Hitotsubashi Journal of Economics 46 (2005), pp.159-175.

LAG AUGMENTATION IN REGRESSION MODELS WITH POSSIBLY INTEGRATED REGRESSORS*

TAKU YAMAMOTO

Graduate School Economics, Hitotsubashi University
Kunitachi, Tokyo 186-8601, Japan
yamamoto@econ.hit-u.ac.jp

AND

EIJI KUROZUMI

Graduate School Economics, Hitotsubashi University
Kunitachi, Tokyo 186-8601, Japan
kurozumi@econ.hit-u.ac.jp

Accepted September 2005

Abstract

This paper is concerned with the Wald test statistic of general restriction in dynamic regression models with possibly integrated regressors. We try to improve the size and power of the Wald statistic through the extended lag augmentation (LA) in the regression model and the bias correction of the OLS estimator. This combination of the extended LA approach and the bias correction is called the modified lag augmented (MLA) approach. We investigate the finite sample properties of the MLA estimator. We find that the MLA estimator is superior to the usual LA approach in view of empirical size and power.

Key Words: Regression model; Integrated process; Wald test; Lag augmentation
JEL classifications: C32, C12

I. *Introduction*

Regressions with integrated and/or cointegrated regressors have been widely discussed. The asymptotic distributions of the OLS estimator and of the Wald statistic to test the hypothesis of restrictions on coefficients have been discussed in Phillips and Durlauf (1986), and Park and Phillips (1988, 1989). It has been shown that they do not necessarily have the standard distributions, namely, normal and chi-square ditributions.

There have been several attempts to modified the model and/or statistics so that the Wald

* An initial stage of Yamamoto's research was partially supported by the Ministry of Education, Science, and Culture under Grants-in-Aids No. 1063001.

statistic has an asymptotic chi-squre distribution or can be approximated by it. See, for example, Phillips and Hansen (1990), Park (1992), Phillips (1995), and more recently, Kitamura and Phillips (1997).

In the case of the vector autoregressive (VAR) model, Toda and Yamamoto (1995) proposed estimating it with one or two intentionally augmented lags. More precisely, if we knew the true lag length of the VAR model to be equal to k and the order of integration to be zero or one, we estimate the $(k+1)$-th order VAR model. Then, the Wald statistic to test the hypothesis has an asymptotic chi-square distribution, i.e., the standard statistical inference can be valid, irrespective of the order of integration and the cointegrating rank. This lag augmented (LA) approach is useful in the sense that we do not have to decide the order of integration, but it suffers from inefficiency because of the artificially augmented lag.Simulation experiments by Yamamda and Toda (1998) confirm that the LA approach is less powerful than those based upon the error correction model (ECM) by Johansen (1988,1991) and the fully modified VAR by Phillips (1995).

Recently, Kurozumi and Yamamoto (2000) proposed a bias correction method for the OLS estimator in the LA approach, which reduces its bias related to terms of $O_p(T)$. The bias corrected OLS estimator based on the LA approach has been called the modified LA (MLA) estimator. By finite sample experiments, the MLA approach was shown to be quite effective in reducing the size distortion of the Wald test statistic.

In this paper, we propose a method to improve the power of the Wald test in the MLA approach. Actually, in order to achieve this, we propose an extended lag aumentation in the OLS estimation. The conventional MLA approach suggests intentionally augmenting the $(k+1)$-th lagged variable in the model, when the true model model contains the k-th lagged variable. Here, we propose to intentionally augment the $(k+p)$-th $(p \geq 2)$ lagged variable. Further, we apply the MLA approach to the usual regression models, which includes a VAR model as a special case, whose regressors are possibly non-stationary, not necessarily specialising to a VAR model.

This paper proceeds as follows: In Section 2 we present the model and fundamental assumptions and propose the extended lag augmented (LA(p) $(p \geq 2)$) approach. The asymptotic theory of this approach is obtained through the transformed model that partitions variables into stationary and nonstationary parts. Section 3 investigates the efficiency of the extended LA(p) approach, while section 4 develops the bias correction theory. The whole sample is divided into two parts and the bias corrected estimator, which is called the modified lag augmented (MLA(p)) estimator, is constructed by estimators in three periods, the whole, the first, and the second periods, respectively. Section 5 investigates the finite sample properties of the MLA(p) approach through the Monte Carlo simulation. Section 6 concludes the paper and its main results.

A summary word on notation. We use $vec(A)$ to stack the *rows* of a matrix A into a column vector, [x] to denote the largest integer $\leq x$, and the inequality ">0" to denote positive definiteness when applied to matrices. The symbols "$\xrightarrow{d}$", "$\xrightarrow{p}$", and "$\equiv$" signify convergence in distribution, convergence in probability, and equality in distribution, respectively. We use BM(Ω) to denote a vector Brownian motion with covariance matrix Ω and we write integrals like $\int_0^1 B(s)dB(s)'$ as simply $\int BdB'$ to achieve notational economy. Also, all integrals are from 0 to 1 except where otherwise stated. All limits in the paper are taken as the sample size T tending to ∞.

II. *The Model, Assumptions, and LA(p) Approach*

1. The Basic Model

Consider the n-vector time series $\{y_t\}$ generated by the following model.

$$(1)\qquad \begin{cases} y_t = J_1 x_{t-1} + \cdots + J_k x_{t-k} + u_t, \\ \Delta x_t = C(L) v_t, \end{cases}$$

where $\{x_t\}$ is an m-variate process, and $C(L) = \Sigma_{j=0}^{\infty} C_j L^j$ $(C_0 = I_m)$, and with $\Sigma_{j=0}^{\infty} j \| C_j \| < \infty$. Suppose we know the true lag length k. The basic assumption for $\underline{u}_t = [u_t', v_t']'$ is as follows, but we will impose further restrictions later.

Assumption 1 :

(*i*) $\{\underline{u}_t\}$ *is independently identically distributed with mean zero and covariance matrix* Σ^0.

$$\underline{u}_t \equiv i.i.d.(0, \Sigma^0),$$

where $\Sigma^0 > 0$, *and* $\Sigma^0 = \begin{bmatrix} \Sigma_0 & \Sigma_{01} \\ \Sigma_{10} & \Sigma_1 \end{bmatrix}$.

(*ii*) *Each element of* $\underline{u}_t$ *has a finite* $2+\delta$*-th moment with* $\delta > 0$.

$$E|\underline{u}_{it}|^{2+\delta} < \infty \text{ for some } \delta > 0 \ (i = 1, \cdots, T).$$

We also assume that $\{x_t\}$ is I(0) or I(1) and may be CI(1, 1).

Suppose our interest is in testing the hypothesis of restrictions on the parameters. We formulate the hypothesis as

$$\mathcal{H}_0 : R vec \underline{J} = q,$$

where R is an $g \times kmn$ matrix with rank$(R) = g$ and $\underline{J} = [J_1, \cdots, J_k]$.

2. The LA(*p*) Approach

Here, we present the LA(p) approach. This is a generalization of Toda and Yamamoto (1995) in two respects. Firstly, as a data generating process, we consider a general regression model which includes a VAR(k) process as a special case. Secondly, as a regression model for estimation, we propose the extended lag augmentation. Namely, we intentionally include the $k+p$-th $(p \geq 2)$ lagged variable rather than the $k+1$-th lagged variable, which is denoted as LA(p). We rewrite D.G.P. (1) with the $k+p$-th lagged variable and a constant.

$$\begin{aligned} (2)\qquad y_t &= J_1 x_{t-1} + \cdots + J_k x_{t-k} + J_{k+1} x_{t-k-p} + \mu \cdot 1 + u_t \\ &= \underline{J} x_{1t} + J_{k+1} x_{t-k-p} + \mu \cdot 1 + u_t, \\ &= \underline{J} x_{1t} + [J_{k+1}, \mu] x_{2t}^{(p)} + u_t, \end{aligned}$$

where $J_{k+1} = 0$, $\mu = 0$, $x_{1t} = [x_{t-1}', \cdots, x_{t-k}']'$, and $x_{2t}^{(p)} = [x_{t-k-p}', 1]'$, and in the matrix form,

$$\begin{aligned} Y' &= \underline{J}X_1' + [J_{k+1}, \mu]X_2^{(p)\prime} + U' \\ &= [\underline{J}, J_{k+1}, \mu]X^{(p)\prime} + U', \end{aligned}$$

where $Y' = [y_1, \cdots, y_T]$, $X_1' = [x_{11}, \cdots, x_{1T}]$, $X_2^{(p)\prime} = [x_{21}^{(p)}, \cdots, x_{2T}^{(p)}]$, $X^{(p)\prime} = [X_1, X_2^{(p)}]'$, and $U' = [u_1, \cdots, u_T]$. Though the constant term is superfluous, it will have an important role in bias correction in the later section and thus we include it here. The OLS estimator of $\underline{J}$ is

$$\hat{\underline{J}}^{(p)} = Y'Q_{X2}^{(p)}X_1(X_1'Q_{X2}^{(p)}X_1)^{-1},$$

where $Q_{X2}^{(p)} = I_T - X_2^{(p)}(X_2^{(p)\prime}X_2^{(p)})^{-1}X_2^{(p)\prime}$.

If $\{x_t\}$ is I(0), the OLS estimator of $\underline{J}$ is well knowun to be asymptotically normally distributed, and the standard Wald statistic is asymptotically chi-square distributed. Therefore, we will consider for a while $\{x_t\}$ is I(1) and may be CI(1,1) with the cointegrating rank r.

Let β be the $m \times r$ cointegrating matrix and $\beta_\perp$ be the $m \times (m-r)$ full rank matrix such that $\beta'\beta_\perp = 0$. Then define a $km \times km$ matrix H_1 and a $(k+1)m \times (k+1)m$ matrix H as

$$H_1 = \begin{bmatrix} I_m & -I_m & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I_m \\ 0 & 0 & \cdots & I_m \end{bmatrix}, \quad H = \begin{bmatrix} H_1 & \begin{matrix} 0 \\ \vdots \\ -I_m \end{matrix} \\ 0 \cdots 0 & \begin{bmatrix} \beta' \\ \beta_\perp' \end{bmatrix} \end{bmatrix},$$

and their inverse matrices are given by

$$H_1^{-1} = \begin{bmatrix} I_m & I_m & \cdots & I_m \\ 0 & I_m & \cdots & I_m \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_m \end{bmatrix}, \quad H^{-1} = \begin{bmatrix} H_1^{-1} & e^k \otimes \begin{bmatrix} \beta' \\ \beta_\perp' \end{bmatrix}^{-1} \\ 0 & \begin{bmatrix} \beta' \\ \beta_\perp' \end{bmatrix}^{-1} \end{bmatrix},$$

where $e^k = [1, \cdots, 1]'$ is a $k \times 1$ vector. Using $H^{-1}H = I$, we rewrite model (2) as

$$\begin{aligned} (3) \qquad y_t &= [J_1, \cdots, J_k, J_{k+1}]H^{-1}H \begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-k} \\ x_{t-k-p} \end{bmatrix} + \mu \cdot 1 + u_t \\ &= J_1^*\Delta x_{t-1} + \cdots + J_{k+1}^*\Delta x_{t-k+1} + J_k^*(x_{t-k} - x_{t-k-p}) \\ &\quad + A_1\beta' x_{t-k-p} + A_2\beta_\perp' x_{t-k-p} + \mu \cdot 1 + u_t \\ &= [\underline{J}^*, A_1]z_{1t}^{(p)} + A_2 z_{2t}^{(p)} + \mu \cdot 1 + u_t, \\ &= [\underline{J}^*, A_1]z_{1t}^{(p)} + [A_2, \mu]z_{3t}^{(p)} + u_t, \end{aligned}$$

where $J_j^* = \Sigma_{i=1}^j J_i$ $(j = 1, \cdots, k)$, $\underline{J}^* = \underline{J}H_1^{-1} = [J_1^*, \cdots, J_k^*]$, $[A_1, A_2] = \Sigma_{j=1}^{k+1} J_j[\beta, \beta_\perp]'^{-1}$, A_1 and A_2 are $n \times r$ and $n \times (n-r)$ matrices, respectively, $z_{1t}^{(p)} = [\Delta x_{t-1}', \cdots, \Delta x_{t-k+1}', (x_{t-k} - x_{t-k-p})', (\beta' x_{t-k-p})']'$, $z_{2t}^{(p)} = \beta_\perp' x_{t-k-p}$, and $z_{3t}^{(p)} = [z_{2t}^{(p)\prime}, 1]'$. Let $w_t^{(p)} = (u_t', v_t', z_{1t}^{(p)\prime}, \Delta z_{2t}^{(p)\prime})'$ and define

$$\Sigma^{(p)} = Ew_t^{(p)}w_t^{(p)\prime}, \ \Lambda^{(p)} = \sum_{j=1}^{\infty} Ew_t^{(p)}w_{t+j}^{(p)\prime}, \ \Omega^{(p)} = \Sigma^{(p)} + \Lambda^{(p)} + \Lambda^{(p)\prime}.$$

We partition $\Sigma^{(p)}$, $\Lambda^{(p)}$, and $\Omega^{(p)}$ comformably with $w_t^{(p)}$. For example,

$$\Sigma^{(p)} = \begin{bmatrix} \Sigma_0 & \Sigma_{01} & \Sigma_{02}^{(p)} & \Sigma_{03}^{(p)} \\ \Sigma_{10} & \Sigma_1 & \Sigma_{12}^{(p)} & \Sigma_{13}^{(p)} \\ \Sigma_{20}^{(p)} & \Sigma_{21}^{(p)} & \Sigma_2^{(p)} & \Sigma_{23}^{(p)} \\ \Sigma_{30}^{(p)} & \Sigma_{31}^{(p)} & \Sigma_{32}^{(p)} & \Sigma_3^{(p)} \end{bmatrix}.$$

Then, we have the following lemma.

Lemma 1 :

(*i*) $$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} w_t^{(p)} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t \otimes z_{1t}^{(p)}) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} B_0(s) \\ B_1(s) \\ B_2^{(p)}(s) \\ B_3^{(p)}(s) \\ \xi^{(p)} \end{bmatrix} \begin{matrix} n \\ m \\ km+r \\ m-r \\ (km+r)n \end{matrix} \quad ,$$

where $B(s)^{(p)} = (B_0(s)', B_1(s)', B_2(s)^{(p)\prime}, B_3(s)^{(p)\prime})'$ *is a* $n+(k+2)m$*-vector Brownian motion with covariance matrix* $\Omega(p)$, $\xi^{(p)}$ *is a* $(km+r)n$*-dimensional normal random vector with mean zero and covariance matrix* $\Sigma_0 \otimes \Sigma_2^{(p)}$, *and* $B(s)^{(p)}$ *and* $\xi^{(p)}$ *are independent.*

(*ii*) $\Omega_0 = \Sigma_0$, Σ_1, $\Sigma_2^{(p)}$, *and* $\Omega^{(p)}$ *are positive definite.*

(*iii*) $$\frac{1}{T} \sum_{t=1}^{T} z_{1t}^{(p)} z_{1t}^{(p)\prime} \xrightarrow{p} \Sigma_2^{(p)}.$$

(*iv*) $$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t z_{1t}^{(p)\prime} \xrightarrow{d} N_0^{(p)}, \text{ where } vecN_0^{(p)} = \xi^{(p)}.$$

(*v*) $$\frac{1}{T} \sum_{t=1}^{T} z_{2t}^{(p)} u_t' \xrightarrow{d} \int B_3 dB_0'.$$

(*vi*) $$\frac{1}{T} \sum_{t=1}^{T} z_{2t}^{(p)} z_{1t}^{(p)} \xrightarrow{d} \int B_3 dB_2' + \Sigma_{32}^{(p)} + \Lambda_{32}^{(p)}.$$

(*vii*) $$\frac{1}{T^2} \sum_{t=1}^{T} z_{2t}^{(p)} z_{2t}^{(p)\prime} \xrightarrow{d} \int B_3 B_3'.$$

Proof: The proofs are obtained as a straightforward generalization of Toda and Phillips (1993) and are omitted.

The OLS estimator of $[\underline{J}^*, A_1]$ is

$$\left[\hat{\underline{J}}^{*(p)}, \hat{\mathrm{A}}_1^{(p)}\right] = Y' Q_3^{(p)} Z_1^{(p)} (Z_1^{(p)\prime} Q_3^{(p)} Z_1^{(p)})^{-1}, \tag{4}$$

where $Q_3^{(p)} = I_T - Z_3^{(p)} (Z_3^{(p)\prime} Z_3^{(p)})^{-1} Z_3^{(p)\prime}$, $Z_1^{(p)\prime} = [z_{11}^{(p)}, \cdots, z_{1T}^{(p)}]$, and $Z_3^{(p)\prime} = [z_{31}^{(p)}, \cdots, z_{3T}^{(p)}]$. Though our interest is in $\hat{\underline{J}}^{(p)}$, it is easier to derive the limiting distribution of $\hat{\underline{J}}^{*(p)}$ with $\hat{A}_1^{(p)}$. By Lemma 1, we have

$$\begin{aligned} \sqrt{T}\left[\hat{\underline{J}}^{*(p)} - \underline{J}^*, \hat{A}_1^{(p)} - A_1\right] &= \left(\frac{1}{\sqrt{T}} U' Q_3^{(p)} Z_1^{(p)}\right)\left(\frac{1}{T} Z_1^{(p)\prime} Q_3^{(p)} Z_1^{(p)}\right)^{-1} \\ &\xrightarrow{d} N_0^{(p)} (\Sigma_2^{(p)})^{-1}. \end{aligned}$$

We partition $\Sigma_2^{(p)}$ comformably with $z_{1t}^{(p)}$,

$$E[z_{1t}^{(p)}z_{1t}^{(p)\prime}]=\Sigma_2^{(p)}=\begin{bmatrix}\Sigma_2^{1(p)} & \Sigma_2^{12(p)}\\ \Sigma_2^{21(p)} & \Sigma_2^{2(p)}\end{bmatrix},$$

where $\Sigma_2^{1(p)}$ is a covariance matrix of $[\Delta x'_{y-1}, \cdots, \Delta x'_{t-k+1}, (x_{t-k}-x_{t-k-p})']'$ and $\Sigma_2^{2(p)}$ is that of $\beta' y_{t-k-p}$. The limiting distribution of $\sqrt{T}\ (\underline{\hat{J}}^{*(p)}-\underline{J}^*)$ is the distribution of the first km columns of $N_0^{(p)}(\Sigma_2^{(p)})^{-1}$ and then it is represented as $N_0^{(p)}(\Sigma_2^{(p)})^{-1}S$, where $S=[I_{km}, 0]'$ is a $(km+r)\times km$ matrix. Then,

$$vec(N_0^{(p)}(\Sigma_2^{(p)})^{-1}S)\equiv N(0, \Gamma(p)), \tag{5}$$

where $\Gamma(p)=\Sigma_0\otimes S'(\Sigma_2^{(p)})^{-1}S$, and we can easily check that $S'(\Sigma_2^{(p)})^{-1}S=(\Sigma_{2\cdot 2}^{(p)})^{-1}$ and $\Sigma_{2\cdot 2}^{(p)}=\Sigma_2^{1(p)}-\Sigma_2^{12(p)}(\Sigma_2^{2(p)})^{-1}\Sigma_2^{21(p)}$. Then,

$$\sqrt{T}\left[\underline{\hat{J}}^{*(p)}-\underline{J}^*\right]\xrightarrow{d} N_0^{*(p)}, \tag{6}$$

where $vecN_0^{*(p)}\equiv N(0, \Sigma_0\otimes(\Sigma_{2\cdot 2}^{(p)})^{-1})$. Noting a relation $\underline{\hat{J}}^{*(p)}=\underline{\hat{J}}^{(p)}H_1^{-1}$, and following the argument in Toda and Yamamoto (1995), we can establish the next proposition.

Proposition 1 (The LA(p) Approach):

The Wald statistic to test the hypothesis $\mathcal{H}_0$ has a chi-square distribution with g degrees of freedom.

$$W^{(p)}=\left\{\sqrt{T}\left(R\ vec\underline{\hat{J}}^{(p)}-q\right)\right\}'\left\{R\left(\hat{\Sigma}_0\otimes T(X_1'Q_{X3}^{(p)}X_1)^{-1}\right)R'\right\}^{-1}\left\{\sqrt{T}\left(R\ vec\underline{\hat{J}}^{(p)}-q\right)\right\}\xrightarrow{d}\chi_g^2,$$

where $\hat{\Sigma}_0=\frac{1}{T}\Sigma_{t=1}^{T}\hat{u}_t\hat{u}_t'$ and $\hat{u}_{1t}$*'s are residuals of the regression.*

By this proposition, we can test the hypothesis $\mathcal{H}_0$ without estimating the order of integration and the cointegrating rank in $\{x_t\}$.

III. *Efficiency of the LA(p) Approach*

In this section, we consider the possibility of enhancing the efficiency of the LA(p) estimator when we increase p. In subsection 3.1, we present the case in which the efficiency always increases as p increases. Alternatively, in subsection 3.2, we present two cases where the efficiency of the LA(p) estimator does not necessarily increase with p.

1. Efficient Case

In this subsection, we present a case in which the efficiency of the OLS estimator $\underline{\hat{J}}^{(p)}$ increases with p.

Proposition 2 (Efficiency of the LA(p) Approach):

Let $\{x_t\}$ be a m-variate VMA$(1, h)$ process. Then we have

$$\Gamma(p+1)\leq\Gamma(p)\qquad(p\geq h+1).$$

Proof: By assumption, we do not have cointegration among $\{x_t\}$. Thus, we have $\Sigma_2^{(p)}=\Sigma_2^{1(p)}$ and $S=I_{km}$. It is sufficient to evaluate $\Sigma_2^{(p)}=E\{z_{1t}^{(p)}z_{1t}^{(p)\prime}\}$ against $\Sigma_2^{(p+1)}$, where $z_{1t}^{(p)}=\left[\Delta x'_{t-1}, \cdots, \Delta x'_{t-k+1}, (x_{t-k}-x_{t-k-p})'\right]'$.

We can decompose $z_{1t}^{(p+1)}$ as

$$z_{1t}^{(p+1)}=z_{1t}^{(p)}+M\Delta x_{t-k-p},$$

where $M=[0, I_m]'(mp\times m)$. Then we have

$$\begin{aligned}\Sigma_2^{(p+1)} &= \Sigma_2^{(p)}+E\left\{z_{1t}^{(p)}\Delta x'_{t-k-p}\right\}M'+ME\left\{\Delta x_{t-k-p}z_{1t}^{(p)}\right\}\\ &\quad +ME\left\{\Delta x_{t-k-p}\Delta x'_{t-k-p}\right\}M'.\end{aligned}$$

Since $E(\Delta x_t\Delta x'_{t-i})=0$ when $i\geq h+1$, the first $k-1$ elements of the second and third terms of the right-hand-side vanish. We have

$$\begin{aligned}\Sigma_2^{(p+1)} &= \Sigma_2^{(p)}+M\Big[E\left\{\Delta x_{t-k-p}\Delta x'_{t-k-p}\right\}\\ &\qquad +E\left\{(x_{t-k}-x_{t-k-p})\Delta x'_{t-k-p}\right\}\\ &\qquad +E\left\{\Delta x_{t-k-p}(x_{t-k}-x_{t-k-p})'\right\}\Big]M'\\ &= \Sigma_2^{(p)}+M\left(\sum_{i=-p}^{p}\Gamma_i\right)M',\end{aligned}$$

where $\Gamma_i=E\left(\Delta x_t\Delta x'_{t-i}\right)$ $(m\times m)$. Noting that $\Sigma_{i=-p}^{p}\Gamma_i$ is proportional to the spectral density matrix of $\{x_t\}$ at frequency 0, which is always positive definite. Thus, we have

$$\Sigma_2^{(p+1)}\geq\Sigma_2^{(p)}.$$

Consequently, $\Gamma(p+1)\leq\Gamma(p)$. Q.E.D.

2. Inefficient Case

In this subsection, we present two illustratrative examples which show that the variance of the LA(p) estimator does not necessarily decrease with p.

First, consider the following simplest case where $m=n=p=1$:

$$\begin{aligned} y_t &= \beta_1 x_{t-1}+u_t,\\ \Delta x_t &= \alpha\Delta x_{t-1}+v_t, \qquad |\alpha|<1,\end{aligned}$$

where $\begin{bmatrix} u_t \\ v_t \end{bmatrix} \equiv i.i.d.\ \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$.

Suppose that we estimate the LA(p) model :

$$y_t=\beta_1 x_{t-1}+\beta_2 x_{t-1-p}+u_t \qquad (p\geq 1),$$

where $\beta_2=0$. In this case, we have $z_{1t}^{(p)}=\Sigma_{i=1}^{p}\Delta x_{t-i}$ $(p=1, 2, \cdots)$. Let $\gamma_i\equiv E(\Delta x_t\Delta x_{t-i})=\alpha^i\gamma_0$ $(i=0, 1, 2, \cdots)$. Then, it is easily seen that $\Sigma_2^{(1)}=\gamma_0$, $\Sigma_2^{(2)}=\gamma_0(2+2\alpha)$, $\Sigma_2^{(3)}=\gamma_0(3+4\alpha+2\alpha^2)$, $\Sigma_2^{(4)}=\gamma_0(4+6\alpha+4\alpha^2+2\alpha^3)$, $\Sigma_2^{(5)}=\gamma_0(5+8\alpha+6\alpha^2+4\alpha^3+2\alpha^4)$, and so forth. Suppose that $0<\alpha<1$. Then, $\Sigma_2^{(p)}$ is an increasing function of p. On the other hand, if $-1<\alpha<0$, $\Sigma_2^{(p)}$ fluctuates with p. If, for example, $\alpha=-0.9$, we have $\Sigma_2^{(1)}=\gamma_0$, $\Sigma_2^{(2)}=0.2\gamma_0$, $\Sigma_2^{(3)}=1.02\gamma_0$, $\Sigma_2^{(4)}=0.382\gamma_0$, and $\Sigma_2^{(5)}=$

$1.0562\gamma_0$, and so forth. Thus, in this case, $\Gamma(p)$ fluctuates with p.

Secondly, consider a little more general model as follows:

$$y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t,$$

and the rest of the model specification are the same as before. The model for estimation is given as

$$y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-2-p} + u_t \qquad (p \geq 1),$$

where $\beta_3 = 0$. It is easily derived that

$$\Sigma_2^{(1)} = \gamma_0 \begin{bmatrix} 2+2\alpha & 1+\alpha \\ 1+\alpha & 1 \end{bmatrix},$$

$$\Sigma_2^{(2)} = \gamma_0 \begin{bmatrix} 3+4\alpha+2\alpha^2 & 2+3\alpha+\alpha^2 \\ 2+3\alpha+\alpha^2 & 2+2\alpha \end{bmatrix},$$

$$\Sigma_2^{(3)} = \gamma_0 \begin{bmatrix} 4+6\alpha+4\alpha^2+2\alpha^3 & 3+5\alpha+3\alpha^2+\alpha^3 \\ 3+5\alpha+3\alpha^2+\alpha^3 & 3+4\alpha+2\alpha^2 \end{bmatrix},$$

and so forth. When $\alpha = 0.9$, we have det $(\Sigma_2^{(1)}) = 0.19\gamma_0$, det $(\Sigma_2^{(2)}) = 0.8759$, det $(\Sigma_2^{(3)}) = 3.68$ and increasing with p. When $\alpha = -0.9$, det $(\Sigma_2^{(1)}) = 0.19$, det $(\Sigma_2^{(2)}) = 0.1919$, det $(\Sigma_2^{(3)}) = 0.5656$, and so forth. Thus, in these cases det $(\Sigma_2^{(p)})$ always increases with p. However, we do not find a case where $\Sigma_2^{(p+1)} - \Sigma_2^{(p)}$ is positive semi-definite. Actually, $\Sigma_2^{(p+1)} - \Sigma_2^{(p)}$ is indefinite in most cases. These examples show that the variance of the LA(p) estimator does not necessarily decrease with p.

IV. *The Modified LA(p) Approach*

1. Motivation

We have proposed the LA(p) approach in section 2. and discussed its relative efficiency with variable p in section 3. In this section, we propose modifying the LA(p) approach by correcting a bias in the OLS estimator and constructing the Wald test statistic with more accurate empirical size. The argument in this section closely follows that of Kurozumi and Yamamoto (2000).

2. The Bias Correction

At first we expand the OLS estimator (4) as

$$\text{(7)} \qquad \left[\hat{\underline{J}}^{*(p)}, \hat{A}_1^{(p)}\right] - [\underline{J}^*, A_1] = \frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}} U' Z_1^{(p)}\right)\left(\frac{1}{T} Z_1^{(p)\prime} Z_1^{(p)}\right)^{-1}$$
$$- \frac{1}{T}\left(U' Z_3^{(p)}\right)\left(Z_3^{(p)\prime} Z_3^{(p)}\right)^{-1}\left(Z_3^{(p)\prime} Z_1^{(p)}\right)\left(\Sigma_2^{(p)}\right)^{-1}$$
$$+ o_p(T^{-1}).$$

We wish to take the expectation of the expanded terms of (7) and, according to Yamamoto and Kunitomo (1984), the expectation of the first term can be expressed explicitly up to $O(T^{-1})$. However, the second term include the products of the unit root processes with dependent innovation and it is difficult to derive the explicit expression of its expectation. Alternatively, following Kurozumi and Yamamoto (2000), we approximate the distribution of the second term by its limiting distribution and define the "*quasi-asymptotic bias*" as the expectation of the first term up to $O(T^{-1})$ plus the expectation of the limiting distribution of the second term. We have the following result. The quasi-asymptotic bias, $QBIAS[\underline{\hat{J}}^{*(p)}, \hat{A}_1^{(p)}]$, is expressed as

$$QBIAS[\underline{\hat{J}}^{*(p)}, \hat{A}_1^{(p)}] = -\frac{1}{T}SB^{(p)} - \frac{1}{T}NB^{(p)}, \tag{8}$$

where both $SB^{(p)}$ and $NB^{(p)}$ are finite valued matrices independent of T, they do not depend on the sample size T and then they are constant for any T. See Kurozumi and Yamamoto (1998) for further detail.

Now we construct the modified lag augmented (MLA) estimator, which can eliminate the quasi-asymptotic bias. Suppose we analyze the regression model with a sample size T, which is an even integer, and regress y_t on $x_{t-1}, \cdots, x_{t-k}, x_{t-k-p}$, and 1 for the whole period ($t=1, \cdots, T$).

$$\begin{aligned} Y' &= \underline{\hat{J}}^{(p)}X_1' + \left[\hat{J}_{k+1}^{(p)}, \hat{\mu}^{(p)}\right] X_3' + \hat{U}^{(p)\prime} \\ &= \left[\underline{\hat{J}}^{(p)}, \underline{\hat{J}}_{k+1}^{(p)}, \hat{\mu}^{(p)}\right] X' + \hat{U}^{(p)\prime} \\ &= \left[\underline{\hat{J}}^{*(p)}, \hat{A}_1^{(p)}\right] Z_1^{(p)\prime} + \left[\hat{A}_2^{(p)}, \hat{\mu}^{(p)}\right] Z_3^{(p)\prime} + \hat{U}^{(p)\prime}, \end{aligned}$$

For the first period ($t=1, \cdots, T/2$) and the second period ($t=T/2+1, \cdots, T$), we write, with subscripts f, s, respectively,

$$\begin{aligned} Y_f' &= \underline{\hat{J}}_f^{(p)}X_{1f}' + \left[\hat{J}_{k+1f}^{(p)}, \hat{\mu}_f^{(p)}\right] X_{3f}' + \hat{U}_f^{(p)\prime} \\ &= \left[\underline{\hat{J}}_f^{(p)}, \hat{J}_{k+1f}^{(p)}, \hat{\mu}_f^{(p)}\right] X_f' + \hat{U}_f^{(p)\prime} \\ &= \left[\underline{\hat{J}}_f^{*(p)}, \hat{A}_{1f}^{(p)}\right] Z_{1f}^{(p)\prime} + \left[\hat{A}_{2f}^{(p)}, \hat{\mu}_f^{(p)}\right] Z_{3f}^{(p)\prime} + \hat{U}_f^{(p)\prime}, \end{aligned}$$

$$\begin{aligned} Y_s' &= \underline{\hat{J}}_s^{(p)}X_{1s}' + \left[\hat{J}_{k+1s}^{(p)}, \hat{\mu}_s^{(p)}\right] X_{3s}' + \hat{U}_s^{(p)\prime} \\ &= \left[\underline{\hat{J}}_s^{(p)}, \hat{J}_{k+1s}^{(p)}, \hat{\mu}_s^{(p)}\right] X_s' + \hat{U}_s^{(p)\prime} \\ &= \left[\underline{\hat{J}}_s^{*(p)}, \hat{A}_{1s}^{(p)}\right] Z_{1s}^{(p)\prime} + \left[\hat{A}_{2s}^{(p)}, \hat{\mu}_s^{(p)}\right] Z_{3s}^{(p)\prime} + \hat{U}_s^{(p)\prime}, \end{aligned}$$

where, e.g., $Y_f' = [y_1, \cdots, y_{T/2}]$ and $Y_s' = [y_{T/2+1}, \cdots, y_T]$.

Using (8), we obtain the following results about the quasi-asymptotic bias in each period.

$$QBIAS\left[\underline{\hat{J}}^{*(p)}, \hat{A}_1^{(p)}\right] = -\frac{1}{T}SB^{(p)} - \frac{1}{T}NB^{(p)}, \tag{9}$$

$$QBIAS\left[\underline{\hat{J}}_f^{*(p)}, \hat{A}_{1f}^{(p)}\right] = -\frac{2}{T}SB^{(p)} - \frac{2}{T}NB^{(p)}. \tag{10}$$

$$QBIAS\left[\underline{\hat{J}}_s^{*(p)}, \hat{A}_{1s}^{(p)}\right] = -\frac{2}{T}SB^{(p)} - \frac{2}{T}NB^{(p)}. \tag{11}$$

Using the estimator in each period, we define the modified estimator of $[\underline{J}^*, A_1]$, which we call the MLA(p) estimator, as

$$\left[\hat{\underline{J}}^{*(p)}_{mla}, \hat{A}^{(p)}_{1mla}\right] = 2\left[\hat{\underline{J}}^{*(p)}, \hat{A}^{(p)}_1\right] - \frac{1}{2}\left(\left[\hat{\underline{J}}^{*(p)}_f, \hat{A}^{(p)}_{1f}\right] + \left[\hat{\underline{J}}^{*(p)}_s, \hat{A}^{(p)}_{1s}\right]\right). \tag{12}$$

We can easily check that this estimator has no quasi-asymptotic bias by substituting the hand side of (12) with (9), (10) and (11).

The MLA(p) estimator of $\underline{J}$ is easily obtained through the relation $\hat{\underline{J}}^{(p)} = \hat{\underline{J}}^{*(p)} H_1$.

$$\begin{aligned} \hat{\underline{J}}^{(p)}_{mla} &= \hat{\underline{J}}^{*(p)}_{mla} H_1 \\ &= 2\hat{\underline{J}}^{(p)} - \frac{1}{2}\left(\hat{\underline{J}}^{(p)}_f + \hat{\underline{J}}^{(p)}_s\right). \end{aligned} \tag{13}$$

We can also show that the asymptotic distribution of this estimator is the same as that of the estimator for the LA(p) approach. We summarize the main results.

(i) The MLA(p) estimator (13) has no quasi-asymptotic bias, irrespective of the order of integration of $\{x_t\}$.

(ii) The MLA(p) estimator (13) is asymptotically normally distributed, irrespective of the order of integration of $\{x_t\}$.

We have the following proposition, which is a direct consequence of the above result.

Proposition 3 :

The Wald statistic to test for the hypothesis $\mathcal{H}_0$ constructed from the MLA(p) estimator, $\mathcal{W}^{(p)}_{mla}$, is asymptotically chi-square distributed with g degrees of freedom irrespective of the order of integration of $\{x_t\}$.

$$\begin{aligned} \mathcal{W}^{(p)}_{mla} &= T\left(Rvec\hat{\underline{J}}^{(p)}_{mla} - q\right)' \left\{R\left(\hat{\Sigma}^{(p)}_{mla}\right)R'\right\}^{-1}\left(Rvec\hat{\underline{J}}^{(p)}_{mla} - q\right) \\ &\xrightarrow{d} \chi^2_g, \end{aligned}$$

where

$$\hat{\Sigma}^{(p)}_{mla} = \hat{\Sigma}^{(p)}_0 \otimes \left(\frac{1}{T} X_1' Q^{(p)}_{X2} X_1\right)^{-1} + T\left(\hat{\beta}^{(p)}_{mla} - \hat{\beta}^{(p)}\right)\left(\hat{\beta}^{(p)}_{mla} - \hat{\beta}^{(p)}\right)', \tag{14}$$

$$\hat{\Sigma}^{(p)}_0 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}^{(p)}_t \hat{u}^{(p)\prime}_t,$$

where $\hat{u}^{(p)\prime}_t$ s are residuals of regression for the whole sample, $\hat{\beta}^{(p)}_{mla} = vec\hat{\underline{J}}^{(p)}_{mla}$, and $\hat{\beta}^{(p)} = vec\hat{\underline{J}}^{(p)}$.

The construction of the covariance matrix $\hat{\Sigma}^{(p)}_{mla}$ is explained in what follows: In theory, we can use any consistent estimator of $\Sigma_0 \otimes \Sigma^{-1}_{2.2}$ and have tried several consistent estimators in Monte Carlo simulations. However, the test statistic using $\hat{\Sigma}^{(p)}_{mla}$ has consistently shown the smallest size distortion among them in the small sample. Thus, we decided to adopt $\hat{\Sigma}^{(p)}_{mla}$ as the estimator of $\Sigma_0 \otimes \Sigma^{-1}_{2.2}$. We can justify $\hat{\Sigma}^{(p)}_{mla}$ by the following argument.

We have

$$T\left(\hat{\beta}^{(p)}_{mla} - \beta\right)\left(\hat{\beta}^{(p)}_{mla} - \beta\right)'$$

$$
\begin{aligned}
&= T\left\{\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)+\left(\hat{\beta}^{(p)}-\beta\right)\right\}\left\{\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)+\left(\hat{\beta}^{(p)}-\beta\right)\right\}' \\
&= T\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)'+T\left(\hat{\beta}^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)' \\
&\quad +T\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)\left(\hat{\beta}^{(p)}-\beta\right)'+T\left(\hat{\beta}^{(p)}-\beta\right)\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)',
\end{aligned}
$$

where $\beta = vec\underline{J}$.

The third term in the hand side of the last equality is given by

$$
\begin{aligned}
T\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)\left(\hat{\beta}^{(p)}-\beta\right)' &= T\left\{\left(\hat{\beta}^{(p)}-\beta\right)-\frac{1}{2}\left(\hat{\beta}_f^{(p)}-\beta\right)-\frac{1}{2}\left(\hat{\beta}_s^{(p)}-\beta\right)\right\}\left(\hat{\beta}^{(p)}-\beta\right)' \\
&= T\left(\hat{\beta}^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)'-\frac{T}{2}\left(\hat{\beta}_f^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)' \\
&\quad -\frac{T}{2}\left(\hat{\beta}_s^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)',
\end{aligned}
$$

where $\hat{\beta}_s^{(p)} = vec\hat{\underline{J}}_s^{(p)}$, and $\hat{\beta}_f^{(p)} = vec\hat{\underline{J}}_f^{(p)}$. It is easily seen that

$$
\begin{aligned}
&T\left(\hat{\beta}^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)' \xrightarrow{p} \Sigma_0 \otimes \Sigma_{2.2}^{-1}, \\
&\frac{T}{2}\left(\hat{\beta}_f^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)' \xrightarrow{p} \frac{1}{2}\Sigma_0 \otimes \Sigma_{2.2}^{-1}, \quad \text{and} \\
&\frac{T}{2}\left(\hat{\beta}_s^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)' \xrightarrow{p} \frac{1}{2}\Sigma_0 \otimes \Sigma_{2.2}^{-1}.
\end{aligned}
$$

Thus, we have

$$
T\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)\left(\hat{\beta}^{(p)}-\beta\right)' \xrightarrow{p} 0.
$$

The fourth term is the transpose of the third and also vanishes. Then,

$$
T\left(\hat{\beta}_{mla}^{(p)}-\beta\right)\left(\hat{\beta}_{mla}^{(p)}-\beta\right)' = T\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)'+T\left(\hat{\beta}^{(p)}-\beta\right)\left(\hat{\beta}^{(p)}-\beta\right)'+o_p(1).
$$

We may construct $\hat{\Sigma}_{mla}^{(p)}$ as

$$
\hat{\Sigma}_{mla}^{(p)} = T\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)\left(\hat{\beta}_{mla}^{(p)}-\hat{\beta}^{(p)}\right)'+\hat{\Sigma}_0^{(p)} \otimes \left(\frac{1}{T}X_1'Q_{X2}^{(p)}X_1\right)^{-1}.
$$

V. *Simulation Experiment*

1. Experimental Design

In the following simulations, we assume an univariate process $\{y_t\}$ generated by

$$
\begin{aligned}
&y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t, \text{ and} \\
&x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + v_t + \theta_1 v_{t-1},
\end{aligned}
\tag{15}
$$

where $\begin{bmatrix} u_t \\ v_t \end{bmatrix} \equiv i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$.

The various specific values of α_1, α_2, β_1, β_2, θ_1, σ_u, σ_v and ρ will be given in each experiment.

The regression models for the LA(p) approach and MLA(p) approach are given by

$$
\begin{aligned}
y_t &= \hat{\beta}_0^{(p)} + \hat{\beta}_1^{(p)} x_{t-1} + \hat{\beta}_2^{(p)} x_{t-2} + \hat{\beta}_3^{(p)} x_{t-2-p} + \hat{u}_t^{(p)}, \\
y_t &= \hat{\beta}_{0f}^{(p)} + \hat{\beta}_{1f}^{(p)} x_{t-1} + \hat{\beta}_{2f}^{(p)} x_{t-2} + \hat{\beta}_{3f}^{(p)} x_{t-2-p} + \hat{u}_{ft}^{(p)}, \qquad (16) \\
y_t &= \hat{\beta}_{0s}^{(p)} + \hat{\beta}_{1s}^{(p)} x_{t-1} + \hat{\beta}_{2s}^{(p)} x_{t-2} + \hat{\beta}_{3s}^{(p)} x_{t-2-p} + \hat{u}_{st}^{(p)}.
\end{aligned}
$$

The MLA(p) estimator is given by

$$\hat{\beta}_{mla}^{(p)} = 2\hat{\beta}^{(p)} - \frac{1}{2}\left(\hat{\beta}_f^{(p)} + \hat{\beta}_s^{(p)}\right),$$

where $\hat{\beta}^{(p)} = \left[\hat{\beta}_1^{(p)}, \hat{\beta}_2^{(p)}\right]'$, $\hat{\beta}_f^{(p)} = \left[\hat{\beta}_{1f}^{(p)}, \hat{\beta}_{2f}^{(p)}\right]'$ and $\hat{\beta}_s^{(p)} = \left[\hat{\beta}_{1s}^{(p)}, \hat{\beta}_{2s}^{(p)}\right]'$.
The estimator of its covariance matrix is

$$\hat{\Sigma}_{mla}^{(p)} = \hat{\sigma}_u^2 \left(\frac{1}{T} X_1' Q_{X2}^{(p)} X_1\right)^{-1} + T\left(\hat{\beta}_{mla}^{(p)} - \hat{\beta}^{(p)}\right)\left(\hat{\beta}_{mla}^{(p)} - \hat{\beta}^{(p)}\right)'. \qquad (17)$$

The null hypothesis to test is given by

$$\mathcal{H}_0 : \beta = \beta_0,$$

where $\beta = [\beta_1, \beta_2]'$ and $\beta_0 = [\beta_{10}, \beta_{20}]'$. Then, the test statistic is given by

$$W_{mla}^{(p)} = \left\{\sqrt{T}\left(\hat{\beta}_{mla}^{(p)} - \beta_0\right)'\right\} \left\{\hat{\Sigma}_{mla}^{(p)}\right\}^{-1} \left\{\sqrt{T}\left(\hat{\beta}_{mla}^{(p)} - \beta_0\right)\right\}, \qquad (18)$$

and $W_{mla}^{(p)}$ is asymptotically chi-square distributed with two degrees of freedom.

2. Simulation Results

In the following simulation experiments, the level of significance is set equal to 5% or 10%. The sample size is set to 100, and the number of replications is 5,000 in all experiments. Computations are performed by the GAUSS matrix programming language.

The Bias and Bias Correction When $p=1$

Tables 1a-1b show the bias of LA(1) and MLA(1) estimators when $T=100$. The detailed model specification is given immediately below the header of the table. The generating process of the $\{x_t\}$ is assumed to be a random walk process. Table 1a shows the bias of the LA(1) and MLA(1) estimators for various values of ρ when $\sigma_u = \sigma_v = 1$. It indicates that the finite sample bias is larger when ρ is larger in absolute value. It also shows that the MLA(1) approach has a uniformly smaller bias than the LA(1) approach, indicating that the bias correction method described in section 4 is quite effective.

Table 1b shows the bias of the LA(1) and MLA(1) estimators for various combinations of σ_u and σ_v when $\rho = 1$. It shows that the bias is larger in absolute value when the the ratio σ_u/σ_v is larger. Again, it shows that the MLA approach has a uniformly smaller bias than the LA approach, indicating that the bias correction method described in section 4 is quite effective in reducing the bias in the LA estimator.

TABLE 1. BIAS OF LA(1) AND MLA(1) ($T=100$)

DGP: $y_t=\beta_1 x_{t-1}+\beta_2 x_{t-2}+u_t$, and $x_t=\alpha_1 x_{t-1}+\alpha_2 x_{t-2}+v_t+\theta_1 v_{t-1}$,
where $\alpha_1=1.0$, $\alpha_2=0.0$, $\theta_1=0.0$, $\beta_1=0.7$ and $\beta_2=0.3$, and

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \text{ i.i.d. N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right).$$

Regression Model: $y_t=\hat{\beta}_0+\hat{\beta}_1 x_{t-1}+\hat{\beta}_2 x_{t-2}+\hat{\beta}_3 x_{t-3}+\hat{u}_t$.

TABLE 1a. CASE OF $\sigma_u=\sigma_v=1$

	LA(1)		MLA(1)	
	$\hat{\beta}_1-\beta_1$	$\hat{\beta}_2-\beta_2$	$\hat{\beta}_1-\beta_1$	$\hat{\beta}_2-\beta_2$
$\rho=-0.9$	0.040	0.008	−0.000	−0.002
−0.6	0.028	0.003	0.001	−0.004
−0.3	0.016	−0.000	0.003	−0.004
0.0	0.003	−0.004	0.003	−0.004
0.3	−0.010	−0.007	0.004	−0.004
0.6	−0.024	−0.010	0.004	−0.003
0.9	−0.038	−0.012	0.003	−0.002

TABLE 1b. CASE OF $\rho=0.9$

		LA(1)		MLA(1)	
		$\hat{\beta}_1-\beta_1$	$\hat{\beta}_2-\beta_2$	$\hat{\beta}_1-\beta_1$	$\hat{\beta}_2-\beta_2$
$\sigma_u=10$	$\sigma_v=1$	−0.376	−0.115	0.032	−0.015
$\sigma_u=5$	$\sigma_v=1$	−0.188	−0.058	0.016	−0.008
$\sigma_u=1$	$\sigma_v=1$	−0.038	−0.012	0.003	−0.002
$\sigma_u=1$	$\sigma_v=5$	−0.008	−0.002	−0.001	−0.000
$\sigma_u=1$	$\sigma_v=10$	−0.004	−0.001	0.000	−0.000

The Effect of Bias Correction on the Empirical Size When $p=1$

Tables 2a-2b show the effect of the bias correction on the empirical size when $p=1$. For the illustrative purposes, the simulation model considered here is exactly the same as in Tables 1a-1b. These tables show that the empirical size of the MLA approach is much closer to the corresponding nominal size than the LA approach. This indicates that the bias correction described in section 4 is also effective in reducing the size distortion in the LA approach.

The Effect of Bias Correction on the Empirical Size When $p\geq 2$

Tables 3a-3c show the effect of the bias correction method in reducing the size distortion in the the extended lag augamented LA(p) ($p\geq 2$) approach. These three tables differ only in the generation process of the $\{x_t\}$. These tables show that the size distortion of the LA(p) approach increases with the order of the extended lag p, and this size distortion is effectively eliminated in the MLA(p) approach.

Table 2. Empirical Size of LA(1) and MLA(1) (T=100)

DGP: $y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$, and $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + v_t + \theta_1 v_{t-1}$,
where $\alpha_1 = 1.0$, $\alpha_2 = 0.0$, $\theta_1 = 0.0$, $\beta_1 = 0.7$ and $\beta_2 = 0.3$, and

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \text{ i.i.d. } \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \right).$$

Regression Model: $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t-1} + \hat{\beta}_2 x_{t-2} + \hat{\beta}_3 x_{t-3} + \hat{u}_t$.

Hypothesis: $\begin{cases} H_0: & \beta_1 = 0.7 \text{ and } \beta_2 = 0.3. \\ H_1: & \text{otherwise.} \end{cases}$

Table 2a. Case of $\sigma_u = \sigma_v = 1$

	LA(1)		MLA(1)	
Nominal Size	5%	10%	5%	10%
ρ=−0.9	7.8	13.9	5.3	10.3
−0.6	6.5	12.1	5.5	10.5
−0.3	6.2	11.2	5.6	10.6
0.0	5.8	10.8	5.6	10.7
0.3	6.1	11.1	5.7	10.2
0.6	6.5	11.6	5.4	10.4
0.9	7.6	13.6	5.3	9.7

Table 2b. Case of ρ=0.9

		LA(1)		MLA(1)	
		5%	10%	5%	10%
σ_u=10	σ_v= 1	7.6	13.6	5.3	9.7
σ_u= 5	σ_v= 1	7.6	13.6	5.3	9.7
σ_u= 1	σ_v= 1	7.6	13.6	5.3	9.7
σ_u= 1	σ_v= 5	7.6	13.6	5.3	9.7
σ_u= 1	σ_v=10	7.6	13.6	5.3	9.7

The Size Adjusted Power of LA(1) and MLA(p)

Tables 4a-4c show the size adjusted power of LA(1) and MLA(p) for various values of p. The simulation models are exactly the same as in Tables 3a-3c. While the pattern of the empirical power of MLA(p) differs for each model, we can find a general pattern. That is, the empirical power of MLA(p) generally increases with p, and then starts to decrease in some cases. It means that by resorting to the MLA(p) approach, we can generally obtain a statistic that is more powerful than the MLA(1) approach.

VI. *Conclusion*

This paper has developed an asymptotic theory for extended lag augmentation for the regression model whose regressors are possibly non-stationary. Theoretically, the extended lag augmentation may or may not improve the efficiency of the estimator. However, the Monte

TABLE 3. EMPIRICAL SIZE OF LA(p) AND MLA(p) (T=100)

DGP: $y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$, and $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + v_t + \theta_1 v_{t-1}$,

where $\beta_2 = 0.3$, and $\begin{pmatrix} u_t \\ v_t \end{pmatrix}$ i.i.d. $\mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}\right)$.

Regression Model: $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t-1} + \hat{\beta}_2 x_{t-2} + \hat{\beta}_3 x_{t-2-p} + \hat{u}_t$.

Hypothesis: $\begin{cases} H_0: & \beta_1 = 0.7 \text{ and } \beta_2 = 0.3. \\ H_1: & \text{otherwise.} \end{cases}$

TABLE 3a. CASE OF $\alpha_1 = 1.0$, $\alpha_2 = 0.0$, and $\theta_1 = 0.0$

	LA(p)		MLA(p)	
Nominal Size	5%	10%	5%	10%
p=1	7.6	13.6	5.3	9.7
2	9.9	15.9	5.8	10.7
3	10.8	18.0	5.5	10.4
5	13.1	20.8	5.8	10.5
8	15.5	29.4	6.2	10.6

TABLE 3b. CASE OF $\alpha_1 = 1.8$, $\alpha_2 = -0.8$, and $\theta_1 = 0.0$

	LA(p)		MLA(p)	
Nominal Size	5%	10%	5%	10%
p=1	9.9	17.2	5.3	10.1
2	11.5	18.7	5.5	9.6
3	12.7	20.4	5.3	9.3
5	15.0	23.4	5.5	9.4
8	18.1	27.8	6.0	10.6

TABLE 3c. CASE OF $\alpha_1 = 0.2$, $\alpha_2 = 0.8$, and $\theta_1 = 0.0$

	LA(p)		MLA(p)	
Nominal Size	5%	10%	5%	10%
p=1	10.4	17.3	5.7	9.8
2	10.8	17.9	6.8	11.3
3	11.7	19.8	5.5	10.3
5	13.6	21.7	6.2	10.7
8	15.3	24.8	6.0	10.8

Carlo simulation revealed that the extended lag augmenttation generally improves the efficiency of the estimator. In particular, when coupled with the bias correction method, the MLA(p) approach is superior to the conventional LA(1) approach in terms of empirical size. Further, the MLA(p) ($p \geqq 2$) estimator is superior to the MLA(1) in terms fo power. In other words, the MLA(p) is quite useful in small samples (say, for example, T=100). However, the optimal choice of the order p in the extended lag augmentation is not easy to find and will be left for the future research.

TABLE 4. SIZE ADJUSTED EMPIRICAL POWER OF LA(1) AND MLA(p) (T=100)

DGP: $y_t=\beta_1 x_{t-1}+\beta_2 x_{t-2}+u_t$, and $x_t=\alpha_1 x_{t-1}+\alpha_2 x_{t-2}+v_t+\theta_1 v_{t-1}$,

where $\beta_2=0.3$, and $\begin{pmatrix} u_t \\ v_t \end{pmatrix}$ i.i.d. $\mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}\right)$.

Regression Model: $y_t=\hat{\beta}_0+\hat{\beta}_1 x_{t-1}+\hat{\beta}_2 x_{t-2}+\hat{\beta}_3 x_{t-2-p}+\hat{u}_t$.

Hypothesis: $\begin{cases} H_0: & \beta_1=0.7 \text{ and } \beta_2=0.3. \\ H_1: & \text{otherwise.} \end{cases}$

TABLE 4a. CASE OF $\alpha_1=1.0$, $\alpha_2=0.0$, and $\theta_1=0.0$ (Significance Level=5%)

β_1 in DGP	0.5	0.6	0.65	0.7	0.75	0.8	0.9
LA(1)	77.2	28.3	12.2	5.0	3.3	7.1	38.1
MLA(1)	52.1	18.8	9.1	5.0	5.1	11.5	47.7
MLA(2)	59.6	21.9	10.3	5.0	5.2	12.7	55.2
MLA(3)	62.0	25.4	11.7	5.0	5.6	14.6	60.6
MLA(5)	60.5	27.6	12.5	5.0	4.2	14.4	60.1
MLA(8)	55.7	28.6	13.5	5.0	3.9	12.6	56.4

TABLE 4b. CASE OF $\alpha_1=1.8$, $\alpha_2=-0.8$, and $\theta_1=0.0$ (Significance Level=5%)

β_1 in DGP	0.5	0.6	0.65	0.7	0.75	0.8	0.9
LA(1)	100.0	89.4	31.2	5.0	8.3	39.4	92.1
MLA(1)	84.2	49.6	19.1	5.0	11.7	42.3	93.0
MLA(2)	92.2	64.9	28.8	5.0	19.4	66.6	99.1
MLA(3)	94.9	72.9	38.5	5.0	29.0	81.7	99.8
MLA(5)	95.9	78.3	47.4	5.0	41.3	92.6	100.0
MLA(8)	95.9	77.5	49.2	5.0	49.9	93.8	99.9

TABLE 4c. CASE OF $\alpha_1=0.2$, $\alpha_2=0.8$, and $\theta_1=0.0$ (Significance Level=5%)

β_1 in DGP	0.5	0.6	0.65	0.7	0.75	0.8	0.9
LA(1)	45.1	15.6	8.9	5.0	3.7	4.6	17.5
MLA(1)	37.5	14.0	7.4	5.0	4.8	8.1	28.3
MLA(2)	55.4	23.1	10.5	5.0	4.1	9.2	47.6
MLA(3)	51.8	19.8	9.5	5.0	4.7	9.3	39.7
MLA(5)	54.3	21.3	10.5	5.0	3.9	8.4	41.7
MLA(8)	55.1	24.8	12.1	5.0	4.2	8.3	43.7

REFERENCES

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, pp. 231-254.

Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, pp. 1551-1580.

Kitamura, Y., and P. C. B. Phillips (1997): "Fully Modified IV, GIVE, and GMM Estimation with Possibly Non-stationary Regressors and Instrumentals," *Journal of Econometrics*, 80, pp. 85-123.

Kurozumi, E., and T. Yamamoto (2000): "Modified Lag Augmented Vector Autoregressions," *Econometric Reviews*, 19, pp. 207-231.

Park, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, pp. 119-143.

Park, J. Y., and P. C. B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, pp. 468-497.

Park, J. Y., and P. C. B. Phillips (1989): "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, pp. 95-131.

Phillips, P. C. B. (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, pp. 1023-1078.

Phillips, P. C. B., and S. N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, pp. 473-495.

Phillips, P. C. B., and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables regression with I(1) processes," *Review of Economic Studies*, 57, pp. 99-125.

Toda, H. Y., and T. Yamamoto (1995): "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes," *Journal of Econometrics*, 66, pp. 225-250.

Yamada, H., and H. Y. Toda (1998): "Inference in Possibly Integrated Vector Autoregressive Models: Finite Sample Evidence," *Journal of Econometrics*, 86, pp. 55-95.

Yamamoto, T., and N. Kunitomo (1984): "Asymptotic Bias of the Least Square Estimator for Multivariate Autoregressive Models," *Annals of the Institute of Statistical Mathematics*, 36, pp. 419-430.