Moment Approximation for Unit Root Models with Nonnormal Errors

Yong Bao*
Department of Economics
Purdue University

Aman Ullah†
Department of Economics
University of California, Riverside

Ru Zhang‡
Department of Economics,
University of California, Riverside

January 22, 2014

ABSTRACT

Phillips (1977a, 1977b) made seminal contributions to time series finite-sample theory, and then, he was among the first to develop the distributions of estimators and forecasts in stationary time series models, see Phillips (1978, 1979), among others. From the mid-eighties Phillips (1987a, 1987b), through his fundamental papers, opened the path of asymptotic (large-sample) theory for the unit root type non-stationary models. This has certainly created a large literature of important papers, including many of Phillips' own papers. However, not much is known about the analytical finite-sample properties of estimators under the unit root, although see Kiviet and Phillips(2005) for the case when the errors are normally distributed. An objective of this paper is to analyze the finite-sample behavior of the estimator in the first-order autoregressive model with unit root and nonnormal errors. In particular, we derive analytical approximations for the first two moments in terms of model parameters and the distribution parameters. Through Monte Carlo simulations, we find that our approximate formula perform quite well across different distribution specifications in small samples. However, when the noise to signal ratio is huge, and bias distortion can be quite substantial, and our approximations do not fare well.

Keywords: unit root, nonnormal, moment approximation

---

*Corresponding Author: Department of Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907. Email: ybao@purdue.edu.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu.

‡Department of Economics, University of California, Riverside, CA 92521. E-mail: ru.zhang@email.ucr.edu

1 Introduction

The autoregressive (AR) model has been a workhorse model in time series econometrics and applied macroeconomics. There has been a huge literature in studying properties of the least-squares (LS) estimator. Hurwicz (1950) was the first to investigate the finite-sample bias issue. Sawa (1978) proposed using the moment generating function approach to numerically evaluating the exact moments. Approximate moments of the LS estimator were developed by Grubb and Symons (1987), Kiviet and Phillips (1993), Bao and Ullah (2007), among others. The asymptotic distribution of the LS estimator was analyzed by White (1958, 1959) and Anderson (1959), whereas the exact distribution was analyzed by Phillips (1977a, 1978), Evans and Savin (1981), and Tsui and Ali (1994), to name just a few.

Along this line of research over the years, Phillips has made seminal contributions to time series finite sampling theory. He was among the first to develop the distributions of estimators and forecasts in stationary autoregressive models. Phillips (1977a) examined the Edgeworth-type approximation to the finite-sample distributions of the LS estimator and the associated t ratio, and compared with the Imhof-approach of numerically obtaining the exact distribution, whereas Phillips (1978) compared the Edgeworth and saddlepoint approximations. Phillips (1977b) provided a general discussion of the validity of the Edgeworth-type expansion to approximate the finite sample distribution of general statistics. In terms of forecasting, Phillips (1979) emphasized that the unbiasedness result (of forecasts) given in Malinvaud (1970) was an unconditional result, and the more interesting analysis should center on the conditional forecasts. The argument was that since the last observation was observed in the sample, then this information should be used in deciding the distribution of the forecasts. To this end, Phillips (1979) developed an approximation to the distribution of the forecast error conditional on the last observed sample data and found that the forecasts were biased in almost all cases. From the mid -eighties, Phillips (1987a, 1987b), through his fundamental papers, opened the path of asymptotic (large-sample) properties for the unit root type non-stationary models. This has certainly created a large literature of important papers to follow, including many of Phillips' own papers.

We note that among these works, not much attention has been paid to the *analytical* finite-sample properties of the LS estimator for the unit root model, though numerically, under normality, the moments and distribution can be evaluated via the techniques of Sawa (1978) and Imhof (1961), respectively. The exceptions are Abadir (1993) and Kiviet and Phillips (2005), where analytical approximations to the moments of the LS estimator for the unit root model were developed with the normality assumption. For stationary autoregressive models, see Bao and Ullah (2007) and Bao (2007) for moment approximations when the normality assumption is relaxed.

In this paper, we exam the moments of the LS estimator in the first-order AR model with an arbitrary number of exogenous regressors (ARX(1)) when the true coefficient of the lagged-dependent variable is unity. We do not make any distribution assumption except the existence of moments of the errors up to order 4. We also discuss the case when the exogenous regressors may be nonstationary. Our main achievements are as follows: (i) derive the approximate bias of the estimated autoregressive coefficient up to order T^{-3} in terms of model parameters and the skewness coefficient of the error distribution; (ii) derive the approximate MSE up to order T^{-4} of this estimator also in terms of model parameters and the skewness coefficient; (iii) obtain approximations to the first two moments of the estimated full coefficient vector for unit root models that contain an arbitrary number of exogenous and possibly nonstationary regressors; (iv) for the random walk with drift model, very neat bias and MSE formulae are presented; (v) provide numerical illustrations of the accuracy of our moment approximations in finite samples.

Our paper proceeds as follows. In the next section, we focus on the moment approximations for the LS estimator of the AR coefficient, and discuss also the special case of random walk with drift. Section 3 presents the results for the full coefficient vector estimator. Section 4 investigates the accuracy of the analytical results developed by comparing the calculated moment approximations with those from simulations under several symmetric and skewed distributions. Section 5 concludes. The appendix contains detailed derivation of the main results in the paper.

2 The Bias and MSE of the Unit Root Coefficient Estimator

In this section, we discuss the approximate moments of the unit root coefficient estimator when an arbitrary number of exogenous variables, with at least a constant term, are included. Throughout, define ι as a $T \times 1$ vector of ones and J as a $T \times T$ strict lower triangular matrix with all the elements below the main diagonal being 1.

We follow Kiviet and Phillips (2005) to consider the following model

$$y = \lambda y_{-1} + X\beta + \sigma\varepsilon, \tag{1}$$

where $y = (y_1, \cdots, y_T)'$ contains a $T \times 1$ vector of observations on the dependent variable, $y_{-1} = (y_0, \cdots, y_{T-1})$ contains the lagged values of the dependent variable, X is a $T \times k$ matrix containing all exogenous regressors that include at least an intercept term, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_T)'$ is the $T \times 1$ vector of iid disturbances.

Assumption 1: *In model (1), the scalar λ and the vector β are unknown coefficients and further: (i) $\lambda = 1$;*

(ii) all elements of β are finite and $\beta \neq 0$; (iii) X is stationary and $X'X = O(T)$; (iv) $Z = (y_{-1}, X)$ has rank $k+1$ with probability one; (v) X is strongly exogenous, that is, X and ε are independent, and it contains at least a constant term; (vi) σ is positive and finite, and the disturbances ε_i are i.i.d., with mean zero, variance 1, and skewness and excess kurtosis coefficients γ_1 and γ_2, respectively; (vii) the start-up value is y_0, either fixed or random, and when y_0 is random, it is independent of ε.

In Assumption 1, we impose stationarity on X, but later this will be relaxed. Regarding the error term ε, we do not make distributional assumption, except the existence of its first four moments.

The LS estimator of λ, which is also the quasi maximum likelihood estimator under nonnormality, is

$$\hat{\lambda} = \frac{y_{-1}' M y}{y_{-1}' M y_{-1}} = \lambda + \sigma \frac{y_{-1}' M \varepsilon}{y_{-1}' M y_{-1}}, \tag{2}$$

where $M = I_T - X(X'X)^{-1}X'$. We may write

$$y_{-1} = y_0 \iota + JX\beta + \sigma J \varepsilon. \tag{3}$$

Note that if X contains a constant term, then obviously $My_{-1} = MJX\beta + \sigma MJ\varepsilon$, so the properties of $\hat{\lambda}$ does not depend on the initial value y_0.

By plugging $My_{-1} = MJX\beta + \sigma MJ\varepsilon$, we have

$$\hat{\lambda} - 1 = \frac{\sigma \varepsilon' MJX\beta + \sigma^2 \varepsilon' MJ\varepsilon}{\beta' X' J' MJX\beta + 2\sigma \varepsilon' J' MJX\beta + \sigma^2 \varepsilon' J' MJ\varepsilon}. \tag{4}$$

Since we assume the regressors are stationary and $X'X = O(T)$, then immediately, $P_X := X(X'X)^{-1}X' = o(1)$ and $P_X J = O(1)$. Also, $X'J'MJX = X'J'JX - X'J'X(X'X)^{-1}X'JX = O(X'J'JX) = O(T^3)$, and $\beta' X' J' MJX\beta = O(T^3)$. This gives $\varepsilon' MJX\beta = O_P(T^{3/2})$. Similarly, we can verify that $\varepsilon' MJ\varepsilon = O_P(T)$, $\varepsilon' J' MJ\varepsilon = O_P(T^2)$, $\varepsilon' J' MJX\beta = O_P(T^{5/2})$. Thus, the dominating term in our expansion is $\beta' X' J' MJX\beta$. By defining $\mu := (\beta' X' J' MJX\beta)^{-1} = O(T^{-3})$, we continue to write

$$\begin{aligned}
\hat{\lambda} - 1 &= \frac{\overbrace{\mu\sigma\varepsilon' MJX\beta}^{O_P(T^{-3/2})} + \overbrace{\mu\sigma^2\varepsilon' MJ\varepsilon}^{O_P(T^{-2})}}{1 + \underbrace{2\mu\sigma\varepsilon' J' MJX\beta}_{O_P(T^{-1/2})} + \underbrace{\mu\sigma^2\varepsilon' J' MJ\varepsilon}_{O_P(T^{-1})}} \\
&= (\mu\sigma\varepsilon' MJX\beta + \mu\sigma^2\varepsilon' MJ\varepsilon) \times [1 - 2\mu\sigma\varepsilon' J' MJX\beta - \mu\sigma^2\varepsilon' J' MJ\varepsilon \\
&\quad + 4\mu^2\sigma^2 \left(\varepsilon' J' MJX\beta\right)^2 + 4\mu^2\sigma^3 \left(\varepsilon' J' MJX\beta\right)\left(\varepsilon' J' MJ\varepsilon\right) - 8\mu^3\sigma^3 \left(\varepsilon' J' MJX\beta\right)^3] + o_P(T^{-3})
\end{aligned}$$

$$
\begin{aligned}
&= \left(\mu\sigma\varepsilon' MJX\beta + \mu\sigma^2\varepsilon' MJ\varepsilon\right)\Big[[1 - 2\mu\sigma\varepsilon' J' MJX\beta - \mu\sigma^2\varepsilon' J' MJ\varepsilon + 4\mu^2\sigma^2\left(\varepsilon' J' MJX\beta\right)^2\Big] \\
&\quad +4\mu^3\sigma^4\left(\varepsilon' MJX\beta\right)\left(\varepsilon' J' MJX\beta\right)\left(\varepsilon' J' MJ\varepsilon\right) - 8\mu^4\sigma^4\varepsilon' MJX\beta\left(\varepsilon' J' MJX\beta\right)^3 + o_P(T^{-3}) \\
&= \underbrace{\mu\sigma\varepsilon' MJX\beta}_{O_P(T^{-3/2})} + \underbrace{\mu\sigma^2\varepsilon' MJ\varepsilon - 2\mu^2\sigma^2\varepsilon' MJX\beta\beta' X' J' MJ\varepsilon}_{O_P(T^{-2})} \\
&\quad + \underbrace{4\mu^3\sigma^3\beta' X' J' M\varepsilon\varepsilon' J' MJX\beta\beta' X' J' MJ\varepsilon - \mu^2\sigma^3\beta' X' J' M\varepsilon\varepsilon' J' MJ\varepsilon - 2\mu^2\sigma^3\beta' X' J' MJ\varepsilon\varepsilon' MJ\varepsilon}_{O_P(T^{-5/2})} \\
&\quad + \underbrace{4\mu^3\sigma^4\varepsilon' J' MJ\varepsilon\varepsilon' MJX\beta\beta' X' J' MJ\varepsilon - 8\mu^4\sigma^4\varepsilon' MJX\beta\beta' X' J' MJ\varepsilon\varepsilon' J' MJX\beta\beta' X' J' MJ\varepsilon}_{O_P(T^{-3})} \\
&\quad + \underbrace{4\mu^3\sigma^4\varepsilon' MJ\varepsilon\varepsilon' J' MJX\beta\beta' X' J' MJ\varepsilon - \mu^2\sigma^4\varepsilon' MJ\varepsilon\varepsilon' J' MJ\varepsilon}_{O_P(T^{-3})} + o_P(T^{-3}). \qquad (5)
\end{aligned}
$$

Then taking expectations term by term in (5), collected in Appendix A, and upon substituting these terms into the expectation of (5), we have the following bias result under nonnormality:

Theorem 1: *Under Assumption 1 and* $M\iota = 0$, *the bias of the least squares estimator of* λ *to the order of* T^{-3} *is given by*

$$
\begin{aligned}
\mathrm{E}(\hat{\lambda} - 1) &= \sigma^2\mu[1 + \mathrm{tr}(MJ)] \\
&\quad +\sigma^3\mu^2\gamma_1\beta' X' J' M[4\mu(I \odot J' MJX\beta\beta' X' J' MJ) - (I \odot J' MJ) - 2J(I \odot MJ)]\iota \\
&\quad -\sigma^4\mu^2\mathrm{tr}(MJ)[\mathrm{tr}(J' MJ) - 4\mu\beta' X' J' MJJ' MJX\beta] + o(T^{-3}). \qquad (6)
\end{aligned}
$$

We see that $\hat{\lambda}$ is unbiased up to $O(T^{-3/2})$, in line with the well-known result that $\hat{\lambda}$ is superconsistent with a convergence rate of $T^{3/2}$. It is also unbiased up to $O(\sigma)$. The leading bias term is of order $O(T^{-2})$, given by

$$
\mathrm{E}(\hat{\lambda} - 1) = -\sigma^2\mu\,\mathrm{tr}((X' X)^{-1} X' JX) + o(T^{-2}), \qquad (7)
$$

which is robust to nonnormality. The distribution effect of the disturbance term on the approximate bias comes through its skewness coefficient γ_1, and it is of order $T^{-3.}$ This stands in contrast to the stationary AR model, where the LS estimator is $\sqrt{T}$-consistent and the "second-order" bias, up to $O(T^{-2})$, is robust to nonnormality, as shown in Bao and Ullah (2007).

For the special case of $X = \iota$, we observe that $MJX = J\iota - T^{-1}\iota' J\iota\iota$ with the t-th element $t - 1 - 0.5(T - 1)$, $J' MJX = J' J\iota - T^{-1}\iota' J\iota J'\iota$ with the t-th element $0.5[T(T - 1) - t(t - 1) - (T - 1)(T - t)]$, $MJ = J - T^{-1}\iota\iota' J$

with the t-th diagonal $-T^{-1}(T-t)$, and $J'MJ = J'J - T^{-1}J'\iota\iota'J$ with the t-th diagonal $T-t-T^{-1}(T-t)^2$. Upon simplification (see Appendix B), the bias result reduces to the following:

Corollary 1: *If in the model of Theorem 1 we have $X = \iota$, then the bias simplifies to*

$$\mathrm{E}(\hat{\lambda}-1) = -6\left(\frac{\sigma}{\beta}\right)^2 \frac{1}{T^2} + 18\left(\frac{\sigma}{\beta}\right)^2 \frac{1}{T^3} + 6\gamma_1\left(\frac{\sigma}{\beta}\right)^3 \frac{1}{T^3} - \frac{84}{5}\left(\frac{\sigma}{\beta}\right)^4 \frac{1}{T^3} + o(T^{-3}). \tag{8}$$

From this, we observe that up to T^{-2}, for the random walk with drift model, the bias of the unit root estimator is negative, and this bias is proportional to the squared noise to signal ratio σ/β. One might immediately predict that in the case of extreme values of σ/β, this bias expression may give poor approximation to the true bias. We might also predict that when the disturbance term is positively skewed, it tends to pull down the magnitude of the bias. (Keep in mind that the leading bias term is negative.)

To approximate the MSE of $\hat{\lambda}$ up to order $O(T^{-4})$, we can directly utilize the expansion of $\hat{\lambda}-1$, and put

$$\begin{aligned}
(\hat{\lambda}-1)^2 \;=\;& (\mu\sigma\varepsilon'MJX\beta)^2 + 2(\mu\sigma\varepsilon'MJX\beta)(\mu\sigma^2\varepsilon'MJ\varepsilon - 2\mu^2\sigma^2\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon) \\
&+(\mu\sigma^2\varepsilon'MJ\varepsilon - 2\mu^2\sigma^2\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon)^2 \\
&+2(\mu\sigma\varepsilon'MJX\beta)[4\mu^3\sigma^3\beta'X'J'M\varepsilon\varepsilon'J'MJX\beta\beta'X'J'MJ\varepsilon \\
&-\mu^2\sigma^3\beta'X'J'M\varepsilon\varepsilon'J'MJ\varepsilon - 2\mu^2\sigma^3\beta'X'J'MJ\varepsilon\varepsilon'MJ\varepsilon] + o_P(T^{-4}) \\
\;=\;& \underbrace{\sigma^2\mu^2\varepsilon'MJX\beta\beta'X'J'M\varepsilon}_{O_P(T^{-3})} + \underbrace{2\sigma^3\mu^2\beta'X'J'M\varepsilon\varepsilon'MJ\varepsilon - 4\mu^3\sigma^3\beta'X'J'M\varepsilon\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon}_{O_P(T^{-7/2})} \\
&+\underbrace{\sigma^4\mu^2(\varepsilon'MJ\varepsilon)^2 - 8\sigma^4\mu^3\varepsilon'MJ\varepsilon\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon - 2\sigma^4\mu^3\varepsilon'J'MJ\varepsilon\varepsilon'MJX\beta\beta'X'J'M\varepsilon}_{O_P(T^{-4})} \\
&+\underbrace{12\sigma^4\mu^4(\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon)^2}_{O_P(T^{-4})} + o_P(T^{-4}).
\end{aligned} \tag{9}$$

So by substituting the expectations (see the appendix) and ignoring terms of smaller orders, we obtain the following results.

Theorem 2: *In the model of Theorem 1, the MSE of the least squares estimator of λ to the order of T^{-4} is given by*

$$\mathrm{E}[(\hat{\lambda}-1)^2] \;=\; \sigma^2\mu + \sigma^4\mu^2\{[\mathrm{tr}(MJ)]^2 + \mathrm{tr}(MJMJ) - \mathrm{tr}(J'MJ)\}$$

$$+4\sigma^4\mu^3(\beta'X'J'MJJ'MJX\beta - \beta'X'J'MJMJMJX\beta)$$

$$+2\sigma^3\mu^2\gamma_1\beta'X'J'M[(I\odot MJ) - 2\mu(I\odot MJX\beta\beta'X'J'MJ)]\iota + o(T^{-4}). \tag{10}$$

We see that the leading term of the MSE is of order T^{-3}. The skewness coefficient contributes to the approximate MSE at order T^{-3}.

Corollary 2: *If $X = \iota$ in the model of Theorem 2, then the MSE and the variance simplify to*

$$\mathrm{E}[(\hat{\lambda}-1)^2] = 12\left(\frac{\sigma}{\beta}\right)^2\frac{1}{T^3} + \frac{48}{5}\gamma_1\left(\frac{\sigma}{\beta}\right)^3\frac{1}{T^3} + \frac{336}{5}\left(\frac{\sigma}{\beta}\right)^4\frac{1}{T^4} + o(T^{-4}), \tag{11}$$

and

$$\mathrm{Var}(\hat{\lambda}) = 12\left(\frac{\sigma}{\beta}\right)^2\frac{1}{T^3} + \frac{48}{5}\gamma_1\left(\frac{\sigma}{\beta}\right)^3\frac{1}{T^3} + \frac{156}{5}\left(\frac{\sigma}{\beta}\right)^4\frac{1}{T^4} + o(T^{-4}). \tag{12}$$

From this corollary, we see that the leading term $12T^{-3}(\sigma/\beta)^2$ gives the asymptotic variance of $T^{3/2}(\hat{\lambda}-1)$. Apparently, it suggests again that for the random walk with drift model, the approximate moments of $\hat{\lambda}$ depend crucially on σ/β. For a given sample size T, the quality of the approximation will go down as $|\sigma/\beta|$ goes up. The distribution affects the approximate MSE through its skewness coefficient γ_1. In Bao (2007), it is found that for the stationary AR model with an intercept, the second-order MSE depends on not only the skewness but also the kurtosis of the disturbance distribution.

3 The Full Coefficient Vector Estimator

Now we consider the full coefficient vector $\alpha := (\lambda, \beta')'$. Define $Z := (y_{-1}, X)$ and the LS estimator is $\hat{\alpha} = (Z'Z)^{-1}Z'y$. We relax Assumption 1 (iii) so that X may contain a linear trend. As $\hat{\lambda}$ and $\hat{\beta}$ might have different convergence rates, we follow Kiviet and Phillips (2005) to define a nonstochastic diagonal matrix $D = \mathrm{diag}(T^{\delta_i}, \cdots, T^{\delta_{k+1}})$, where $\delta_i <= 0$, $i = 1, \cdots, k+1$, such that $DZ'ZD = O_P(T)$.

Let $\bar{Z} = \mathrm{E}(Z)$, $\tilde{Z} = Z - \bar{Z}$, $W = ZD$, $\bar{W} = \mathrm{E}(W) = \bar{Z}D$, $\tilde{W} = W - \bar{W} = \tilde{Z}D$. Conditional on X and y_0, we can write $y_{-1} - \mathrm{E}(y_{-1}) = \sigma J\varepsilon$ and thus $\tilde{Z} = \sigma J\varepsilon e_1'$, where $e_1 = (1, 0, \cdots, 0)'$ is $(k+1)\times 1$. Then $\tilde{W} = \tilde{Z}D = \sigma J\varepsilon e_1'D$. One can verify that $\bar{W}'\bar{W} = O(T)$, $\bar{W}'\tilde{W} = O_P(T^{1/2})$, $\tilde{W}'\tilde{W} = O_P(1)$, $\bar{W}'\varepsilon = O_P(T^{1/2})$, and $\tilde{W}'\varepsilon = O_P(1)$.

The rescaled estimation error may be written as

$$\begin{aligned} D^{-1}(\hat{\alpha}-\alpha) &= \sigma(DZ'ZD)^{-1}DZ'\varepsilon \\ &= \sigma(\bar{W}'\bar{W}+\bar{W}'\tilde{W}+\tilde{W}'\bar{W}+\tilde{W}'\tilde{W})^{-1}(\bar{W}+\tilde{W})'\varepsilon \\ &= \sigma R(I+P+S)^{-1}(\bar{W}+\tilde{W})'\varepsilon, \end{aligned}$$

where $R=(\bar{W}'\bar{W})^{-1}=O(T^{-1})$, $P=\bar{W}'\tilde{W}R+\tilde{W}'\bar{W}R=O_P(T^{-1/2})$, $S=\tilde{W}'\tilde{W}R=O_P(T^{-1})$. Thus, for the rescaled estimator, up to order $O_P(T^{-1})$, we have the following expansion,

$$\begin{aligned} D^{-1}(\hat{\alpha}-\alpha) &= \sigma R(I-P)(\bar{W}+\tilde{W})'\varepsilon \\ &= \underbrace{\sigma R\bar{W}'\varepsilon}_{O_P(T^{-1/2})} + \underbrace{\sigma R\tilde{W}'\varepsilon-\sigma RP\bar{W}'\varepsilon}_{O_P(T^{-1})} + o_P(T^{-1}). \end{aligned} \tag{13}$$

Obviously, by substituting $\tilde{W}=\sigma J\varepsilon e_1'D$, we put $\tilde{W}'\varepsilon=\sigma D'e_1\varepsilon'J'\varepsilon$ and $P\bar{W}'\varepsilon=(\bar{W}'\tilde{W}+\tilde{W}'\bar{W})R\bar{W}'\varepsilon=\sigma\bar{W}'J\varepsilon e_1'DR\bar{W}'\varepsilon+\sigma D'e_1\varepsilon'J'\bar{W}R\bar{W}'\varepsilon$, where the random terms have expectations $\mathrm{E}(\varepsilon'J'\varepsilon)=0$, $\mathrm{E}(\varepsilon e_1'DR\bar{W}'\varepsilon)=\bar{W}RDe_1$, and $\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'\varepsilon)=\mathrm{tr}(J'\bar{W}R\bar{W}')$, regardless of the distribution of ε_t. Thus, Theorem 3 of Kiviet and Phillips (2005) regarding the bias of $\hat{\alpha}$, up to order $O(T^{-1+\delta_i})$, is robust to nonnormality. For completeness, we reproduce here.

Theorem 3: *In the first-order dynamic regression model, where the coefficient of the lagged dependent variable λ is equal to unity, the regressor matrix $Z=(y_{-1},X)$, and the scaling matrix $D=$diag$(T^{\delta_1},\cdots,T^{\delta_{k+1}})$ such that conditional on X and y_0, $DZ'ZD=O_P(T)$, the bias of the the least squares estimator of the coefficient vector $\alpha=(\lambda,\beta')'$, conditional on X and y_0, can be approximated, provided that $\bar{Z}=(y_0\iota+JX\beta,\ X)$ has full column rank and β is finite and non-zero, as follows:*

$$\mathrm{E}(\hat{\beta}_i-\beta)=-\sigma^2e_{i+1}'\left[(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}+\frac{1}{2}(T-k-1)I\right](\bar{Z}'\bar{Z})^{-1}e_1+o(T^{-1+\delta_{i+1}}), \tag{14}$$

and

$$\mathrm{E}(\hat{\lambda}-1)=-\frac{1}{2}(T-k)\sigma^2e_1'(\bar{Z}'\bar{Z})^{-1}e_1+o(T^{-1+\delta_1}). \tag{15}$$

One can check immediately that $\mathrm{E}(\hat{\lambda}-1)$ in (15) corresponds to (7).To approximate the MSE of $\hat{\alpha}$, we can further expand

$$\begin{aligned} D^{-1}(\hat{\alpha}-\alpha) &= \sigma R(I-P-S+PP)\bar{W}'\varepsilon + \sigma R(I-P)\tilde{W}'\varepsilon + o_P(T^{-3/2}) \\ &= \underbrace{\sigma R\bar{W}'\varepsilon}_{O_P(T^{-1/2})} + \underbrace{\sigma R\tilde{W}'\varepsilon - \sigma RP\bar{W}'\varepsilon}_{O_P(T^{-1})} + \underbrace{\sigma R(PP-S)\bar{W}'\varepsilon - \sigma RP\tilde{W}'\varepsilon}_{O_P(T^{-3/2})} + o_P(T^{-3/2}). \end{aligned} \tag{16}$$

So the MSE of $D^{-1}(\hat{\alpha}-\alpha)$, up to $O(T^{-2})$, can be written as

$$\begin{aligned} \mathrm{E}[D^{-1}(\hat{\alpha}-\alpha)(\hat{\alpha}-\alpha)'D^{-1}] &= \sigma^2 R + \sigma^2 R\mathrm{E}[(\tilde{W}'\varepsilon - P\bar{W}'\varepsilon)(\tilde{W}'\varepsilon - P\bar{W}'\varepsilon)']R \\ &\quad + \sigma^2 R\bar{W}'\mathrm{E}[\varepsilon(\tilde{W}'\varepsilon - P\bar{W}'\varepsilon)']R + \sigma^2 R\mathrm{E}[(\tilde{W}'\varepsilon - P\bar{W}'\varepsilon)\varepsilon']\bar{W}R \\ &\quad + \sigma^2 R\bar{W}'\varepsilon[R(PP-S)\bar{W}'\varepsilon - \sigma RP\tilde{W}'\varepsilon]' \\ &\quad + \sigma^2[R(PP-S)\bar{W}'\varepsilon - \sigma RP\tilde{W}'\varepsilon]\varepsilon'\bar{W}R + o(T^{-2}) \\ &= \sigma^2 R + \sigma^2 R\mathrm{E}[\tilde{W}'\varepsilon\varepsilon'\tilde{W} + P\bar{W}'\varepsilon\varepsilon'\bar{W}P' - P\bar{W}'\varepsilon\varepsilon'\tilde{W} - \tilde{W}'\varepsilon\varepsilon'\bar{W}P' \\ &\quad + \bar{W}'\varepsilon\varepsilon'\tilde{W} - \bar{W}'\varepsilon\varepsilon'\bar{W}P' + \tilde{W}'\varepsilon\varepsilon'\bar{W} - P\bar{W}'\varepsilon\varepsilon'\bar{W} \\ &\quad + \bar{W}'\varepsilon\varepsilon'\bar{W}P'P' - \bar{W}'\varepsilon\varepsilon'\bar{W}S' - \bar{W}'\varepsilon\varepsilon'\tilde{W}P' \\ &\quad + PP\bar{W}'\varepsilon\varepsilon'\bar{W} - S\bar{W}'\varepsilon\varepsilon'\bar{W} - P\tilde{W}'\varepsilon\varepsilon'\bar{W}]R + o(T^{-2}), \end{aligned} \tag{17}$$

By by substitution, simplification, and omission of terms of smaller orders, see Appendix A, we obtain

$$\begin{aligned} \mathrm{E}[D^{-1}(\hat{\alpha}-\alpha)(\hat{\alpha}-\alpha)'D^{-1}] &= \sigma^2 R + \sigma^4 R\{[-\mathrm{tr}(J'J) + \mathrm{tr}^2(\bar{W}R\bar{W}'J) + \mathrm{tr}(\bar{W}'JJ'\bar{W}R) \\ &\quad + \mathrm{tr}(\bar{W}R\bar{W}'J\bar{W}R\bar{W}'J) - 2\mathrm{tr}(\bar{W}R\bar{W}'J'J')]De_1e_1'D \\ &\quad + \bar{W}'(JJ' - J'J' - J'J)\bar{W}RDe_1e_1'D \\ &\quad + De_1e_1'DR\bar{W}'(JJ' - J'J' - J'J)\bar{W} \\ &\quad + e_1'DRDe_1\bar{W}'(JJ' - J'J' - J'J)\bar{W} \\ &\quad + e_1'DRDe_1(\bar{W}'J\bar{W}R\bar{W}'J\bar{W} + \bar{W}'J'\bar{W}R\bar{W}'J'\bar{W}) \\ &\quad + [e_1'DR\bar{W}'J\bar{W}RDe_1 + \mathrm{tr}(\bar{W}R\bar{W}'J)e_1'DRDe_1]\bar{W}'(J+J')\bar{W}\}R \\ &\quad - \sigma^3\gamma_1 R[\bar{W}'\left(\bar{W}RDe_1\iota' \odot I\right)J'\bar{W} + \bar{W}'J\left(\bar{W}RDe_1\iota' \odot I\right)\bar{W}]R + o(T^{-2}), \end{aligned}$$

which leads to the following theorem:

Theorem 4: *In the model of Theorem 3, the elements of the MSE($\hat{\alpha}$) matrix, that is,* $\mathrm{E}[(\hat{\alpha}_i-\alpha_i)(\hat{\alpha}_j-\alpha_j)]$,

$i, j = 1, \cdots, k+1$, are given by

$$\begin{aligned}
&\sigma^2 e_i' Q e_j + \sigma^4[\mathrm{tr}(Q\bar{Z}' J J' \bar{Z}) - 2\mathrm{tr}(Q\bar{Z}' J J \bar{Z}) - \mathrm{tr}(J'J) \\
&\quad + \mathrm{tr}(Q\bar{Z}' J \bar{Z} Q \bar{Z}' J \bar{Z}) + \mathrm{tr}^2(Q\bar{Z}' J \bar{Z})] e_i' D e_1 e_1' D e_j \\
&+\sigma^4 e_1' Q e_1 e_i' Q \bar{Z}'[JJ' - J'J' - J'J + J\bar{Z} Q \bar{Z}' J + J'\bar{Z} Q \bar{Z}' J'] \bar{Z} Q e_j \\
&+\sigma^4 e_1' D e_j e_i' Q \bar{Z}'(JJ' - J'J' - J'J) \bar{Z} Q e_1 \\
&+\sigma^4 e_1' Q e_i e_1' Q \bar{Z}'(JJ' - J'J' - J'J) \bar{Z} Q e_j \\
&+\sigma^4[e_1' Q \bar{Z}' J \bar{Z} Q e_1 + \mathrm{tr}(Q\bar{Z}' J \bar{Z}) e_1' Q e_1] e_i' Q \bar{Z}'(J + J') \bar{Z} Q e_j \\
&-\sigma^3 \gamma_1 e_i' Q \bar{Z}'[\left(\bar{Z} Q e_1 \iota' \odot I\right) J' + J\left(\bar{Z} Q e_1 \iota' \odot I\right)] \bar{Z} Q e_j + o(T^{-2+\delta_i+\delta_j}),
\end{aligned} \tag{18}$$

where $Q = (\bar{Z}'\bar{Z})^{-1}$, $\bar{Z} =$E(Z), and e_i is the i-th unit vector.

From this theorem we can obtain, for example, when $k = 2$ with the first column of X being the constant and the second being a linear trend, the approximate MSE of $\hat{\beta}_1$ up to $O(T^{-2})$, that of $\hat{\beta}_2$ up to $O(T^{-4})$, and that of $\hat{\lambda}$ up to $O(^{-6})$ (if $\beta_2 \neq 0$ and up to $O(T^{-4})$ otherwise). For a model with stationary X, the orders of approximations to the MSE of $\hat{\lambda}$ and β_j, $j = 1, \cdots, k$, are $O(T^{-4})$ and $O(T^{-2})$, respectively. Compared with Kiviet and Phillips (2005), we see that the in (18) the approximate MSE has an additional term that involves the skewness coefficient γ_1.

4 Approximation Accuracy

In this section, we investigate how our moment approximations fare under different distributions, sample sizes, and the noise to signal ratio. We focus on the random walk with drift model $y_t^* = \lambda y_{t-1}^* + \beta^* + \varepsilon_t$, where $y_0^* = 0$, $\lambda = 1$, $\beta^* = \beta/\sigma \neq 0$. Recall that the approximate moments are invariant to the starting value. Since the approximate moments are related to the ratio β/σ, but not to β and σ separately, we rescale the model so that the error term has unit variance. We consider five nonnormal distributions for ε, including both symmetric and asymmetric case. These five nonnormal distributions are: standard normal, uniform on $[0, 1]$, exponential distribution exp(1), student-t distribution with 5 degrees of freedom, mixture of two normals of N$(-3, 1)$ and N$(3, 1)$, with probability equal to 0.2, 0.8 respectively, and log normal distribution lnN$(0, 1)$. The corresponding γ_1 are 0, 2, 0, -1.1798, and 8.1073, respectively. For comparison purpose, we also include the results under a standard normal distribution. Our estimates of the true bias, MSE, and variance are based on 10,000,000 simulations and we also provide the estimated standard errors of our Monte Carlo estimates, denoted by MCSE.

We experiment with $T = 10, 15, 20, 40, 80$, $\beta^* = 10, 5, 2, 1, 0.5, 0.2, 0.1$. As indicated in the previous sections, for small values of β^* (equivalently, large values of the noise to signal ratio), the accuracy of our approximations is expected to deteriorate. Similar to Kiviet and Phillips (2005), we employ an expansion in the orders of β^* and the approximate bias can be defined as

$$\mathrm{E}(\hat{\lambda} - 1) = g_0(T) - g_{1/2}(T)\beta^* - g_1(T)\beta^{*2} + o(\beta^{*2}) \tag{19}$$

where

$$g_0(T) = \mathrm{E}\left(\frac{\varepsilon' J' A \varepsilon}{\varepsilon' J' A J \varepsilon}\right), \qquad g_{1/2}(T) = 2\mathrm{E}\left[\frac{\iota' J' A J \varepsilon \varepsilon' J' A \varepsilon}{(\varepsilon' J' A J \varepsilon)^2}\right],$$
$$g_1(T) = \mathrm{E}\left[\frac{\iota' J' A J \iota \varepsilon' J' A \varepsilon + 2\varepsilon' A J \iota \iota' J' A J \varepsilon}{(\varepsilon' J' A J \varepsilon)^2} - \frac{4\varepsilon' J' A \varepsilon \varepsilon' J' A J \iota \iota' J' A J \varepsilon}{(\varepsilon' J' A J \varepsilon)^3}\right],$$

in which $A = I_T - T^{-1}\iota\iota'$. Under normality, $g_0(T)$, $g_{1/2}(T)$, and $g_1(T)$ might be numerically calculated via the moment-generating function approach of Sawa (1978). With a general nonnormal distribution, neither analytical nor numerical approach can be straightforwardly applied. So we follow Kiviet and Phillips (2005) to obtain them by simulations.

To save space, we report only the results pertaining to $\beta^* = 5, 1, 0.2$, presented in Tables 1-6, whereas the results under other values of β^* are available upon request from the corresponding author.

We observe that in general our analytical formulae (6), (11), (12), (14), and (18) capture well the true moments of the LS estimator for large and moderate β/σ, across different distributions. However, on occasions where there is substantial bias, and these are the cases when the signal to noise ratio is small (0.2), our formulae approximate poorly the true moments, and sometimes the order T^{-3} bias approximations are even worse than the order T^{-2} approximations. For large values of β/σ, the higher-order approximations are typically better. The difference between $O(\sigma^2)$ ($O(\sigma^3)$) and $O(T^{-2})$ ($O(T^{-3})$) bias approximations is in general small. Under small T and small β^*, the bias in $\hat{\lambda}$ can be quite substantial, and only the $O(\beta^{*2})$ approximation captures the bias well.

The bias of the estimated intercept increases when the intercept decreases and it is very substantial when β/σ is small. And it seems to decrease with T not as quickly as the bias of $\hat{\lambda}$. When β/σ is not small, even for a small $T = 10$, our analytical formulae perform reasonably well in approximating the true moments of $\hat{\lambda}$ and $\hat{\beta}$, regardless of the disturbance distribution.

Note that the effect of nonnormality on the finite-sample moments of the LS estimator is not as strong as other aspects of the model, especially the value of β/σ. In particular, the bias, variance, and MSE approximations developed under normality seem quite robust to the distributions we have experimented with, though we see more variations in those for $\hat{\beta}$. Of course, a more comprehensive conclusion regarding the effects of nonnormality has yet to be drawn, given that in reality there are countless possibilities of nonnormal distributions.

5 Concluding Remarks

We have derived analytical moment approximation for the LS estimator in ARX(1) with unit root, where an arbitrary number of possibly nonstationary exogenous regressors may be included. In this framework, no distribution assumption is made except that we assume the existence of moments of the errors up to order 4. In the special case of random walk with drift, very neat bias and MSE formulae are presented. Numerical illustrations of the accuracy of our moment approximations in finite samples are also provided.

Note that we have restrained ourselves from deriving moment approximation for the random walk model $y = \lambda y_{-1} + \sigma\varepsilon$, $\lambda = 1$. One can show that $\hat{\lambda} - 1 = \sigma(y_0\iota'\varepsilon + \sigma\varepsilon' J\varepsilon)/(y_0^2 T + 2\sigma y_0 \iota' J\varepsilon + \sigma^2\varepsilon' J' J\varepsilon)$, where $y_0\iota'\varepsilon = O_P(T^{1/2})$, $\varepsilon' J\varepsilon = O_P(T)$, $\iota' J\varepsilon = O_P(T^{3/2})$, and $\varepsilon' J' J\varepsilon = O_P(T^2)$. Thus, a possible expansion may be as follows:

$$\begin{aligned}
\hat{\lambda} - 1 &= \left(\underbrace{\frac{\sigma y_0 \iota'\varepsilon}{\sigma^2\varepsilon' J' J\varepsilon}}_{O_P(T^{-3/2})} + \underbrace{\frac{\sigma^2\varepsilon' J\varepsilon}{\sigma^2\varepsilon' J' J\varepsilon}}_{O_P(T^{-1})}\right)\left(1 + \underbrace{\frac{2\sigma y_0\iota' J\varepsilon}{\sigma^2\varepsilon' J' J\varepsilon}}_{O_P(T^{-1/2})} + \underbrace{\frac{y_0^2 T}{\sigma^2\varepsilon' J' J\varepsilon}}_{O_P(T^{-1})}\right)^{-1} \\
&= \left(\frac{\varepsilon' J\varepsilon}{\varepsilon' J' J\varepsilon} + \frac{y_0\iota'\varepsilon}{\sigma\varepsilon' J' J\varepsilon}\right)\left[1 - \frac{2_0\iota' J\varepsilon}{\sigma\varepsilon' J' J\varepsilon} - \frac{y_0^2 T}{\sigma^2\varepsilon' J' J\varepsilon} + \left(\frac{2y_0\iota' J\varepsilon}{\sigma\varepsilon' J' J\varepsilon} + \frac{y_0^2 T}{\sigma^2\varepsilon' J' J\varepsilon}\right)^2 + o_P(T^{-1})\right] \\
&= \underbrace{\frac{\varepsilon' J\varepsilon}{\varepsilon' J' J\varepsilon}}_{O_P(T^{-1})} + \underbrace{\frac{y_0}{\sigma}\frac{\iota'\varepsilon}{\varepsilon' J' J\varepsilon} - 2\frac{y_0}{\sigma}\frac{\iota' J\varepsilon\varepsilon' J\varepsilon}{(\varepsilon' J' J\varepsilon)^2}}_{O_P(T^{-3/2})} \\
&\quad \underbrace{+4\left(\frac{y_0}{\sigma}\right)^2\frac{\varepsilon' J\varepsilon\varepsilon' J'\iota\iota' J\varepsilon}{(\varepsilon' J' J\varepsilon)^3} - \left(\frac{y_0}{\sigma}\right)^2\frac{T\varepsilon' J\varepsilon}{(\varepsilon' J' J\varepsilon)^2} - 2\left(\frac{y_0}{\sigma}\right)^2\frac{\varepsilon'\iota\iota' J\varepsilon}{(\varepsilon' J' J\varepsilon)^2}}_{O_P(T^{-2})} + o_P(T^{-2}).
\end{aligned}$$

Under nonnormality, it is not obvious how to evaluate directly the above expectations of ratios. We may attempt to implement further expansions for the random ratios. But unfortunately,

$$\frac{\varepsilon' J\varepsilon}{\varepsilon' J' J\varepsilon} = \frac{\varepsilon' J\varepsilon}{\mathrm{E}\left(\varepsilon' J' J\varepsilon\right)}\left[1 + \frac{\varepsilon' J' J\varepsilon - \mathrm{E}\left(\varepsilon' J' J\varepsilon\right)}{\mathrm{E}\left(\varepsilon' J' J\varepsilon\right)}\right]^{-1},$$

is not going to work, as $\mathrm{E}(\varepsilon' J' J\varepsilon) = O(T^2)$ and $\varepsilon' J' J\varepsilon - \mathrm{E}(\varepsilon' J' J\varepsilon) = O(T^2)$, and we are not aware of straightforward expansions that allow us to calculate analytically the moments under nonnormality, although Phillips (2012) derived bias expansions under normality with the help of moment generation function of ratio of normal quadratic forms.

One area of future research is to explore the finite-sample properties of predictive regressions, where the lagged regressor follows a unit-root process. Phillips (2013) considered the asymptotic results when the regressor has local-to-unit behavior and Phillips and Lee (2013) generalized this to the multivariate framework.

Given the analytical bias and MSE formulae developed in this paper, one might follow Phillips (2012) to consider the indirect inference estimator and Abadir (1995) to consider the minimum MSE estimator. We leave these for our future research.

References

Abadir, K.M. (1993). OLS bias in a nonstationary regression. *Econometric Theory* 9, 81–93.

Abadir, K.M. (1995). Unbiased estimation as a solution to testing for random walks. *Economics Letters* 47, 263–268.

Anderson, T.W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676–687.

Bao, Y. (2007). The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Econometric Theory* 23, 1013–1021.

Bao. Y. and A. Ullah (2007). The second-order bias and mean squared error of estimators in time series models. *Journal of Econometrics* 140, 650–669.

Evans, G.B.A. and N.E. Savin (1981). Testing unit roots: I. *Econometrica* 49, 753–779.

Grubb, D. and J. Symons (1987). Bias in regressions with a lagged-dependent variable. *Econometric Theory* 3, 371–86.

Hurwicz, L. (1950). Least-squares bias in time series. In T. C. Koopmans (ed.) *Statistical Inference in Dynamic Economic Models*, pp 365–383. New York: Wiley.

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.

Kiviet, J.F. and G.D.A. Phillips (1993). Alternative bias approximations in regressions with a lagged dependent variable. *Econometric Theory* 9, 62–80.

Kiviet, J.F. and G.D.A. Phillips (2005). Moment approximation for least-squares estimators in dynamic regression models with a unit root. *The Econometrics Journal* 8, 115–142.

Malinvaud, E. (1970). *Statistical Methods of Econometrics (2nd ed.).* Amsterdam: North Holland.

Phillips, P.C.B. (1977a). Approximations to some finite sample distributions associated with a first order stochastic difference equation. *Econometrica* 45, 463–485.

Phillips, P.C.B. (1977b). A general theorem in the theory of asymptotic expansions as approximations to finite sample distributions of econometric estimators. *Econometrica* 45, 1517–1534.

Phillips, P.C.B. (1978). Edgeworth and saddlepoint approximations in a first order non-circular autoregression. *Biometrik*a 65, 91–98.

Phillips, P.C.B. (1979). The sampling distribution of forecasts from a first order autoregression. *Journal of Econometrics* 9, 241–261.

Phillips, P.C.B. (1987a) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (2012). Folklore theorems, implicit maps, and indirect inference. *Econometrica* 80, 425–454.

Phillips, P.C.B. (2013). On Confidence intervals for autoregressive roots and predictive regression. *Econometrica*, forthcoming.

Phillips, P.C.B. and J.H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics* 177, 250–264.

Sawa, T. (1978). The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics* 8, 159–72.

Tsui, A.K. and M.M. Ali (1994). Exact distributions, density functions and moments of the least squares estimator in a first-order autoregressive model. *Computational Statistics and Data Analysis* 17, 433–454.

Ullah, A. (2004) *Finite Sample Econometrics.* New York: Oxford University Press.

White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197

White, J.S. (1959). The Limiting distribution of the serial correlation coefficient in the explosive case II. *Annals of Mathematical Statistics* 30, 831–834

Appendix A: Expectations of Quadratic Forms

To evaluate the expectations of quadratic forms under nonnormality, we can use the following results (see, for example, Ullah (2004)):

$$\mathrm{E}(\varepsilon\varepsilon' A\varepsilon) = \gamma_1 (I \odot A)\iota, \quad \mathrm{E}(\varepsilon' A_1 \varepsilon\varepsilon' A_2 \varepsilon) = \mathrm{tr}(A_1)\mathrm{tr}(A_2) + \mathrm{tr}(A_1 A_2) + \mathrm{tr}(A_1 A_2') + \gamma_2 \mathrm{tr}(A_1 \odot A_2). \tag{20}$$

In addition,

$$\mathrm{E}(\varepsilon\varepsilon' A\varepsilon\varepsilon') = [\mathrm{tr}(A)I + A + A' + \gamma_2 (A \odot I)]. \tag{21}$$

Based on (20) and (21), and observing that $MJ = O(1)$, $JJ' = O(T)$, $MJX = O(T)$, $J'MJ = O(T)$, $J'MJX = O(T^2)$, and $J + J' = 0.5(\iota\iota' - I)$, we collect the following:

$\mathrm{E}(\varepsilon' MJ\varepsilon) = \mathrm{tr}(MJ) = 0.5\mathrm{tr}(M(J + J')) = 0.5\mathrm{tr}(M(\iota\iota' - I)) = -0.5(T - k) = O(T)$,

$\mathrm{E}(\varepsilon' MJX\beta\beta' X' J' MJ\varepsilon) = \beta' X' J' MJMJX\beta = -0.5\beta' X' J' MJX\beta = -0.5\mu^{-1} = O(T^3)$,

$\beta' X' J' M\mathrm{E}(\varepsilon\varepsilon' J' MJX\beta\beta' X' J' MJ\varepsilon) = \gamma_1 \beta' X' J' M(I \odot J' MJX\beta\beta' X' J' MJ)\iota = O(T^6)$,

$\beta' X' J' M\mathrm{E}(\varepsilon\varepsilon' J' MJ\varepsilon) = \gamma_1 \beta' X' J' M(I \odot J' MJ)\iota = O(T^3)$,

$$\beta'X'J'MJ\mathrm{E}(\varepsilon\varepsilon'MJ\varepsilon) = \gamma_1\beta'X'J'MJ(I \odot MJ)\iota = O(T^3),$$

$$\begin{aligned}
\mathrm{E}(\varepsilon'J'MJ\varepsilon\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon) =& \mathrm{tr}(J'MJ)\mathrm{tr}(MJX\beta\beta'X'J'MJ) + 2\mathrm{tr}(J'MJMJX\beta\beta'X'J'MJ)\\
&+\gamma_2\mathrm{tr}(J'MJ \odot MJX\beta\beta'X'J'MJ)\\
=&\mathrm{tr}(J'MJ)\beta'X'J'MJMJX\beta + 2\beta'X'J'MJJ'MJMJX\beta + +\gamma_2\mathrm{tr}(J'MJ \odot MJX\beta\beta'X'J'MJ)\\
=& O(T^5),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon\varepsilon'J'MJX\beta\beta'X'J'MJ\varepsilon) =& \mathrm{tr}(MJX\beta\beta'X'J'MJ)\mathrm{tr}\big(J'MJX\beta\beta'X'J'MJ\big)\\
&+2\mathrm{tr}\big(MJX\beta\beta'X'J'MJJ'MJX\beta\beta'X'J'MJ\big) + \gamma_2\mathrm{tr}(MJX\beta\beta'X'J'MJ \odot J'MJX\beta\beta'X'J'MJ)\\
=& 3\beta'X'J'MJMJX\beta\beta'X'J'MJJ'MJX\beta + \gamma_2\mathrm{tr}(MJX\beta\beta'X'J'MJ \odot J'MJX\beta\beta'X'J'MJ)\\
=& -1.5\beta'X'J'MJX\beta\beta'X'J'MJJ'MJX\beta + \gamma_2\mathrm{tr}(MJX\beta\beta'X'J'MJ \odot J'MJX\beta\beta'X'J'MJ)\\
=& O(T^8),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\varepsilon'MJ\varepsilon\varepsilon'J'MJX\beta\beta'X'J'MJ\varepsilon) =& \mathrm{tr}(MJ)\mathrm{tr}\big(J'MJX\beta\beta'X'J'MJ\big) + 2\mathrm{tr}\big(MJJ'MJX\beta\beta'X'J'MJ\big)\\
&+\gamma_2\mathrm{tr}(MJ \odot J'MJX\beta\beta'X'J'MJ)\\
=&\mathrm{tr}(MJ)\beta'X'J'MJJ'MJX\beta + 2\beta'X'J'MJMJJ'MJX\beta + \gamma_2\mathrm{tr}(MJ \odot J'MJX\beta\beta'X'J'MJ)\\
=&\mathrm{tr}(MJ)\beta'X'J'MJJ'MJX\beta + o(T^6),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\varepsilon'MJ\varepsilon\varepsilon'J'MJ\varepsilon) =& \mathrm{tr}(MJ)\mathrm{tr}(JMJ) + 2\mathrm{tr}(MJJ'MJ) + \gamma_2\mathrm{tr}(MJ \odot J'MJ)\\
=&\mathrm{tr}(MJ)\mathrm{tr}(J'MJ) - \mathrm{tr}(J'MJ) + \gamma_2\mathrm{tr}(MJ \odot J'MJ)\\
=&\mathrm{tr}(MJ)\mathrm{tr}(J'MJ) + o(T^3),
\end{aligned}$$

$$\mathrm{E}(\varepsilon'MJX\beta\beta'X'J'M\varepsilon) = \beta'X'J'MJX\beta = \mu^{-1} = O(T^3),$$

$$\beta'X'J'M\mathrm{E}(\varepsilon\varepsilon'MJ\varepsilon) = \gamma_1\beta'X'J'M(I \odot MJ)\iota = O(T^2),$$

$$\beta'X'J'M\mathrm{E}(\varepsilon\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon) = \gamma_1\beta'X'J'M(I \odot MJX\beta\beta'X'J'MJ)\iota = O(T^5),$$

$$\begin{aligned}
\mathrm{E}[(\varepsilon'MJ\varepsilon)^2] =& [\mathrm{tr}(MJ)]^2 + \mathrm{tr}(MJMJ) + \mathrm{tr}(J'MJ) + \gamma_2\mathrm{tr}(MJ \odot MJ)\\
=& [\mathrm{tr}(MJ)]^2 + \mathrm{tr}(MJMJ) + \mathrm{tr}(J'MJ) + o(T^2),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\varepsilon'MJ\varepsilon\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon) =& \mathrm{tr}(MJ)\beta'X'J'MJMJX\beta + \mathrm{tr}(MJMJX\beta\beta'X'J'MJ) + \mathrm{tr}(J'MJX\beta\beta'X'J'MJ)\\
&+\gamma_2\mathrm{tr}(MJ \odot MJX\beta\beta'X'J'MJ)\\
=& -0.5\mu^{-1}\mathrm{tr}(MJ) + \beta'X'J'MJMJMJX\beta + \beta'X'J'MJJ'MJX\beta + \gamma_2\mathrm{tr}(MJ \odot MJX\beta\beta'X'J'MJ)\\
=& \beta'X'J'M(JMJ + JJ')MJX\beta + o(T^5),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\varepsilon'J'MJ\varepsilon\varepsilon'MJX\beta\beta'X'J'M\varepsilon) =& \mu^{-1}\mathrm{tr}(J'MJ) + 2\beta'X'J'MJ'MJMJX\beta + \gamma_2\mathrm{tr}(J'MJ \odot MJX\beta\beta'X'J'M)\\
=& \mu^{-1}\mathrm{tr}(J'MJ) + 2\beta'X'J'MJ'MJMJX\beta + o(T^5),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[(\varepsilon'MJX\beta\beta'X'J'MJ\varepsilon)^2] =& (\beta'X'J'MJMJX\beta)^2 + \mathrm{tr}(MJX\beta\beta'X'J'MJMJX\beta\beta'X'J'MJ)\\
&+\mathrm{tr}(MJX\beta\beta'X'J'MJJ'MJX\beta\beta'X'J'M) + \gamma_2\mathrm{tr}(MJX\beta\beta'X'J'MJ \odot MJX\beta\beta'X'J'MJ)\\
=& 2(\beta'X'J'MJMJX\beta)^2 + \beta'X'J'MJJ'MJX\beta\beta'X'J'MJX\beta + \gamma_2\mathrm{tr}(MJX\beta\beta'X'J'MJ \odot MJX\beta\beta'X'J'MJ)\\
=& 0.5\mu^{-2} + \mu^{-1}\beta'X'J'MJJ'MJX\beta + \gamma_2\mathrm{tr}(MJX\beta\beta'X'J'MJ \odot MJX\beta\beta'X'J'MJ)\\
=& \mu^{-1}\beta'X'J'MJJ'MJX\beta + o(T^8),
\end{aligned}$$

$$\mathrm{E}(\tilde{W}'\varepsilon\varepsilon'\tilde{W}) = \sigma^2 D'e_1\mathrm{E}(\varepsilon'J'\varepsilon\varepsilon'J\varepsilon)e_1'D = \sigma^2\mathrm{tr}(J'J)De_1e_1'D,$$

$$\begin{aligned}
\mathrm{E}(P\bar{W}'\varepsilon\varepsilon'\bar{W}P') =& \sigma^2\bar{W}'J\mathrm{E}(\varepsilon e_1'DR\bar{W}'\varepsilon\varepsilon'\bar{W}RDe_1\varepsilon')J'\bar{W} + \sigma^2 De_1\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon'\bar{W}RDe_1\varepsilon')J'\bar{W}\\
&+\sigma^2\bar{W}'J\mathrm{E}(\varepsilon e_1'DR\bar{W}'\varepsilon\varepsilon'\bar{W}R\bar{W}'J\varepsilon)e_1'D + \sigma^2 De_1\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon'\bar{W}R\bar{W}'J\varepsilon)e_1'D\\
=& \sigma^2\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'\bar{W}RDe_1e_1'DR\bar{W}'\varepsilon\varepsilon')J'\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'\mathrm{E}(\varepsilon\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon')J'\bar{W}\\
&+\sigma^2\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'\bar{W}R\bar{W}'J\varepsilon\varepsilon')\bar{W}RDe_1e_1'D + \sigma^2\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon'\bar{W}R\bar{W}'J\varepsilon)De_1e_1'D
\end{aligned}$$

$$
\begin{aligned}
&= \sigma^2 e_1' DRDe_1 \bar{W}'JJ'\bar{W} + 2\sigma^2 \bar{W}'J\bar{W}RDe_1e_1'DR\bar{W}'J'\bar{W} + \sigma^2\gamma_2 \bar{W}'J(\bar{W}RDe_1e_1'DR\bar{W}' \odot I)J'\bar{W} \\
&\quad + \sigma^2 \mathrm{tr}(\bar{W}R\bar{W}'J)De_1e_1'DR\bar{W}'J'\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'J'\bar{W}R\bar{W}'J'\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'JJ'\bar{W} \\
&\quad + \sigma^2\gamma_2 De_1e_1'DR\bar{W}'(\bar{W}R\bar{W}'J \odot I)J'\bar{W} + \sigma^2 \mathrm{tr}(\bar{W}R\bar{W}'J)\bar{W}'J\bar{W}RDe_1e_1'D \\
&\quad + \sigma^2 \bar{W}'J\bar{W}R\bar{W}'J\bar{W}RDe_1e_1'D + \sigma^2\bar{W}'JJ'\bar{W}RDe_1e_1'D + \sigma^2\gamma_2\bar{W}'J(\bar{W}R\bar{W}'J \odot I)\bar{W}RDe_1e_1'D \\
&\quad + \sigma^2 \mathrm{tr}^2(\bar{W}R\bar{W}'J)De_1e_1'D + \sigma^2\mathrm{tr}(\bar{W}R\bar{W}'J\bar{W}R\bar{W}'J)De_1e_1'D + \sigma^2\mathrm{tr}(\bar{W}'JJ'\bar{W}R)De_1e_1'D \\
&\quad + \sigma^2\gamma_2 \mathrm{tr}(\bar{W}R\bar{W}'J \odot \bar{W}R\bar{W}'J)De_1e_1'D,
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(P\bar{W}'\varepsilon\varepsilon'\tilde{W}) &= \sigma^2\bar{W}'J\mathrm{E}(\varepsilon e_1'DR\bar{W}'\varepsilon\varepsilon'J\varepsilon)e_1'D + \sigma^2 De_1\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon'J\varepsilon)e_1'D \\
&= \sigma^2\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'J\varepsilon\varepsilon')\bar{W}RDe_1e_1'D + \sigma^2\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon'J\varepsilon)De_1e_1'D \\
&= \sigma^2\bar{W}'J(J+J')\bar{W}RDe_1e_1'D + \sigma^2\mathrm{tr}(J'\bar{W}R\bar{W}'(J+J'))De_1e_1'D,
\end{aligned}
$$

$$
\mathrm{E}(\bar{W}'\varepsilon\varepsilon'\tilde{W}) = \sigma\bar{W}'\mathrm{E}(\varepsilon\varepsilon'J\varepsilon)e_1'D = 0,
$$

$$
\begin{aligned}
\mathrm{E}(\bar{W}'\varepsilon\varepsilon'\bar{W}P') &= \sigma\bar{W}'\mathrm{E}(\varepsilon\varepsilon'\bar{W}R\bar{W}'J\varepsilon)e_1'D + \sigma\bar{W}'\mathrm{E}(\varepsilon\varepsilon'\bar{W}RDe_1\varepsilon')J'\bar{W} \\
&= \sigma\gamma_1\bar{W}'(I \odot \bar{W}R\bar{W}'J)\iota e_1'D + \sigma\gamma_1\bar{W}'\left(\bar{W}RDe_1\iota' \odot I\right)J'\bar{W},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(P\tilde{W}'\varepsilon\varepsilon'\bar{W}) &= \sigma^2\bar{W}'J\mathrm{E}(\varepsilon e_1'DRDe_1\varepsilon'J'\varepsilon\varepsilon')\bar{W} + \sigma^2 De_1\mathrm{E}(\varepsilon'J'\bar{W}RDe_1\varepsilon'J'\varepsilon\varepsilon')\bar{W} \\
&= \sigma^2 e_1'DRDe_1\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'J'\varepsilon\varepsilon')\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'J'\varepsilon\varepsilon')\bar{W} \\
&= \sigma^2 e_1'DRDe_1\bar{W}'J(J+J')\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'J(J+J')\bar{W},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(PP\bar{W}'\varepsilon\varepsilon'\bar{W}) &= \sigma^2\bar{W}'J\mathrm{E}(\varepsilon e_1'DR\bar{W}'J\varepsilon e_1'DR\bar{W}'\varepsilon\varepsilon')\bar{W} + \sigma^2 De_1\mathrm{E}(\varepsilon'J'\bar{W}R\bar{W}'J\varepsilon e_1'DR\bar{W}'\varepsilon\varepsilon')\bar{W} \\
&\quad + \sigma^2\bar{W}'J\mathrm{E}(\varepsilon e_1'DRDe_1\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon')\bar{W} + \sigma^2 De_1 E(\varepsilon'J'\bar{W}RDe_1\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon')\bar{W} \\
&= \sigma^2\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'J'\bar{W}RDe_1e_1'DR\bar{W}'\varepsilon\varepsilon')\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'\mathrm{E}(\varepsilon\varepsilon'J'\bar{W}R\bar{W}'J\varepsilon\varepsilon')\bar{W} \\
&\quad + \sigma^2 e_1'DRDe_1\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon')\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'J\mathrm{E}(\varepsilon\varepsilon'J'\bar{W}R\bar{W}'\varepsilon\varepsilon')\bar{W} \\
&= \sigma^2 e_1'DR\bar{W}'J\bar{W}RDe_1\bar{W}'J\bar{W} + \sigma^2\bar{W}'JJ'\bar{W}RDe_1e_1'D + \sigma^2\bar{W}'J\bar{W}RDe_1e_1'DR\bar{W}'J\bar{W} \\
&\quad + \sigma^2\gamma_2\bar{W}'J(J'\bar{W}RDe_1e_1'DR\bar{W}' \odot I)\bar{W} + \sigma^2\mathrm{tr}(\bar{W}R\bar{W}'JJ')De_1e_1'D + 2\sigma^2 De_1e_1'DR\bar{W}'J'\bar{W}R\bar{W}'J\bar{W} \\
&\quad + \sigma^2\gamma_2 De_1e_1'DR\bar{W}'(J'\bar{W}R\bar{W}'J \odot I)\bar{W} + \sigma^2\mathrm{tr}(\bar{W}R\bar{W}'J)e_1'DRDe_1\bar{W}'J\bar{W} + \sigma^2 e_1'DRDe_1\bar{W}'JJ'\bar{W} \\
&\quad + \sigma^2 e_1'DRDe_1\bar{W}'J\bar{W}R\bar{W}'J\bar{W} + \sigma^2\gamma_2 e_1'DRDe_1\bar{W}'J(\bar{W}R\bar{W}'J \odot I)\bar{W} + \sigma^2\mathrm{tr}(\bar{W}R\bar{W}'J)De_1e_1'DR\bar{W}'J\bar{W} \\
&\quad + \sigma^2 De_1e_1'DR\bar{W}'JJ'\bar{W} + \sigma^2 De_1e_1'DR\bar{W}'J\bar{W}R\bar{W}'J\bar{W} + \sigma^2\gamma_2 De_1e_1'DR\bar{W}'J(\bar{W}R\bar{W}'J \odot I)\bar{W},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\bar{W}'\varepsilon\varepsilon'\bar{W}S') &= \sigma^2\bar{W}'\mathrm{E}(\varepsilon\varepsilon'\bar{W}RDe_1\varepsilon'J'J\varepsilon)e_1'D = \sigma^2\bar{W}'\mathrm{E}(\varepsilon\varepsilon'J'J\varepsilon\varepsilon')\bar{W}RDe_1e_1'D \\
&= \sigma^2\mathrm{tr}(J'J)De_1e_1'D + 2\sigma^2\bar{W}'J'J\bar{W}RDe_1e_1'D + \sigma^2\gamma_2\bar{W}'(J'J \odot I)\bar{W}RDe_1e_1'D.
\end{aligned}
$$

Note that since $\bar{W} = \bar{Z}D$ contains a constant term, then $\bar{W}'(\bar{W}'\bar{W})^{-1}\bar{W}'\iota = \iota$. Also, ι is the second column of $\bar{W}$, so $(\bar{W}'\bar{W})^{-1}\bar{W}'\iota = e_2$, where $e_2 = (0,1,0,\cdots,0)'$ is $(k+1)\times 1$. We can verify that $De_1e_1'D = O(T^{2\delta_1}) = O(T^{-2})$, $e_1'DRDe_1 = O(T^{-1+2\delta_1}) = O(T^{-3})$, $\bar{W}'J = O(T)$, $\bar{W}R = O(T^{-1})$, $\bar{W}'J'\bar{W} = O(T^2)$, $\bar{W}'JJ' = O(T^2)$, $\bar{W}'JJ'\bar{W} = O(T^3)$, $\bar{W}'J'J\bar{W} = O(T^3)$, $\bar{W}'J\bar{W}R = O(T)$, $\bar{W}R\bar{W}'J = O(1)$, $\bar{W}'JJ'\bar{W}R = O(T^2)$, $\mathrm{tr}(\bar{W}R\bar{W}'J) = O(T)$, $\mathrm{tr}(\bar{W}R\bar{W}'J\bar{W}R\bar{W}'J) = O(T^2)$, and $\mathrm{tr}(\bar{W}'JJ'\bar{W}R) = O(T^2)$. We further observe that $\sigma^2\gamma_2\bar{W}'J(\bar{W}RDe_1e_1'DR\bar{W}' \odot I)J'\bar{W}$, $\sigma^2\gamma_2 De_1e_1'DR\bar{W}'(\bar{W}R\bar{W}'J \odot I)J'\bar{W}$, $\sigma^2\gamma_2\bar{W}'J(\bar{W}R\bar{W}'J \odot I)\bar{W}RDe_1e_1'D$, $\sigma^2\gamma_2\mathrm{tr}(\bar{W}R\bar{W}'J \odot \bar{W}R\bar{W}'J)De_1e_1'D$, $\sigma\gamma_1\bar{W}'(I \odot \bar{W}R\bar{W}'J)\iota e_1'D$, $\sigma^2\gamma_2\bar{W}'J(J'\bar{W}RDe_1e_1'DR\bar{W}' \odot I)\bar{W}$, $\sigma^2\gamma_2 De_1e_1'DR\bar{W}'(J'\bar{W}R\bar{W}'J \odot I)\bar{W}$, $\sigma^2\gamma_2 e_1'DRDe_1\bar{W}'J(\bar{W}R\bar{W}'J \odot I)\bar{W}$, $\sigma^2\gamma_2 De_1e_1'DR\bar{W}'J(\bar{W}R\bar{W}'J \odot I)\bar{W}$, and $\sigma^2\gamma_2\bar{W}'(J'J \odot I)\bar{W}RDe_1e_1'D$ are all of order $o(1)$.

Appendix B: Special Case of $X = \iota$

When $X = \iota$, we collect the following results:

$$
\begin{aligned}
\beta'X'J'M(I \odot J'MJX\beta\beta'X'J'MJ)\iota &= \beta^3 \textstyle\sum_{t=1}^{T}[t-1-0.5(T-1)]\{0.5[T(T-1)-t(t-1)-(T-1)(T-t)]\}^2 \\
&= -\tfrac{\beta^3}{240}(T^5 - T),
\end{aligned}
$$

$$
\beta'X'J'M(I \odot J'MJ)\iota = \beta\textstyle\sum_{t=1}^{T}[t-1-0.5(T-1)][T-t-T^{-1}(T-t)^2] = -\tfrac{\beta}{12}(T^2-1),
$$

$$\beta' X' J' MJ(I \odot MJ)\iota = \beta \textstyle\sum_{t=1}^{T} 0.5[T(T-1) - t(t-1) - (T-1)(T-t)][-T^{-1}(T-t)] = -\frac{\beta}{24}(T^3 - T),$$

$$\beta' X' J' MJX\beta = \beta^2 \textstyle\sum_{t=1}^{T} [t - 1 - 0.5(T-1)]^2 = \frac{\beta^2}{12} T(T^2 - 1),$$

$$\mathrm{tr}(MJ) = -T^{-1} \textstyle\sum_{t=1}^{T} (T-t) = -\frac{1}{2}(T-1),$$

$$\mathrm{tr}(J'MJ) = \textstyle\sum_{t=1}^{T} [T - t - T^{-1}(T-t)^2] = \frac{1}{6}(T^2 - 1),$$

$$\beta' X' J' MJJ' MJX\beta = \beta^2 \textstyle\sum_{t=1}^{T} \{0.5[T(T-1) - t(t-1) - (T-1)(T-t)]\}^2 = \frac{\beta^2}{120}(T^5 - T),$$

$$\mathrm{tr}(MJMJ) = \mathrm{tr}(JJ) - \frac{2}{T}\iota' JJ\iota + \frac{1}{T^2}(\iota' J\iota)^2 = -\frac{1}{T} \textstyle\sum_{t=1}^{T} (t-1)(t-2) + \frac{1}{T^2} \textstyle\sum_{t=1}^{T} (t-1) = -\frac{1}{12}T^2 + \frac{1}{2}T - \frac{5}{12},$$

$$\beta' X' J' M(I \odot MJ)\iota = \beta \textstyle\sum_{t=1}^{T} [t - 1 - 0.5(T-1)][-T^{-1}(T-t)] = \frac{\beta}{12}(T^2 - 1),$$

$$\begin{aligned}\beta' X' J' M(I \odot MJX\beta\beta' X' J' MJ)\iota &= \beta^3 \textstyle\sum_{t=1}^{T} [t - 1 - 0.5(T-1)]^2 \{0.5[T(T-1) - t(t-1) - (T-1)(T-t)]\} \\ &= \frac{\beta^3}{240}(T^5 - T),\end{aligned}$$

$$\begin{aligned}\beta' X' J' MJMJMJX\beta = \beta^2[&\iota' J' JJJ\iota - \tfrac{1}{T}\iota' J' J\iota\iota' JJ\iota - \tfrac{1}{T}\iota' J\iota\iota' JJJ\iota - \tfrac{1}{T}\iota' J\iota\iota' J' JJ\iota + \tfrac{2}{T^2}(\iota' J\iota)^2(\iota' JJ\iota) \\ &+ \tfrac{1}{T^2}(\iota' J\iota)^2(\iota' J' J\iota) - \tfrac{1}{T^3}(\iota' J\iota)^4] = \beta^2[\tfrac{1}{2} \textstyle\sum_{t=1}^{T} (t-1)(t-2)(t-1)(T - \tfrac{t}{2}) - \tfrac{1}{2T} \textstyle\sum_{t=1}^{T} (t-1)^2 \textstyle\sum_{t=1}^{T} (t-1)(t-2) \\ &- \tfrac{1}{2T} \textstyle\sum_{t=1}^{T} (t-1) \textstyle\sum_{t=1}^{T} (T-t)(t-1)(t-2) - \tfrac{1}{2T} \textstyle\sum_{t=1}^{T} (t-1) \textstyle\sum_{t=1}^{T} (t-1)^2(t-2) \\ &+ \tfrac{1}{T^2}(\textstyle\sum_{t=1}^{T} (t-1))^2 \textstyle\sum_{t=1}^{T} (t-1)(t-2) + \tfrac{1}{T^2}(\textstyle\sum_{t=1}^{T} (t-1))^2 \textstyle\sum_{t=1}^{T} (t-1)(T-t) - \tfrac{1}{T^3}(\textstyle\sum_{t=1}^{T} (t-1))^4] \\ = -\tfrac{\beta^2}{720}&(T^5 + 10T^3 - 11T).\end{aligned}$$

Table 1: Moments of the LS Estimator with $\beta^* = 5$ and $T = 10, 20$

	$T = 10$						$T = 20$					
	Normal	Uniform	Exp(1)	t_5	Mixture	logN	Normal	Uniform	Exp(1)	t_5	Mixture	logN
Bias($\hat{\lambda}$)	−0.00170	−0.00170	−0.00156	−0.00172	−0.00183	−0.00133	−0.00051	−0.00051	−0.00049	−0.00051	−0.00053	−0.00046
(MCSE)	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$O(T^{-2})$	−0.00240	−0.00240	−0.00240	−0.00240	−0.00240	−0.00240	−0.00060	−0.00060	−0.00060	−0.00060	−0.00060	−0.00060
$O(\sigma^2)$	−0.00168	−0.00168	−0.00168	−0.00168	−0.00168	−0.00168	−0.00051	−0.00051	−0.00051	−0.00051	−0.00051	−0.00051
$O(T^{-3})$	−0.00170	−0.00170	−0.00161	−0.00170	−0.00176	−0.00131	−0.00051	−0.00051	−0.00050	−0.00051	−0.00052	−0.00046
$O(\sigma^3)$	−0.00168	−0.00168	−0.00158	−0.00168	−0.00173	−0.00129	−0.00051	−0.00051	−0.00049	−0.00051	−0.00051	−0.00046
Var($\hat{\lambda}$)	0.00049	0.00049	0.00048	0.00049	0.00050	0.00046	0.00006	0.00006	0.00006	0.00006	0.00006	0.00006
$O(T^{-4})$	0.00049	0.00049	0.00064	0.00049	0.00039	0.00111	0.00006	0.00006	0.00008	0.00006	0.00005	0.00014
MSE($\hat{\lambda}$)	0.00049	0.00049	0.00048	0.00050	0.00050	0.00047	0.00006	0.00006	0.00006	0.00006	0.00006	0.00006
$O(T^{-4})$	0.00049	0.00049	0.00064	0.00049	0.00040	0.00111	0.00006	0.00006	0.00008	0.00006	0.00005	0.00014
Bias($\hat{\beta}/\sigma$)	0.05815	0.05810	0.05605	0.05829	0.06004	0.05190	0.03448	0.03431	0.03361	0.03412	0.03489	0.03279
(MCSE)	0.00019	0.00019	0.00020	0.00019	0.00018	0.00021	0.00014	0.00014	0.00014	0.00014	0.00013	0.00015
$O(T^{-1})$	0.05818	0.05818	0.05818	0.05818	0.05818	0.05818	0.03429	0.03429	0.03429	0.03429	0.03429	0.03429
Var($\hat{\beta}/\sigma$)	0.34412	0.34371	0.38275	0.34533	0.32115	0.45897	0.18557	0.18540	0.19686	0.18578	0.17894	0.22183
$O(T^{-2})$	0.34119	0.34119	0.35462	0.34119	0.33326	0.39565	0.18454	0.18454	0.18841	0.18454	0.18226	0.20022
MSE($\hat{\beta}/\sigma$)	0.34750	0.34708	0.38590	0.34872	0.32476	0.46166	0.18676	0.18658	0.19799	0.18695	0.18015	0.22290
$O(T^{-2})$	0.34457	0.34457	0.35801	0.34457	0.33665	0.39903	0.18571	0.18571	0.18958	0.18571	0.18343	0.20140

Table 2: Moments of the LS Estimator with $\beta^* = 5$ and $T = 40, 80$

	$T = 40$						$T = 80$					
	Normal	Uniform	Exp(1)	t_5	Mixture	logN	Normal	Uniform	Exp(1)	t_5	Mixture	logN
Bias($\hat{\lambda}$)	−0.00014	−0.00014	−0.00013	−0.00014	−0.00014	−0.00013	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003
(MCSE)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$O(T^{-2})$	−0.00015	−0.00015	−0.00015	−0.00015	−0.00015	−0.00015	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003
$O(\sigma^2)$	−0.00013	−0.00013	−0.00013	−0.00013	−0.00013	−0.00013	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003
$O(T^{-3})$	−0.00013	−0.00013	−0.00013	−0.00013	−0.00014	−0.00013	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003
$O(\sigma^3)$	−0.00013	−0.00013	−0.00013	−0.00013	−0.00014	−0.00013	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003	−0.00003
Var($\hat{\lambda}$)	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$O(T^{-4})$	0.00001	0.00001	0.00001	0.00001	0.00001	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
MSE($\hat{\lambda}$)	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$O(T^{-4})$	0.00001	0.00001	0.00001	0.00001	0.00001	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Bias($\hat{\beta}/\sigma$)	0.01860	0.01858	0.01832	0.01861	0.01872	0.01811	0.00964	0.00961	0.00958	0.00950	0.00960	0.00951
(MCSE)	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007
$O(T^{-1})$	0.01854	0.01854	0.01854	0.01854	0.01854	0.01854	0.00963	0.00963	0.00963	0.00963	0.00963	0.00963
Var($\hat{\beta}/\sigma$)	0.09631	0.09637	0.09942	0.09642	0.09450	0.10620	0.04906	0.04909	0.04989	0.04909	0.04863	0.05156
$O(T^{-2})$	0.09603	0.09603	0.09706	0.09603	0.09542	0.10022	0.04899	0.04899	0.04926	0.04899	0.04884	0.05008
MSE($\hat{\beta}/\sigma$)	0.09666	0.09671	0.09976	0.09676	0.09486	0.10652	0.04915	0.04918	0.04998	0.04918	0.04872	0.05165
$O(T^{-2})$	0.09637	0.09637	0.09741	0.09637	0.09577	0.10056	0.04909	0.04909	0.04935	0.04909	0.04893	0.05017

Table 3: Moments of the LS Estimator with $\beta^* = 1$ and $T = 10, 20$

	$T = 10$						$T = 20$					
	Normal	Uniform	Exp(1)	t_5	Mixture	logN	Normal	Uniform	Exp(1)	t_5	Mixture	logN
Bias($\hat{\lambda}$)	−0.06075	−0.05969	−0.03486	−0.05957	−0.07699	−0.02257	−0.01516	−0.01496	−0.01183	−0.01543	−0.01757	−0.00912
(MCSE)	0.00004	0.00004	0.00004	0.00005	0.00005	0.00003	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
$O(T^{-2})$	−0.06000	−0.06000	−0.06000	−0.06000	−0.06000	−0.06000	−0.01500	−0.01500	−0.01500	−0.01500	−0.01500	−0.01500
$O(\sigma^2)$	−0.04200	−0.04200	−0.04200	−0.04200	−0.04200	−0.04200	−0.01275	−0.01275	−0.01275	−0.01275	−0.01275	−0.01275
$O(T^{-3})$	−0.05880	−0.05880	−0.04680	−0.05880	−0.06587	−0.01015	−0.01485	−0.01485	−0.01335	−0.01485	−0.01573	−0.00876
$O(\sigma^3)$	−0.04200	−0.04200	−0.03000	−0.04200	−0.04907	0.00664	−0.01275	−0.01275	−0.01125	−0.01275	−0.01363	−0.00666
Var($\hat{\lambda}$)	0.01908	0.01827	0.01399	0.01981	0.02344	0.01120	0.00183	0.00179	0.00164	0.00197	0.00207	0.00149
$O(T^{-4})$	0.01512	0.01512	0.03432	0.01512	0.00379	0.09295	0.00170	0.00170	0.00410	0.00170	0.00028	0.01142
MSE($\hat{\lambda}$)	0.02277	0.02184	0.01521	0.02336	0.02937	0.01171	0.00206	0.00202	0.00178	0.00221	0.00238	0.00158
$O(T^{-4})$	0.01872	0.01872	0.03792	0.01872	0.00739	0.09655	0.00192	0.00192	0.00432	0.00192	0.00050	0.01165
Bias($\hat{\beta}/\sigma$)	0.29577	0.30054	0.25253	0.28036	0.31061	0.20131	0.17667	0.17650	0.16167	0.17480	0.18503	0.14240
(MCSE)	0.00018	0.00018	0.00022	0.00019	0.00016	0.00027	0.00014	0.00013	0.00016	0.00014	0.00012	0.00019
$O(T^{-1})$	0.29091	0.29091	0.29091	0.29091	0.29091	0.29091	0.17143	0.17143	0.17143	0.17143	0.17143	0.17143
Var($\hat{\beta}/\sigma$)	0.32393	0.31068	0.50238	0.35720	0.24459	0.73438	0.18121	0.17884	0.24081	0.18924	0.15136	0.36687
$O(T^{-2})$	0.23874	0.23874	0.30592	0.23874	0.19912	0.51104	0.15629	0.15629	0.17563	0.15629	0.14487	0.23471
MSE($\hat{\beta}/\sigma$)	0.41141	0.40100	0.56615	0.43580	0.34107	0.77491	0.21242	0.20999	0.26695	0.21979	0.18560	0.38715
$O(T^{-2})$	0.32337	0.32337	0.39055	0.32337	0.28375	0.59567	0.18567	0.18567	0.20502	0.18567	0.17426	0.26410

Table 4: Moments of the LS Estimator with $\beta^* = 1$ and $T = 40, 80$

	$T = 40$						$T = 80$					
	Normal	Uniform	Exp(1)	t_5	Mixture	logN	Normal	Uniform	Exp(1)	t_5	Mixture	logN
Bias($\hat{\lambda}$)	−0.00375	−0.00373	−0.00335	−0.00378	−0.00401	−0.00287	−0.00093	−0.00093	−0.00088	−0.00093	−0.00096	−0.00081
(MCSE)	0.00001	0.00000	0.00000	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$O(T^{-2})$	−0.00375	−0.00375	−0.00375	−0.00375	−0.00375	−0.00375	−0.00093	−0.00093	−0.00093	−0.00093	−0.00093	−0.00093
$O(\sigma^2)$	−0.00346	−0.00346	−0.00346	−0.00346	−0.00346	−0.00346	−0.00090	−0.00090	−0.00090	−0.00090	−0.00090	−0.00090
$O(T^{-3})$	−0.00373	−0.00373	−0.00354	−0.00373	−0.00384	−0.00297	−0.00093	−0.00093	−0.00091	−0.00093	−0.00094	−0.00084
$O(\sigma^3)$	−0.00346	−0.00346	−0.00328	−0.00346	−0.00357	−0.00270	−0.00090	−0.00090	−0.00087	−0.00090	−0.00091	−0.00080
Var($\hat{\lambda}$)	0.00020	0.00020	0.00020	0.00021	0.00021	0.00019	0.00002	0.00002	0.00002	0.00002	0.00003	0.00002
$O(T^{-4})$	0.00020	0.00020	0.00050	0.00020	0.00002	0.00142	0.00002	0.00002	0.00006	0.00002	0.00000	0.00018
MSE($\hat{\lambda}$)	0.00022	0.00022	0.00021	0.00022	0.00023	0.00020	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002
$O(T^{-4})$	0.00021	0.00021	0.00051	0.00021	0.00004	0.00143	0.00003	0.00003	0.00006	0.00003	0.00000	0.00018
Bias($\hat{\beta}/\sigma$)	0.09429	0.09425	0.09029	0.09435	0.09669	0.08423	0.04857	0.04853	0.04756	0.04846	0.04911	0.04582
(MCSE)	0.00010	0.00010	0.00011	0.00010	0.00009	0.00012	0.00007	0.00007	0.00007	0.00007	0.00007	0.00008
$O(T^{-1})$	0.09268	0.09268	0.09268	0.09268	0.09268	0.09268	0.04815	0.04815	0.04815	0.04815	0.04815	0.04815
Var($\hat{\beta}/\sigma$)	0.09539	0.09514	0.11139	0.09689	0.08679	0.15184	0.04885	0.04884	0.05292	0.04906	0.04662	0.06291
$O(T^{-2})$	0.08857	0.08857	0.09373	0.08857	0.08552	0.10951	0.04707	0.04707	0.04841	0.04707	0.04629	0.05248
MSE($\hat{\beta}/\sigma$)	0.10428	0.10402	0.11954	0.10579	0.09613	0.15894	0.05121	0.05120	0.05518	0.05141	0.04903	0.06501
$O(T^{-2})$	0.09716	0.09716	0.10232	0.09716	0.09411	0.11810	0.04939	0.04939	0.05072	0.04939	0.04861	0.05480

Table 5: Moments of the LS Estimator with $\beta^* = 0.2$ and $T = 10, 20$

	$T = 10$						$T = 20$					
	Normal	Uniform	Exp(1)	t_5	Mixture	logN	Normal	Uniform	Exp(1)	t_5	Mixture	logN
Bias($\hat{\lambda}$)	−0.37197	−0.38119	−0.37504	−0.36073	−0.35895	−0.36407	−0.19126	−0.19404	−0.19358	−0.18593	−0.18634	−0.18528
(MCSE)	0.00010	0.00010	0.00010	0.00010	0.00010	0.00011	0.00006	0.00006	0.00006	0.00006	0.00006	0.00006
$O(T^{-2})$	−1.50000	−1.50000	−1.50000	−1.50000	−1.50000	−1.50000	−0.37500	−0.37500	−0.37500	−0.37500	−0.37500	−0.37500
$O(\sigma^2)$	−1.05000	−1.05000	−1.05000	−1.05000	−1.05000	−1.05000	−0.31875	−0.31875	−0.31875	−0.31875	−0.31875	−0.31875
$O(T^{-3})$	−11.55000	−11.55000	−10.05000	−11.55000	−12.43480	−5.46949	−1.63125	−1.63125	−1.44375	−1.63125	−1.74185	−0.87118
$O(\sigma^3)$	−1.05000	−1.05000	0.45000	−1.05000	−1.93484	5.03050	−0.31875	−0.31875	−0.13125	−0.31875	−0.42935	0.44131
$O(\beta^{*2})$	−0.36943	−0.37915	−0.38799	−0.35698	−0.34991	−0.40050	−0.18589	−0.18911	−0.19981	−0.17944	−0.17666	−0.20976
Var($\hat{\lambda}$)	0.09672	0.09693	0.09775	0.09821	0.09538	0.12419	0.03112	0.03138	0.03049	0.03085	0.03093	0.03308
$O(T^{-4})$	2.25000	2.25000	4.65000	2.25000	0.83425	11.97880	0.15938	0.15938	0.45938	0.15938	−0.01759	1.37548
MSE($\hat{\lambda}$)	0.23508	0.24224	0.23840	0.22833	0.22423	0.25674	0.06770	0.06903	0.06797	0.06542	0.06566	0.06741
$O(T^{-4})$	4.50000	4.50000	6.90000	4.50000	3.08425	14.22880	0.30000	0.30000	0.60000	0.30000	0.12303	1.51610
Bias($\hat{\beta}/\sigma$)	0.21413	0.22174	0.29437	0.20451	0.14695	0.30957	0.23955	0.24344	0.29650	0.23174	0.20067	0.31889
(MCSE)	0.00026	0.00026	0.00028	0.00028	0.00026	0.00032	0.00020	0.00020	0.00022	0.00021	0.00019	0.00027
$O(T^{-1})$	1.45455	1.45455	1.45455	1.45455	1.45455	1.45455	0.85714	0.85714	0.85714	0.85714	0.85714	0.85714
Var($\hat{\beta}/\sigma$)	0.69958	0.67149	0.80646	0.75682	0.66138	1.03012	0.39989	0.38790	0.47785	0.43303	0.37780	0.71659
$O(T^{-2})$	−2.32231	−2.32231	−1.98644	−2.32231	−2.52044	−0.96081	−0.55000	−0.55000	−0.45326	−0.55000	−0.60706	−0.15787
MSE($\hat{\beta}/\sigma$)	0.74543	0.72066	0.89312	0.79864	0.68297	1.12596	0.45727	0.44716	0.56577	0.48673	0.41806	0.81828
$O(T^{-2})$	−0.20661	−0.20661	0.12926	−0.20661	−0.40473	1.15489	0.18469	0.18469	0.28143	0.18469	0.12763	0.57682

Table 6: Moments of the LS Estimator with $\beta^* = 0.2$ and $T = 40, 80$

	$T = 40$						$T = 80$					
	Normal	Uniform	Exp(1)	t_5	Mixture	logN	Normal	Uniform	Exp(1)	t_5	Mixture	logN
Bias($\hat{\lambda}$)	−0.08464	−0.08536	−0.08498	−0.08267	−0.08353	−0.07894	−0.03075	−0.03092	−0.02993	−0.03025	−0.03097	−0.02625
(MCSE)	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
$O(T^{-2})$	−0.09375	−0.09375	−0.09375	−0.09375	−0.09375	−0.09375	−0.02343	−0.02343	−0.02343	−0.02343	−0.02343	−0.02343
$O(\sigma^2)$	−0.08671	−0.08671	−0.08671	−0.08671	−0.08671	−0.08671	−0.02255	−0.02255	−0.02255	−0.02255	−0.02255	−0.02255
$O(T^{-3})$	−0.25078	−0.25078	−0.22734	−0.25078	−0.26460	−0.15577	−0.04306	−0.04306	−0.04013	−0.04306	−0.04479	−0.03119
$O(\sigma^3)$	−0.08671	−0.08671	−0.06328	−0.08671	−0.10054	0.00829	−0.02255	−0.02255	−0.01962	−0.02255	−0.02428	−0.01068
$O(\beta^{*2})$	−0.07417	−0.07497	−0.08428	−0.07092	−0.06951	−0.09295	−0.01215	−0.01244	−0.01795	−0.01033	−0.00969	−0.02475
Var($\hat{\lambda}$)	0.00815	0.00819	0.00806	0.00805	0.00813	0.00832	0.00158	0.00159	0.00153	0.00157	0.00161	0.00144
$O(T^{-4})$	0.01231	0.01231	0.04981	0.01231	−0.00981	0.16432	0.00106	0.00106	0.00575	0.00106	−0.00170	0.02006
MSE($\hat{\lambda}$)	0.01531	0.01547	0.01528	0.01489	0.01511	0.01455	0.00253	0.00255	0.00243	0.00248	0.00257	0.00213
$O(T^{-4})$	0.02109	0.02109	0.05859	0.02109	−0.00102	0.17311	0.00161	0.00161	0.00630	0.00161	−0.00115	0.02061
Bias($\hat{\beta}/\sigma$)	0.22779	0.22969	0.25950	0.22264	0.20793	0.27512	0.18287	0.18366	0.19444	0.18003	0.17538	0.19670
(MCSE)	0.00014	0.00014	0.00015	0.00014	0.00013	0.00020	0.00009	0.00009	0.00009	0.00009	0.00009	0.00013
$O(T^{-1})$	0.46342	0.46342	0.46342	0.46342	0.46342	0.46342	0.24074	0.24074	0.24074	0.24074	0.24074	0.24074
Var($\hat{\beta}/\sigma$)	0.18814	0.18448	0.22553	0.20138	0.17759	0.38902	0.07590	0.07506	0.08909	0.07951	0.07159	0.15671
$O(T^{-2})$	−0.09802	−0.09802	−0.07220	−0.09802	−0.11326	0.00666	−0.00095	−0.00095	0.00571	−0.00095	−0.00488	0.02606
MSE($\hat{\beta}/\sigma$)	0.24003	0.23724	0.29287	0.25095	0.22083	0.46472	0.10934	0.10879	0.12689	0.11193	0.10235	0.19540
$O(T^{-2})$	0.11672	0.11672	0.14255	0.11672	0.10149	0.22142	0.05700	0.05700	0.06367	0.05700	0.05307	0.08402