

# Measurement Error in Linear Autoregressive Models

BY JOHN STAUDENMAYER AND JOHN P. BUONACCORSI

October 24, 2004

John Staudenmayer (jstauden@math.umass.edu) is an Assistant Professor and John P. Buonaccorsi is a Professor in the Department of Mathematics and Statistics, Lederle Graduate Research Tower, University of Massachusetts, Amherst MA, 01003-9305, USA. This research was supported by NSF-DMS 0306227. We thank two anonymous referees, an Associate Editor, and the Editor for comments and suggestions that improved many aspects of this paper.

## ABSTRACT

Time series data are often subject to measurement error, usually the result of needing to estimate the variable of interest. While it is often reasonable to assume the measurement error is additive, that is, the estimator is conditionally unbiased for the missing true value, the measurement error variances often vary as a result of changes in the population/process over time and/or changes in sampling effort. In this paper we address estimation of the parameters in linear autoregressive models in the presence of additive and uncorrelated measurement errors, allowing heteroskedasticity in the measurement error variances. The asymptotic properties of naive estimators that ignore measurement error are established, and we propose an estimator based on correcting the Yule-Walker estimating equations. We also examine a pseudolikelihood method, which is based on normality assumptions and computed using the Kalman filter. Other techniques that have been proposed, including two that require no information about the measurement error variances are reviewed as well. The various estimators are compared both theoretically and via simulations. The estimator based on corrected estimating equations is easy to obtain, readily accommodates (and is robust to) unequal measurement error variances, and asymptotic calculations and finite sample simulations show that it is often relatively efficient.

*Keywords: Estimating equation, heteroskedasticity, error in variables, time series, state-space, Kalman filter.*

# 1 Introduction

Measurement error is a ubiquitous in many statistical problems and has received considerable attention in various regression contexts (e.g., Fuller, 1987 and Carroll, Ruppert, and Stefanski, 1995). Time series data is no exception when it comes to the presence of measurement error; the variable of interest is often estimated rather than observed exactly. Ecological examples include the estimation of animal densities through capture/recapture techniques or the estimation of food abundance over a large area by spatial subsampling. In environmental studies chemical concentrations in the air or water at a particular time are typically from collections at various locations, while many economic, labor and public health variables are estimated on a regular basis over time often through a complex sampling scheme. Examples of the latter include medical data indices and results of a national opinion poll (Scott and Smith, 1974, Scott, Smith and Jones, 1977), estimation of the number of households in Canada and labor force statistics in Israel (Pfeffermann, Feder, and Signorelli, 1998) and retail sales figures (Bell and Hillmer, 1990, Bell and Wilcox, 1993).

To frame the general problem, consider a times series  $\{Y_t\}$ , where  $Y_t$  is a random variable and  $t$  indexes time. The realized true values are denoted  $\{y_t, t = 1, \dots, T\}$ , but instead of  $y_t$  one observes the outcome of  $W_t$ , where  $E(W_t|y_t) = y_t$ , with  $|y_t$  shorthand for “given  $Y_t = y_t$ ”. Hence,  $W_t$  is a conditionally unbiased estimator of  $y_t$ , or equivalently,  $W_t = y_t + u_t$ , where  $u_t$ , which represents measurement/survey error, has  $E(u_t|y_t) = 0$ . In Section 2, we show that this last assumption implies  $\text{Cov}(u_t, Y_t) = 0$  but allows the conditional measurement error variance to depend on  $Y_t$ .

The focus of this paper is estimation of the parameters in an autoregressive model for  $Y_t$  from observations of  $W_t$  when the measurement error is additive, uncorrelated, and possibly heteroskedastic. Autoregressive models are a

popular choice in many disciplines for the modeling of time series. This is especially true of population dynamics where AR(1) or AR(2) models are often employed with  $Y$  equal to the log of population abundance (or density), see for example Williams and Liebhold (1995), Dennis and Taper (1994, equation 7) and references therein. That work does not consider measurement error.

To motivate some aspects of the problem, Table 1 presents estimated mouse densities (mice per hectare on the log scale to which autoregressive models are usually fit for abundance data) and associated standard errors over a nine year period from one stand in the Quabbin Reservoir, located in Western Massachusetts. This is one part of the data used in Elkinton et al. (1996). The stand density and associated standard errors were obtained using estimates from subplots. The sampling effort consists of two subplots in 87, 89 and 90 and three in the other years. The estimated standard errors for the log densities were obtained using the delta method. The purpose of this example is to illustrate that the estimated standard errors vary considerably (with 1989 and 1990 values differing by a factor of about 11). This can occur for a number of reasons, including the changes in sampling effort and changes of the nature of the population being sampled at each time point, and indicate the need to consider a heteroskedastic measurement error model. Of course, if the estimated density is unbiased for true density then (by Jensen's inequality) the same will not be true after a log transformation. In this example estimates of the approximate biases were negligible, and the assumption of additive measurement error is reasonable. If this were not the case then a non-linear time-series / measurement model would be required.

Table 1 about here.

The following describes the paper's organization and highlights our contributions. First, the models are spelled out in Section 2, with particular interest

in allowing for heteroskedasticity in the measurement error variances and the case when estimates of the measurement error variances are available. This problem has not received careful attention in the literature. (Existing methods are reviewed and discussed in Sections 4.2-4.4.) Following that, in Section 3 we derive the asymptotic bias of estimators of parameters in the AR( $p$ ) model when the measurement error is ignored; the bias of so-called naive estimators. These results do not appear to be in the literature currently. For the AR(1) model the bias has a simple, explicit, and somewhat surprising “attenuation” form. For  $p > 1$ , the biases have simple matrix forms, and we illustrate the pernicious effects of measurement error in the AR(2) case. The naive estimator of the error variance has positive asymptotic bias for all  $p$ . Next, Section 4.1 presents a simple new estimator for the problem of fitting autoregressive models in the presence of measurement error. This new estimator is based on corrected asymptotic estimating equations and does not require the specification of a likelihood or the assumption of constant measurement error variance. In Section 4.1 we also develop the asymptotic properties of that estimator under various assumptions. We believe that this is the first paper to develop asymptotic results for this problem when the measurement error is heteroskedastic. After that, Section 4.2 discusses a pseudo-maximum likelihood approach that assumes normality and uses estimated measurement error variances. Sections 4.3 and 4.4 review two other existing methods that do not require knowledge of the measurement error variance. In Section 5 we compare our new method to the existing methods both asymptotically and with a small simple simulation study. Discussion follows in Section 6.

In addition to estimating the parameters in the dynamic model, a number of authors have addressed forecasting  $Y_t$  for  $t > T$  or estimating  $y_t$  for  $t \leq T$  in the presence of measurement error. The latter has been the focus of much

of the previous work in this area (e.g. Scott and Smith (1974), Scott, Smith and Jones (1977), Bell and Hillmer (1990), Pfeffermann (1991), Pfeffermann, Feder and Signorelli (1998), Feder (2001) and references therein). Certain aspects of forecasting are addressed in Scott and Smith (1974), Scott, Smith and Jones (1977), Wong and Miller (1990) and Koreisha and Fang (1999). We will not address forecasting or estimation of  $y_t$  in this paper directly.

## 2 Model for Observed Data

Assuming  $Y_t$  is centered to have mean 0, the autoregressive model specifies that  $Y_t$  is a linear combination of previous values, plus noise:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t, t = p + 1, \dots, T, \quad (1)$$

where the  $e_t$  are *iid* with mean zero and variance  $\sigma_e^2$ . Let  $\mathbf{Y}$  refer to the entire series. We focus on the case where  $\phi_1, \dots, \phi_p$  are such that the process  $Y_t$  is stationary (e.g. Box, Jenkins, and Reinsel, 1994 Chapter 3).

As in the Introduction,  $\mathbf{Y}$  is unobserved; we observe  $W_t = Y_t + u_t$  with  $E(u_t|y_t) = 0$  or equivalently  $E(W_t|y_t) = y_t$ . The  $u_t$  represents additive measurement or survey error. We assume the measurement errors are uncorrelated, that is  $\text{Cov}(u_t, u_{t'}|y_t, y_{t'}) = 0$  for  $t \neq t'$ . Note that the assumption  $E(u_t|y_t) = 0$  is enough by itself to imply  $\text{Cov}(u_t, Y_t) = 0$  since  $\text{Cov}(u_t, Y_t) = E\{\text{Cov}(u_t, Y_t|Y_t)\} + \text{Cov}\{E(u_t|Y_t), E(Y_t|Y_t)\} = E(0) + \text{Cov}(0, Y_t) = 0$ . Interestingly, this occurs even if  $\text{Var}(u_t|y_t)$  is a function of  $Y_t$  in which case  $u_t$  and  $Y_t$  are dependent but uncorrelated.

One of the techniques discussed later relies on the following well known result (see Box *et al.*, 1994, Section A4.3, Morris and Granger, 1976, or Pagano, 1974). **Lemma 1:** If  $Y_t$  is an  $\text{AR}(p)$  process with autoregressive parameters  $\phi$

and  $U_t$  is white noise with constant variance  $\sigma_u^2$ , then  $W_t = Y_t + U_t$  follows an  $ARMA(p, p)$  process. This can be written as  $\phi(B)W_t = \theta(B)b_t$  where  $b_t$  is white noise with mean 0 and variance  $\sigma_b^2$ ,  $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ ,  $\theta(z) = 1 + \theta_1 z + \dots + \theta_p z^p$ , and  $B$  is the backward shift operator with  $B^j W_t = W_{t-j}$ . The autoregressive parameters ( $\phi_k$ s) in this model are the same as in the model for  $\mathbf{Y}$ . The moving average ( $\theta_k$ s) and conditional variance ( $\sigma_b^2$ ) parameters can be related to  $\phi$ ,  $\sigma_\epsilon^2$ , and  $\sigma_u^2$  by solving  $\sigma_\epsilon^2 + \phi(z)\phi(z^{-1})\sigma_u^2 = \theta(z)\theta(z^{-1})\sigma_b^2$ . These equations equate the covariance generating functions of  $\mathbf{W}$  and  $\mathbf{Y} + \mathbf{U}$ . There may be multiple solutions, only some of which result in a stationary and invertible process.

## 2.1 Heteroskedastic Measurement Error

We allow for heteroskedasticity in the measurement error variances and distinguish between the conditional (given  $Y_t = y_t$ ) and the unconditional measurement error variance. Consider first  $\text{Var}(u_t|y_t)$ , denoted by  $\sigma_{ut}^2(y_t)$ , which may depend on  $y_t$  through a function  $h(y_t)$  and may depend on the sampling effort at time  $t$ . We do not require a specific form for  $h(\cdot)$ , but a power model is perhaps common. Unconditionally,  $\text{Var}(u_t) = \sigma_{ut}^2 = \text{E}[\text{Var}(u_t|Y_t)] = \text{E}[\sigma_{ut}^2(Y_t)]$ . For asymptotic purposes it is assumed that

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T \sigma_{ut}^2 / T = \sigma_u^2. \quad (2)$$

The unconditional variance  $\sigma_{ut}^2$  can change over  $t$  if a fixed sampling effort changes over time. In general, one can view the conditional variance,  $\sigma_{ut}^2(y_t)$ , as a function of two components; one describes the “variability” of the population sampled at time  $t$  (the meaning of this “variability” depends on the sampling scheme), and the other is sampling effort. Unconditionally, the “vari-

ability" will average out.

For example, in many applications  $Y_t$  is the mean of measurements on  $N_t$  subunits (e.g. individuals or spatial areas). Let  $\mathbf{D}_t = \{Y_{t1}, \dots, Y_{tN_t}\}$  denote the collection of random variables defined over the subunits, with  $Y_t = \sum_{j=1}^{N_t} Y_{tj}/N_t$ . For simplicity, suppose that a simple random sample of  $n_t$  subunits is chosen, and  $Y_t$  is estimated by  $W_t$ , the sample mean of those  $n_t$  selected subunits. Let  $\mathbf{d}_t$  and  $y_t$  denote realized values. The conditional variance of  $W_t$  given the subunit measurements is  $\text{Var}(W_t|\mathbf{d}_t) = f_t s_t^2/n_t$ , where  $f_t = 1 - (n_t/N_t)$  and  $s_t^2 = \sum_{j=1}^{N_t} (y_{tj} - y_t)^2 / (N_t - 1)$ . If the sample size  $n_t$  is fixed, then  $\sigma_{ut}^2(y_t) = f_t E(s_t^2|y_t)/n_t$ . As a result,  $\sigma_{ut}^2 = f_t E[E(s_t^2|Y_t)]/n_t = f_t E(S_t^2)/n_t$ . Notice that in this example unconditional heteroskedasticity comes from unequal sampling effort, while conditional heteroskedasticity is the result of the sampling effort and among unit variability. Also, if the sampling effort is either equal throughout, or unequal but random with a common distribution that does not depend on  $t$ , then unconditionally  $\sigma_{ut}^2 = \sigma_u^2$ , a constant.

The sampling that leads to  $W_t$  will almost always yield an estimate,  $\hat{\sigma}_{ut}^2$ , of the conditional measurement error variance, where it is assumed that  $E(\hat{\sigma}_{ut}^2|y_t) = \sigma_{ut}^2(y_t)$ , or unconditionally  $E(\hat{\sigma}_{ut}^2) = \sigma_{ut}^2$ . This last assumption typically follows from the nature of the sampling leading to  $W_t$  and  $\hat{\sigma}_{ut}^2$ .

### 3 Models and Properties of the Naive Estimators

We do two things in this section. First, we write down the asymptotic estimating equations for the parameters in the  $AR(p)$  model without measurement error. Following that, we derive the bias of the resulting estimators when measurement error is present but ignored.



### 3.1 Naive Estimation Methods and Asymptotic Estimating Equations

The unknown parameters in model (1) are  $\phi = [\phi_1, \dots, \phi_p]^\top$  and  $\sigma_e^2$ . Without measurement error, four common estimation methods are Maximum Likelihood (ML) assuming Gaussian Errors, Least Squares (LS), Approximate Maximum Likelihood (aML), and Yule-Walker (YW). Letting  $\gamma_k$  be the lag  $k$  autocovariance for a process with an AR( $p$ ) structure (without measurement error), we denote the  $p$  by  $p$  autocovariance matrix for an AR( $p$ ) process (without mea-

$$\text{surement error) as } \Gamma = \begin{pmatrix} \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_0 \end{pmatrix}. \text{ Next, let } \hat{\mathbf{G}} = \begin{pmatrix} \hat{\gamma}_0 & -\hat{\gamma}_1 & -\hat{\gamma}_2 & \cdots & -\hat{\gamma}_p \\ -\hat{\gamma}_1 & \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ -\hat{\gamma}_p & \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} := \begin{pmatrix} \hat{\gamma}_0 & -\hat{\gamma}_1 & \cdots & -\hat{\gamma}_p \\ -\hat{\gamma}_1 & & & \\ \vdots & & \hat{\mathbf{\Gamma}} & \\ -\hat{\gamma}_p & & & \end{pmatrix}, \text{ where } \hat{\gamma}_k =$$

$\sum_{t=1}^{T-k} Y_t Y_{t+k} / T$ . (Remember that  $Y_t$  is centered.) Also, let  $\phi_u = [1, \phi^\top]^\top$ . As discussed in Box *et al.* (1994, Chapter 7) for instance, without measurement error the various estimation methods listed above lead to asymptotically equivalent estimators. They all asymptotically solve the following estimating equations for  $\phi$  and  $\sigma_e^2$ :

$$(\hat{\gamma}_1 \dots \hat{\gamma}_p)^\top - \hat{\mathbf{\Gamma}}\phi = \mathbf{0}, \text{ and } \frac{-1}{\sigma_e} + \frac{\phi_u^\top \hat{\mathbf{G}} \phi_u}{\sigma_e^3} = 0. \quad (3)$$

The naive estimators of  $\phi$  and  $\sigma_e^2$  calculate  $\hat{\mathbf{G}}$  and its components based on  $\mathbf{W}$  rather than  $\mathbf{Y}$ . In the sequel, we denote  $\hat{\mathbf{\Gamma}}$  calculated from  $\mathbf{W}$  as  $\hat{\mathbf{\Gamma}}_W$  with components  $\hat{\gamma}_{W,k}$ .

### 3.2 Asymptotic Bias of $\widehat{\phi}_{naive}$

Next, we consider estimates of the elements of  $\widehat{\Gamma}_W$  using the measurement error model discussed in Section 2 and under regularity conditions to ensure the convergence of sample autocovariances (e.g. Brockwell and Davis, 1996, Chapter 7). The key observation is that for  $t \neq s$ ,  $\text{Cov}(W_t, W_s) = \text{Cov}(Y_t, Y_s) + \text{Cov}(U_t, U_s) = \gamma_{|t-s|}$ , so  $\lim_{T \rightarrow \infty} \widehat{\gamma}_{W,|t-s|} \stackrel{P}{=} \gamma_{|t-s|}$ , but  $\lim_{T \rightarrow \infty} \widehat{\gamma}_{W,0} \stackrel{P}{=} \gamma_0 + \sigma_u^2$  (see Appendix A.1), where  $\stackrel{P}{=}$  denotes equal in probability. Note that, as discussed in Section 2.1, when the measurement error is heteroskedastic,  $\sigma_u^2$  is defined as the limit average of  $\sigma_{ut}^2$ . As  $T \rightarrow \infty$  the estimating equation for  $\phi$  from the previous section (Equation 3) converges in probability to:

$$\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix} - \begin{pmatrix} \gamma_0 + \sigma_u^2 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{p-1} & \cdots & \gamma_1 & \gamma_0 + \sigma_u^2 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \mathbf{0}, \text{ or } \boldsymbol{\gamma} - (\boldsymbol{\Gamma} + \sigma_u^2 \mathbf{I}_p) \boldsymbol{\phi} = \mathbf{0}.$$

Hence, the naive estimator,  $\widehat{\phi}_{naive}$ , will converge in probability to the solution of this equation:  $\lim_{T \rightarrow \infty} \widehat{\phi}_{naive} \stackrel{P}{=} (\boldsymbol{\Gamma} + \sigma_u^2 \mathbf{I}_p)^{-1} \boldsymbol{\gamma} = (\boldsymbol{\Gamma} + \sigma_u^2 \mathbf{I}_p)^{-1} \boldsymbol{\Gamma} \boldsymbol{\phi}_{true}$  where  $\boldsymbol{\phi}_{true}$  contains the true autoregressive parameters. As  $T \rightarrow \infty$ , the asymptotic bias of the naive estimator is

$$\{(\boldsymbol{\Gamma} + \sigma_u^2 \mathbf{I}_p)^{-1} - \boldsymbol{\Gamma}^{-1}\} \boldsymbol{\gamma}. \quad (4)$$

When  $p = 1$ , the naive estimator converges in probability to an attenuated version of the true parameter:  $\lim_{T \rightarrow \infty} \widehat{\phi}_{1,naive} \stackrel{P}{=} \frac{\gamma_1}{\gamma_0 + \sigma_u^2} = \lambda \phi_{true}$ , where  $\lambda = \frac{\gamma_0}{\gamma_0 + \sigma_u^2}$ . We find it a bit surprising that the current model leads to the familiar attenuation result of covariate measurement error in simple linear regression. Specifically, the current model can be seen as a linear regression model

$Y_t = \phi_1 X_t + \epsilon_t$  (where  $X_t = Y_{t-1}$ ) with measurement error in both the response ( $W_t = Y_t + u_t$ ) and in the covariate ( $V_t = X_t + u_{t-1}$ ). While the two measurement errors within an observation ( $u_t$  and  $u_{t-1}$ ) are uncorrelated, the measurement error in  $Y$  at time  $t$  is correlated with the measurement error in  $X$  at time  $t-1$ . In fact, the two are the same measurement error. This falls outside of the traditional treatments of measurement error in regression, which almost always assume independence of measurement errors across observations. Still, the attenuation is the same as the type seen in simple linear regression with measurement error in the predictor only (e.g. Fuller, 1987, page 3).

Similar to the case of linear regression with covariate measurement error (e.g. Carroll, Ruppert, and Stefanski, 1995, pages 25-26), when  $p > 1$ , the relation between the large sample limits of the naive estimators become more complicated and a simple result that relates  $\hat{\phi}_{naive}$  and  $\phi_{true}$  does not exist. For instance, when  $p = 2$  we can write the estimators  $\hat{\phi}_1$  and  $\hat{\phi}_2$  in terms of the the estimated lag one and two correlations ( $\hat{\rho}_1$  and  $\hat{\rho}_2$ ). Since we know that the naive estimate of the correlation is attenuated,  $\lim_{T \rightarrow \infty} \hat{\rho}_{j,naive} \stackrel{P}{=} \lambda \rho_j, j > 1$ ,  $\lim_{T \rightarrow \infty} \begin{pmatrix} \hat{\phi}_{1,naive} \\ \hat{\phi}_{2,naive} \end{pmatrix} \stackrel{P}{=} \begin{pmatrix} \frac{\lambda \rho_1 - \lambda^2 \rho_1 \rho_2}{1 - \lambda^2 \rho_1^2} \\ \frac{\lambda \rho_2 - \lambda^2 \rho_1^2}{1 - \lambda^2 \rho_1^2} \end{pmatrix}$ . The asymptotic bias in either element of  $\hat{\phi}_{naive}$  can be either attenuating (smaller in absolute value) or accentuating (larger in absolute value), and the type of asymptotic bias depends on both the amount of measurement error and the true values for  $\phi_1$  and  $\phi_2$ . We illustrate this phenomenon in Figure 1. The top two rows of Figure 1 use Equation (4) to calculate the asymptotic biases in  $\hat{\phi}_{naive,1}$  and  $\hat{\phi}_{naive,2}$  as a function of  $\lambda$  when the true series follows one of four different AR(2) models. The bottom row of Figure 1 shows the type of measurement error asymptotic bias as a function of the true  $\phi$  over the range of possible values that result in a stationary model.

### 3.3 Asymptotic Bias of $\hat{\sigma}_{e,naive}$

Using notation from Section 3.2  $\lim_{T \rightarrow \infty} \hat{\sigma}_{naive,e}^2 \stackrel{P}{=} \lim_{T \rightarrow \infty} \gamma_0 + \sigma_u^2 - 2\hat{\phi}_{naive}^\top \gamma + \hat{\phi}_{naive}^\top (\mathbf{\Gamma} + \sigma_u^2 \mathbf{I}) \hat{\phi}_{naive} \stackrel{P}{=} \gamma_0 + \sigma_u^2 - \gamma^\top (\mathbf{\Gamma} + \sigma_u^2 \mathbf{I})^{-1} \gamma$ . As a result, since the true  $\lim_{T \rightarrow \infty} \hat{\sigma}_e^2 \stackrel{P}{=} \gamma_0 + \gamma^\top \mathbf{\Gamma}^{-1} \gamma$ , the asymptotic bias of  $\hat{\sigma}_{e,naive}^2$  is

$$\sigma_u^2 + \gamma^\top \{ \mathbf{\Gamma}^{-1} - (\mathbf{\Gamma} + \sigma_u^2 \mathbf{I})^{-1} \} \gamma. \quad (5)$$

Appendix A.2 shows that  $\hat{\sigma}_{e,naive}^2$  has positive asymptotic bias.

## 4 Estimation Methods That Address Measurement Error

In this Section we develop new estimators and review existing estimators that account for measurement error. It is worth noting that the existing literature that proposes techniques for estimating  $\phi$  and  $\sigma_\epsilon^2$  in the presence of measurement error is scattered over essentially three disciplines: econometrics, survey sampling, and traditional mathematical statistical time series. Among the existing methods are two simple techniques that neither use nor require  $\hat{\sigma}_{ut}^2$ ; a method based on the ARMA result in Lemma 1 (which requires homoskedastic measurement error) and a modified Yule-Walker approach. These are reviewed in Sections 4.3 and 4.4. Other authors (e.g., Pagano, 1974, Miazaki and Dorea, 1993, Lee and Shin 1997), also working under a homoskedastic measurement error model, begin with an ARMA fit and then obtain slightly more efficient estimators by enforcing the restrictions in (2).

Section 4.1 presents an approach based on correcting the asymptotic estimating equations which is quite simple, explicitly uses the estimated measurement error variances, accommodates heteroskedasticity and does not require

full likelihood assumptions. In Section 4.2 pseudo-likelihood based methods are discussed. We compare the methods in Section 5.

#### 4.1 Corrected Asymptotic Estimating Equations

As discussed in Section 3.2, the problem with the naive estimating equations in large samples is that the diagonal of  $\widehat{\Gamma}_W$  is too large. This suggests an estimator ( $\widehat{\phi}_{CEE}$ ) based on a corrected version of Equation (3):

$$\widehat{\phi}_{CEE} = (\widehat{\Gamma}_W - \widehat{\sigma}_u^2 \mathbf{I}_p)^{-1} (\widehat{\gamma}_{W,1}, \dots, \widehat{\gamma}_{W,p})^\top. \quad (6)$$

where  $\widehat{\sigma}_u^2 = \sum_{t=1}^T \widehat{\sigma}_{ut}^2 / T$ . When  $p = 1$  for instance  $\widehat{\phi}_{CEE} = \widehat{\gamma}_{W,1} / (\widehat{\gamma}_{W,0} - \widehat{\sigma}_u^2)$ .

Like estimators described in Fuller (1987, e.g. Section 1.2), the sampling moments of  $\widehat{\phi}_{CEE}$  as defined above often do not exist since the sampling density of  $\widehat{\gamma}_{W,0} - \widehat{\sigma}_u^2$  often has mass at zero and values less than zero (e.g. Fuller, 1987, Section 2.5). In order to fix this problem, we use a modified version of this estimator: if  $\widehat{\phi}_{CEE}$  results in a non-stationary model or if  $\widehat{\Gamma}_W - \widehat{\sigma}_u^2 \mathbf{I}_p$  is non-positive definite, then  $\widehat{\phi}_{CEE} := \widehat{\phi}_{naive}$ . When  $\lim_{T \rightarrow \infty} \widehat{\sigma}_u^2 = \sum_{t=1}^T \widehat{\sigma}_{ut}^2 / T \stackrel{p}{=} \sigma_u^2$ , this adjustment will not effect the asymptotic properties of the estimator since  $\lim_{T \rightarrow \infty} (\widehat{\Gamma}_W - \widehat{\sigma}_u^2 \mathbf{I}_p) \stackrel{p}{=} \Gamma$ , which is assumed to positive definite in that case (see Proposition 1 below).

The following three results describe the asymptotic behavior of  $\widehat{\phi}_{CEE}$ . First we state a consistency result. An asymptotic variance result for  $p = 1$  follows that. Finally, we state an asymptotic normality result which includes an asymptotic variance for general  $p \geq 1$ . We emphasize that none of these results require either the measurement error or the unobserved time series to have a normal distribution, but the asymptotic variances have slightly simpler forms when the measurement error is normally distributed.

**Proposition 1** *Under the model described in Section 2, the estimator  $\widehat{\phi}_{CEE}$  is consistent if*

$$\lim_{T \rightarrow \infty} \widehat{\sigma}_u^2 = \sum_{t=1}^T \widehat{\sigma}_{ut}^2 / T \stackrel{p}{=} \sigma_u^2. \quad (7)$$

Proof: Assuming non-singularity,  $\lim_{T \rightarrow \infty} \widehat{\phi}_{CEE} \stackrel{p}{=} (\mathbf{\Gamma} + \sigma_u^2 \mathbf{I}_p - \sigma_u^2 \mathbf{I}_p)^{-1} \boldsymbol{\gamma} = (\mathbf{\Gamma})^{-1} \boldsymbol{\gamma} = (\mathbf{\Gamma})^{-1} \mathbf{\Gamma} \boldsymbol{\phi}_{true} = \boldsymbol{\phi}_{true}$ . We assume that  $E(\widehat{\sigma}_{ut}^2) = \sigma_{ut}^2$  and that  $\sum_t \sigma_{ut}^2 / T \rightarrow \sigma_u^2$  (see (2)). Hence  $E(\widehat{\sigma}_{ut}^2) \rightarrow \sigma_{ut}^2$ , and (7) holds if  $\text{Var}(\widehat{\sigma}_{ut}^2) = \sum_{t=1}^T \text{Var}(\widehat{\sigma}_{ut}^2) / T^2 + \sum_t \sum_{t' \neq t} \text{Cov}(\widehat{\sigma}_{ut}^2, \widehat{\sigma}_{ut'}^2) / T^2 \rightarrow 0$ . Using conditioning,  $\text{Cov}(\widehat{\sigma}_{ut}^2, \widehat{\sigma}_{ut'}^2) = E[\text{Cov}(\widehat{\sigma}_{ut}^2, \widehat{\sigma}_{ut'}^2 | Y_1, \dots, Y_T)] + \text{Cov}[E(\widehat{\sigma}_{ut}^2 | Y_1, \dots, Y_T), E(\widehat{\sigma}_{ut'}^2 | Y_1, \dots, Y_T)]$ . Since the conditional covariance is 0, this becomes  $\text{Cov}[E(\widehat{\sigma}_{ut}^2 | Y_t), E(\widehat{\sigma}_{ut'}^2 | Y_{t'})]$ . So, sufficient conditions for (7) are that  $\sum_{t=1}^T \text{Var}(\widehat{\sigma}_{ut}^2) / T$  and  $\sum_t \sum_{t' \neq t} \text{Cov}[\sigma_{ut}^2(Y_t), \sigma_{ut'}^2(Y_{t'})] / T$  converge to constants.

The asymptotic variance of  $\widehat{\phi}_{CEE}$  is needed to assess efficiency. We first provide a result for the case with  $p = 1$ . Further detail for  $p > 1$  is provided in Proposition 3.

**Proposition 2** *For  $p = 1$  under the model described in Section 2 plus the assumption of Proposition 1 and the assumptions that the  $U_t$ s and  $\widehat{\sigma}_{ut}^2$ s are independent of the  $Y_t$ s, and that the limits below exist, the asymptotic variance of  $\widehat{\phi}_{CEE}$  is*

$$a\text{Var}(\widehat{\phi}_{CEE}) = \frac{1}{T} \left[ \frac{(1 - \phi^2)(2 - \lambda)}{\lambda} + \frac{(1 - \phi^2)\overline{\sigma_u^4} + \phi^2 \overline{E(u^4)}}{\gamma_0^2} + \frac{\phi^2 \sigma_{\sigma_u^2}^2}{\gamma_0^2} \right],$$

with  $\overline{E(u^4)} = \lim_{T \rightarrow \infty} \sum_t E(u_t^4) / T$ ,  $\overline{\sigma_u^4} = \lim_{T \rightarrow \infty} \sum_t \sigma_{ut}^4 / T$ ,  $\sigma_u^2 = \lim_{T \rightarrow \infty} \sum_t \sigma_{ut}^2 / T$ , and  $\sigma_{\sigma_u^2}^2 = \lim_{T \rightarrow \infty} T \text{Var}(\widehat{\sigma}_{ut}^2)$ . If we assume further constant variance  $\sigma_u^2$  and  $E(u_t^4) = \delta \sigma_u^4$ , then

$$a\text{Var}(\widehat{\phi}_{CEE}) = \frac{1}{T} \left[ \frac{(1 - \phi^2)(2 - \lambda)}{\lambda} + \frac{(1 - \lambda)^2 (1 + (\delta - 1)\phi^2)}{\lambda^2} + \frac{\phi^2 \sigma_{\sigma_u^2}^2}{\gamma_0^2} \right].$$

When the measurement error is normally distributed ( $\delta = 3$ ) this reduces to

$$a\text{Var}(\hat{\phi}_{CEE}) = \frac{1}{T} \left[ \frac{1 + 2\phi^2 - 3\phi^2\lambda(2 - \lambda)}{\lambda^2} + \frac{\phi^2\sigma_u^2}{\gamma_0^2} \right].$$

If the covariances between  $\hat{\sigma}_u^2$  and  $\hat{\gamma}_{W,k}$ ,  $k = 0, 1$  are non-zero then the following terms are added:  $-2\phi^2\text{Cov}(\hat{\gamma}_{W,0}, \hat{\sigma}_u^2)/\gamma_0^2 + 2\phi\text{Cov}(\hat{\gamma}_{W,1}, \hat{\sigma}_u^2)/\gamma_0^2$ .

Finally, we give an asymptotic normality (and asymptotic variance) result for any  $p$ . First, let  $\hat{\gamma}_W = (\hat{\gamma}_{W,0}, \hat{\gamma}_{W,1}, \dots, \hat{\gamma}_{W,p})^\top$  and  $\gamma_W = (\gamma_0 + \sigma_u^2, \gamma_1, \dots, \gamma_p)^\top$ .

If  $\lim_{T \rightarrow \infty} T\text{Var} \begin{bmatrix} \hat{\gamma}_W - \gamma_W \\ \hat{\sigma}_u^2 - \sigma_u^2 \end{bmatrix} = \mathbf{Q}$  then from the multivariate delta method

$\lim_{T \rightarrow \infty} T\text{Var}(\hat{\phi}_{CEE}) = \mathbf{G}\mathbf{Q}\mathbf{G}^\top$  where  $\mathbf{G}$  is the matrix of derivatives of  $\hat{\phi}_{CEE}$

with respect to the components of  $\hat{\gamma}_W$  and  $\hat{\sigma}_u^2$ , evaluated at  $\gamma_W$  and  $\sigma_u^2$ . Ana-

lytical expressions for  $\mathbf{Q}$  depend on assumptions about the  $u_{ts}$  and the  $\hat{\sigma}_{ut}^2$ s, in particular whether they are dependent on the  $Y_{ts}$  or not. Decomposing  $\mathbf{Q}$  as

$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_\gamma & \mathbf{Q}_{\gamma, \sigma_u^2} \\ \mathbf{Q}_{\gamma, \sigma_u^2}^\top & Q_{\sigma_u^2} \end{bmatrix}$ , let  $q_{\gamma rs}$  denote the  $(r, s)$ th element of  $\mathbf{Q}_\gamma$  and  $q_{rs}$  the

corresponding value if there were no measurement error (e.g. Brockwell and

Davis, 1996, equation 7.3.13 with their  $\mathbf{V}$  as our  $\mathbf{Q}$ ). The following is proven in

Appendix A.3.

**Proposition 3** *Under the model described in Section 2, the assumption of Proposition 1, and the assumptions that the  $U_{ts}$  and  $\hat{\sigma}_{ut}^2$ s are independent of the  $Y_{ts}$ ,*

- $q_{\gamma 00} = q_{00} + 4\gamma_0\sigma_u^2 + \overline{E(u^4)} - \overline{\sigma_u^4}$ , where  $\overline{E(u^4)}$  and  $\overline{\sigma_u^4}$  are defined in Proposition 2.
- $q_{\gamma 0r} = q_{0r} + 4\gamma_r\sigma_u^2$ ; for  $r \neq 0$ ,
- $q_{\gamma rs} = q_{rs} + 2\sigma_u^2[\gamma_{r-s} + \gamma_{r+s}]$ ; for  $r \neq 0, s \neq 0, r \neq s$
- $q_{\gamma rr} = q_{rr} + 2\sigma_u^2[\gamma_0 + \gamma_{2r}] + \lim_{T \rightarrow \infty} \sum_t \sigma_{ut}^2 \sigma_{u,t+r}^2 / T$ , for  $r \neq 0$ .

- $\mathbf{Q}_{\gamma, \sigma_u^2} = \mathbf{0}$
- $\mathbf{Q}_{\sigma_u^2} = \sigma_{\sigma_u^2}^2$ .

If we assume further that the  $(u_t, \hat{\sigma}_{u_t}^2)$  are i.i.d. with  $E(u_t^2) = \sigma_u^2, E(u_t^4) = \delta \sigma_u^4$ , then

$$T^{1/2}(\hat{\phi}_{CEE} - \phi) \Rightarrow N(\mathbf{0}, \mathbf{G}\mathbf{Q}\mathbf{G}^\top), \quad (8)$$

where the components of  $\mathbf{Q}$  are as above with some simplification:  $q_{\gamma 00} = q_{00} + 4\gamma_0 \sigma_u^2 + \sigma_u^4(\delta - 1)$ , ( $\delta = 3$  under normality of  $u_t$ ), and  $q_{\gamma rr} = q_{rr} + 2\sigma_u^2[\gamma_0 + \gamma_{2r}] + \sigma_u^4$ , for  $p \neq 0$ .

In this case, assuming normality of  $u_t$ , the asymptotic variance of  $\hat{\phi}_{CEE}$  can be estimated using  $\hat{\sigma}_u^2, \hat{\gamma}_0 = \hat{\gamma}_{W,0} - \hat{\sigma}_u^2, \hat{\gamma}_p = \hat{\gamma}_{W,p}$  for  $p \neq 0$  and estimating the variance of  $\hat{\sigma}_u^2$  using the (sample variance of the  $\hat{\sigma}_{u_t}^2$ s)/ $T$ . Additional discussion on estimating the variance of  $\hat{\phi}_{CEE}$  is in Appendix A.3.

## 4.2 Likelihood Based Methods and the Kalman Filter

If  $\mathbf{Y}$  and  $\mathbf{u}$  are both normal and independent of each other then  $\mathbf{W} = \mathbf{Y} + \mathbf{u}$  is normal with covariance matrix  $\Sigma_Y + \Sigma_u$  and maximum likelihood (ML) techniques can be employed. Our interest is in cases where estimates of  $\Sigma_u$  are available. Treating these as fixed and maximizing the Gaussian likelihood over the rest of the parameters leads to a pseudo-ML estimation. Bell and Hillmer (1990) and Bell and Wilcox (1993) took this approach where  $Y_t$  and  $u_t$  follow ARIMA models. Wong, Miller and Shrestha (2001) formulate and estimate this approach using a state space model and related techniques. We implement this model using an EM algorithm with the Kalman filter to compute the required conditional expectations (e.g. Shumway and Stoffer, 1982). Details are in Appendix A.4. The following describes the asymptotic behavior of  $\hat{\phi}_{pML}$ .



**Proposition 4** Let  $W_t = Y_t + U_t, t = 1, \dots, T$  where  $Y_t$  is iid Gaussian and AR(1) with parameter  $|\phi| < 1$  and  $U_t \stackrel{i.i.d.}{\sim} N(0, \sigma_u^2)$ . Let  $\hat{\phi}_{pML}$  be a pseudo-maximum likelihood estimate of  $\phi$  from the series of length  $T$ , estimated by the Kalman filter and the EM algorithm for instance (see Appendix A.4). Let  $\hat{\phi}_{pML}$  use  $\hat{\sigma}_u^2$ , an estimate of the measurement error variance that is consistent as  $T \rightarrow \infty$  with  $\lim_{T \rightarrow \infty} T \text{Var}(\hat{\sigma}_u^2) = \sigma_{\sigma_u}^2$ . As  $T \rightarrow \infty$ ,  $\hat{\phi}_{pML}$  is normally distributed with mean  $\phi$  and

$$\begin{aligned}
a\text{Var}(T^{1/2}\hat{\phi}_{pML}) &= \frac{1}{\mathcal{I}_{\phi,\phi}} + \sigma_{\sigma_u}^2 \left( \frac{\mathcal{I}_{\phi,\sigma_u^2}}{\mathcal{I}_{\phi,\phi}} \right)^2, \\
&\text{where } g(\omega) = \sigma_u^2 + \frac{\sigma_e^2}{1 + \phi^2 - 2\phi \cos \omega}, \\
\mathcal{I}_{\phi,\phi} &= \frac{1}{2\pi} \int_0^\pi \left\{ g(\omega)^{-1} \frac{\partial g(\omega)}{\partial \phi} \right\}^2 d\omega \\
&= \frac{1}{2\pi} \int_0^\pi \left\{ \frac{2\phi - 2 \cos \omega}{1 + \phi^2 - 2\phi \cos \omega} \right\}^2 \\
&\quad \times \left\{ \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2(1 + \phi^2 - 2\phi \cos \omega)} \right\}^2 d\omega, \\
\text{and } \mathcal{I}_{\phi,\sigma_u^2} &= \frac{1}{2\pi} \int_0^\pi g(\omega)^{-2} \frac{\partial g(\omega)}{\partial \phi} \frac{\partial g(\omega)}{\partial \sigma_u^2} d\omega \\
&= \frac{1}{2\pi} \int_0^\pi \frac{-\sigma_e^2(2\phi - 2 \cos \omega)}{\{\sigma_u^2(1 + \phi^2 - 2\phi \cos \omega) + \sigma_e^2\}^2} d\omega.
\end{aligned}$$

Proof: With  $\hat{\sigma}_u^2$  held fixed,  $T^{1/2}\hat{\phi}_{pML}$  is asymptotically normal with variance  $\mathcal{I}_{\phi,\phi}^{-1}$  (e.g. Harvey, 1990, section 4.5.1, page 210). The additional variability produced by estimating  $\sigma_u^2$  is accounted for using Gong and Samaniego, 1981, equations 2.5 and 2.6. We note also that the spectral generating function for the Gaussian AR(1) process plus independent homoskedastic measurement error,  $g(\omega)$ , is given in Harvey, 1990, example 2.4.6, page 60. The integrals have evaded analytic solution, and equivalent time domain representations of the asymptotic information matrix involve a matrix inverse that is analytically intractable (although the likelihood and its derivatives can be computed efficiently using the Kalman filter). Nonetheless, these one-dimensional definite

integrals can be evaluated numerically.

Although the pseudo-ML / Kalman filter approach appears to be straightforward, it has subtle potential problems when the measurement error is heteroskedastic. First, if the measurement error variances are treated as unknown parameters without additional structure (as in Wong et al., 2001), then the number of parameters increases with the sample size (e.g. Neyman and Scott, 1948). We are unaware of any work that explores the asymptotic properties of the estimator in this specific time series case, and we expect this problem to be delicate. As an Associate Editor pointed out, related work by Carroll and Cline (1988) considers the problem of linear regression with independent heteroskedastic errors in a replicated response and treats the variances as unknown parameters. In that simpler case, consistency and asymptotic normality depend on the number of replicates from which the unknown variances are estimated.

An alternative is to model the variance with a simple parametric function, treat  $\{W_t, \hat{\sigma}_{u_t}^2\}$  as data (including assuming a distributional model for  $\hat{\sigma}_{u_t}^2$ ), and use the flexible Kalman filter / state space approach to fit the model (e.g. Harvey, 1990, chapter 6). One way to implement this would be to model the  $\sigma_{u_t}^2$ s a simple function of  $t$ . This approach is sometimes susceptible to misspecification though, and it does not address fact that the measurement error at time  $t$  is often proportional to  $y_t$ . A second implementation could model the heteroskedasticity parametrically as a simple function of  $y_t$ . This approach has the problem that the marginal density of  $\mathbf{W}$  typically will not be normal though since  $\text{Var}(W_t|y_t) = \text{Var}(u_t|y_t)$  depends on  $y_t$ . However, if we expand the density of  $W|y$  as a function of  $\text{Var}(W_t|y_t)$  around  $\sigma_{u_t}^2$  then a first order approximation gives an approximately normal  $\mathbf{W}$  with covariance  $\Sigma_Y + \Sigma_u$ . In our problem  $\Sigma_u = \text{diag}[E\{\text{Var}(W_1|Y_1)\}, \dots, \{\text{Var}(W_T|Y_T)\}]$ . The quality of this approximation will be situation dependent. We do not pursue these

approaches in this paper.

### 4.3 Method based on the ARMA( $p, p$ ) model

From Lemma 1 in Section 2  $W_t$  follows an ARMA( $p, p$ ) model when the measurement error is homoskedastic. As a result, a simple approach to estimating  $\phi$  (part of the hybrid approach in Wong and Miller, 1990), is to fit an ARMA( $p, p$ ) to the observed  $w$  series and to use the resulting estimate of the autoregressive parameters, which we call  $\hat{\phi}_{ARMA}$ . We perform this estimation using the “off the shelf” `arima()` function in the `ts` package in R. This function uses the Kalman filter to compute the Gaussian likelihood and its partial derivatives and finds maximum likelihood estimates using quasi-Newton methods. The asymptotic properties of  $\hat{\phi}_{ARMA}$  can be computed using existing results (e.g. Brockwell and Davis, 1996, Chapter 8). This will be a function of the  $\phi$ 's and the  $\theta$ 's and can be expressed in terms of the original parameters by obtaining the  $\theta$ 's as functions of  $\phi$ ,  $\sigma_u^2$  and  $\sigma_\epsilon^2$ ; see Appendix A.5.

For instance, when  $p = 1$   $\hat{\phi}_{ARMA} \sim AN \left\{ \phi, \frac{(1+\phi_1\theta_1)^2(1-\phi_1)^2}{(\phi_1+\theta_1^2)T} \right\}$ , where  $AN$  denotes approximately normal. Note that this result does not require normality of the observed series. There are two solutions for  $\theta_1$  resulting from (2), but only one leads to a stationary process. Defining  $k = (\sigma_\epsilon^2 + \sigma_u^2(1 + \phi_1^2))/(\sigma_u^2\phi_1)$ , then,  $\theta_1 = (-k + (k^2 - 4)^{1/2})/2$  if  $0 < \phi_1 < 1$  and  $\theta_1 = (-k - (k^2 - 4)^{1/2})/2$  when  $-1 < \phi_1 < 0$ .

### 4.4 Modified Yule-Walker

This approach assumes uncorrelated measurement errors and requires no knowledge of the measurement error variances. For a fixed  $j > p$  define  $\hat{\gamma}_{W(j)} =$

$$(\widehat{\gamma}_{W,j}, \dots, \widehat{\gamma}_{W,j+p-1})^\top \text{ and } \widehat{\Gamma}_{W(j)} = \begin{pmatrix} \widehat{\gamma}_{W,j-1} & \dots & \widehat{\gamma}_{W,j-p} \\ \vdots & \ddots & \vdots \\ \widehat{\gamma}_{W,j+p-2} & \dots & \widehat{\gamma}_{W,j-1} \end{pmatrix}. \text{ For } j > 0,$$

recall that the lag  $j$  covariance for  $\mathbf{W}$  is equivalent to that in the  $\mathbf{Y}$  series, and  $\widehat{\gamma}_{W,j}$  estimates  $\gamma_j$  consistently under regularity conditions (Section 3.2). The modified Yule-Walker estimator  $\widehat{\phi}_{MYW_j}$  solves  $\widehat{\Gamma}_{W(j)}\phi - \widehat{\gamma}_{W(j)} = \mathbf{0}$ , or  $\widehat{\phi}_{MYW_j} = \widehat{\Gamma}_{W(j)}^{-1}\widehat{\gamma}_{W(j)}$ . For the AR(1) model,  $\widehat{\phi}_{MYW_j} = \widehat{\gamma}_{W,j}/\widehat{\gamma}_{W,j-1}$ , for  $j \geq 2$ . These estimators are consistent, and one can use Proposition 3 to arrive at the covariance and more general results for the modified Yule-Walker estimates. Note however that the denominator of  $\widehat{\phi}_{MYW_j}$  can have positive probability around 0, leading to an undefined mean for the estimator. In Section 5, we found this estimator to be impractical for moderate sample sizes.

The modified Yule-Walker estimate was first discussed in detail by Walker (1960) (his Method B) and subsequently by Sakai et al. (1979). For  $p = 1$  and constant measurement error variance, Walker shows the asymptotic variance of  $T^{1/2}\widehat{\phi}_{MYW_2}$  is  $\frac{1-\phi^2}{\phi^2} \left\{ 1 + 2\frac{1-\lambda}{\lambda} + \frac{(1-\lambda)^2(1+\phi^2)}{\lambda^2(1-\phi^2)} \right\}$  and that this is smaller than the asymptotic variance of  $\widehat{\phi}_{MYW_j}$ , for  $j \geq 3$ . Sakai et al. (1979) provide the asymptotic covariance matrix of  $\widehat{\phi}_{MWY_{p+1}}$  for general  $p$ . Our Proposition 3 can also be used to arrive at these, and more general, results for the modified Yule-Walker estimators.

Chanda (1996) proposed a related estimator, in which  $j$  increases as a function of  $n$  and derived the asymptotic behavior of that estimator. Sakai and Arase (1979) also discuss this estimator, extensions, and asymptotic properties. Neither the fact that this estimator often has no finite sampling mean nor corrections for that deficiency appear to be discussed in the literature.

## 5 Mean Squared Error Comparisons of the Estimators

First, we compare the asymptotic variances of the estimators of  $\phi$  ( $p = 1$ ) in Section 5.1. After that, Section 5.2 summarizes two Monte Carlo simulations that evaluate the small sample performance of the estimators in the presence of homoskedastic and heteroskedastic measurement error.

### 5.1 Comparison of the Asymptotic Variances

Section 4 presented four estimation methods to address the problem of an autoregressive model plus additive measurement error: corrected estimating equations (CEE), modified Yule-Walker (MYW), ARMA, and pseudo-Maximum Likelihood (pML). The following proposition compares the pML estimator with the CEE estimator. The result is an immediate consequence of Propositions 2 and 4.

**Proposition 5** *Under the assumptions of Proposition 4 (which include normality of  $\mathbf{Y}$  and the measurement error and homoskedastic measurement error), the asymptotic relative efficiency of  $\hat{\phi}_{CEE}$  to  $\hat{\phi}_{pML}$ , with ( $p = 1$ ), is  $ARE(\hat{\phi}_{CEE}/\hat{\phi}_{pML}) = \frac{\mathcal{I}_{\phi, \phi}^2 \left[ \gamma_0^2 \{1 + 2\phi^2 - 3\phi^2 \lambda(2 - \lambda)\} + \lambda^2 \phi^2 \sigma_u^2 \right]}{\lambda^2 \gamma_0^2 \left( \mathcal{I}_{\phi, \phi} + \sigma_u^2 \mathcal{I}_{\phi, \sigma_u}^2 \right)}$ .*

Although we do not have simple analytic expressions for the information terms in this equation, they are easy to compute. Figure 2 calculates the ARE as a function of  $0.1 < \phi < 0.9$  and  $0.5 < \lambda \leq 1$  when  $\sigma_{\sigma_u}^2 = 2\sigma_u^4$  (left panel) and  $\sigma_{\sigma_u}^2 = 0$  (right panel). The variance  $\sigma_{\sigma_u}^2 = 2\sigma_u^4$  would occur if each  $\hat{\sigma}_{ut}^2$  were computed from two independent Gaussian replicates for instance. For moderate amounts of measurement error ( $\lambda$  greater than about 0.6) and as  $\phi$  gets closer to zero (typically the null),  $\hat{\phi}_{CEE}$  is nearly as efficient as  $\hat{\phi}_{pML}$ . A

comparison of the two panels shows that a known  $\sigma_u^2$  ( $\sigma_{\sigma_u}^2 = 0$ ) often does not cause a large gain in efficiency relative to an estimator based on two replicates per time point. This suggests that a joint model for  $(\hat{\sigma}_{ut}^2, W_t)$  often would not result in a much more efficient estimator.

We note that the difference between the asymptotic variances of the pML and CEE estimators is a departure from the case of linear regression with covariate measurement error. This can be explained simply by observing that in contrast to linear regression with covariate measurement error, in the time series problem the corrected estimating equation is not the same as the pML estimating equation.

We also have simple expressions for the asymptotic variance of  $\hat{\phi}_{MYW}$  and  $\hat{\phi}_{ARMA}$ , estimators that do not require or use an estimate of  $\sigma_u^2$ . Again, although an analytical comparison of these asymptotic variances appears to be intractable, Figure 3 compares the asymptotic variances as a function of  $\lambda$  and  $\phi$  for the AR(1) case with homoskedastic measurement errors. The asymptotic variances of the CEE and pML estimators use  $\sigma_{\sigma_u}^2 = 2\sigma_u^4$ . Three observations about this figure follow. First, unless  $\phi$  is close to one, the MYW method is the least asymptotically efficient of all the estimators. Next, similarly, the ARMA estimator is less efficient asymptotically than either the CEE estimator or the pML estimator when  $\phi$  is less than about 0.5. Finally, the simulations in the next Section suggest that when the time series is relatively short (20, 50, or 200) then the pML and CEE estimators (which perform quite similarly in the simulations) result in a much lower MSE than the ARMA estimator over a range of  $\phi$ s and  $\lambda$ s.

## 5.2 Short Time Series Simulations

We performed four sets of Monte Carlo simulations to compare the small sample performance of the CEE, pML, and ARMA estimators presented in Section 4. We do not include the MYW estimator since it often does not have a sampling mean, and, as a result, Monte Carlo estimates of its bias were (not surprisingly) very unstable. As benchmarks, we also include the “naive” estimator that ignores measurement error and the “gold standard” that applies the maximum likelihood estimator to  $\mathbf{Y}$ .

The four sets of simulations are (1) Gaussian time series errors and homoskedastic Gaussian measurement errors, (2) Gaussian time series errors and heteroskedastic Gaussian measurement errors, (3) Gaussian time series errors and homoskedastic recentered and rescaled chi-squared measurement errors, and (4) recentered and rescaled chi-squared time series errors and homoskedastic recentered and homoskedastic Gaussian measurement errors. The chi-squared errors had three degrees of freedom and were recentered to have mean zero and rescaled so that the resulting variance was the same as in the Gaussian simulations. The heteroskedastic measurement error uses a power model for the measurement error variance:  $\sigma_{ut}^2 = \beta y_t^2$  and with equal sampling effort. As in the homoskedastic case,  $\lambda = \text{var}(Y_t)/\text{var}(W_t)$ . In the heteroskedastic case,  $\lambda = \{\sigma_e^2(1 - \phi^2)\} / [\beta \{1 + \sigma_e^2/(1 - \phi^2)\}]$  and  $\beta = (1 - \lambda)/\lambda$ . In the homoskedastic case,  $\hat{\phi}_{CEE}$  and  $\hat{\phi}_{pML}$  use  $\hat{\sigma}_u^2$  estimated from two normal replicates per observation, but using the true  $\sigma_u^2$  yielded nearly identical results. In the heteroskedastic simulations,  $\hat{\phi}_{CEE}$  uses  $\hat{\sigma}_u^2 = \sum_{t=1}^T \hat{\sigma}_{ut}^2/T$  and  $\hat{\phi}_{pML}$  uses  $\hat{\sigma}_{ut}^2, t = 1, \dots, T$ .

Each set of simulations used a replicated (Monte Carlo sample size = 500) full factorial combination of 3 factors: sample size ( $T = 20, 50, 200$ ), measurement error ( $\lambda = 0.5, 0.75, 0.9$ ), and autoregressive parameter ( $\phi = 0.1, 0.5, 0.9$ ).

Note that we parameterize the measurement error in terms of  $\lambda$  rather than  $\sigma_u^2$  since a value for  $\sigma_u^2$  only has meaning relative to  $\text{var}(y_t)$  which is a function of  $\phi$ .

Since each of the four sets of simulations told a nearly identical story about the relative performances of these estimators, we only include a summary of the first set of simulations (Figure 4). This figure depicts the average bias and standard error of the five estimators as a function of either  $T$ ,  $\phi$  or  $\lambda$  with the average computed over combinations of the other two factors. The established deficiencies of the naive approach are evident. The CEE and PML estimators have very similar performance and as we move towards extreme values of the factors ( $T = 200$ ,  $\phi = 0.1$  and  $\lambda = 0.9$ ) their performance approaches that of the gold standard; e.g., estimators based on the true values  $Y_1, \dots, Y_T$ .

## 6 Discussion

We have done three things in this paper. First, we derived the biases caused by ignoring measurement error in autoregressive models. Next, we proposed an easy to compute estimator to correct for the effects of measurement error and explored its asymptotic properties. An important and novel feature of this estimator is that it can be proven to be consistent even when faced with heteroskedastic measurement error. Finally, we reviewed and critiqued some of the disparate literature proposing estimators for autoregressive models with measurement error and compared our new estimator to some of the available estimators both asymptotically and with a designed simulation study. These comparisons suggest that the new estimator is often quite efficient relative to existing estimators.

Following up on this work, we have identified two areas that we feel are ripe for future study. First is further development of procedures for valid in-



ferences about the autoregressive parameters in both large and small samples in the presence of measurement error. An important consideration with small samples will be the fact that the current estimators all have substantial small sample bias (as does the gold standard). We would like to explore the use of REML type estimators and second order bias corrections (e.g. Cheang and Reinsel, 2000) in the presence of measurement error. Additionally, further investigation is needed into estimation of the asymptotic variance of  $\hat{\phi}_{CEE}$  (see the end of Section 4.1 and Appendix A.3 for some discussion) and the performance of inferences that make use of these. Also, while Section 4.1 provides some results that accommodate heteroskedasticity, general determination of the asymptotic distribution and how to estimate the asymptotic covariance matrix of the estimators under various scenarios of measurement error heteroskedasticity is a challenging problem that needs more study.

A second area for future study that may benefit from an estimating equations based approach is where the measurement errors are correlated, a common occurrence with the use of block resampling techniques (see, for example, Pfeiffermann, Feder, and Signorelli, 1998 or and Wilcox, 1993). Related to that work, we believe that the corrected estimating equations approach can be used for estimation of the unobserved responses and forecasting. This approach may prove to be especially useful when the measurement error is heteroskedastic.

## APPENDIX: TECHNICAL DETAILS

### A.1 Sample moments based on $\mathbf{W}$ .

Writing  $W_t = Y_t + u_t$ ,  $\hat{\gamma}_{W,0} = \sum_{t=1}^T W_t^2/T = \sum_{t=1}^T Y_t^2/T + \sum_{t=1}^T u_t Y_t/T + \sum_{t=1}^T u_t^2/T$ . The first term converges in probability to  $\gamma_0$ . The third will converge to  $\sigma_u^2 = \lim \sum_{t=1}^T \sigma_{ut}^2/T = \lim \sum_{t=1}^T E(u_t^2)/T$  as long as  $\sum \text{Var}(u_t^2)/T$  converges to a constant. The middle term converges to 0 in probability since it

has mean 0 (recall  $\text{Cov}(Y_t, u_t) = 0$ ) and variance  $(1/T^2)\{\sum_t \text{E}[\text{Var}(Y_t u_t | u_t)] + \text{Var}[\text{E}(Y_t u_t | u_t)]\} = (1/T^2)\{\sum_t \text{E}[u_t^2 \gamma_0]\} = (\gamma_0/T) \sum_t \sigma_{ut}^2/T$ , which converges to 0. Hence  $\hat{\gamma}_{W,0}$  converges in probability to  $\gamma_0 + \sigma_u^2$ .

## A.2 Asymptotic Bias of naive estimate of variance

We prove that the naive estimate  $\hat{\sigma}_{e,naive}^2$  has positive bias as  $T \rightarrow \infty$ . It suffices to show that  $\{\Gamma^{-1} - (\Gamma + \sigma_u^2 \mathbf{I})^{-1}\}$  is positive definite which results from showing that its eigenvalues are all positive. Since  $\Gamma$  is symmetric positive definite,  $\Gamma = \mathbf{U}\mathbf{D}\mathbf{U}^\top$  where  $\mathbf{U}$  is an orthogonal matrix, and  $\mathbf{D} = \text{diag} d_j, d_j > 0, j = 1, \dots, p+1$ . The  $d_j$ s are the eigenvalues of  $\Gamma$ . Similarly,  $\Gamma^{-1} = \mathbf{U}^\top \mathbf{D}^{-1} \mathbf{U}$ ,  $\Gamma + \sigma_u^2 \mathbf{I} = \mathbf{U}(\mathbf{D} + \sigma_u^2 \mathbf{I})\mathbf{U}^\top$ , and  $(\Gamma + \sigma_u^2 \mathbf{I})^{-1} = \mathbf{U}^\top (\mathbf{D} + \sigma_u^2 \mathbf{I})^{-1} \mathbf{U}$ . This means that  $\{\Gamma^{-1} - (\Gamma + \sigma_u^2 \mathbf{I})^{-1}\} = \mathbf{U}^\top \mathbf{D}^{-1} \mathbf{U} - \mathbf{U}^\top (\mathbf{D} + \sigma_u^2 \mathbf{I})^{-1} \mathbf{U} = \mathbf{U}^\top \{\mathbf{D}^{-1} - (\mathbf{D} + \sigma_u^2 \mathbf{I})^{-1}\} \mathbf{U}$ . Hence, the eigenvalues of  $\{\Gamma^{-1} - (\Gamma + \sigma_u^2 \mathbf{I})^{-1}\}$  are  $1/d_j - 1/(d_j + \sigma_u^2), j = 1, \dots, p+1$  which are all positive.

## A.3: Asymptotic Properties of $\hat{\phi}_{CEE}$ .

Using developments similar to those in Brockwell and Davis, 1996, Chapter 7), the asymptotic behavior of  $T^{1/2} \begin{bmatrix} \hat{\gamma}_W - \gamma_W \\ \hat{\sigma}_u^2 - \sigma_u^2 \end{bmatrix}$  is equivalent to that of  $T^{1/2} \bar{\mathbf{Z}}$ , where  $\bar{\mathbf{Z}} = \sum_t \mathbf{Z}_t/T$ , with  $\mathbf{Z}'_t = [W_t^2, W_t W_{t+1}, \dots, W_t W_{t+p}, \hat{\sigma}_{ut}^2]$ . Hence, assuming the limit exists,  $\mathbf{Q} = \lim_{T \rightarrow \infty} \text{Cov}(T^{1/2} \bar{\mathbf{Z}})$ . To evaluate the components of  $\mathbf{Q}_{\gamma_r}$ , one needs  $\lim_{T \rightarrow \infty} T \text{Cov}(\hat{\gamma}_{W,p}, \hat{\gamma}_{W,r})$   
 $= \lim_{T \rightarrow \infty} (1/T) \text{E}(\sum_t W_t W_{t+p} \sum_s W_s W_{s+r}) - \gamma_{W,p} \gamma_{W,s}$ . This requires  
 $\text{E}(W_t W_{t+p} W_s W_{s+r}) = \text{E}(Y_t Y_{t+p} Y_s Y_{s+r}) + [\text{E}(Y_t Y_{t+p} U_s^2) + \text{E}(Y_t U_{t+p}^3)$   
 $+ \text{E}(Y_{t+p} U_t^3)] I(r = 0) + [\text{E}(Y_s Y_{s+r} U_t^2) + \text{E}(Y_s U_t^3) + \text{E}(Y_{s+r} U_t^3)] I(p = 0) +$   
 $\text{E}(Y_t Y_s U_{t+p}^2) I(s = t+p-r) + \text{E}(Y_t Y_{s+r} U_{t+p}^2) I(s = t+p) + \text{E}(Y_{t+p} Y_s U_t^2) I(s = t-r)$   
 $+ \text{E}(Y_{t+p} Y_{s+r} U_t^2) I(s = t) + \text{E}(U_t^2 U_s^2) I(p = r = 0) + \text{E}(U_t^2 U_{t+p}^2) I(p = r \neq 0)$ ,  
where  $I(\cdot)$  is the indicator function. In general this requires modeling the dependence of the  $U_t$  on the true  $Y_t$ . Assuming independence, the expectations can be split into the product of a term involving  $U$ s and one involving  $Y$ s. Do-

ing this, straightforward but lengthy calculations lead to the expressions for  $\mathbf{Q}_\gamma$  in Section 4.1.

Similarly, the limiting value of  $TCov(\hat{\gamma}_{W,p}, \hat{\sigma}_u^2)$  is

$\lim_{T \rightarrow \infty} (1/T)E(\sum_t W_t W_{t+p} \sum_r \hat{\sigma}_r^2) - \gamma_{W,p} \sigma_u^2$ . Assuming the  $\hat{\sigma}_t^2$ s are independent of the  $Y_t$ s leads, after some simplification, to  $\lim_{T \rightarrow \infty} (1/T)E(\sum_t W_t W_{t+p} \sum_r \hat{\sigma}_r^2) = \gamma_{W,p} \sigma_u^2$  and  $\mathbf{Q}_{\lambda, \sigma_u^2} = \mathbf{0}$ .

For the AR(1), using the delta method, the asymptotic variance of  $\hat{\phi}_{CEE}$  is  $a\text{Var}(\hat{\phi}_{CEE}) = \frac{1}{\gamma_0^2} [\text{Var}(\hat{\gamma}_{W,1} - \phi \hat{\gamma}_{W,0}) + \phi^2 \text{Var}(\hat{\sigma}_u^2) - 2\phi^2 \text{Cov}(\hat{\gamma}_{W,0}, \hat{\sigma}_u^2) + 2\phi \text{Cov}(\hat{\gamma}_{W,1}, \hat{\sigma}_u^2)]$ . Using  $\mathbf{Q}_\gamma$  in Proposition 3, writing  $q_{\gamma rs} = q_{rs} + c_{rs}$ , and recognizing that the asymptotic variance of  $\hat{\phi}$  without measurement error, which is  $(1 - \phi^2)/T$ , can be expressed as  $(1/T\gamma_0^2)(q_{11} - 2\phi q_{01} + \phi^2 q_{00})$ , then  $a\text{Var}(\hat{\phi}_{CEE}) = (1 - \phi^2)/T + (1/T\gamma_0^2) [c_{11} - 2\phi c_{01} + \phi^2 c_{00}]$ . Obtaining  $c_{rs} = q_{\gamma rs} - q_{rs}$  from Proposition 3 and simplifying leads to the expressions in Proposition 2.

If

$$T^{1/2} \begin{bmatrix} \hat{\gamma}_W - \gamma_W \\ \hat{\sigma}_u^2 - \sigma_u^2 \end{bmatrix} \Rightarrow N(\mathbf{0}, \mathbf{Q}), \quad (9)$$

then the asymptotic normality of  $\hat{\phi}_{CEE}$  in (8) follows. If the  $u_t$  are i.i.d. then  $W_t$  follows an ARMA process and the asymptotic normality of  $\hat{\gamma}_W$  follows immediately from known results (e.g., Proposition 7.3.4 in Brockwell and Davis, 1996). If the  $\hat{\sigma}_{ut}^2$ s are *iid* and independent of the  $W_t$ s then then (9) follows immediately from use of the standard central limit theorem for  $\hat{\sigma}_u^2$  and the independence of  $\hat{\sigma}_{ut}^2$  and  $\hat{\gamma}_W$ .

If the  $u_t$ s and  $\hat{\sigma}_{ut}^2$ s are independent of the  $Y_t$ s but the moments of  $u_t$  and  $\hat{\sigma}_{ut}^2$  change with  $t$ , then generalizing the proofs in Chapter 7 of Brockwell and Davis (1996) requires use of a central limit theorem for m-dependent sequences with heteroskedasticity (see for example Theorem 6.3.1 in Fuller, 1996). Suitable

conditions on the moments of  $u_t$  and  $\hat{\sigma}_{ut}^2$  will lead to (9) and hence (8)

Estimation of the asymptotic variance of  $\hat{\phi}_{CEE}$  requires an estimate of  $\mathbf{Q}$ . If  $\text{Cov}(\mathbf{Z}_t, \mathbf{Z}_s)$  depends only on  $|t - s|$  (as it would if the  $u_t$ s and  $\hat{\sigma}_{ut}^2$ s were i.i.d.) and is denoted  $C(|t - s|)$  then  $\mathbf{Q} = \lim_{T \rightarrow \infty} \sum_{|h| < T} \{1 - (|h|/T)\} C(h)$ . As pointed out by an anonymous referee,  $\mathbf{Q}$  can be estimated by  $\hat{\mathbf{Q}} = \sum_{|h| < M} \{1 - (|h|/M)\} \hat{C}(h)$ , where  $\hat{C}(h) = \sum_{t=1}^{n-h} (\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_{t+h} - \bar{\mathbf{Z}})^\top / T$ ,  $M \rightarrow \infty$ , and  $M/T \rightarrow 0$  as  $T \rightarrow \infty$ . This is the Bartlett kernel estimator that truncates for  $h \geq M$ . Under general conditions, a  $\hat{Q}$  with  $T$  instead of  $M$  in the above would converge to a random variable that is proportional to  $Q$  not equal to  $Q$  in probability as  $T \rightarrow \infty$ . For instance, see Bunzel, Kiefer, and Vogelsang (2001) or Kiefer and Vogelsang (2002).

#### A.4 State space formulation, EM algorithm, and the Kalman filter

We assume  $\mathbf{Y}$  is Gaussian and  $\text{AR}(p)$ , the measurement error is independent and Gaussian with a fixed variance that is held fixed at  $\hat{\sigma}_u^2$ , and estimate  $\phi$  and  $\sigma_e^2$  by “pseudo-maximum likelihood.” A straightforward way to implement this procedure is to use the Kalman filter combined with the EM algorithm (e.g. Shumway and Stoffer, 1982). Specifically, let  $\mathbf{y}_t = (y_t, \dots, y_{t-p+1})^\top$ ,  $\mathbf{M} = (1, \mathbf{0}_{p-1}^\top)$  where  $\mathbf{0}_{p-1}$  is a vector of all zeros with length  $p - 1$ ,  $\Phi = \begin{pmatrix} \phi_1 & \dots & \dots & \phi_p \\ \mathbf{1}_{p-1} & \mathbf{0}_{p-1} & \dots & \mathbf{0}_{p-1} \end{pmatrix}$  where  $\mathbf{1}_{p-1}$  is a vector of all ones with length  $p - 1$ , and  $\mathbf{e}_t = (e_t, \mathbf{0}_{p-1}^\top)^\top$ . The *observation equation* is  $w_t = \mathbf{M}\mathbf{y}_t + u_t$ , and the *state equations* are  $\mathbf{y}_t = \Phi\mathbf{y}_{t-1} + \mathbf{e}_t$ . Further,  $\mathbf{R} = \text{var}(u_t) = \hat{\sigma}_u^2$  and  $\mathbf{Q} =$

$$\text{Var}(\mathbf{e}_t) = \begin{pmatrix} \sigma_e^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \text{ Using this formulation, the }^{(s+1)}\text{th update in}$$

an EM algorithm follows. Note that  $\llbracket_{(i,j)}$  denotes the  $i, j$ th element in a matrix and  $\llbracket_{(i,:)}$  denotes the  $i$ th row.  $\left\{ \hat{\phi}^{(s+1)} \right\}^\top = \left[ \mathbf{B}^{(s)} \{ \mathbf{A}^{(s)} \}^{-1} \right]_{(1,:)}'$

$\hat{\sigma}_e^{2(s+1)} = \left( T^{-1} \left[ \mathbf{C}^{(s)} - \mathbf{B}^{(s)} \{ \mathbf{A}^{(s)} \}^{-1} \mathbf{B}^{(s)\top} \right] \right)_{(1,1)}$ , where  
 $\mathbf{A}^{(s)} = \sum_{t=1}^T \left( \mathbf{P}_{t-1,t-1}^{T(s)} + \mathbf{y}_{t-1}^{T(s)} \mathbf{y}_{t-1}^{T(s)\top} \right)$ ,  $\mathbf{B}^{(s)} = \sum_{t=1}^T \left( \mathbf{P}_{t,t-1}^{T(s)} + \mathbf{y}_t^{T(s)} \mathbf{y}_{t-1}^{T(s)\top} \right)$ ,  
and  $\mathbf{C}^{(s)} = \sum_{t=1}^T \left( \mathbf{P}_{t,t}^{T(s)} + \mathbf{y}_t^{T(s)} \mathbf{y}_t^{T(s)\top} \right)$ , with  $\mathbf{y}_t^{v(s)} = \mathbb{E}(\mathbf{y}_t | w_1, \dots, w_v)$  and  
 $\mathbf{P}_{t,u}^{v(s)} = \mathbb{E} \left\{ (\mathbf{y}_t - \mathbf{y}_t^{v(s)}) (\mathbf{y}_u - \mathbf{y}_u^{v(s)})^\top | w_1, \dots, w_v \right\}$ . The expectations are evalu-  
ated using the  $(s)$ th set of parameter updates. The last two quantities can be  
computed efficiently using the Kalman filter. Suppressing the notation for the  
EM iteration and listing the equations as they are in the Appendix of Shumway  
and Stoffer (1982), the *forward equations* are: (for  $t = 1, \dots, T$ )  $\mathbf{y}_t^{t-1} = \Phi \mathbf{y}_{t-1}^{t-1}$ ,  
 $\mathbf{P}_{t,t}^{t-1} = \Phi \mathbf{P}_{t-1,t-1}^{t-1} \Phi^\top + \mathbf{Q}$ ,  $\mathbf{K}_t = \mathbf{P}_{t,t}^{t-1} \mathbf{M}^\top (\mathbf{M} \mathbf{P}_{t,t}^{t-1} \mathbf{M}^\top + \mathbf{R})^{-1}$ ,  $\mathbf{y}_t^t = \mathbf{y}_t^{t-1} +$   
 $\mathbf{K}_t (w_t - \mathbf{M} \mathbf{y}_t^{t-1})$ , and  $\mathbf{P}_{t,t}^t = \mathbf{P}_{t,t}^{t-1} - \mathbf{K}_t \mathbf{M} \mathbf{P}_{t,t}^{t-1}$ , where  $\mathbf{y}_0^0 = \mathbf{0}$  and  $\mathbf{P}_{0,0}^0 = \mathbf{0}$ . The  
*backward equations* are: (for  $t = T, \dots, 1$ ),  $\mathbf{J}_{t-1} = \mathbf{P}_{t-1,t-1}^{t-1} \Phi^\top (\mathbf{P}_{t,t}^{t-1})^{-1}$ ,  $\mathbf{y}_{t-1}^T =$   
 $\mathbf{y}_{t-1}^{t-1} + \mathbf{J}_{t-1} (\mathbf{y}_t^T - \Phi \mathbf{y}_{t-1}^{t-1})$ , and  $\mathbf{P}_{t-1,t-1}^T = \mathbf{P}_{t-1,t-1}^{t-1} + \mathbf{J}_{t-1} (\mathbf{P}_{t,t}^T - \mathbf{P}_{t,t}^{t-1}) \mathbf{J}_{t-1}^\top$ . Ad-  
ditionally, for  $t = T, \dots, 2$ ,  $\mathbf{P}_{t-1,t-2}^T = \mathbf{P}_{t-1,t-1}^{t-1} \mathbf{J}_{t-2}^\top + \mathbf{J}_{t-1} (\mathbf{P}_{t,t-1}^T - \Phi \mathbf{P}_{t-1,t-1}^{t-1} \mathbf{J}_{t-2}^\top)$   
with  $\mathbf{P}_{T,T-1}^T = (\mathbf{I} - \mathbf{K}_T \mathbf{M}) \Phi \mathbf{P}_{T-1,T-1}^{T-1}$ .

#### A.5 Asymptotic variance of ARMA estimator

The asymptotic variance of  $\hat{\phi}_{ARMA}$  is available in general form (see for ex-  
ample Chapter 8 of Brockwell and Davis, 1996), but it involves  $\phi$  as well as  $\theta$ .  
For the expression to be useful in our context,  $\theta$  must be expressed in terms of  
 $\phi$ ,  $\sigma_e^2$  and  $\sigma_u^2$ .

**The AR(1) case:** Applying equation (2) for  $p = 1$  leads to  $\sigma_e^2 + \sigma_u^2(1 -$   
 $\phi_1 z - \phi_1 z^{-1} + \phi_1^2) = \sigma_b^2(1 + \theta_1 z + \theta_1 z + \theta_1^2)$ . Equating coefficients and making  
substitutions leads to  $\sigma_b^2 = -\sigma_u^2 \phi_1 / \theta_1$  and  $\theta_1$  is a solution to the quadratic  
equation  $1 + \theta_1^2 + \theta_1 k = 0$  where  $k = (\sigma_e^2 + \sigma_u^2(1 + \phi_1^2)) / (\sigma_u^2 \phi_1)$ . It can be  
shown that  $k^2 - 4 = (k - 2)(k + 2)$  is positive since  $k > 2$  is equivalent to  
 $\sigma_e^2 + \sigma_u^2(1 - \phi_1^2) > 0$ . The induced model is invertible if  $|\theta_1| \leq 1$  which after some  
algebra is shown to be true for the root  $(-k + (k^2 - 4)^{1/2})/2$  when  $0 < \phi_1 < 1$   
and for the root  $(-k - (k^2 - 4)^{1/2})/2$  when  $-1 < \phi_1 < 0$ .

**AR(2) Model:** Applying equation (2) when  $p = 2$  leads to the moving average parameters  $\theta_1$  and  $\theta_2$  solving  $\theta_1(\theta_2 + 1) - c_1(1 + \theta_1^2 + \theta_2^2) = 0$  and  $\theta_2 + c_2(1 + \theta_1^2 + \theta_2^2) = 0$ , where with  $M = \sigma_e^2 + \sigma_u^2(1 + \phi_1^2 + \phi_2^2)$ ,  $c_1 = \sigma_u^2(\phi_1\phi_2 - \phi_1)/M$  and  $c_2 = -\sigma_u^2\phi_2/M$ . These equations have multiple solutions and the one leading to a stationary and invertible process would be used.

## 7 References

- Bell, W. R. , and Hillmer, S. C. (1990), "The time series approach to estimation for repeated surveys," *Survey Methodology*, 16 , 195-215.
- Bell, W. R. , and Wilcox, D. W. (1993), "The effect of sampling error on the time series behavior of consumption data," *Journal of Econometrics*, 55 , 235-265.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. (1994), *Time series analysis: forecasting and control; Third Edition*, New York: Prentice Hall.
- Brockwell, P. J. and Davis, R. A. (1996). *Introduction to time series and forecasting; Second Edition*, New York: Springer.
- Bunzel, H., Kiefer, N.M., and Vogelsang, T.J. (2001), "Simple robust testing of hypotheses in non-linear models," *Journal of the American Statistical Association*, 96, 1088-1096.
- Carroll, R., Ruppert, D., and Stefanski, L. (1995), *Measurement Error in Non-linear Models*, London: Chapman Hall.
- Chanda, K. C. (1996), "Asymptotic properties of estimators for autoregressive models with errors in variables," *The Annals of Statistics*, 24 , 423-430.
- Chanda, K. C. (1995), "Large sample analysis of autoregressive moving-average models with errors in variables," *Journal of Time Series Analysis*, 16, 1-15.
- Cheang, W-K, and Reinsel, G. C. (2000), "Bias reduction of autoregressive estimates in time series regression model through restricted maximum likeli-

- hood", *Journal of the American Statistical Association*, 95, 1173-1184.
- Dennis, B., and Taper, M. L. (1994), "Density dependence in time series observations of natural populations: estimation and testing," *Ecology*, 205-224.
- Elkinton, J., Healy, W., Buonaccorsi, J., Boettner, G., Hazzard, A., Smith, H., and Liebhold, A. (1996), "Interactions among gypsy moths, white-footed mice and acorns," *Ecology*, 77, 2332-2342.
- Feder, M. (2001), "Time series analysis of repeated surveys: The state-space approach," *Statistica Neerlandica*, 55 (2), 182-199.
- Fuller, W. A. (1996). *Introduction to Statistical Time Series, Second Edition*, New York: Wiley.
- Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley.
- Gong, G. and Samaniego, F. G. (1981), "Pseudo Maximum Likelihood Estimation: Theory and Applications," *The Annals of Statistics*, 9, 861-869.
- Harvey, A. C. (1990), "Forecasting, structural time series models, and the Kalman filter," Cambridge: Cambridge University Press.
- Granger, C. W. J., and Morris, M. J. (1976), "Time series modeling and interpretation," *Journal of the Royal Statistical Society-Series A*, 139, 246-257.
- Kiefer, N.M., and Vogelsang, T.J. (2002), "Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation," *Econometrica*, 70, 2093-2095.
- Koreisha, S. G., and Fang, Y. (1999), "The impact of measurement errors on ARMA prediction," *Journal of Forecasting*, 18, 95-109. Lee, J. H. , and Shin, D. W. (1997), "Maximum likelihood estimation for ARMA models in the presence of ARMA errors", *Communications in Statistics, Part A – Theory and Methods*, 26, 1057-1072.
- Miazaki, E. S., and Dorea, C. C. Y. (1993), "Estimation of the parameters of a

- time series subject to the error of rotation sampling", *Communications in Statistics, Part A – Theory and Methods*, 22, 805-825.
- Neyman, J. and Scott, E. L. (1948), "Consistent estimates based on partially consistent observations", *Econometrika*, 16, 1-32.
- Pagano, M. (1974), "Estimation of models of autoregressive signal plus white noise," *The Annals of Statistics*, 2 , 99-108.
- Pfeffermann, D., Feder, M., and Signorelli, D. (1998), "Estimation of autocorrelations of survey errors with application to trend estimation in small areas," *Journal of Business and Economic Statistics*, 16, 339-348.
- Pfeffermann, D. (1991), "Estimation and seasonal adjustment of population means using data from repeated surveys," *Journal of Business and Economic Statistics*, 9, 163-175.
- Sakai, H., and Arase, M. (1979), "Recursive parameter estimation of an autoregressive process disturbed by white noise," *International Journal of Control*, 30, 949-966.
- Sakai, H., Soeda, T., and Hidekatsu, T. (1979), "On the relation between fitting autoregression and periodogram with applications," *The Annals of Statistics*, 7, 96-107.
- Scott, A. J., Smith, T. M. F., and Jones, R. G. (1977), "The application of time series methods to the analysis of repeated surveys," *International Statistical Review*, 45, 13-28.
- Scott, A. J., and Smith, T. M. F. (1974), "Analysis of repeated surveys using time series methods," *Journal of the American Statistical Association*, 69, 674-678.
- Shumway, R. H. and Stoffer, D. S. (1982), "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, 3, 253-264.



Walker, A. M. (1960), "Some consequences of superimposed error in time series analysis," *Biometrika*, 47, 33-43.

Williams D. W., and A. M. Liebhold. (1995). "Influence of weather on the synchrony of gypsy moth (Lepidoptera: Lymantriidae) outbreaks in New England." *Environmental Entomology*, 24, 987-995.

Wong, Wing-keung, and Miller, R. B. (1990), "Repeated time series analysis of ARIMA-noise models," *Journal of Business and Economic Statistics*, 8, 243-250.

Wong, Wing-keung, Miller, R. B., and Shrestha, K. (2001), "Maximum likelihood estimation of ARMA models with error processes for replicated observations," *Journal of Applied Statistical Science*, 10, 287-297.

Table 1:  $w_t = \log(\text{estimated density})$  and  $\hat{\sigma}_{ut} = \text{estimated standard error of } w_t$  for mouse population over a nine year period. The estimated density is number of mice per hectare.

Year	$w_t$	$\hat{\sigma}_{ut}$
86	1.67391	0.17626
87	0.69315	0.50000
88	1.88510	0.53484
89	3.17388	0.80335
90	3.28091	0.06015
91	2.81301	0.13613
92	2.82891	0.32262
93	2.03510	0.08533
94	3.37461	0.20515

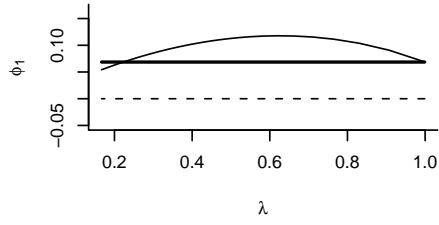
Figure 1: This figure illustrates the pernicious nature of the bias in the naive estimates in an AR(2) model. The first two rows show that the direction of the bias depends on both the autoregressive parameters  $(\phi_1, \phi_2)$  and the amount of measurement error ( $\lambda = \gamma_0 / (\gamma_0 + \sigma_u^2)$ ). The thick lines are the true parameters; the thin lines are the naive estimates; and the dotted lines are at zero. Note that measurement error increases as  $\lambda$  decreases. The two cases are  $(\phi_1, \phi_2) = (0.0686, 0.8178)$  and  $(0.7346, 0.0669)$ ;  $\hat{\phi}_{1,naive}$  is accentuated in case 1, and  $\hat{\phi}_{2,naive}$  is accentuated in case 2. The last row illustrates the potential direction of the measurement error bias in the naive AR(2) model as a function of the true autoregressive parameters over the range of possible values that result in stationary model.

Figure 2: This figure displays the log efficiency of  $\hat{\phi}_{CEE}$  relative to  $\hat{\phi}_{pML}$  as a function of  $\phi$  (different lines) and  $\lambda$  ( $x$ -axis) when  $p = 1$ . The left panel considers an estimated measurement error variance when  $\sigma_{\sigma_u^2} = 2\sigma_u^4$  and the right panel considers the variance when the measurement error is known ( $\sigma_{\sigma_u^2} = 0$ ). When  $\phi$  is close to the null and the amount of measurement error is moderate ( $\lambda > 0.7$ ),  $\hat{\phi}_{CEE}$  is nearly as asymptotically efficient as  $\hat{\phi}_{pML}$ . In small sample simulations (Section 5.2), the CEE and pML estimators resulted in nearly identical mean squared errors over a range of parameter values. Note that this figure is based on asymptotic calculations, not simulation.

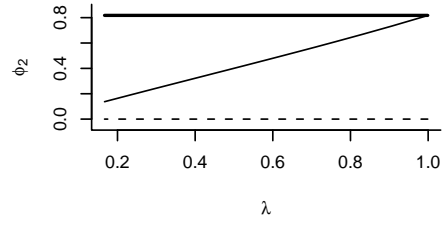
Figure 3: This figure presents the log relative asymptotic variances of four asymptotically unbiased estimators of the autoregression parameters in the AR(1) model discussed in Section 4. The log relative asymptotic variances are shown as functions of  $\phi$  (different lines) and  $\lambda$  (x-axes). Note that this figure is based on asymptotic calculations, not simulation.

Figure 4: Homoskedastic simulation results.

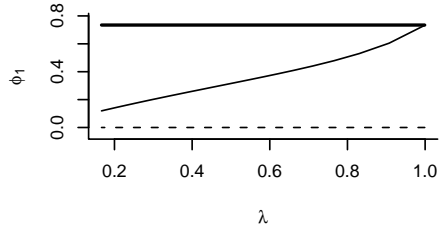
**Case 1**



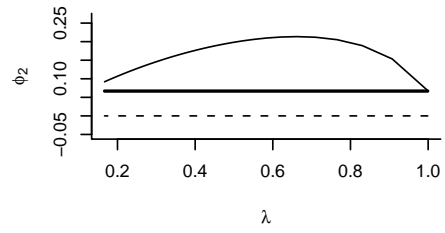
**Case 1**



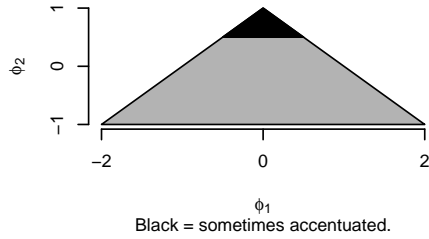
**Case 2**



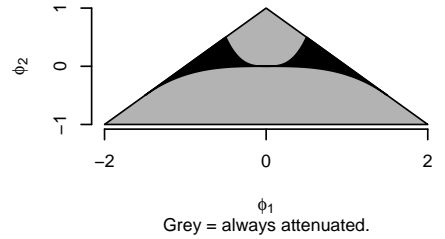
**Case 2**



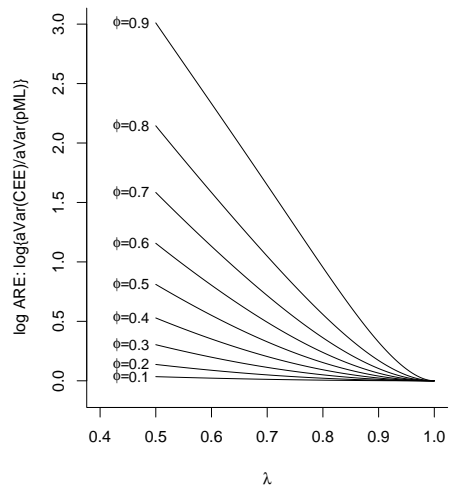
Direction of bias in  $\hat{\phi}_1$



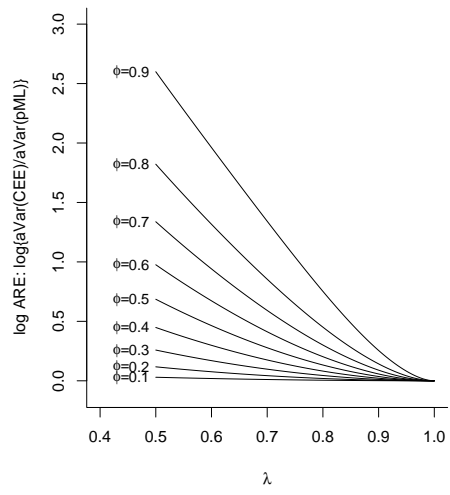
Direction of bias in  $\hat{\phi}_2$



$$\sigma_{\phi}^2 = 2\sigma_u^4$$

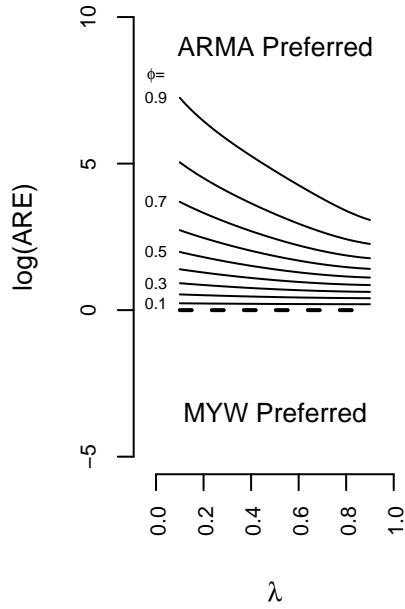


$$\sigma_{\phi}^2 = 0$$

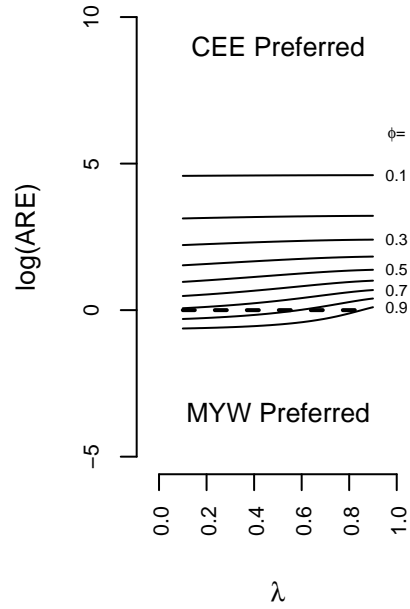




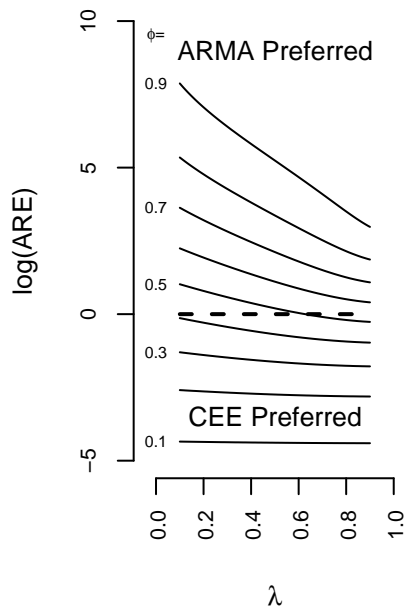
$\log[\text{var}(\text{MYW}) / \text{var}(\text{ARMA})]$



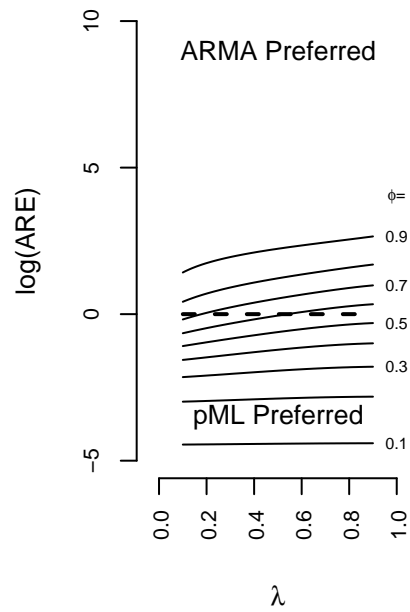
$\log[\text{var}(\text{MYW}) / \text{var}(\text{CEE})]$



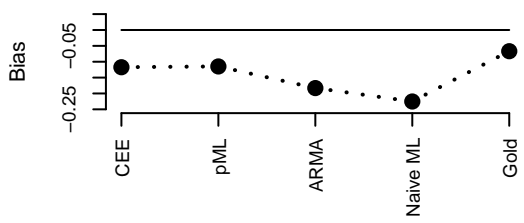
$\log[\text{var}(\text{CEE}) / \text{var}(\text{ARMA})]$



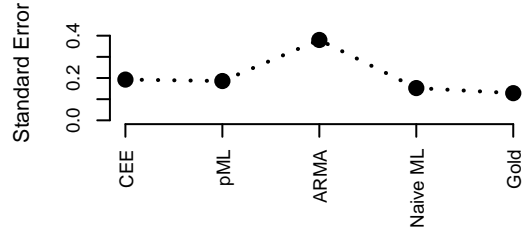
$\log[\text{var}(\text{pML}) / \text{var}(\text{ARMA})]$



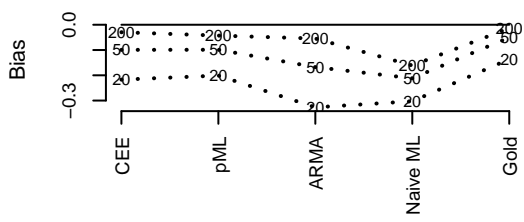
Average Bias



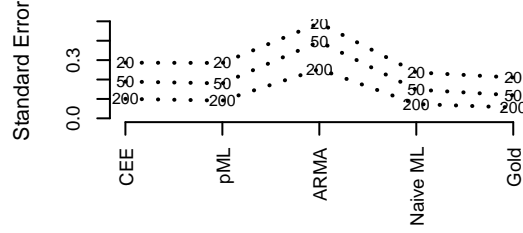
Average S.E.



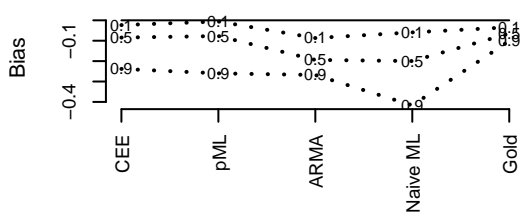
Average Bias by T



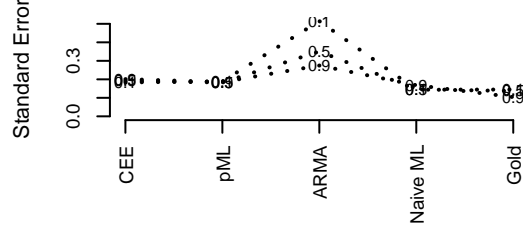
Average S.E. by T



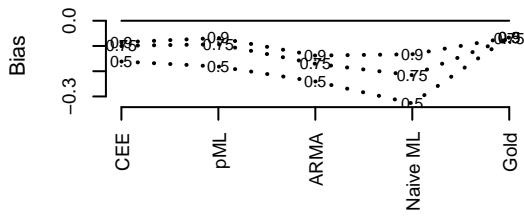
Average Bias by  $\phi$



Average S.E. by  $\phi$



Average Bias by  $\lambda$



Average S.E. by  $\lambda$

