Variable Lag Augmentation in Regression Models with Possibly Integrated Regressors: Some Experimental Results

Taku Yamamoto Eiji Kurozumi

Abstract

This paper is concerned with the Wald test statistic of general restrictions in dynamic regression models with possiobly 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 instrumental variable (IV) estimator. It has been known that the extended lag augmentation is generally, but not always, useful in increasing the finite sample power of the Wald statistic. In this papper we propose a new approach, called the variable lag augmentation approach, which selects an appropriate lag length. The finite sample experiments show that the proposed approach produces higher power of the test than the conventional LA estimator.

1. 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 don not necessarily have the standard asymptotic distribution, namely, the normal or the chi-square ditribution.There have been several attempts to modified the model and/or statistics so that the Wald statistic has a chi-squre distribution or can be approximated by a chi-square distribution. See, for example, Phillips and Hansen (1990), Park (1992), Phillips (1995), and more recently, Kitamura and Phillips (1997). It should be noted that all forementioned approaches are based upon the correct model specification.

In the case of the vector autoregressive (VAR) model, Toda and Yamamoto (1995) proposed to estimate the model with an intentionally augmented lag. More precisely, if we know that the true lag length of the VAR model is equal to k and the order of integration is either zero or one, we intentionally estimate the $(k+1)$-th order VAR model. We call it as the *ordinary* lag augmented (LA) approach in this paper. Then, the Wald statistic to test the hypothesis has an asymptotic chi-square distribution. That is, the standard statistical inference can be valid irrespective of the order of integration or the cointegrating rank. This ordinary LA approach is useful in the sense that we do not have to decide the order of integration or the cointegration rank before testing the hypothesis. However, it has been known that it suffers from inefficiency because of the artificially augmented lagged variable. Further, it also suffers from the size ditortion of the test in finite samples.

There have been a few attempts to ovecome the above mentioned drawbacks of the ordinary LA approach. Kurozumi and Yamamoto (2000) gave a procedure in order to reduce the size distortion. Specifically, they proposed a bias correction method for the OLS estimator in the ordinary LA approach, which reduces its bias related to terms of $O_p(T^{-1})$, where T is the sample size. The bias corrected OLS estimator based on the LA approach has been called the *modified* lag augmented (MLA) estimator. By finite sample experiments, it has been shown that the MLA approach is quite effective in reducing the size distortion of the Wald test statistic.

Yamamoto and Kurozumi (2005) extended the work of Kurozumi and Yamamoto (2000) in two directions.

First, the model is generalized to a usual regression model whose regressors are possibly non-stationary, which includes a VAR model as a special case. However, note that in their paper the regressors are confined to be lagged variables and comtempraneously uncorrelated with the error term. Thus, the OLS estimation can be used for estimating coefficient parameters. Second, in order to improve the power of the Wald test, the *extended* MLA approach (denoted as MLA(p)) with $p \geq 2$ is proposed. Note that the ordinary MLA approach suggests to intentionally augment the $(k+1)$-th lagged variable to the model, when the true model contains the k-th lagged variable. Here, the MLA(p) approach intentionally augments the $(k+p)$-th ($p \geq 2$) lagged variable. Oboviously, when $p=1$, the MLA(p) approach reduces to the ordinary MLA approach. They showed that the MLA(p) approach generally improves the finite sample performance of the power of the Wald test without affecting that of the size.

This paper is a sequel of Yamamoto and Kurozumi (2005) and genrralize in two directions. First, in this paper we consider a model where regressors and the error term are contemporaneously correlated. Thus, the instrumental variable (Ⅳ) estimator instead of the OLS estimator is called for. Second, we propose a practical method to choose a suitable lag length p in the MLA(p) approach. Yamamoto and Kurozumi (2005) showed that the MLA(p) ($p \geq 2$) approach generally, but not always, improves the finite sample power of the Wald test in comparison with the ordinary MLA approach. But it has not been clear which lag length p is approppriate for a certain model and data. In this paper we propose the *variable* modified lag augmented (VMLA) approach which selects a suitable lag length p for each model and data. The term "variable" comes from the fact that the chosen lag length p varies depending upon each model and data. The experiments in section 4 below show that the VMLA possess always higher power of the test than the ordinary MLA approach.

This paper proceeds as follows: In Section 2 presents the model and fundamental assumptions, and propose the extended lag augmented (LA(p) ($p \geq 2$)) approach based upon the Ⅳ estimation. The asymptotic theory of this approach is obtained through the transformed model that partitions variables into stationary parts and nonstationary parts. Section 3 explains the modification based upon the bias correction method. The whole sample is divided into two parts and the bias corrected estimator, which is called the modified extended lag augmented (MLA(p)) estimator, is constructed by estimators in three periods, the whole, the first, and the second periods. Further, an approach to select an appropriate lag length is proposed. In conjunction with the above mentioned modification, it is called the variable modified LA (VMLA) approach. Section 4 gives experimental results of the MLA(p) and the VMLA approaches through Monte Carlo simulations. Section 5 concludes the paper.

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 definite 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, and all integrals are from 0 to 1 except where otherwise noted. All limits in the paper are taken as the sample size T tends to ∞.

2. THE MODEL, ASSUMPTIONS, AND EXTENDED LA(p) APPROACH}

2.1. THE BASIC MODEL

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

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

where $\{w_t\}$ is an m-variate process, and $C(L) = \Sigma_{j=0}^{\infty} C_j L^j$ $(C_0 = I_m)$, and with $\Sigma_{j=0}^{\infty} j \| C_j \| < \infty$. The model is a generalization of Yamamoto and Kurozumi (2005) in the sense it now includes a contemporaneous regressor w_t in addition to lagged ones. Suppose we know the true lag length k. The basic assumption for $\underline{\varepsilon}_t=[\varepsilon_t', v_t']'$ is as follows, although we will impose further restrictions later.

Assumption 1 :

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

$$\underline{\varepsilon}_t \equiv i.i.d.(0, \Sigma^0), \quad where \ \Sigma^0 > 0,$$

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

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

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

We further assume that $\{w_t\}$ 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

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

where R is a $g\times(k+1)\,n^2$ matrix with rank $(R)=g$, q is a $g\times 1$ vector, and $\underline{J}=[J_1,\cdots,J_k]$.

2.2. THE EXTENDED LA APPROACH

Here, we present the extended LA approach. Following Yamamoto and Kurozumi (2005), we consider the extended lag augmentation for a regression model for estimation. Namely, we intentionally include $(k+p)$-th $(p\geq 2)$ lagged variable, which is denoted as LA(p), rather than the $(k+1)$-th lagged variable, which is denoted as LA(1). We rewrite D.G.P. (1) with the $(k+p)$-th lagged variable and a constant:

$$(2) \qquad y_t=J_0w_t+\cdots+J_kw_{t-k}+J_{k+1}w_{t-k-p}+\mu\cdot 1+\varepsilon_t$$
$$=[\underline{J}, J_{k+1}, \mu]\,w_t^{(p)}+\varepsilon_t,$$

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

$$Y'=[\underline{J}, J_{k+1}, \mu]\,W^{(p)\prime}+E',$$

where $Y'=[y_1,\cdots,y_T]$, $W^{(p)\prime}=[w_1^{(p)},\cdots, w_T^{(p)}]$, $E'=[\varepsilon_1, \cdots, \varepsilon_T]$. The superscript ($p$) siginifies that $(k+p)$-th lagged variable is augmented to the original model. Let $W_{IV}^{(p)}$ be the appropriate instrumental variables for $W^{(p)}$ such that

$$\frac{1}{T^2}W_{IV}^{(p)\prime}W_{IV}^{(p)} \ \xrightarrow{p}\ \text{positive definite},$$
$$\frac{1}{T^2}W_{IV}^{(p)\prime}W^{(p)} \ \xrightarrow{d}\ O_p(1), \text{ and}$$
$$\frac{1}{T^2}W_{IV}^{(p)\prime}E \ \xrightarrow{p}\ 0.$$

The predictive value of $W^{(p)}$, denoted as $X^{(p)}$ is given by

$$X^{(p)}\equiv[x_1^{(p)},\cdots, x_T^{(p)}]'=W^{(p)}Y'W_{IV}^{(p)}[W_{IV}^{(p)\prime}W_{IV}^{(p)}]^{-1}.$$

Based on the predicted value $X^{(p)}$, the model is rewritten as

$$(3) \qquad y_t=J_0x_t+\cdots+J_kx_{t-k}+J_{k+1}x_{t-k-p}+\mu\cdot 1+u_t$$
$$=\underline{J}x_{1t}+J_{k+1}x_{t-k-p}+\mu\cdot 1+u_t,$$
$$=\underline{J}x_{1t}+[J_{k+1}, \mu]\,x_{2t}+\mu_t,$$

where $u_t=\varepsilon_t+[\underline{J}, J_{k+1}, \mu](w_t^{(p)}-x_t^{(p)})$, $x_t^{(p)}=[x_{1t}', x_{2t}^{(p)\prime}]'$, $x_{1t}=[x_t',\cdots, x_{t-k}']'$, and $x_{2t}^{(p)}=[x_{t-k-p}', 1]'$, and in the matrix form,

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

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

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

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

If $\{w_t\}$ is I(0), it is well known that the Ⅳ estimator of $\underline{J}$ is asymptotically normally distributed and the standard Wald statistic is asymptotically chi-square distributed. Therefore, we will consider for a while $\{w_t\}$ is I(1) and may be CI(1,1) with the cointegration 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 $(k+1)\,m\times(k+1)\,m$ matrix H_1 and a $(k+2)\,m\times(k+2)\,m$ matrix H as

$$H_1 = \begin{bmatrix} I_m & -I_m & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I_m \\ 0 & 0 & \cdots & I_m \end{bmatrix},$$

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

and their inverse matrices are given by

$$H_1^{-1}=\begin{bmatrix} I_m & I_m & \cdots & I_m \\ 0 & I_m & \cdots & I_m \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_m \end{bmatrix},$$

$$H^{-1}=\begin{bmatrix} H_1^{-1} & e^k\otimes\begin{bmatrix}\beta' \\ \beta'_\perp\end{bmatrix}^{-1} \\ 0 & \begin{bmatrix}\beta' \\ \beta'_\perp\end{bmatrix}^{-1}\end{bmatrix},$$

where $e^k=[1,\cdots,1]'$ is a $(k+1)\times 1$ vector. Using $H^{-1}H=I$, we rewrite model (3) as

$$(4)\quad y_t=[J_0,\cdots,J_k,J_{k+1}]\,H^{-1}H\begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-k} \\ x_{t-k-p}\end{bmatrix}+\mu\cdot 1+u_t$$

$$=J_0^*\Delta x_{t-1}+\cdots+J_{k-1}^*\Delta x_{t-k+1}+J_k^*(x_{t-k}-x_{t-k-p})$$
$$+A_1\beta' x_{t-k-p}+A_2\beta'_\perp x_{t-k-p}+\mu\cdot 1+u_t$$
$$=[\underline{J}^*,A_1]\,z_{1t}^{(p)}+A_2 z_{2t}^{(p)}+\mu\cdot 1+u_t,$$
$$=[\underline{J}^*,A_1]\,z_{1t}^{(p)}+[A_2,\ \mu]\,z_{3t}^{(p)}+u_t,$$

where $J_j^*=\sum_{i=0}^{j}J_i\,(j=0,\cdots,k)$, $\underline{J}^*=\underline{J}H_1^{-1}=[J_0^*,\cdots,J_k^*]$, $[A_1,A_2]=\sum_{j=0}^{k+1}J_i[\beta,\beta_\perp]'^{-1}$, A_1 and A_2 are $n\times r$ and $n\times(n-r)$ matrices, respectively, $z_{1t}^{(p)}=[\Delta x'_t,\cdots,\Delta x'_{t-k+1},(x_{t-k}-x_{t-k-p})',(\beta' x_{t-k-p})']'$, $z_{2t}^{(p)}=\beta'_\perp x_{t-k-p}$, and $z_{3t}^{(p)}=[z_{2t}^{(p)\prime},1]'$. Let $\eta_t^{(p)}=(u'_t,v'_t,z_{1t}^{(p)\prime},\Delta z_{2t}^{(p)\prime})'$ and define

$$\Sigma^{(p)}=E\eta_t^{(p)}\eta_t^{(p)\prime},\ \Lambda^{(p)}=\sum_{j=1}^{\infty}E\eta_t^{(p)}\eta_{t+j}^{(p)\prime},$$

$$\Omega^{(p)}=\Sigma^{(p)}+\Lambda^{(p)}+\Lambda^{(p)\prime}.$$

We partition $\Sigma^{(p)},\Lambda^{(p)}$, and $\Omega^{(p)}$ comformably with $\eta_t^{(p)}$. 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)
$$\begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\eta_t^{(p)} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t\otimes z_{1t}^{(p)})\end{bmatrix} \xrightarrow{d} \begin{bmatrix} B_0(s) \\ B_1(s) \\ B_2^{(p)}(s) \\ B_3^{(p)}(s) \\ \xi^{(p)}\end{bmatrix} \begin{matrix} n \\ m \\ (k+1)m+r \\ m-r \\ ((k+1)m+r)n\end{matrix},$$

where $B(s)^{(p)}=(B_0(s)',B_1(s)',B_2(s)^{(p)\prime},B_3(s)^{(p)\prime})'$ is a $n+(k+3)m$-vector Brownian motion with covariance matrix $\Omega(p)$, $\xi^{(p)}$ is a $((k+1)m+r)n$-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) $\Omega_0=\Sigma_0,\ \Sigma_1,\ \Sigma_2^{(p)}$, *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) $$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t z_{1t}^{(p)\prime}\xrightarrow{d}N_0^{(p)},$$

where $vecN_0^{(p)}=\xi^{(p)}$.

(v) $$\frac{1}{T}\sum_{t=1}^{T}z_{2t}^{(p)}u'_t\xrightarrow{d}\int B_3dB'_0.$$

(vi) $$\frac{1}{T}\sum_{t=1}^{T}z_{2t}^{(p)}z_{1t}^{(p)\prime}\xrightarrow{d}\int B_3dB'_2+\Sigma_{32}^{(p)}+\Lambda_{32}^{(p)}.$$

(vii) $$\frac{1}{T^2}\sum_{t=1}^{T}z_{2t}^{(p)}z_{2t}^{(p)\prime}\xrightarrow{d}\int B_3B'_3.$$

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

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

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

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 $\underline{\tilde{J}}^{(p)}$, it is easier to derive the limiting distribution of $\underline{\tilde{J}}^{*(p)}$ with $\tilde{A}_1^{(p)}$. By Lemma 1, we have

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

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 $[\Delta x_t', \cdots, \Delta x_{t-k+1}', (x_{t-k} - x_{t-k-p})']'$ and $\Sigma_2^{2(p)}$ is that of $\beta' x_{t-k-p}$. The limiting distribution of $\sqrt{T}(\underline{\tilde{J}}^{*(p)} - \underline{J}^{*})$ 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,

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

where $\Gamma(p) = \Sigma_0 \otimes S'(\Sigma_2^{(p)})^{-1}S$, and we can easily check that $S'(\Sigma_2^{(p)})^{-1}S = (\Sigma_{2\cdot2}^{(p)})^{-1}$ and $\Sigma_{2\cdot2}^{(p)} = \Sigma_2^{1(p)} - \Sigma_2^{12(p)}(\Sigma_2^{2(p)})^{-1}\Sigma_2^{21(p)}$. Then,

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

where $vecN_0^{*(p)} \equiv N(0, \Sigma_0 \otimes (\Sigma_{2\cdot2}^{(p)})^{-1})$. Noting a relation $\underline{\tilde{J}}^{*(p)} = \underline{\tilde{J}}^{(p)}H_1^{-1}$, and following the argument in Toda and Yamamoto (1995), we can establish the next proposition.

Ppropos i t ion 1 (The LA(p) Approach):

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

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

where $\tilde{\Sigma}_0 = \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_t\tilde{u}_t'$ and $\tilde{u}_{1t}$'s are residuals of the IV estimation.

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

3. MODIFICATION AND LAG SELECTION

3.1. THE FINITE SAMPLE MODIFICATION

We have proposed the extended LA(p) approach in section 2 based upon the IV estimation. In this subsection, we propose to modify the LA(p) approach by correcting a bias of the IV estimator and to modify its variance covariance materix in order to obtain an accurate empirical size of the Wald test statistic.

At first we expand the OLS estimator (5) as

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

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. The quasi-asymptotic bias, $QBIAS[\underline{\tilde{J}}^{*(p)}, \tilde{A}_1^{(p)}]$, is expressed as

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

where both $SB^{(p)}$ and $NB^{(p)}$ are finite valued matrices independent of T, and then they are constant for any large T. The above result is straightforward from the similar result in Kurozumi and Yamamoto (2000). Its rigorous poof is currently under study.

Now we construct the modified lag augmented (MLA(p)) 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 ($t=1,\cdots,T$).

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

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,

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

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

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

Using (9), we have the following results about the quasi-asymptotic bias in each period.

$$(10)\quad QBIAS\left[\tilde{\underline{J}}^{*(p)}, \tilde{A}^{(p)}_1\right] = -\frac{1}{T}SB^{(p)} - \frac{1}{T}NB^{(p)},$$

$$(11)\quad QBIAS\left[\tilde{\underline{J}}^{*(p)}_f, \tilde{A}^{(p)}_{1f}\right] = -\frac{2}{T}SB^{(p)} - \frac{2}{T}NB^{(p)}.$$

$$(12)\quad QBIAS\left[\tilde{\underline{J}}^{*(p)}_s, \tilde{A}^{(p)}_{1s}\right] = -\frac{2}{T}SB^{(p)} - \frac{2}{T}NB^{(p)}.$$

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

$$(13)\quad \left[\tilde{\underline{J}}^{*(p)}_{mla}, \tilde{A}^{(p)}_{1mla}\right] = 2\left[\tilde{\underline{J}}^{*(p)}, \tilde{A}^{(p)}_1\right] - \frac{1}{2}\left(\left[\tilde{\underline{J}}^{*(p)}_f, \tilde{A}^{(p)}_{1f}\right] + \left[\tilde{\underline{J}}^{*(p)}_s, \tilde{A}^{(p)}_{1s}\right]\right).$$

We can easily check that this estimator has no quasi-asymptotic bias by substituting the right hand side of (13) with (10), (11) and (12).

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

$$\begin{aligned} (14)\quad \tilde{\underline{J}}^{(p)}_{mla} &= \tilde{\underline{J}}^{*(p)}_{mla}H_1 \\ &= 2\tilde{\underline{J}}^{(p)} - \frac{1}{2}\left(\tilde{\underline{J}}^{(p)}_f + \tilde{\underline{J}}^{(p)}_s\right). \end{aligned}$$

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 (14) has no quasi-asymptotic bias irrespective of the order of integration of $\{w_t\}$.

(ii) The MLA(p) estimator (14) is asymptotically normally distributed irrespective of the order of integration of $\{w_t\}$.

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

Proposition 2 :

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 $\{w_t\}$.

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

where

$$(15)\quad \tilde{\Sigma}_{mla}^{(p)} = \tilde{\Sigma}_0^{(p)} \otimes \left(\frac{1}{T} X_1' Q_{X2}^{(p)} X_1\right)^{-1}$$

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

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

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

The explanation of the covariance matrix $\tilde{\Sigma}_{mla}^{(p)}$ is in order: In theory, we can use any consistent estimator of $\Sigma_0 \otimes \Sigma_{22}^{-1}$. We have tried several consistent estimators in Monte Carlo simulations. The test statistic using $\tilde{\Sigma}_{mla}^{(p)}$ has generally shown the smallest size distortion among them in the small sample. Thus, we decided to adopt $\tilde{\Sigma}_{mla}^{(p)}$ as the estimator of $\Sigma_0 \otimes \Sigma_{22}^{-1}$. See Yamamoto and Kurozumi (2005) for detail. Its basic idea is to intentionally inflate the conventional estimate of $\Sigma_0^{(p)}$ a little bit in order to reduce the size ditortion of the test, since the conventional estimate of $\Sigma_0^{(p)}$ appears to be underestimated in finite samples. See also Chigira and Yamamoto (2007) for a similar attempt.

3.2. SELECTION OF LAG LENGTH

The Monte Carlo experiments in Yamamoto and Kurozumi (2005) revealed that the extended MLA(p) is generally more powerful than the ordinary MLA(1) estimator. On the other hand, they analytically showed, in a slightly different model specification, that there could be a case where the extended MLA(p) ($p \geq 2$) approach is less efficient than the ordinary MLA, i.e., MLA(1), approach. Thus, the unconditional use of the extended MLA(p) ($p \geq 2$) approach should be avoided. The issue is that we have to select a suitable lag length p in the extended MLA(p) approach including the case of $p=1$.

In this paper, we propose a practical method to select a suitable p. We propose to select a suitable p^* which gives the minimum $det(\tilde{\Sigma}_0^{(p)})$ ($p = 1, 2, \cdots, p_{max}$), where p_{max} is a priori specified not-so-small number, say, $p_{max}=10$. Mathematically, it is written as

$$p^* = \underset{1 \leq p \leq p_{max}}{argmin} \left\{ det\left(\tilde{\Sigma}_{mla}^{(p)}\right)\right\}$$

The use of $det(\tilde{\Sigma}_0^{(p)})$ as a criteria is motivated the fact that it is a generalized variance of $\tilde{\beta}_{mla}^{(p)}$. We call the MLA($p$) with the selected p^* as the variable modified LA (VMLA) approach.

4. SIMULATION EXPERIMENT

4.1. EXPERIMENTAL DESIGN

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

$$(16)\qquad \begin{aligned} y_t &= \beta_1 w_t + \beta_2 w_{t-1} + \varepsilon_t, \text{ and} \\ w_t &= \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + v_t, \end{aligned}$$

where

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

Three models for data generation of $\{w_t\}$ are considered.

Model Ⅰ : $\alpha_1=1.8$, and $\alpha_2=-0.8$.
Model Ⅱ : $\alpha_1=0.2$, and $\alpha_2=0.8$.
Model Ⅲ : $\alpha_1=1.6$, and $\alpha_2=-0.64$.

The first two models are non-stationary, that is, they have a unit root. The first two models can alternatively expressed as

Model Ⅰ : $\Delta w_t = 0.8\Delta w_{t-1} + v_t$, and
Model Ⅱ : $\Delta w_t = -0.8\Delta w_{t-1} + v_t$.

Thus, the stationary component of model Ⅰ exhibits strong positive autocorrelations, whereas that of model Ⅱ a strong negative first order autocorrelation. Model Ⅲ is a purely stationary one with strong positive autocorrelations. Since most sconomic data exibit positive serial correlations, we are more interested in the results of Models Ⅰ and Ⅲ than those of Model Ⅱ.

The three models for the Ⅳ estimation of the LA(p) approach are given by

$$y_t = \tilde{\beta}_0^{(p)} + \tilde{\beta}_1^{(p)} w_t + \tilde{\beta}_2^{(p)} w_{t-1} + \tilde{\beta}_3^{(p)} w_{t-1-p} + \tilde{u}_t^{(p)},$$

$$(17) \quad y_t = \tilde{\beta}_{0f}^{(p)} + \tilde{\beta}_{1f}^{(p)} w_t + \tilde{\beta}_{2f}^{(p)} w_{t-1} + \tilde{\beta}_{3f}^{(p)} w_{t-1-p} + \tilde{u}_{ft}^{(p)},$$

$$y_t = \tilde{\beta}_{0s}^{(p)} + \tilde{\beta}_{1s}^{(p)} w_t + \tilde{\beta}_{2s}^{(p)} w_{t-1} + \tilde{\beta}_{3s}^{(p)} w_{t-1-p} + \tilde{u}_{st}^{(p)}.$$

Throughout the experimet, the instrumental variables are $w_{t-1}, w_{t-2}, \cdots, w_{t-2-p}$. The bias corrected MLA(p) estimator is given by

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

where

$$\tilde{\beta}^{(p)} = \left[\tilde{\beta}_1^{(p)}, \tilde{\beta}_2^{(p)}\right]', \ \tilde{\beta}_f^{(p)} = \left[\tilde{\beta}_{1f}^{(p)}, \tilde{\beta}_{2f}^{(p)}\right]' \text{ and } \tilde{\beta}_s^{(p)} = \left[\tilde{\beta}_{1s}^{(p)}, \tilde{\beta}_{2s}^{(p)}\right]'.$$

The estimator of its covariance matrix is obtained as

$$(18) \quad \tilde{\Sigma}_{mla}^{(p)} = \tilde{\sigma}_u^2 \left(\frac{1}{T} X_1' Q_{X2}^{(p)} X_1\right)^{-1} + T\left(\tilde{\beta}_{mla}^{(p)} - \tilde{\beta}^{(p)}\right)\left(\tilde{\beta}_{mla}^{(p)} - \tilde{\beta}^{(p)}\right)'.$$

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

$$(19) \quad W_{mla}^{(p)} = \left\{\sqrt{T}\left(\tilde{\beta}_{mla}^{(p)} - \beta_0\right)'\right\} \tilde{\Sigma}_{mla}^{(p)-1} \left\{\sqrt{T}\left(\tilde{\beta}_{mla}^{(p)} - \beta_0\right)\right\},$$

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

4.2. SIMULATION RESULTS

In the following simulation experiments, the number of replication is 1,000 in all experiments. Computations are performed by the GAUSS matrix programming language.

4.2.1. Effects of Bias Correction and Adjustment of Variance Covariance Matrix

We first examine how the modifications proposed in Section 3.1 work in finite samples. Specifically, Tables 1a and 1b reveal the effects of bias correction and modification of the variance covariance matrix on empirical size of the test. Table 1a gives the empirical size of of the ordinary Wald test based upon the IV estimation without any adjustment, LA(p) (p=1, 2,$\cdots$, 8). On the ther hand, Table 1b shows the corresponding empirical size of MLA(p) (p=1, 2,$\cdots$, 8) which incorporates the modifications. We find that, while LA(p) shows a large size distortion, MLA(p) a relatively small distortion which is acceptable for practical purposes. Thus, we have confirmed that bias correction and modification of the variance covariance matrix proposed in Section 3.1 are useful for reducing size distortion of the original test statistic, LA(p) (p= 1, 2,$\cdots$, 8), in finite samples.

4.2.2. Empirical Size and Power when T=100

Tables 2a-2e show the empirical size and power of MLA(p) (p=1,2,$\cdots$,8) and VMLA for T=100.

Table 2a shows the empirical size of MLA(p) (p= 1, 2,$\cdots$, 8) which is replication of Table 1b in additon to VMLA for T= 100. The case of MLA(1) corresponds to the ordinary MLA approach, and MLA(p) (p=2,$\cdots$, 8) to the extended MLA approach. Table 2a shows that the emprical size of MLA(p) generally decreases with p. For VMLA, the empirical size slightly overestimates the corresponding nominal one for Models I and II. But the size distortion is relatively large for Model III.

Tables 2b-2e show the (size unadjusted) empirical power of the test of MLA(p) (p=1, 2,$\cdots$, 8) and VMLA for T=100. As in Yamamoto and Kurozumi (2005), MLA(p) (p=2, 3,$\cdots$,8) has generally higher power than MLA(1). But, in some cases, MLA(1) has higher power than some of MLA(p) (p=1, 2,$\cdots$, 8). See, for example, model II in Table 2b, where MLA(1) is larger than MLA(2). Turning our attention to VMLA, it is evident that VMLA generally has higher power than MLA(p) ($p=1, 2,\cdots, 8$), and it is *consitently* more powerful than MLA(1). That is, the lag selection scheme proposed in section 3.2 is quite useful for the improvement for the power of the ordinary lag augmention MLA, that is, MLA(1).

4.2.3. Effect of Sample Size

Tables 3a-3e show the empirical size and power of MLA(p) ($p=1, 2,\cdots, 8$) and VMLA for $T=200$. These tables exacly correspond to Tables 2a-2e. The only difference is the sample size. Table 3a shows the empirical size of MLA(p) ($p=1, 2,\cdots, 8$) and VMLA for $T=200$. Generally, the size distortion of MLA(p) ($p=1, 2,\cdots,8$) decreases in comparison with Table 2a. The size distortion of VMLA also decreases for models Ⅰ and Ⅲ, but not so for Model Ⅱ. There are still rooms for improvement for the size of the test, These will be left to the future research.

Tables 3b-3e show the (size unadjusted) emprical power of of MLA(p) ($p=1, 2,\cdots, 8$) and VMLA for $T=200$. They correspond to Tables 2b-2e. The power smoothly increases as T increases.

5. CONCLUSION

This paper has consider the Wald type test for a regression model whose regressors are possibly non-stationary and contemporaneously correlated the error term.

The present paper has extended our previous works, Kurozumi and Yamamoto (2000) and Yamamoto and Kurozumi (2005), in two directions. First, it has dealt with Ⅳ estimator instead of OLS estimator. Socond, it has proposed a method to select a suitable lag length p in the MLA(p) approach, which is called the variable MLA (VMLA) approach. The Monte Carlo simulation in this paper has revealed that the VMLA always gives higher power of the test than the MLA, that is, MLA(1). In other words, the VMLA has been shown to be quite useful in improving the power of the Wald type test in finite samples.

REFERENCES

[1] Chigira, H., and T. Yamamoto (2007): "Finite Sample Modification of the Granger Non-Causality Test in Cointegrated Vector Autoregressions," *Communications in Statistics, Theory and Methods*, Vol. 36, 2007, 981-1003.

[2] 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, 85-123.

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

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

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

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

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

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

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

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

[11] Yamamoto, T., and E. Kurozumi (2005): "Lag Augmentation in Regression Models with Possibly Integrated Regressors," *Hitotsubashi Journal of Economics*, 46, 159-175.

Table 1. Effects of Bias Correction and Modification on Empirical Size (T=100)

DGP : $y_t = \beta_1 w_t + 0.3 w_{t-1} + \varepsilon_t$, and $w_t = \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + v_t$,

$$\text{where} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \; i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \right).$$

Model Ⅰ : $\alpha_1 = 1.8$, and $\alpha_2 = -0.8$.

Model Ⅱ : $\alpha_1 = 0.2$, and $\alpha_2 = 0.8$.

Model Ⅲ : $\alpha_1 = 1.6$, and $\alpha_2 = -0.64$.

Model for Estimation : $y_t = \tilde{\beta}_0 + \tilde{\beta}_1 w_t + \tilde{\beta}_2 w_{t-1} + \tilde{\beta}_3 w_{t-1-p} + \tilde{u}_t$.

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

Table 1a. No Adjustment LA(p) : β_1=0.7 and β_2=0.3 in DGP

Model	Ⅰ		Ⅱ		Ⅲ	
Significance Level	5%	10%	5%	10%	5%	10%
p of LA (p)						
1	11.2	15.3	4.2	8.9	9.1	11.6
2	10.2	15.9	6.3	12.5	8.5	13.0
3	10.1	16.3	4.6	8.9	9.0	13.5
4	9.3	17.5	6.8	13.9	9.7	14.5
5	10.9	20.0	5.3	10.7	10.3	16.7
6	13.1	20.8	7.8	14.3	11.9	19.3
7	14.1	21.4	6.7	13.4	13.6	20.4
8	15.0	24.3	7.8	15.6	15.8	23.3

Table 1b. With Modification MLA(p) : β_1=0.7 and β_2=0.3 in DGP

Model	Ⅰ		Ⅱ		Ⅲ	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	8.2	11.0	3.1	6.1	9.6	12.9
2	5.3	8.5	5.0	8.4	7.3	11.1
3	4.3	6.7	2.5	5.6	7.1	10.0
4	3.5	6.0	4.4	8.2	5.9	9.2
5	3.5	6.8	3.5	6.2	6.2	9.7
6	3.5	6.3	4.4	9.2	5.4	9.3
7	3.6	6.6	3.9	7.5	5.5	8.6
8	3.5	6.8	4.0	8.5	5.8	10.1

Table 2. Empirical Size and Power of MLA(p) and VMLA (T=100)

For detailed description of the model and the hypothesis, see Table 1.

Table 2a. Empirical Size : β_1=0.7 and β_2=0.3 in DGP

Model	I		II		III	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	8.2	11.0	3.1	6.1	9.6	12.9
2	5.3	8.5	5.0	8.4	7.3	11.1
3	4.3	6.7	2.5	5.6	7.1	10.0
4	3.5	6.0	4.4	8.2	5.9	9.2
5	3.5	6.8	3.5	6.7	6.2	8.7
6	3.5	6.3	4.4	9.2	5.4	9.3
7	3.6	6.6	3.9	7.5	5.5	8.6
8	3.5	6.8	4.0	8.5	5.8	10.1
VMLA	7.6	11.4	6.0	10.7	13.5	18.9

Table 2b. Empirical Power : β_1=0.5 and β_2=0.3 in DGP

Model	I		II		III	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	83.6	85.9	26.1	36.6	86.1	89.0
2	91.3	93.2	15.7	26.7	93.5	95.1
3	93.2	94.4	37.0	46.8	95.0	96.1
4	93.8	95.2	32.9	46.0	95.8	96.9
5	94.1	95.9	48.2	58.0	94.9	96.5
6	94.0	95.4	45.3	57.8	94.5	96.7
7	94.4	95.6	54.5	64.8	95.6	97.1
8	94.3	95.8	53.7	66.2	95.2	96.8
VMLA	96.7	97.8	40.5	53.0	97.6	98.7

Table 2c. Empirical Power : β_1=0.65 and β_2=0.3 in DGP

Model	I		II		III	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	26.5	33.2	3.3	7.7	22.3	28.4
2	33.5	40.9	4.8	7.9	24.8	32.1
3	43.3	50.0	3.0	8.1	28.8	37.0
4	48.4	54.9	5.3	9.7	29.8	39.5
5	49.6	56.8	4.6	10.5	31.7	41.4
6	51.6	57.9	5.6	11.6	32.6	41.8
7	51.3	58.6	5.5	12.5	31.9	43.0
8	51.7	58.3	6.0	12.6	32.0	41.0
VMLA	58.2	64.1	5.9	9.8	45.4	54.8

Table 2d. Empirical Power : β_1=0.75 and β_2=0.3 in DGP

Model	I		II		III	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (*p*)						
1	19.2	28.4	4.8	8.9	24.7	32.9
2	25.4	34.5	7.1	11.2	27.8	35.6
3	29.7	40.4	4.5	9.0	29.2	40.0
4	33.7	44.6	6.9	11.2	30.0	40.9
5	37.5	49.2	5.4	9.1	32.2	41.1
6	39.8	51.6	6.7	12.1	33.3	40.8
7	42.8	52.7	6.1	10.2	33.6	41.5
8	43.3	52.5	6.5	11.6	34.2	41.8
VMLA	45.4	55.7	10.3	16.6	39.7	48.6

Table 2e. Empirical Power : β_1=0.9 and β_2=0.3 in DGP

Model	I		II		III	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (*p*)						
1	92.6	95.3	29.7	38.6	53.9	62.3
2	97.4	98.9	23.7	33.9	63.5	72.2
3	98.8	99.6	31.1	41.8	68.0	77.5
4	99.0	99.8	30.7	39.9	70.8	80.1
5	98.7	99.7	34.1	45.5	72.4	81.3
6	99.4	99.7	35.3	45.5	73.9	81.3
7	99.4	99.6	37.2	46.5	74.5	82.3
8	99.1	99.6	38.1	48.7	74.6	80.6
VMLA	99.3	99.9	47.7	59.0	98.8	99.4

Table 3. Empirical Size and Power of MLA(p) and VMLA (T=200)
For detailed description of the model and the hypothesis, see Table 1.

Table 3a. Empirical Size : β_1=0.7 and β_2=0.3 in DGP

Model	I		II		III	
Significance Level	5%	10%	5%	10%	5%	10%
p of MLA (*p*)						
1	5.2	9.1	4.7	8.6	6.8	11.6
2	4.1	8.0	6.1	10.9	5.6	10.0
3	4.5	8.6	4.5	9.4	5.9	11.0
4	4.5	8.7	5.4	10.5	6.6	10.1
5	4.7	10.1	4.7	9.2	6.6	11.1
6	5.2	9.6	5.4	10.2	7.0	11.2
7	4.9	9.4	4.9	10.1	6.5	10.8
8	4.8	8.7	6.5	11.0	7.1	11.5
VMLA	6.0	10.4	7.1	11.4	10.0	16.0

Table 3b. Empirical Power : β_1=0.5 and β_2=0.3 in DGP

Model	I		II		III	
Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	98.8	99.2	68.7	74.7	99.3	99.5
2	99.7	99.8	52.1	63.7	99.7	99.8
3	99.8	99.8	73.2	81.0	100.0	100.0
4	99.7	100.0	71.8	79.0	100.0	100.0
5	99.8	100.0	78.7	84.3	100.0	100.0
6	99.8	100.0	77.3	82.9	100.0	100.0
7	99.9	100.0	80.9	85.7	100.0	100.0
8	100.0	100.0	82.0	85.8	100.0	100.0
VMLA	100.0	100.0	79.0	83.8	100.0	100.0

Table 3c. Empirical Power : β_1=0.65 and β_2=0.3 in DGP

Model	I		II		III	
Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	52.4	61.8	6.3	11.9	35.0	43.8
2	69.9	76.1	6.6	12.6	48.8	61.1
3	77.1	83.0	7.5	14.8	60.2	70.2
4	82.4	86.0	8.9	14.9	67.7	76.9
5	85.0	88.2	9.6	17.8	71.2	80.0
6	86.7	89.4	10.2	17.3	73.2	81.8
7	87.9	90.2	11.4	19.7	73.8	81.0
8	88.1	90.9	11.9	19.4	73.5	81.4
VMLA	89.3	91.7	8.6	16.1	77.9	85.1

Table 3d. Empirical Power : β_1=0.75 and β_2=0.3 in DGP

Model	I		II		III	
Level	5%	10%	5%	10%	5%	10%
p of MLA (p)						
1	50.6	63.9	8.7	15.5	39.3	50.1
2	70.4	81.2	10.5	15.9	52.3	62.4
3	82.8	88.8	11.2	17.1	59.4	71.0
4	87.2	92.2	11.4	18.1	65.5	75.5
5	90.1	94.3	11.1	17.9	68.3	78.3
6	92.6	95.8	13.2	19.4	70.6	80.2
7	93.9	96.9	12.5	20.0	72.0	80.3
8	94.1	96.8	13.3	21.9	72.4	80.2
VMLA	92.9	95.5	17.6	25.3	70.8	78.9

Table 3e. Empirical Power : β_1=0.90 and β_2=0.3 in DGP

Model	I		II		III	
Level	5%	10%	5%	10%	5%	10%
p of MLA (*p*)						
1	100.0	100.0	69.2	78.4	99.8	100.0
2	100.0	100.0	54.1	61.6	100.0	100.0
3	100.0	100.0	75.5	82.4	100.0	100.0
4	100.0	100.0	71.8	79.4	100.0	100.0
5	100.0	100.0	79.2	84.6	100.0	100.0
6	100.0	100.0	76.9	84.4	100.0	100.0
7	100.0	100.0	81.3	86.6	100.0	100.0
8	100.0	100.0	82.7	88.5	100.0	100.0
VMLA	100.0	100.0	85.4	91.0	100.0	100.0