also considered. Since PPI is often presumed to be the leading index of CPI, we also consider 4 sub-categories of PPI: capital goods (ca), consumer goods (co), light manufacturing (lm) and heavy manufacturing (hm). Year-on-year growth rate data for these 11 state variables are obtained from Wind Database. Since all variables are non-stationary, first-order differenced series are used.

First, we conduct the model specification test to detect the state variables whose corresponding coefficient functions differ from a constant significantly. For each variable, we include lagged variables starting from the 2-nd order to the 12-th order. The 1-st order lagged variables are excluded for identification requirements. Among all of 121 state variables (11 variables with 11 lags for each), we find that 20 of them are significant at $\alpha = 0.05$. These findings are summarized in Table 3.

Due to the space limit, we plot the estimate of $\theta(\cdot)$ with respect to 6 of the significant variables, $M0_{t-12}$, $M2_{t-8}$, Ex_{t-11}, Im_{t-12}, co_{t-12} and lm_{t-12}, as illustrations in Figure 8, which displays strong departure from constancy.

Table 3: Significant State Variables

z_{t-d}	d	z_{t-d}	d
M0	12	PPI	4, 11
M1	9	ca	11
M2	8, 9	co	12
Ex	2, 11, 12	lm	7, 8, 9, 12
Im	2, 12	hm	11
RS	11, 12		

The forecast RMSE with respect to all variables are summarized in Table 4. Among these 20 variables, over 85% of them outperforms the MA(1) model in terms of the forecast RMSE. Among all significant variables, we find that the 12-th lag of import leads to the best forecasts. The forecasts RMSE reaches 0.463, which is a 21.4% improvement compared to that of the MA(1) model.

Table 4: Forecasting RMSE of FMA(1) with Various Variables

z_t	$\mathrm{M0}_{t-12}$	$\mathrm{M1}_{t-9}$	$\mathrm{M2}_{t-8}$	$\mathrm{M2}_{t-9}$
	0.524	0.572	0.638	0.525
z_t	Ex_{t-2}	Ex_{t-11}	Ex_{t-12}	Im_{t-2}
	0.532	0.505	0.563	0.549
z_t	Im_{t-12}	RS_{t-11}	RS_{t-12}	PPI_{t-4}
	0.463	0.521	0.575	0.607
z_t	PPI_{t-11}	ca_{t-11}	co_{t-12}	lm_{t-7}
	0.519	0.528	0.494	0.577
z_t	lm_{t-8}	lm_{t-9}	lm_{t-12}	hm_{t-11}
	0.513	0.502	0.508	0.543

5 Conclusion

This paper extends the moving averaging models by allowing the MA coefficients to adapt with a covariate. Under parameter identification, we proposed to estimate the functional coefficient by a ratio of two conditional moment estimators derived from local linear least squares. The consistency and asymptotic distribution of the proposed estimators are established. A Härdle and Mammen type adequacy test of the constancy of the functional coefficient is also proposed. Both simulation and empirical exercises show that our proposed method perform well in finite samples.

The FMA(1) framework can be extended to the general ARFMA(p,q). Let us outline how the extension can be made via ARFMA(1,2)

$$x_t - \alpha x_{t-1} = \epsilon_t + \theta_1(z_t, z_{t-1})\epsilon_{t-1} + \theta_2(z_t, z_{t-1})\epsilon_{t-2} \tag{12}$$

where α is the AR coefficient, and $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are two MA nonparametric coefficient functions which depends on (z_t, z_{t-1}) as suggested by a referee. We have assume in (12) the mean of x_t is zero to simplify the notation. After algebraic manipulation similar to those exhibated in (3)-(4), it can be shown that

$$\begin{aligned} &Var(x_t|z_t, z_{t-1}) - 2\alpha Cov(x_t, x_{t-1}|z_t, z_{t-1}) + \alpha^2 Var(x_{t-1}|z_t, z_{t-1}) \\ &= \sigma^2\{1 + \theta_1^2(z_t, z_{t-1}) + \theta_2^2(z_t, z_{t-1})\}, \end{aligned} \tag{13}$$

$$Cov(x_t, x_{t-1}|z_t, z_{t-1}, z_{t-2}) - \alpha Var(x_{t-1}|z_t, z_{t-1}, z_{t-2})$$
$$= \sigma^2\{\theta_1^{(}z_t, z_{t-1}) + \theta_1(z_{t-1}, z_{t-2})\theta_2(z_t, z_{t-1})\}, \quad (14)$$
$$Cov(x_t, x_{t-2}|z_t, z_{t-1}) - \alpha\, Cov(x_{t-1}, x_{t-2}|z_t, z_{t-1}) = \sigma^2\theta_2(z_t, z_{t-1}) \quad \textbf{and} \quad (15)$$
$$Cov(x_t, x_{t-3}|z_t, z_{t-1}) - \alpha\, Cov(x_{t-1}, x_{t-3}|z_t, z_{t-1}) = 0. \quad (16)$$

Let $g_j(z_1, z_2) = Cov(x_t, x_{t-j}|z_t = z_1, z_{t-1} = z_2)$ **for** $j = 0, 1, 2, 3$**,** $g_{3+j}(z_1, z_2) = Cov(x_{t-1}, x_{t-j}|z_t = z_1, z_{t-1} = z_2)$ **for** $j = 1, 2, 3$**, and** $g_{7+j}(z_1, z_2, z_3) = Cov(x_{t-j}, x_{t-1}|z_t = z_1, z_{t-1} = z_2, z_{t-2} = z_3)$**. Carrying out the local linear estimation to these functions, and denote the estimator as** $\hat{g}_k(z_1, z_2)$ **for** $k = 0, 1, \cdots$ **and** 8**. Then, estimators for** α **is**

$$\frac{\hat{\alpha} = n^{-1}\sum_{t=1}^{n}\hat{g}_3(z_t, z_{t-1})}{\hat{g}_6(z_t, z_{t-1}),}$$

which should be more efficient than having the estimation based on a single or a few (z_t, z_{t-1})**. The estimators for** $\theta_1(z_1, z_2)$ **and** $\theta_2(z_1, z_2)$ **can be obtained by solving the estimating equations based on (13) to (16). The conditions assumed for FMA(1) given in Assumptions A.2-A.5 Section 2.2 need to be updated by replacing** z_t **by the pair** (z_t, z_{t-1}, z_{t-2})**.**

We can see that as the order of the ARFMA increases, the estimation procedure involves more functions. Hence, ARFMA(p,q) models with shorter order are more useful. Indeed, one criteron one should adapt in choosing the co-state covariable z_t **is that it would allows shorter orders in the ARFMA(p,q). There are certainly more to reserach on in future on this topics.**

Acknowledgement

The authors thank the Co-Editor, Qiwei Yao, the Associate Editor, two anonymous referees, and seminar participants at School of Statistics and Management, Shanghai University of Finance and Economics, China Meeting of Econometric Society 2014, and Financial Engineering and Risk Management 2014 for helpful comments and suggestions. We acknowledge support from National

Statistics Bureau, Center for Statistical Science and LMEQF at Peking University. Chen was partially supported by National Natural Science Foundation of China Grants 11131002 and G0113. Tu acknowledges support from NSFC # ... and that from the Guanghua School of Management and . This work grows out of the weekly discussion of Yandong-School of Data and an earlier draft of this paper has been Lei's undergraduate thesis, under the supervision of Chen and Tu.

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.

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)∼(A6), it holds that

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

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.

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

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

$$\sqrt{Th}(\hat{a}^*(z; v) - E(y_t(v)|z_t = z) - \frac{1}{2}\sigma^2h^2\sigma_K^2\ v^T\begin{pmatrix} \theta''(z) \\ 2(\theta(z)\theta''(z) + \theta'^2(z)) \end{pmatrix}) \xrightarrow{d} N(0, \frac{v^T\Gamma(z)v}{p(z)}\sigma^4 R(K)). \tag{17}$$

Then Lemma 2 will be proved by Cramér Device. Now we prove (17).

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

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

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

$$\sum_{l\ge1} l^\lambda(\alpha(l))^{1-\frac{2}{\delta}} \le A^{1-\frac{2}{\delta}} \sum_{l\ge1} l^{-(1-\frac{2}{\delta})(\frac{\beta}{2}+\frac{1}{\delta-2})} < \infty.$$

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, (17) is proved by Lemma 1 and hence the lemma is proved by Cramér Device. □

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

$$|\hat{a}_j(z) - \hat{a}_j^*(z)| = O_p(\frac{1}{\sqrt{T}}) \tag{18}$$

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 $M_T = T^{-1}\sum_{t=1}^{T} K_h(z_t - z)$,

$$\hat{g}\{\theta(z)\} - g\{\theta(z)\} = \frac{\hat{a}_1(z)}{\hat{a}_0(z)} - \frac{\theta(z)}{1+\theta^2(z)} = \frac{\hat{a}_1^*(z) + O_p(T^{-\frac{1}{2}})}{\hat{a}_0^*(z) + O_p(T^{-\frac{1}{2}})} - \frac{\theta(z)}{1+\theta^2(z)}$$

$$= \frac{\theta(z)\sigma^2 + \frac{1}{2}\sigma_K^2\theta''(z)\sigma^2h^2 + (Th)^{-\frac{1}{2}}A_1 + O_p(T^{-\frac{1}{2}})}{(1+\theta^2(z))\sigma^2 + \sigma_K^2(\theta(z)\theta''(z) + \theta'^2(z))\sigma^2h^2 + (Th)^{-\frac{1}{2}}A_0 + O_p(T^{-\frac{1}{2}})} - \frac{\theta(z)}{1+\theta^2(z)}$$

$$= G(z)h^2 + (Th)^{-\frac{1}{2}}\frac{(1+\theta^2(z))A_1 - \theta(z)A_0 + O_p(1)}{(1+\theta^2(z))^2 + o_p(1)}$$

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 $\epsilon > 0$,

$$P(\sqrt{Th/\nu(z)}(\tilde{g}\{\theta(z)\} - \hat{g}\{\theta(z)\}) > \epsilon) \leq P(\tilde{g}\{\theta(z)\} \neq \hat{g}\{\theta(z)\})$$

$$=P(|\hat{g}\{\theta(z)\}| > \frac{1}{2}) \leq P(|\hat{g}\{\theta(z)\} - g\{\theta(z)\}| > \frac{1}{2} - |g\{\theta(z)\}|) \to 0.$$

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

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

$$=P\left\{\sqrt{Th}[g(\hat{\theta}(z)) - \frac{1}{2}] \leq \sqrt{Th}[g(1 - \frac{r\sqrt[4]{\nu(z)}}{\sqrt[4]{Th}}) - g(1)]\right\}$$

$$=P\left\{\sqrt{Th}[g(\hat{\theta}(z)) - \frac{1}{2}] \leq \sqrt{Th}[-\frac{r^2\sqrt{\nu(z)}}{4\sqrt{Th}} + o(\frac{1}{\sqrt{Th}})]\right\}$$

$$=P\left\{\sqrt{Th/\nu(z)}(g(\hat{\theta}(z)) - \frac{1}{2}) \leq -\frac{r^2}{4} + o(1)\right\}$$

$$\to\Phi(-\frac{r^2}{4}).$$

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

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

□

References

Barsky, R.B. (1987). The Fisher Hypothesis and the Forecastability and Persistence of Inflation. *Journal of Monetary Economics* **19**, 3-24.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* **31**, 307-327.

Bosq, D.(1998). Nonparametric Statistics for Stochastic Processes: Estimation and Prediction(2nd ed.). Springer-Verlag, Berlin.

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control (revised edition), Holden Day, San Francisco.

Box, G. E. P. and Tiao, G. C. (1975) Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association* **70**, 70-79.

Cai, Z. and Li, Q. (2008). Nonparametric Estimation of Varying Coefficient Dynamic Panel Data Models. *Econometric Theory* **24**, 1321-1342.

Caner, M. and B. Hansen (2001). Threshold autoregression with a unit root. *Econometrica* **69**, 1555-1596.

Chan, K.S. and H. Tong (1986). On Estimating Thresholds in Autoregressive Models. *Journal of Time Series Analysis* **7**, 179-190.

Chen, R. and Tsay, R. S. (1993). Functional-Coefficient Autoregressive Models. *Journal of American Statistical Association* **88**, 298-308.

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

Cleveland, W. P. and Tiao, G. C. (1976) Decomposition of Seasonal Time Series - A Model for the Census X-11 Program. *Journal of the American Statistical Association* **71**, 581-587

Dahlhaus, R. (1989), Efficient Parameter Estimation for Self-Similar Processes. *The Annals of Statistics* **17**, 1749-1766.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of U.K. Inflation. *Econometrica* **55**, 251-276.

Fan, J. and I. Gijbels (1995). Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation. *Journal of the Royal Statistical Society* **57**, 371-394.

Fan, J. and I. Gijbels (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.

Fan, J. and Yao, Q. (2006). Nonlinear Time Series–Nonparametric and Parametric Methods. Springer-Verlag, Berlin.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric function forms. *Econometrica* **64**, 865-90.

Granger, C. W. J.,and Joyeux, R. (1980), An Introductionto Long- Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* **1**, 15-29.

Haggan, V. and Ozaki, T. (1981). Modeling Nonlinear Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model. *Biometrika* **68**, 189-196.

Hamilton, J.D. (1994) Time Series Analysis. Princeton University Press.

Hannan, E. J. and Deistler, M. (1988), The Statistical Theory of Linear Systems. NewYork:Wiley.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *The Annals of Statistics* **21**, 1926-1947.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics* **58**, 71-120.

Masry, E. and Fan, J.(1997). Local polynomial estimation for stationary stable processes. *Stochastic Processes and Their Applications* **18**, 1-31.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and its Applications* **9**(1), 141-142.

Nelson, C.R. and G.W. Schwert (1977), Short-Term Interest Rates as Predictors of Inflation: On Testing the Hypothesis that the Real Rate of Interest Is Constant. *American Economic Review* **67**, 478-486.

Ozaki,T.(1982). The Statistical Analysis of Perturbed Limit Cycle Processes Using Nonlinear Time Series Models. *Journal of Time Series Analysis* **3**, 29-41.

Priestley, M.B. (1980). State-dependent Models: A General Appraoch to Nonlinear Time Series Analysis . *Journal of Time Series Analysis*, 1(1), 47–71.

Stock, J.H. and M.W. Watson (2006). Why Has U.S. Inflation Become Harder to Forecast? *NBER Working Paper no. 12324*.

Schwert, G.W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics* **20**, 73-103.

Tiao, G. C. and Tsay,R. S. (1994). Some Advances in Nonlinear and Adaptive Modeling in Time Series. *Journal of Forecasting* **13**, 109-131.

Tong, H. (1983). Threshold autoregression, limit cycles and cyclical data(with discussion). *Journal of the Royal Statistical Society, Series B* **42**, 245-292.

Tong, H. (1990). Nonlinear Time Series: A Dynamical System Approach. Oxford,U.K.: Oxford University Press.

Wang, H.B. (2008). Nonlinear ARMA models with functional MA coefficients. *Journal of Time Series Analysis*, 19(6), 1032–1056.

Watson, G. S. (1964). Smooth regression analysis. *The Indian Journal of Statistics, Series A* **26**, 359-372.

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

M012.pdf

M28.pdf

Export11.pdf

Import12.pdf

PPIco12.pdf

PPIlm12.pdf

Figure 8: Estimates of $\theta(z_t)$ where (a) z_t=ΔM0_{t-12}; (b) z_t=ΔM2_{t-8}; (c) z_t=Ex_{t-11}; (d) z_t=Im_{t-12}; (e) z_t=co_{t-12}; (f) z_t=lm_{t-12}