Measurement Error in Dynamic Models

John P. Buonaccorsi

Abstract Many time series contain measurement (often sampling) error and the problem of assessing the impacts of such errors and accounting for them has been receiving increasing attention of late. This paper provides a survey of this problem with an emphasis on estimating the coefficients of the underlying dynamic model, primarily in the context of fitting linear and nonlinear autoregressive models. An overview is provided of the biases induced by ignoring the measurement error and of methods that have been proposed to correct for it, and remaining inferential challenges are outlined.

1 Introduction

Measurement error is a commonly occurring problem and is especially prominent in many time series, where the variable of interest often has to be estimated rather than observed exactly. There is a fairly diverse statistical literature which has addressed the problem of measurement error in time series as well as a burgeoning ecological literature, where the problem of modeling population dynamics in the presence of the so-called observation error has garnered considerable attention. Included among the many papers addressing this problem with real data are ones that account for errors in series involving population abundances of waterfowl (Lillegard et al. 2008; Saether et al. 2008; Viljugrein et al. 2005), voles (Stenseth et al. 2003), grouse (Ives et al. 2003) as well as labor force statistics (Pfeffermann et al. (1998), retail sales (Bell and Wilcox 1993), the number of households in Canada (Feder 2001), and disease rates (Burr and Chowell 2006).

The main ingredients here are a dynamic model for the true (but unobserved) values and a measurement error model. The time series of interest is denoted by

J.P. Buonaccorsi (✉)
University of Massachusetts, Amherst, MA, USA
e-mail: johnpb@math.umass.edu

B.C. Sutradhar (ed.), *ISS-2012 Proceedings Volume On Longitudinal Data Analysis Subject to Measurement Errors, Missing Values, and/or Outliers*, Lecture Notes in Statistics 211,
DOI 10.1007/978-1-4614-6871-4_3,

$\{Y_t, t = 1, \ldots, T\}$, where Y_t is random and t indexes time. The realized true value at time t is denoted by y_t. Measurement error occurs where instead of y_t we observe the outcome of W_t, where W_t is an estimator or general index of y_t. A particular point of emphasis in our coverage is to allow the behavior of the measurement error to depend on the underlying true value and/or sampling effort, as commonly occurs in practice.

The two main questions are 1. What happens if we ignore the measurement error? and 2. How can we correct for the measurement error? Of course, there are many possible objectives of a time series analysis and we need to limit our scope. The primary focus in this paper is on estimation of the parameters in the dynamic model, autoregressive models in particular. This is a logical first step as these parameters provide the building blocks for other objectives including forecasting or estimating probabilities about the process in the future. Also, in many of the ecological problems we discuss the estimation of the coefficients of the underlying dynamic process is the main thing of interest (and for this reason our discussion has a heavy ecological orientation to it).

Because of space limitations, there are a number of important related problems that we mostly ignore. These include repeated sample surveys where the main objective is updated estimation of the current true value, predicting and/or filtering in the presence of measurement error (e.g., Berliner (1991); Tripodis and Buonaccorsi (2009)), direct estimation of trends, model identification and problems where there are other variables in addition to the dynamic model. There are many examples of the latter; see, e.g., Ives et al. (2003), De Valpine and Hilborn (2005), and references therein for access to an extensive fisheries literature, Burr and Chowell (2006) in fitting SIR model to disease dynamics and Schmid et al. (1994).

Within the above stated focus, the objective is to provide a broad survey of modeling considerations, the effects of, and ways to correct for, measurement error and some of the challenges in carrying out estimation and inference. The intent is not to look at any particular model in great detail, although we do illustrate some key concepts with linear autoregressive models and the Ricker model and present a few new results. Dynamic models for the true values and measurement error models are discussed in Sects. 2.1 and 2.2, respectively. This is followed by a discussion about the performance of naive analyses that ignore measurement error in Sect. 3 and then a survey of correction methods in Sect. 4. Concluding remarks appear in Sect. 5.

2 Models

2.1 Dynamic Models for True Values

There is of course a very rich class of dynamic models that can be used for time series. As noted in the introduction the main interest here is in autoregressive

Table 1 Some dynamic models

Name	Model
AR(p)	$Y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$
Ricker	$Y_t = \phi_0 + y_{t-1} + \phi_1 e^{y_{t-1}} + \varepsilon_t$
Beaverton-Holt	$Y_t = y_{t-1} + log(\phi_0) - log(1 + \phi_1 n_{t-1}) + \varepsilon_t$
Theta-logistic	$Y_t = y_{t-1} + \phi_0 + (\phi_1 n_{t-1})^{\theta} + \varepsilon_t$
Logistic	$Y_t = y_{t-1} + \phi_0 + log(1 - e^{yt-1}) + \varepsilon_t$.

models with $E(Y_t|\mathbf{y}_{t-1}) = m(\mathbf{y}_{t-1}, \boldsymbol{\phi})$ and $V(Y_t|\mathbf{y}_{t-1}) = v(\mathbf{y}_{t-1}, \boldsymbol{\phi}, \boldsymbol{\sigma})$, where $\mathbf{y}_{t-1} = (\ldots, y_{t-2}, y_{t-1})$ indicates past values and $\boldsymbol{\sigma}$ contains additional variance parameters. Alternatively we can write $Y_t|\mathbf{y}_{t-1} = m(\mathbf{y}_{t-1}, \boldsymbol{\phi}) + \varepsilon_t$, where $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = v(\mathbf{y}_{t-1}, \boldsymbol{\phi}, \boldsymbol{\sigma})$.

We allow the mean function to be linear or nonlinear in Y with some examples given in Table 1. All of these models, with the last four being nonlinear in Y, arise in population ecology with $Y_t = log(N_t)$ where N_t is abundance, or density, at time t. The ε_t is typically taken to be normal with mean 0 and $V(\varepsilon_t) = \sigma^2$, referred to as process error. There are, of course, numerous other models that can be considered that we won't discuss in any detail. For example, as noted briefly later, measurement error has received some attention in an autoregressive integrated moving average (ARIMA) and basic structural models (BSMs) used to model labor variables, medical indices, and other variables over time. Another important model in the economics literature is to extend the linear autoregressive model in Table 1 to allow conditional heteroscedasticity in the process errors; the so-called ARCH model.

With $Y_t = log(N_t)$ where N_t is a count or scaled count, the usual assumption on ε_t may not be appropriate and an alternate nonlinear models arise by working explicitly if we work with N_t. These lead to another set of mean-variance models. (It is worth noting that one can only move directly from a model for $N_t|\mathbf{n}_{t-1}$ to one for $Y_t|\mathbf{y}_{t-1}$ in certain special cases, e.g., when δ_t is log-normal.) For example with the Ricker model, $N_t = n_{t-1} e^{(\phi_0 + \phi_1 n_{t-1})} \delta_t$. If we assume δ_t has mean 1 and constant variance σ_δ^2, then $E(N_t|n_{t-1}) = m(n_{t-1}, \boldsymbol{\phi}) = n_{t-1} e^{(\phi_0 + \phi_1 n_{t-1})}$ and $V(N_t|n_{t-1}) = m(n_{t-1}, \boldsymbol{\phi})^2 \sigma_\delta^2$. For the multiplicative version of the AR(1) model we might just work with $E(N_t|n_{t-1}) = e^{\phi_0} n_{t-1}^{\phi_1}$. In general, if $N_t|n_{t-1}$ is distributed Poisson, then $V(N_t|n_{t-1}) = E(N_t|n_{t-1})$. For a full discussion of fitting dynamic models with count data, possibly with the inclusion of time varying covariates, see Mallick and Sutradhar (2008).

Turning to the linear $AR(p)$ models we will assume the model is stationary (e.g., Box et al. 1994, Chap. 3), which for AR(1) model means $|\phi_1| < 1$. In general for stationary models we denote $E(Y_t) = \mu_Y$ and $V(Y_t) = \sigma_Y^2$, both constant in t, and $Cov(Y_t, Y_{t+k}) = \gamma_k$, a function only of the lag k. In the population ecology literature the AR models are referred to as the Gompertz model. There, based on the multiplicative versions for the AR(1) and AR(2), the primary objective is estimation of ϕ_1 and/or ϕ_2, usually interpreted as measures of density dependence and delayed density dependence, respectively; see Stenseth et al. (2003); Solow (2001), and

references therein. The AR(p) models can also be extended to multivariate autoregressive (MAR) models (also called vector autoregressive models or VAR); see Ives et al. (2003) and Aigner et al. (1984), both which accommodate measurement error.

All of the models in Table 1 have the random walk model $Y_t = y_{t-1} + \mu + \varepsilon_t$ as a special case. If the ε_t's are assumed to be independent and identically distributed (iid) with mean 0 and variance σ^2, this is a moving average of order 1, which is non-stationary. In the ecological context, once again with $Y = log(N)$, it represents a density independent model and is commonly used in population viability analysis and related trend analysis (Morris and Doak (2002)). There is a fairly large literature on measurement error in this problem. While there are some methodological connections to our discussion here we will not discuss it in detail both for space reasons and the fact that with the difference there is no dynamic piece left. See Buonaccorsi and Staudenmayer (2009) for a comprehensive treatment and references.

For likelihood-based approaches, the joint density of $Y_1, \ldots, Y_T$ is denoted $f(\mathbf{y}; \boldsymbol{\phi}, \boldsymbol{\sigma})$, with the term density applying for either the continuous or discrete case. For conditional likelihood approaches, assuming an autoregressive model (see Box et al. (1994, Chap. 7) for extension) suppose the distribution of $Y_t|\mathbf{y}_{t-1}$ depends on the past p values. Then partition $\mathbf{y}$ into $\mathbf{y}' = (\mathbf{y}_1^{*'}, \mathbf{y}_2^{*'})$, where $\mathbf{y}_1^* = (y_1, \ldots, y_p)'$ and $\mathbf{y}_2^* = (y_{p+1}, \ldots, y_T)'$. The conditional density of $\mathbf{Y}_2^*$ given $\mathbf{y}_1^*$ is

$$f_2(\mathbf{y}_2^*; \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathbf{y}_1^*) = \prod_{t=p+1}^{T} f(y_t|\mathbf{y}_{t-1}, \boldsymbol{\phi}, \boldsymbol{\sigma}), \tag{1}$$

where $f(y_t|\mathbf{y}_{t-1}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ is the density of $\mathbf{Y}_t$ given $\mathbf{y}_{t-1}$. Here, $\mathbf{y}_1^*$ is treated as fixed.

Notice that we can also write $f(\mathbf{y}; \boldsymbol{\phi}, \boldsymbol{\sigma}) = \int_{\mathbf{y}_1^*} f_2(\mathbf{y}_2^*; \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathbf{y}_1^*) f(\mathbf{y}_1^*) d\mathbf{y}_1^*$, where $f(\mathbf{y}_1^*)$ is the density of $\mathbf{Y}_1^*$, which can depend on some parameters (suppressed in the notation). If the Y's are discrete, then integration is replaced by summation above.

2.2 Measurement Error Models

The measurement error model describes the conditional behavior of the observable random variables $\mathbf{W} = (W_1, \ldots, W_T)'$ given $\mathbf{y} = (y_1, \ldots, y_T)'$. Given the huge number of sampling methods that can be used to estimate the y_t's, there are many measurement error models that can be entertained here. The bulk of the literature assumes that the measurement errors are conditionally uncorrelated and, frequently, additive. The assumption of conditionally uncorrelated measurement errors is reasonable when there is independent sampling at each time point. There are settings, however, where some common sampling units occur over time, leading to correlated measurement errors. This can occur, for example, in biological and climatological monitoring and is also a key feature in large national repeated sample surveys which use block resampling, where fairly general dynamic models have

been used for the measurement error itself; see, for example, Pfeffermann et al. (1998); Feder (2001), and references therein.

The conditional mean and covariance are $E(\mathbf{W}|\mathbf{y})$ and $Cov(\mathbf{W}|\mathbf{y})$, which usually depend on unknown parameters, contained in $\boldsymbol{\theta}$. A more convenient representation is

$$\mathbf{W} = \mathbf{y} + \mathbf{u}, \quad E(\mathbf{u}|\mathbf{y}) = \mathbf{B}_c, \quad \text{and} \quad Cov(\mathbf{u}|\mathbf{y}) = Cov(\mathbf{W}|\mathbf{y}) = \boldsymbol{\Sigma}_{uc}. \tag{2}$$

The $\mathbf{B}_c' = (\theta_{1c}, \ldots, \theta_{Tc})$ contains conditional biases, while $\boldsymbol{\Sigma}_{uc}$ is the conditional covariance matrix. The c here is a reminder that these are conditional on $\mathbf{y}$ and this conditional behavior could be a function of the underlying true values or sampling effort/design. This is elaborated on in more detail below when discussing the measurement error variances.

The frequently used additive model assumes $E(\mathbf{W}|\mathbf{y}) = \mathbf{y}$ or $E(\mathbf{u}|\mathbf{y}) = \mathbf{0}$, but models with bias have been considered including constant bias, $E(W_t|y_t) = y_t + \boldsymbol{\theta}$, and proportional bias, $E(W_t|y_t) = \theta y_t$. The constant bias model arises in various ways. For example, if $\hat{N}_t$ is the estimated abundance at time t, $W_t = log(\hat{N}_t)$, $Y_t = log(N_t)$, and $E(\hat{N}_t|n_t) = an_t$ with constant coefficient of variation, then $E(W_t|y_t) = y_t + \theta$, exactly or approximately (see Buonaccorsi et al. 2006). Often, a constant bias term can be easily absorbed. The proportional bias model arises from Poisson-type sampling where y_t is abundance, W_t is a count, adjusted for sampling effort, and θ is an unknown representing "catchability"; e.g., Stenseth et al. (2003). There are certainly even richer bias models that can be considered. For example, in calibrating aerial counts W versus ground counts Y of waterfowl, Lillegard et al. (2008) build a model where $W_t^{1/2} \sim N(\theta_0 + \theta_1 y_t^{1/2}, \tau^2)$, leading to $E(W_t|y_t) = \tau^2 + \theta_0^2 + 2\theta_0\theta_1 y_t^{1/2} + \theta_1^2 y_t^2$.

Much of our coverage is around settings assuming conditionally uncorrelated measurement errors with

$$Cov(\mathbf{W}|\mathbf{y}) = \boldsymbol{\Sigma}_{uc} = diag(\sigma_{u1c}^2, \ldots, \sigma_{uTc}^2), \tag{3}$$

a diagonal matrix with (t,t) element $V(W_t|y_t) = \sigma_{utc}^2$. Note that this allows for heteroscedastic measurement errors where the conditional variance, σ_{utc}^2, possibly depending on y_t, or on sampling effort, although the functional nature of that dependence need not be specified. The unconditional (over random Y) variance is denoted by σ_{ut}^2. Suppose, for example, that $\sigma_{utc}^2 = h(y_t, \boldsymbol{\theta})$ and there is additive error or constant bias. Then, unconditionally, $\sigma_{ut}^2 = V(u_t) = E[V(u_t|Y_t)] + V[E(u_t|Y_t)] = E[h(Y_t, \boldsymbol{\theta})]$. An important point here is that if the conditional variance only changes over t as a result of y_t and the process is stationary, then $h(Y_t, \boldsymbol{\theta})$ is stationary and unconditionally $\sigma_{ut}^2 = \sigma_u^2$. Hence, we can have conditional heteroscedasticity but unconditional homoscedasticity. As noted above, however, the conditional and unconditional variance may also change as a function of sampling effort.

For assessing the properties of naive estimators and corrected estimators under the additive model and (3) we assume that

$$Lim_{T\to\infty}\frac{\sum_{t=1}^{T}\sigma_{ut}^2}{T}=\sigma_u^2 \tag{4}$$

exists, where the limit is in probability if σ_{utc}^2 depends on Y_t.

Returning to the general setting, the unconditional moments of the observable $\mathbf{W}$ are given by $E(\mathbf{W})=E[E(\mathbf{W}|\mathbf{Y}))]$ and $Cov(\mathbf{W})=E[Cov(\mathbf{W}|\mathbf{Y}))]+Cov[E(\mathbf{W}|\mathbf{Y})]$. For additive measurement error this becomes $E(\mathbf{W})=E(\mathbf{Y})(=\mu\mathbf{1}$ under stationarity) and $Cov(\mathbf{W})=E[\boldsymbol{\Sigma}_{uc}]+Cov[\mathbf{Y}]=\boldsymbol{\Sigma}_u+\boldsymbol{\Sigma}_Y$.

For likelihood methods $\mathbf{W}|\mathbf{y}$ is assumed to have density $f(\mathbf{w}|\mathbf{y},\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ includes any measurement error parameters. When $\boldsymbol{\theta}$ is a parameter vector of fixed length (e.g., σ_u^2, or $(\theta_0,\theta_1,\tau^2)$ in the model of Lillegard et al. (2008)) there is no difficulty in interpreting and using this density in standard fashion. The handling of $\boldsymbol{\theta}$ is more delicate if we allow the conditional measurement error parameters to change over time in some unspecified manner. For example, if we assume $W_t|y_t\sim N(y_t,\sigma_{utc}^2)$ where no structure is given to σ_{utc}^2, then $\boldsymbol{\theta}=(\sigma_{u1c}^2,\ldots,\sigma_{uTc}^2)'$, which increases in size with T. Further, if σ_{utc}^2 involves y_t, then unconditionally $\boldsymbol{\theta}$ is random. There are two densities for $\mathbf{W}$ that will be used later,

$$f_W(\mathbf{w};\phi,\sigma,\boldsymbol{\theta})=\int_y f(\mathbf{w}|\mathbf{y},\boldsymbol{\theta})f(\mathbf{y};\boldsymbol{\phi},\boldsymbol{\sigma})d\mathbf{y} \tag{5}$$

and

$$f_W^*(\mathbf{w};\phi,\sigma,\boldsymbol{\theta},\mathbf{y}_1^*)=\int_{\mathbf{y}_2^*} f(\mathbf{w}|\mathbf{y},\boldsymbol{\theta})f_2(\mathbf{y}_2^*;\boldsymbol{\phi},\boldsymbol{\sigma},\mathbf{y}_1^*)d\mathbf{y}_2^*, \tag{6}$$

where the second conditions on $\mathbf{y}_1^*$ with $f_2(\mathbf{y}_2^*;\boldsymbol{\phi},\boldsymbol{\sigma},\mathbf{y}_1^*)$ as given in (1). As before, integrals are replaced by sums for the discrete case.

3 Performance of Naive Estimators

An important question, especially in the absence of any specific information about the measurement error, is the performance of the so-called naive analyses, which ignore the measurement error and treat W_t as if it is Y_t. Based on what we know from regression models it is not surprising that the measurement error will lead to biases in estimated coefficients as well as in the process variance parameters. However, we will see that obtaining analytical expressions for asymptotic or approximate biases is generally difficult here.

Of course the first question to ask here is "which naive analysis?", since there are a plethora of approaches to estimation in time series. For stationary normal models, maximum likelihood (ML) and (more recently) restricted maximum likelihood (REML) are the most popular. Conditional maximum likelihood (CML), which maximizes $L(\boldsymbol{\phi},\boldsymbol{\sigma},\mathbf{y}_1^*|\mathbf{y})$ arising from (1), is another choice. This drops making an assumption about the marginal distribution of $\mathbf{Y}_1^*$ and usually leads to easier

Table 2 Selected papers assessing the performance of naive estimators via simulation

Author(s)	Dynamic Model	ME Model	Method
Solow (1998)	Logistic	Poisson	LS
De Valpine and Hastings (2002)	Ricker and	$N(0,\sigma_u^2)$	ML
De Valpine (2002)	Beaverton-Holt	〃 〃	〃 〃
SB (2005)	AR(1)	$N(0,\sigma_{utc}^2)$	ML
Lele (2006)	AR(1)	Poisson	ML
Hovestadt and Nowicki (2008)	Modified Ricker	$N(0,\sigma_u^2)$	LS
Barker and Sibly (2008)	theta-logistic	$N(0,\sigma_u^2)$	LS
Ives et al. (2010)	ARMA	$N(0,\sigma_u^2)$	REML
Resendes (2011)	Ricker	$N(0,\sigma_u^2)$	LS

computing. For a normal autoregressive model with the ε_t assumed iid $N(0,\sigma^2)$ this leads to least squares, where $\hat{\boldsymbol{\phi}}_{CLS}$ minimizes $\sum_{t=p+1}^{T}(y_t - m(\mathbf{y}_{t-1},\boldsymbol{\phi}))^2$, and p is the number of previous y terms in the m function. Obviously if $m(\mathbf{y}_{t-1},\boldsymbol{\phi})$ is linear in the ϕ's, then this leads to simple least squares. More generally these lead to nonlinear least squares with estimating equations of the form

$$\sum_{t=p+1}^{t}(y_t - m(\mathbf{y}_{t-1},\boldsymbol{\phi}))\Delta(\mathbf{y}_{t-1},\boldsymbol{\phi}) = \mathbf{0}, \tag{7}$$

where the jth element of the $\Delta(\mathbf{y}_{t-1},\boldsymbol{\phi})$ is $\partial m(\mathbf{y}_{t_1},\boldsymbol{\phi})/\partial\phi_j$. Similar estimating equations can arise from conditional maximum likelihood in other situations. For example, if $Y_t|y_{t-1}$ is assumed to be Poisson with mean $e^{\phi_0+\phi_1 log(y_{t-1})}$, then CML leads to (7) with $\Delta(\mathbf{y}_{t-1},\boldsymbol{\phi})' = [1, log(y_{t-1})]$.

A number of general strategies have been tried to investigate the performance of naive estimators in the presences of measurement error. These include:

1. Use an explicit expression for the estimators, e.g., $\hat{\boldsymbol{\phi}} = g(\mathbf{W})$), and determine the limiting or approximate properties analytically. This is mainly used for the linear autoregressive models, discussed in the next section.
2. View the naive estimators as solutions to estimating equations $S(\mathbf{W},\boldsymbol{\phi},\boldsymbol{\sigma}) = \mathbf{0}$. In this case the naive estimators (under some conditions) will converge to $\boldsymbol{\phi}^*$ and $\boldsymbol{\sigma}^*$ which satisfy $E_{\boldsymbol{\phi},\boldsymbol{\sigma}}[S(\mathbf{W},\boldsymbol{\phi}^*,\boldsymbol{\sigma}^*)]/T \to \mathbf{0}$.
3. Find an "induced" model, or an approximation, for the behavior of $W_t|\mathbf{w}_{t-1}$. If this is in the same class as the original with $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$ replaced by ϕ^* and $\boldsymbol{\sigma}^*$, then the naive methods are consistent (or approximately consistent) for ϕ^* and $\boldsymbol{\sigma}^*$.
4. Simulate the performance of the naive estimators. While it is difficult to gain a good understanding of the nature of the biases from simulations alone, this is often the only option. Even when analytical asymptotic properties are available, as in linear AR models, simulations are needed to assess "small" sample behavior; see the next section. Table 2 summarizes some of the papers that have utilized simulation. SB (2005) refers to Staudenmayer and Buonaccorsi (2005), an abbreviation used throughout.

Table 3 Simulation means for naive least squares estimators from the Ricker model with $\varepsilon_t \sim N(0, \sigma^2)$ with $K = 100$ and $\sigma = 0.2$. From Resendes (2011), with permission

	True ϕ_0					
	0.2	0.75	1.5	2.4	2.6	σ_u^2
$\hat{\phi}_0$	0.413	0.822	1.444	2.381	2.575	0.05
	0.593	0.939	1.266	2.206	2.427	0.2
	0.729	0.975	1.182	1.988	2.231	0.3
	0.813	0.983	1.053	1.628	1.855	0.5
$\hat{K}$	97.537	100.093	100.007	100.188	100.035	0.05
	94.771	102.044	102.324	101.516	101.668	0.2
	96.631	104.456	104.028	102.375	104.261	0.3
	100.686	113.058	113.294	111.585	112.367	0.5
$\hat{\sigma}$	0.202	0.206	0.205	0.226	0.231	0.05
	0.301	0.291	0.294	0.456	0.536	0.2
	0.387	0.371	0.379	0.617	0.720	0.3
	0.586	0.566	0.573	0.906	1.069	0.5

While methods 2 and 3 have proved fruitful in many regression problems (see Carroll et al. (2006) and Buonaccorsi (2010)) they are less useful in dynamic situations. A simple example that illustrates the shortcomings of method 3 is the linear AR(p) model with additive normal measurement error with constant variance. This leads to an induced model which is an Autoregressive Moving Average (ARMA) (p,p) (see Sect. 4.2.1), so the problem is one of model misspecification. Method 2, an approach first used in regression contexts by Stefanski (1985), is the only recourse for assessing bias analytically when the estimators do not have a closed form. Expanding the estimating equations leads to $\boldsymbol{\phi}^* \approx \boldsymbol{\phi} - (Lim_{T\to\infty} \sum_t E[\dot{S}(W_t, \boldsymbol{\phi})]/T)^{-1} Lim_{t\to T} \sum_t E(S(W_t, \boldsymbol{\phi}))/T$, assuming the limits exist and where $\dot{S}$ denotes partial derivatives with respect to the parameters. The problem here is finding the expected values and limits. The best potential for this method is with using conditional ML/LS approaches leading to estimating equations as given in (7), but even there, the analysis is not straightforward. To illustrate we consider the Ricker model, which was investigated by Resendes (2011). Although nonlinear in Y, it is linear in the parameters, leading to simple linear least squares. Defining, $D_t = W_{t+1} - W_t$ and $C_t = e^{W_t}$ $S_{DC} = \sum_{t=1}^{T-1}(D_t - \bar{D})(C_t - \bar{C})/T$ and $S_{CC} = \sum_{t=1}^{T-1}(C_t - \bar{C})^2/T$, then the naive estimators of ϕ_1 is $\hat{\phi}_{1,naive} = S_{DC}/S_{CC}$ and asymptotically $\hat{\phi}_{1,naive} \Rightarrow LimE(S_{DC})/LimE(S_{CC})$, provided the two limits exists. We faced two problems here; the first in finding the expected values involving nonlinear functions, the second in determining whether the sums converge, and to what. If the process is assumed stationary, then these limits generally exist but one needs to determine the stationary moments for the two series involved. Resendes (2011) examined the biases of the naive estimators in the Ricker model extensively, both via simulation and via the estimating equation just described (on both additive and multiplicative scales). While clean final expressions for the approximate biases proved elusive, his approximations did show how the direction of the bias can change with the values of the parameters. This is also seen in his simulations, a very small portion of which appear in Table 3 and highlight his main

conclusions. These are based on parameter values similar to those in De Valpine and Hastings (2002) but with some larger measurement error variances. The magnitude and direction of the biases depend heavily on ϕ_0 with naive estimators of ϕ_0 tending to be overestimates at small ϕ_0 and underestimates as ϕ_0 increases. The same thing happens for the naive estimator of the carrying capacity $K = -\phi_0/\phi_1$ (the Ricker model can be parameterized instead in terms of ϕ_0 and K), but it is less sensitive to the measurement error than naive estimators of ϕ_0 are. It also shows overestimation of σ, although modest at small measurement error variances.

3.1 Linear Autoregressive Models

Here we summarize and illustrate some results for the stationary linear AR models under the additive error model with (2). This is one of the simpler settings, allowing exactly analytical bias expressions, and is useful for illustrating some key points. We concentrate on the coefficients, but it also can be shown that σ^2 is overestimated. The standard method of analysis here is traditionally maximum likelihood, but there is increasing support for the use of REML given that it reduces small sample bias (e.g., Cheang and Reinsel (2000)), REML estimators for the AR(1) model can be easily obtained using most mixed models software. SB (2005) extended earlier work and developed a general expression for the limiting values of the Yule-Walker (YW) estimates allowing changing measurement error variances and under the assumption in (4). For the AR(1) model this leads to $\hat{\phi}_{1,naive} \Rightarrow \kappa\phi_1$, where $\Rightarrow$ denotes convergence in probability and

$$\kappa = \sigma_Y^2/(\sigma_Y^2 + \sigma_u^2). \tag{8}$$

This shows asymptotic attenuation (bias towards 0) in the estimator of ϕ_1. If, as in ecological applications, the focus is on $\beta_1 = \phi_1 - 1$, then $\hat{\beta}_{1,naive} \Rightarrow \kappa\beta_1 + (\kappa - 1)$, which is greater than β_1 if $\beta_1 < -1$ but less than β_1 if $\beta_1 > -1$.

For the AR(2), we have

$$\begin{pmatrix} \hat{\phi}_{1,naive} \\ \hat{\phi}_{2,naive} \end{pmatrix} \Rightarrow \begin{pmatrix} (\kappa\rho_1 - \kappa^2\rho_1\rho_2)/(1 - \kappa^2\rho_1^2) \\ (\kappa\rho_2 - \kappa^2\rho_1^2)/(1 - \kappa^2\rho_1^2) \end{pmatrix},$$

where $\rho_j = \gamma_j/\sigma_X^2$ for $j = 1$ and 2 with $\gamma_1 = (\phi_1\sigma_X^2)/(1 - \phi_2)$ and $\gamma_2 = (\phi_1^2 + \phi_2 - \phi_2^2)\sigma_X^2/(1 - \phi_2)$. This leads to some more interesting results, as the bias in either element of $\hat{\boldsymbol{\phi}}_{naive}$ can be either attenuating or accentuating (larger in absolute value), depending on both the amount of measurement error and the other parameters.

It is well known that YW and ML/REML estimators have the same asymptotic properties without measurement error. With measurement error the asymptotic properties of the ML (equivalently REML) can be examined through their associated estimating equations, see the Appendix. This leads to the same asymptotic behavior

Table 4 Performance of maximum likelihood (ML), Yule-Walker (YW), and restricted maximum likelihood (REML) estimators of ϕ_1, using true and mismeasured values in the AR(1) model. $\phi_1^* = \kappa\phi_1$ = limiting value of naive estimator

			True (using Y)				Naive (using W)		
ϕ_1	σ_u^2	n	YW	ML	REML	ϕ_1^*	YW	ML	REML
0.2	0.15	10	0.023	0.029	0.176	0.169	0.006	0.009	0.152
0.2	0.15	30	0.140	0.146	0.187	0.169	0.116	0.120	0.161
0.2	0.15	50	0.161	0.164	0.188	0.169	0.136	0.139	0.163
0.2	0.15	100	0.179	0.180	0.192	0.169	0.149	0.150	0.162
0.2	0.6	10	0.031	0.040	0.185	0.116	−0.022	−0.020	0.120
0.2	0.6	30	0.151	0.157	0.199	0.116	0.074	0.077	0.116
0.2	0.6	50	0.171	0.174	0.198	0.116	0.090	0.092	0.115
0.2	0.6	100	0.185	0.187	0.199	0.116	0.101	0.102	0.113
0.8	0.15	10	0.356	0.435	0.623	0.749	0.290	0.354	0.536
0.8	0.15	30	0.648	0.686	0.750	0.749	0.586	0.619	0.681
0.8	0.15	50	0.710	0.731	0.768	0.749	0.650	0.669	0.705
0.8	0.15	100	0.756	0.765	0.783	0.749	0.701	0.709	0.727
0.8	0.6	10	0.357	0.444	0.641	0.630	0.195	0.235	0.410
0.8	0.6	30	0.652	0.689	0.753	0.630	0.463	0.487	0.542
0.8	0.6	50	0.709	0.729	0.766	0.630	0.526	0.540	0.572
0.8	0.6	100	0.756	0.766	0.785	0.630	0.574	0.582	0.598

as the YW estimators when $\boldsymbol{\Sigma}_u = \sigma_u^2\mathbf{I}$. Note that the $\boldsymbol{\Sigma}_u$ is the unconditional covariance of $\mathbf{u}$ and can have conditional heteroscedasticity arising through Y_t; see the discussion in Sect. 2.2.

One question is how useful the bias expressions are for "small" samples. Even without measurement error, the issue of bias in small samples is an important one with time series. To illustrate data was generated from the AR(1) model with $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$ and $W_t = y_t + u_t$, where the ε_t are iid $N(0, \sigma^2)$ and the ε_t are iid $N(0, \sigma_u^2)$. The process variance was held to $\sigma^2 = .8$, while $\phi_1 = .2, .5$, or .8 and $\sigma_u^2 = .15, .4$ and .8. The case with $\phi_1 = .5$ and $\sigma_u^2 = .15$ is based roughly on an analysis of mouse dynamics given in Buonaccorsi (2010, Chap. 12). For each combination 1,000 simulations were run and YW, ML, and REML estimators obtained, using the true Y's and the error prone W's. Partial results appear in Table 4. The analysis with true values is given for two reasons. First it shows the clear superiority of REML to ML estimation, especially at small sample sizes, with the ML only being modestly better than the YW estimator. Notice that even using true values all of the estimators are attenuated towards zero, sometimes dramatically with $n = 10$. Second it gives a baseline to compare the performance of the naive estimators to. The measurement error leads to further attenuation, increasing in σ_u^2, as it should. The REML estimator obviously provides some extra protection against measurement error compared to ML and YW, especially at small sample sizes. The variable ϕ_1^* is the limiting value of the naive estimator (whether YW, ML, or REML). The asymptotic bias associated with this limiting value can be significantly off compared to the simulated bias, sometimes even with samples of size 50.

4 Correcting for Measurement Error

There are three general contexts within which measurement error corrections are carried out. 1. Using only the observed $W_1, \ldots, W_T$; 2. Using the observed W's and estimated measurement error parameters leading to pseudo-methods. 3. Using richer data than just W_t' from each time point t. We first comment on each of these three.

1. Unlike many other measurement error problems it has been shown that with some assumptions on the measurement error (e.g., that the errors are independent with mean 0 and common variance) all of the parameters can be identified (and estimated) from the $\mathbf{W}$ data alone for a variety of dynamic models; see, e.g., some of the references in Sect. 4.2.1 and Aigner et al. (1984). The majority of the work on correcting for measurement error in dynamic settings has attacked the problem from this perspective. In the likelihood context these are state-space models and the approach is standard in principle but can face computational challenges as discussed later. The shortcoming of this approach is the potential restrictive nature of the measurement error model and the fact that identifiability does not guarantee good estimators. However, without any information about the measurement error process the only option is to use this approach to estimate the dynamic and measurement error parameters simultaneously.
2. There is often data that allows for estimating the measurement error parameters, contained in $\hat{\boldsymbol{\theta}}$, say. This may include estimated variances as well as biases and/or correlations if they are part of the model. With additive uncorrelated measurement errors, allowing changing variance over time then $\hat{\boldsymbol{\theta}}$ contains an estimate of measurement error variances, $\hat{\sigma}^2_{ut}$, at each time point t, typically arising from the same data that produces W_t. The pseudo-methods set $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and then estimate the parameters in the dynamic model. While it seems natural to exploit estimates about the measurement error parameters, this strategy has been seriously underutilized. Note that $\hat{\boldsymbol{\theta}}$ may be of fixed size or contain separate estimates at each time point, depending on the assumed structure of $\boldsymbol{\theta}$; see the discussion in Sect. 2.2. We subsume under pseudo-methods approaches that simply assume that the measurement error parameters are known. The difference between viewing $\hat{\boldsymbol{\theta}}$ as estimated or known and fixed will come in trying to account for uncertainty $\hat{\boldsymbol{\theta}}$; see Sects. 4.1 and 4.2.2.
3. The third approach has some richer "data," denoted $\mathbf{Q}_t$, at time t with a model for $\mathbf{Q}_t$ given y_t. The simplest example is where $\mathbf{Q}_t$ contains replicate measures of y_t; e.g, Wong et al. (2001); Dennis et al. (2010), and Knape et al. (2011). The analyses here connect to the previous two approaches. If $\mathbf{Q}_t|\mathbf{y}_t$ depends on a finite collection of parameters $\boldsymbol{\theta}$, then this is like approach 1 but with W_t replaced by $\mathbf{Q}_t$. Or, the richer data can be used to first estimate the measurement error parameters, in which case this reduces to a pseudo-method. In some cases those two strategies concur. With these connections we won't mention this method in detail in the later sections.We do note that one advantage of using the full $\mathbf{Q}_t$ is that the uncertainty from estimating the measurement error parameters is accounted for. It does, however, require that $\boldsymbol{\theta}$ be fixed and that a distribution is

Table 5 Selected papers correcting for measurement error

Methods using W only			
Author(s)	Model for Y	ME Model	Method
De Valpine and Hilborn (2005)	General,AR(1),	$N(0,\sigma_u^2)$	ML
De Valpine and Hastings (2002)	Ricker and		
De Valpine (2002)	Beaverton-Holt		
Calder et al. (2003)	Gen	Gen	Bayesian
Clark and Bjornstad (2004)	Gen	Gen	Bayesian
SB (2005)	AR(1)	$N(0,\sigma_u^2)$	ML/ARMA
	AR(p)	$(0,\sigma_{ut}^2)$	modified YW
Wang et al. (2006)	RW, AR(1), AR(2)	$N(0,\sigma_u^2)$	ML
Dennis et al. (2006)	RW, AR(1)	$N(0,\sigma_u^2)$	ML/REML
Lele (2006)	Gen	Gen	ComML
Knape (2008)	AR(1)	$N(0,\sigma_u^2)$	ML
Ponciano et al. (2009)	Gen	Gen	ML
Knape et al. (2011)	AR(1)	varied	ML
Pseudo-methods			
Solow (1998)	"logistic"	Poisson	SIMEX
Williams et al. (2003)	AR(1)+ trend	$N(0,\sigma_{ut}^2)$	ML
Ives et al. (2003)	MAR(1)	$N(0,\sigma_u^2)$	ML
Clark and Bjornstad (2004)	Gen	$(0,\sigma_{ut}^2)$	Bayesian
SB (2005)	AR(p)	$(0,\sigma_{ut}^2)$	CEE/ML
Wang (2007)	Theta-logistic	$N(0,\sigma_u^2)$	ML/Bayesian
Lillegard et al. (2008)	MAR(1)	(* below)	Bayesian
Ives et al. (2010)	ARMA	$N(0,\sigma_u^2)$	ML/REML
Dennis et al. (2010)	AR(1)	$N(0,\sigma_u^2)$(reps)	REML
Knape et al. (2011)	AR(1)	varied	ML
Resendes (2011)	Ricker	varied	SIMEX, MEE

$* \sqrt{(W_t)} \sim N(a + b\sqrt{exp(Y_t)}, \sigma_u^2)$

specified for the within time data. In many applications the within time sampling can be very complex and the results will be reduced to W_t and $\hat{\sigma}_{ut}^2$, leading us back to the pseudo-methods.

Table 5 contains a partial listing of papers addressing connection techniques, many containing simulations evaluating the methods. A number of these will be referred to in the later discussion, along with additional papers.

Within each of the three contexts described above, moment-based, maximum likelihood, and Bayesian methods are all options. The pseudo-methods also open the door for other methods including Simex, modifying the estimating equations and what is known as "regression calibration" all of which have received limited, or no, attention, in dynamic settings. Our survey below is categorized by the correction technique (moment, likelihood, Simex, etc.) with the last subsection addressing the use of bootstrapping.

4.1 Moment Methods

Moment-based corrections are for the most part limited to linear problems including the random walk model and linear autoregressive models, which we concentrate on here. Working only with the W_t's, if the measurement error is additive and conditionally uncorrelated but with possibly changing variances, modified Yule-Walker estimators for $\boldsymbol{\phi}$ are available. These take advantage of the fact that the lag covariances, which involve the ϕ's, are estimated consistently; see Walker (1960); Sakai et al. (1979); Chanda (1996), and comments in SB (2005). While these estimators are consistent and robust to changing measurement error variances, in practice they are often very ill behaved even for moderate sample sizes and cannot be recommended.

For a pseudo approach with uncorrelated additive measurement errors, but possibly unequal variances, define

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_0 \end{pmatrix} \text{ and } \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \dots \\ \gamma_p \end{bmatrix},$$

where γ_k is the lag k covariance and $\gamma_0 = \sigma_Y^2$. SB (2005) proposed the simple estimator $\hat{\boldsymbol{\phi}}_{CEE} = (\hat{\boldsymbol{\Gamma}}_W - \hat{\sigma}_u^2 \mathbf{I})^{-1} \hat{\boldsymbol{\gamma}}_W$, where $\hat{\boldsymbol{\Gamma}}_W$ and $\hat{\boldsymbol{\gamma}}_W$ are naive estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{\gamma}$ using the sample variances and covariances of the observed W_t's and $\hat{\sigma}_u^2 = \sum_{t=1}^{T} \hat{\sigma}_{ut}^2 / T$. The estimator can be viewed as arising from either correcting the naive estimating equations so they have mean 0 (hence CEE for corrected estimating equation) or from a simple correction based on the fact that the sample variance of the observed W_t's estimates $\sigma_Y^2 + \sigma_u^2$. They show that $\hat{\boldsymbol{\phi}}_{CEE}$ is consistent as long as (4) holds and in addition $\hat{\sigma}_u^2$ converges in probability to σ_u^2. They also establish asymptotic normality and provide the asymptotic covariance of $\hat{\boldsymbol{\phi}}_{CEE}$ but only under certain assumptions. An extension of their result is the following:

Proposition 1. *Let* $G_1 = \partial \hat{\boldsymbol{\phi}}_{CEE} / \partial \hat{\boldsymbol{\gamma}}_W |_*$, $G_2 = \partial \hat{\boldsymbol{\phi}}_{CEE} / \partial \hat{\sigma}_u^2 |_*$, *(with* $|_*$ *denoting evaluation at* $\hat{\gamma}_{Wj} = \gamma_{Wj}$ *and* $\hat{\sigma}_u^2 = \sigma_u^2$*),* $\mathbf{X}_t' = [W_t^2, W_t W_{t+1}, \dots, W_t W_{t+p}]$, *and* $\bar{\mathbf{X}} = \sum_t \mathbf{X}_t / T$. *Assuming the following limits exist,* $\mathbf{Q}_1 = LimCov(\bar{\mathbf{X}})T$, $Q_{12} = LimCov(\bar{\mathbf{X}}, \hat{\sigma}_u^2)T$ *and* $Q_{22} = LimV(\hat{\sigma}_u^2)T$, *then the asymptotic/approximate covariance of* $\hat{\boldsymbol{\phi}}_{CEE}$ *is*

$$ACov(\hat{\boldsymbol{\phi}}_{CEE}) = (G_1 Q_1 G_1 + + G_1 Q_{12} G_2' + G_1' Q_{12}' G_2)/T + G_2 G_2' V(\hat{\sigma}_u^2). \quad (9)$$

The first term $G_1 Q_1 G_1 / T = C_K$, say, is the approximate covariance of $\hat{\boldsymbol{\phi}}_{CEE}$ if the $\hat{\sigma}_{ut}^2$'s are treated as known. The terms involved in (9) get complicated, as does estimation of them. We omit details (given in Buonaccorsi and Staudenmayer (2012)) but the general result is useful for highlighting the difficulty in accounting for the uncertainty in the measurement error parameters by needing to handle the terms involving $\hat{\sigma}_{ut}^2$. The problem simplifies considerably when the $\hat{\sigma}_{ut}^2$'s are

assumed independent of the W_t's, as would hold if the W_t and $\hat{\sigma}^2_{ut}$ arise from using normal replicate measures at time t. In this case $ACov(\hat{\boldsymbol{\phi}}_{CEE}) = C_K + G_2 G_2' V(\hat{\sigma}^2_u)$ (where C_k was defined above) so only $\eta_4 = \sum_t E(u_t^4)/T (= 3\sigma^4_{ut}$ if u_t is normal) and $V(\hat{\sigma}^2_u)$ and estimates of them are needed. To illustrate (see SB (2005)), for the AR(1) model, the approximate variance of $\hat{\phi}_{1,CEE}$ is

$$\frac{1}{T}\left[\frac{(1-\phi_1^2)(2-\kappa)}{\kappa} + \frac{(1-\phi_1^2)(\sum_t \sigma^4_{ut}/T) + \phi_1^2\eta_4}{\sigma^2_Y}\right] + \frac{\phi_1^2 V(\hat{\sigma}^2_u)}{\sigma^2_Y},$$

where κ is given in (8).

More broadly, the challenge is to accommodate the case where the conditional variance (or higher moments) of $u_t|y_t$ may depend on y_t. Unconditionally this leads to the $\hat{\sigma}^2_{ut}$'s being correlated with the W_t's and expressions and estimates for Q_1 and Q_{12} are needed. It appears impossible to do this robustly using just the W_t and $\hat{\sigma}^2_{ut}$ in the most general setting, without some assumptions on the measurement error variances. Buonaccorsi and Staudenmayer (2012) develop a strategy for estimating the covariance matrix under the assumptions that either i) the sampling effort is constant so the heteroscedasticity arises through the Y_t only or ii) the measurement error variance is inversely proportional to known sampling effort. The methodology used considers the joint series $(W_t, \hat{\sigma}^2_{ut})$ and exploits time series methods for estimating covariance structures robustly using consistent estimators of the spectral density of a multivariate stationary process (e.g., the modified Bartlett kernel estimator); see, e.g., Fuller (1996, Chap. 7).

4.2 Likelihood Methods

Likelihood and related Bayesian methods have dominated the correction approaches. There are two-likelihoods of interest, the full and conditional likelihoods $L(\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\theta}|\mathbf{w}) = f(\mathbf{w}|\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\theta})$ and $L^*(\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \mathbf{y}_1^*|\mathbf{w}) = f^*(\mathbf{w}|\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \mathbf{y}_1^*)$, based on the densities in (5) and (6), respectively. For stationary normal linear models, an alternative is to use the REML likelihood in place of $L(\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\theta}|\mathbf{w})$.

4.2.1 Using W Values Only

Assuming $\boldsymbol{\theta}$ is of fixed size, these methods maximize either the full or conditional likelihood, or their REML versions. Of course, they are only used when all of the parameters are identifiable. This is a classical state-space formulation and there is a fairly large literature on fitting these models, whether linear or nonlinear using maximum likelihood, and in a few cases, REML. While this is straightforward in principle, there are a number of challenges in using these techniques, including needing to dealing with local maxima and/or the maximum occurring on the boundary of the parameter space and, general difficulty in computing the likelihood

function and an associated covariance matrix of the estimates for use in inference. However, as discussed below, there have been recent advances tackling some of the computational challenges.

Normal linear stationary models: Assuming $\mathbf{u}|\mathbf{y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ where $\boldsymbol{\Sigma}_u$ depends on a fixed number of parameters, the full likelihood approach can be used based on $\mathbf{W} \sim N(\mu\mathbf{1}, \boldsymbol{\Sigma}_Y + \boldsymbol{\Sigma}_u)$. (It is important to note that if the measurement error variances (or covariances if present) are changing with $\mathbf{y}$, then even if $\mathbf{u}$ is conditionally normal and unconditionally $Cov(\mathbf{u}) = \boldsymbol{\Sigma}_u$, $\mathbf{W} = \mathbf{Y} + \mathbf{u}$ is not normal since it involves mixtures of normals with changing variances. This same comment applies for the pseudo likelihood methods in the next section.) Often $\boldsymbol{\Sigma}_u$ is taken to be $\sigma_u^2\mathbf{I}$, while $\boldsymbol{\Sigma}_y$ is a function of the parameters, depending on the specific model for the true values. Computational methods here typically use the Kalman filter or some variation on it; see, e.g., Harvey (1990), Brockwell and Davis (2002), Ives et al. (2010, 2003) and, for the AR(1) models, Dennis et al. (2006) and Knape (2008). These last two papers touch on the important problem that even in the simple AR(1) model there may be issues with local maxima and/or the maximum occurring on the boundary. Knape (2008) also zeros in more on the fact that, not surprisingly, it can be difficult to separate the process error variance and the measurement error variance. These models can also be cast as mixed models which may employ other computational techniques (e.g., the EM algorithm and modifications of it), although only the AR(1) model is typically available in canned mixed model routines (e.g, proc Mixed in SAS and lmm in R).

For normal autoregressive (and more generally ARMA models) there is another quick and easy option. If the model for true values is ARMA(p,p) and the measurement error model is an $MA(q)$ process ($q = 0$ corresponding to u_t's being iid $N(0, \sigma_u^2)$), then it is well known that the model for W is an ARMA(p,p + q) and the autoregressive parameters are unchanged (Box et al. 1994, Sect. A4.3, Ives et al. 2010). This means we could estimate the autoregressive coefficients in $\boldsymbol{\phi}$ by simply fitting an ARMA(p,q) model. This is the hybrid approach of Wong and Miller (1990) used in an ARIMA setting. For the $AR(p)$ models with measurement errors being iid $N(0, \sigma_u^2)$, this means we can estimate the autoregressive coefficients in simply fitting an ARMA(p,p) model. In the case of $p = 1$ this yields the ML estimator of $\boldsymbol{\phi}_1$ and using results on fitting ARMA models (see Brockwell and Davis (2002)) an exact, but somewhat complicated, expression for the asymptotic variance of $\hat{\phi}_{1,ML}$ can be obtained; see SB (2005). When $p > 1$ this does not yield the ML estimate for $\boldsymbol{\phi}$, since there are restrictions involving the moving average parameters. There is an older literature that attacked this problem by first fitting the ARMA model and then bringing in the restrictions in various ways to get approximate MLEs (e.g., Lee and Shin (1997); Pagano (1974)).

General Models. The computational challenges are more severe for nonlinear models both in the integration required to obtain the likelihood, and in getting the information matrix or other quantities for use in inference. Note that the full likelihood function $L(\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\theta}|\mathbf{w})$ also requires specifying a marginal distribution for Y_1, while in working with the conditional likelihood $\mathbf{y}_1^*$ is treated as

another parameter. This would seem to argue more for the use of the conditional approach. For overviews and references, see De Valpine and Hastings (2002), De Valpine (2002), De Valpine and Hilborn (2005), and Wang (2007) (who assumes the measurement error parameters are known but the basic algorithms are the same). Recently, some new methods have been developed to try and overcome some of the computational challenges. These include the Monte Carlo Kernel Likelihood (MCKL) (De Valpine (2004)), a method called data cloning, which borrows from Bayesian computing (Lele et al. (2007); Ponciano et al. (2009)) and composite maximum likelihood estimation (ComML) (Lele 2006).

4.2.2 Pseudo Likelihood Methods

The pseudo ML estimates maximize either $L(\boldsymbol{\phi},\boldsymbol{\sigma},\hat{\boldsymbol{\theta}}|\mathbf{w})$ or $L_2(\boldsymbol{\phi},\boldsymbol{\sigma},\hat{\boldsymbol{\theta}},\mathbf{y}_1^*|\mathbf{w})$, or an REML modified version, where $\hat{\boldsymbol{\theta}}$ contains estimated, or assumed known, measurement error parameters. These methods also have a long history of use in modeling with repeated samples surveys where the true values are ARIMA (and special cases thereof) or follow a basic structural model (BSM) and the sampling error also may follow a dynamic model (AR, MA, etc.), which is first estimated and then held fixed. See Bell and Wilcox (1993); Koons and Foutz (1990); Wong and Miller (1990); Miazaki and Dorea (1993); Lee and Shin (1997); Feder (2001), and references therein. Many pseudo-approaches treat stationary normal models assuming $\mathbf{W} \sim N(\mu\mathbf{1}, \boldsymbol{\Sigma}_Y + \hat{\boldsymbol{\Sigma}}_u)$, but see the caution in the previous section about non-normality of $\mathbf{W}$ if the measurement error covariance change with Y_t.

For the most part, the pseudo-likelihood approaches face the same computational demands as the non-pseudo-likelihood methods. An added challenge lies in finding the covariance matrix for the estimated dynamic parameters which accounts for uncertainty in $\hat{\boldsymbol{\theta}}$. To illustrate, let $\omega' = (\boldsymbol{\phi}', \boldsymbol{\sigma}')$ for the collection of dynamic parameters and $I(\boldsymbol{\omega})$ the corresponding information matrix, with submatrices I_ϕ, etc. If $\hat{\boldsymbol{\theta}}$ is treated as fixed, the asymptotic covariance matrix of $\hat{\boldsymbol{\omega}}_{ML}$ is $I(\boldsymbol{\omega})^{-1}$, leading to an asymptotic covariance matrix of $\hat{\boldsymbol{\phi}}_{ML}$ of $Acov(\hat{\boldsymbol{\phi}}_{ML,K}) = (I_\phi - I_{\phi,\sigma} I_\sigma^{-1} I'_{\phi,\sigma})^{-1}$, K for known. This covariance can be estimated in standard fashion, computational issues aside.

What about accounting for the uncertainty in $\hat{\boldsymbol{\theta}}$, say with covariance $\boldsymbol{\Sigma}_{\hat{\theta}}$? This part has been essentially ignored, an exception being SB (2005). If $\boldsymbol{\theta}$ is of fixed dimension and $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal with covariance matrix $\boldsymbol{\Sigma}_{\hat{\theta}}$, then as shown by Parke (1986), often

$$Acov(\hat{\boldsymbol{\omega}}_{ML}) = I(\boldsymbol{\omega})^{-1} + I(\boldsymbol{\omega})^{-1} I_{\omega,\theta} \boldsymbol{\Sigma}_{\hat{\theta}} I'_{\omega,\theta} I(\boldsymbol{\omega})^{-1}. \tag{10}$$

Hence, $Acov(\hat{\boldsymbol{\phi}}_{ML}) = Acov(\hat{\boldsymbol{\phi}}_{ML,K}) + \mathbf{Q}$, where $\mathbf{Q}$ is the upper left $p \times p$ block of the second matrix in (10) and p is the size of $\boldsymbol{\phi}$. If $\hat{\theta} = \hat{\sigma}_u^2$, then $\boldsymbol{\Sigma}_{\hat{\theta}}$ is the exact or approximate variance of $\hat{\sigma}_u^2$. SB (2005) used this result for the AR(1) model with additive constant measurement error variance and compared the asymptotic variance

of the pseudo-MLE and that of the CEE estimator in Sect. 4.1. As expected, under normality the pseudo-MLE is more efficient but in many cases the moment-based CEE did not lose much and in finite sample simulations the performances were similar.

The expression in (10) depends on $\hat{\boldsymbol{\theta}}$ being asymptotically uncorrelated with $\hat{\boldsymbol{\omega}}$ (if computed at the true $\boldsymbol{\theta}$). This certainly holds if $\hat{\boldsymbol{\theta}}$ is independent of $\mathbf{W}$. Besides possibly violating these conditions on $\hat{\boldsymbol{\theta}}$, we also need to worry about the case where the pseudo method uses the individual measurement error parameters (e.g., the $\hat{\sigma}^2_{ut}$'s) and not a simple function of them, such as the mean, so $\hat{\boldsymbol{\theta}}$ increases in dimension as T increases. Suppose that $\hat{\boldsymbol{\theta}}$ can be written as solving equations $S_2(\boldsymbol{\theta}) = \mathbf{0}$ and the $\boldsymbol{\omega}$ arises from solving $S_1(\boldsymbol{\omega}, \hat{\boldsymbol{\theta}}) = \mathbf{0}$; e.g, score equations. Note that both S_1 and S_2 depend on random quantities which have been suppressed in the notation. Use of a standard first order expansion of the estimating equations and results on the inverse of a partitioned matrix lead to an approximate covariance matrix: $Cov(\hat{\boldsymbol{\omega}}) \approx Acov(\hat{\boldsymbol{\omega}}_{ML,K}) + \mathbf{H}_{11}^{-1}\mathbf{H}_{12}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\mathbf{H}_{12}'\mathbf{H}_{11}^{-1'} - \mathbf{P}$, where $\mathbf{H}_{11} = E(\partial \mathbf{S}_1(\boldsymbol{\omega}, \boldsymbol{\theta})/\partial \boldsymbol{\omega})$, $\mathbf{H}_{12} = E(\partial \mathbf{S}_1(\boldsymbol{\omega}, \boldsymbol{\theta})/\partial \boldsymbol{\theta})$, $\mathbf{H}_{22} = E(\partial \mathbf{S}_2(\boldsymbol{\theta})/\partial \boldsymbol{\theta})$, $\mathbf{C}_{12} = E(\mathbf{S}_1\mathbf{S}_2')$, and $\mathbf{P} = \mathbf{H}_{11}^{-1}(\mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{C}_{12}' + \mathbf{C}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{12}')\mathbf{H}_{11}^{-1}$. This is a bit daunting and we are faced with many of the same issues faced in using the CEE estimator in the linear autoregressive models; see Sect. 4.1. This is a case where bootstrapping will help. Also, the approximate covariance above comes from simply using a first order approximation to the estimating equations. Work remains to be done, however, to carefully examine asymptotics in this setting where the size of $\hat{\boldsymbol{\theta}}$ is increasing with T. Notice that treating the asymptotics for the CEE estimator was easier since only the average estimated measurement variance was used.

4.3 Bayesian Methods

Bayesian methods begin with the same structure as the likelihood methods above but utilize priors for the parameters ($\boldsymbol{\phi}, \boldsymbol{\sigma}$, and $\boldsymbol{\theta}$) and base inferences on the posterior distribution of the parameters with the main focus being on estimating $\boldsymbol{\phi}$. The formulas are standard so we won't repeat them here. Some of the computational challenges are similar to those in the likelihood setting, but there are some formulations that can be easily fit using Winbugs (see, e.g., Bolker (2008, Section 11.6.2) and Viljugrein et al. (2005)) General discussion of the Bayesian approach can be found in Calder et al. (2003); Clark and Bjornstad (2004); Wang (2007), and Jungbacker and Koopman (2007). Viljugrein et al. (2005) use a pseudo-approach by using estimated measurement error variances from each point in time. Clark and Bjornstad (2004) is notable for its treatment of unequal measurement error variances by incorporating different priors on each of the σ^2_{utc} to reflect the amount of information about them. Additional applications of the Bayesian method can be found in Stenseth et al. (2003); Saether et al. (2008), and Lillegard et al. (2008), among others.

4.4 SIMEX, MEE, and RC

Here we briefly discuss three other correction methods, Simex (simulation-extrapolation), modifying estimating equations and regression calibration, all of which have been successful in treating standard regression settings; see Carroll et al. (2006) and Buonaccorsi (2010) for background and details. All of three use information about the measurement error parameters and are designed to both ease the computational burden and, more importantly, relax some of the distributional assumptions underlying likelihood and Bayesian techniques.

SIMEX. Originally due to Stefanski and Cook (1995), it has been used mainly with additive measurement error but can also accommodate multiplicative errors (e.g, Solow (1998); Resendes (2011)). Briefly it proceeds by simulating different amounts of additional measurement error to the observed W's, estimating the mean behavior of the naive approach at each of these levels of measurement error (by simulating multiple samples at that level of measurement error), then fitting a curve relating the mean behavior to the level of measurement error and projecting back to the case of no measurement error. It has seen some, but rather, limited use in dynamic settings; see Solow (1998); Ellner et al. (2002), and Bolker (2008, Chap. 11). While certainly not bullet proof, Simex has proven itself to perform quite well across a variety of regression models. Its great advantage is the need to only have to be able to fit the naive estimator. Getting analytical standard errors is more challenging and has not been examined in dynamic contexts. This is another place where the bootstrap will come in handy. Resendes (2011) evaluated the performance of SIMEX in fitting the Ricker model via simulation and obtained bootstrap standard errors and confidence intervals. He found that except for large (and unreasonable) levels of measurement error, SIMEX was quite successful in removing bias and bootstrap-based inferences performed fairly well. One problem with SIMEX in combination with the bootstrap was the huge number of fits that need to be done. This was easy for the version of the Ricker model that leads to linear least squares but was more problematic when needing to use root finding methods for solving nonlinear equations for the multiplicative version.

Modified estimating equations. This is closely related to finding corrected scores and is also motivated by minimizing distributional assumptions. The idea is to use the estimated measurement error parameters and try to modify the naive estimating equations so the corrected equations have asymptotic mean $\mathbf{0}$. For linear autoregressive models modifying the Yule-Walker equation leads to the CEE estimator while under normality modifying the score equations leads to pseudo-ML estimators. For many other cases however, it is difficult to implement this method for the same reasons associated with using the estimating equations to assess bias (Sect. 3), although approximate corrections can be found for least squares-type estimators. An advantage of an explicit set of corrected estimating equations is the ability to build off of them to get analytical expressions for the approximate covariance matrix of the estimators. While a promising approach in

general, an extensive investigation into its use for the Ricker model (the easiest of the "nonlinear" models) by Resendes (2011) found that the resulting estimator could be erratic, was outperformed by SIMEX and it was difficult to get good standard errors, either analytically or via the bootstrap. Further fine-tuning of the correction term might alleviate some of these issues.

Regression calibration. Finally we briefly speculate on regression calibration, an extremely popular method in regression contexts, which has not yet explored at all in the dynamic contexts. For an autoregressive model, consider

$$\begin{aligned} E(W_t|\mathbf{W}_{t-1}) &= E[E(W_t|\mathbf{W}_{t-1},\mathbf{Y}_{t-1}] = E[E(W_t|\mathbf{Y}_{t-1}|\mathbf{W}_{t-1})] \\ &= E[m(\boldsymbol{\phi},\mathbf{Y}_{t-1})|\mathbf{W}_{t-1}] \approx m(\boldsymbol{\phi},E(\mathbf{Y}_{t-1}|\mathbf{W}_{t-1})), \end{aligned}$$

where the approximation is exact if the model is linear in the Y's. (The last step of running the expectation through the m function is what motivated RC methods in regression.) This suggests finding an estimate $\hat{\mathbf{Y}}_{t-1}$ of $E(\mathbf{Y}_{t-1}|\mathbf{W}_{t-1})$ and then estimating $\boldsymbol{\phi}$ by regressing W_t on $\hat{\mathbf{Y}}_{t-1}$. Notice that this is *not* the same as running the usual naive analysis but replacing W_t with $\hat{Y}_t$ since we are leaving W_t as is when it is the "outcome" but modifying "predictors" by using $\hat{\mathbf{Y}}_{t-1}$ in place of $\mathbf{W}_{t-1}$. For normal stationary models and using estimated best linear predictors, it can be shown that this leads essentially to the CEE estimator in Sect. 4.1. A fruitful line of future work would be to examine the procedure above for nonlinear but stationary models and also to consider modifications to handle non-stationary models without an explicit expression for $E(Y_t)$ and $V(Y_t)$ which enter the best linear predictor of Y_t.

4.5 Bootstrapping

The preceding discussions provide a number of reasons why the bootstrap will be useful, both for getting standard errors and for assessing bias. The parametric bootstrap, based on an assumed distribution for both the measurement errors and the true values is relatively easy to implement. For an autoregressive model depending on the past p values, for each bootstrap sample b (= 1 to B), we can set $y_{b1} = (y_{b1},\ldots,y_{bp})'$, where $y_{bj} = W_j$ for $j = 1$ to p. We would then generate (sequentially) $Y_{bt} = m(\hat{\boldsymbol{\phi}},\mathbf{y}_{b,t-1}) + e_{bt}$, where e_{bt} is based on the distribution of ε_t with estimated parameters. Measurement error is then added to generate $\mathbf{W}_b$ from $\mathbf{y}_b$ according to the estimated model for $\mathbf{W}|\mathbf{y}$. If a pseudo-method is being used and we want to account for uncertainty from estimating the measurement error parameters, then we also would generate $\hat{\boldsymbol{\theta}}_b$, based on a distributional assumption. For each bootstrap sample, the corrected estimators are obtained and standard bootstrap inferences obtained using the B bootstrap values.

Notice that we need to resample from the dynamic model explicitly. (There are some methods that bootstrap via block resampling, e.g., but these are of limited

value with short series and their use needs to be carefully considered if there are measurement errors present with changing properties over time). The difficulty in using a nonparametric bootstrap is getting an estimate of the process error distribution; i.e., the distribution of the ε_t. To see the problem suppose we knew the dynamic parameters exactly and examine residuals $r_t = W_t - m(\boldsymbol{\phi}, W_{t-1}) = Y_t + u_t - m(\boldsymbol{\phi}, \mathbf{Y}_{t-1}) + (m(\boldsymbol{\phi}, \mathbf{Y}_{t-1}) - m(\boldsymbol{\phi}, W_{t-1}) = \varepsilon_t + u_t + \varepsilon_t + (m(\boldsymbol{\phi}, \mathbf{Y}_{t-1}) - m(\boldsymbol{\phi}, W_{t-1})$. This is contaminated by the measurement errors and some type of "unmixing" (related to deconvolution) is needed to get a nonparametric estimate of the distribution of ε_t. This is not an easy problem and is especially challenging with short series. The problem is even more difficult if the last term is nonlinear in Y and exacerbated further, of course, when an estimated $\hat{\boldsymbol{\phi}}$ is used. There is no work on unmixing in this particular context and so the nonparametric bootstrap remains undeveloped here, as it still does for many regression problems with measurement error.

5 Discussion

The main goal here was a broad overview of modeling and methodological issues when accounting for measurement error in fitting dynamic models. Much of the work in this area has tended to focus on likelihood methods involving distributional assumptions under fairly limited measurement error models. While these may provide good approximations in some settings, in general methods that drop the distributional assumptions and/or allow for richer measurement error models are often required; as are methods that explicitly exploit estimated measurement error parameters, which may be changing with time. While some important strides have been made in addressing these problems, the only problem with a somewhat complete solution is for additive errors in the linear autoregressive models. Of course one question still to be answered thoroughly is whether using the estimated measurement error parameters always improves the situation. More generally, a number of possible approaches to correcting for measurement error were described in Sect. 4. While a few papers have compared a couple of techniques (e.g., SB (2005) and Resendes (2011)) the majority of papers assess the performance of a single method often in comparison with a naive approach which ignores measurement error. A more comprehensive understanding of the pros and cons of the different methods is still needed. Future work is needed to explore the performance and robustness of the various procedures under a variety of assumptions; in particular the robustness of likelihood-based methods to distributional violations.

We identified a number of other problems that need further attention. One is in the treatment of nonlinear models without distributional assumptions. This turns out to be a challenge both for assessing bias and correcting for measurement error, even with fairly simple measurement error models; see Sects. 3 and 4.4 and Resendes (2011). Simex in combination with the bootstrap appears to be the best option here,

but as noted in Sect. 4.4 currently only the parametric bootstrap is available. The development of the nonparametric bootstrap would be helpful, although this will be very difficult with short series. Modifying the estimating equations does not seem all that promising here, but the regression calibration approach is worthy of investigation.

Two other areas needing attention, even when working under distributional assumptions, are allowing for richer measurement error models (such as letting the measurement error variance be different at each point in time) and accounting from uncertainty from the estimated measurement error parameters. The latter is of added difficulty when the number of estimated measurement error parameters is changing in time. In that context some avenues worth pursuing are just using the average measurement error variance for each t (there is theoretical justification for this if the process is stationary and the variance is changing as just a function of Y_t) or smoothing the variances in some way.

Lastly we note that in addition to other problems put aside for space reasons in the introduction, there is the problem of simultaneously fitting models to multiple series. This is an important topic where the use of series from many different locations which exploit the spatial structure or other assumptions about common dynamic parameters can help with the ever present problem of short series. There has been some work on measurement error in these contexts (e.g., Lillegard et al. (2008), Ives et al. (2003)), but a number of the issues raised above in treating a single series remain of interest.

Appendix

Assessing bias via estimating equations. Suppose $\mathbf{Y} \sim N(\mu\mathbf{1}, \boldsymbol{\Sigma}_Y)$. The estimating equations for the ML estimators of the parameters in $\boldsymbol{\Sigma}_Y$ (say $\psi_1, \ldots, \psi_J$) can be written McCulloch et al. (2008, p. 165) as $tr(\boldsymbol{\Sigma}_Y^{-1}\mathbf{G}_j) - (\mathbf{y} - \mu\mathbf{1})'\boldsymbol{\Sigma}_Y^{-1}\mathbf{G}_j\boldsymbol{\Sigma}_Y^{-1}(\mathbf{y} - \mu\mathbf{1}) = 0$, for $j = 1$ to J, where J is the number of parameters in $\boldsymbol{\Sigma}_Y$ and $\mathbf{G}_j = \partial\boldsymbol{\Sigma}_Y/\partial\psi_j$. Replacing $\mathbf{y}$ with $\mathbf{W}$ and taking the expected value, but denoting the arguments of the estimating equations denoted with a $*$ leads to an expected value of the jth estimating equation of $E_j = tr(\boldsymbol{\Sigma}_Y^{*-1}\mathbf{G}_j^*) - tr(\boldsymbol{\Sigma}_Y^{*-1}\mathbf{G}_j^*\boldsymbol{\Sigma}_Y^{*-1}\boldsymbol{\Sigma}_W)$. If we can find Σ_Y^* of the same form as $\boldsymbol{\Sigma}_Y$ so that each E_j is 0, then the naive estimators of the parameters in $\boldsymbol{\Sigma}_Y$ are consistent for the parameters in $\boldsymbol{\Sigma}_Y^*$. Obviously E_j is 0 if $\boldsymbol{\Sigma}_Y^* = \boldsymbol{\Sigma}_W$ but this only provides the asymptotic bias immediately if $\boldsymbol{\Sigma}_W$ is of the same form as $\boldsymbol{\Sigma}_Y$. If the measurement errors are additive with constant (unconditional) variance σ_u^2, then $\boldsymbol{\Sigma}_W = \boldsymbol{\Sigma}_Y + \sigma_u^2\mathbf{I}$, and we can take $\boldsymbol{\Sigma}_Y^* = \boldsymbol{\Sigma}_W$. This means the naive estimator or σ_Y^2 asymptotically estimates $\sigma_Y^2 + \sigma_u^2$ while the naive estimators of the off-diagonal covariance terms in Y are correct and the ML estimators are asymptotically like the YW estimators.

Allowing unequal unconditional variances (as can occur with changing sampling effort) $\boldsymbol{\Sigma}_W = \boldsymbol{\Sigma}_Y + Diag(\sigma_{u1}^2, \ldots, \sigma_{uT}^2)$ the question, which we have not investigated, is whether $E_j \to 0$ as T increases if we take $\boldsymbol{\Sigma}_Y^* = \boldsymbol{\Sigma}_Y + \sigma_u^2$, where $\sigma_u^2 = \sum_{t=1}^T \sigma_{ut}^2/T$.

Acknowledgments I wish to express my appreciation to Brajendra Sutradhar for organizing the International Symposium in Statistics on Longitudinal Data Analysis and for serving as editor for the written contributions to this volume. I would also like to thank the members of the audience at the conference for helpful feedback and particularly thank the two referees whose comments and corrections helped improve the initial draft of this paper. Portions of this work grew out of earlier joint work with John Staudenmayer, which was supported by NSF-DMS-0306227, and with David Resendes. I am grateful to David for allowing the use of some results here from his dissertation. This paper also grew out of a seminar presented at the Center for Ecological and Evolutionary Synthesis (CEES), at the University of Oslo, in spring of 2010. I am very grateful to Nils Stenseth and CEES for support during my stay there.

References

Aigner, D., Hsiao, C., Kapteyn, A., Wansbeek, T.: Latent variable models in econometrics. In: Griliches, Z., Intriligator, M.D. (eds.) Handbook of Econometrics, pp. 1321–1393. Elsevier, Amsterdam (1984)

Barker, D., Sibly, R.M.: The effects of environmental perturbation and measurement error on estimates of the shape parameter in the theta-logistic model of population regulation. Ecol. Model. **219**, 170–177 (2008)

Bell, W.R., Wilcox, D.W.: The effect of sampling error on the time series behavior of consumption data. J. Econometrics. **55**, 235–265 (1993)

Berliner L.M.: Likelihood and Bayesian prediction of chaotic systems. J. Am. Stat. Assoc. **86**, 938–952 (1991)

Bolker, B.: Ecological Models and Data in R. Princeton University Press, Princeton, NJ (2008)

Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis, Forecasting and Control, 3rd edn. Prentice Hall, Englewood Cliffs, NJ (1994)

Brockwell, P., Davis, R.: Introduction to Time Series and Forecasting. Springer, New York (2002)

Buonaccorsi, J.P.: Measurement Error, Models, Methods, and Applications. Chapman and Hall, London (2010)

Buonaccorsi, J.P., Staudenmayer, J., Carreras, M.: Modeling observation error and its effects in a random walk/extinction model. Theor. Popul. Biol. **70**, 322–335 (2006)

Buonaccorsi J,P., Staudenmayer, J.: Statistical methods to correct for observation error in a density-independent population model. Ecol. Monogr. **79**, 299–324 (2009)

Buonaccorsi, J.P., Staudenmayer, J.: Measurement error in linear autoregressive models II: further results and inferences. Working Paper, University of Massachusetts (2012)

Burr, T., Chowell, G.: Observation and model error effects on parameter estimates in susceptible-infected-recovered epidemic model. Far East J. Theor. Stat. **19**, 163–183 (2006)

Calder, C., Lavine, M., Muller, P., Clark, J.: Incorporating multiple sources of stochasticity into dynamic population models. Ecology **84**, 1395–1402 (2003)

Carroll, R.J., Stefanski, L.A., Ruppert, D., Crainiceanu C.M.: Measurement Error in Nonlinear Models, 2nd edn. Chapman and Hall, London (2006)

Chanda, K.C.: Asymptotic properties of estimators for autoregressive models with errors in variables. Ann. Stat. **24**, 423–430 (1996)

Cheang, W., Reinsel, G.C.: Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. J. Am. Stat. Assoc. **95**, 1173–1184 (2000)

Clark, J.S., Bjornstad, O.N.: Population time series, process variability, observation errors, missing values, lags, and hidden states. Ecology. **85**, 3140–3150 (2004)

Dennis, B., Ponciano, J., Lele, S., Taper, M., Staples, D.: Estimating density dependence, process noise, and observation error. Ecol. Monogr. **76**, 323–341 (2006)

Dennis, B., Ponciano, J.M., Taper, M.L.: Replicated sampling increases efficiency in monitoring biological populations. Ecology **91**, 610–620 (2010)

De Valpine, P.: Review of methods for fitting time-series models with process and observation error and likelihood calculations for nonlinear non-Gaussian state-space models. Bull. Mar. Sci. **70**, 455–471 (2002)
De Valpine, P.: Monte-Carlo state-space likelihoods by weighted posterior kernel density estimation. J. Am. Stat. Assoc. **99**, 523–536 (2004)
De Valpine, P., Hastings, A.: Fitting population models incorporating process noise and observation error. Ecol. Monogr. **72**, 57–76 (2002)
De Valpine, P., Hilborn, R.: State-space likelihoods for nonlinear fisheries time series. Can. J. Fish. Aquat. Sci. **62**, 1937–1952 (2005)
Ellner, S., Yodit, S., Smith, R.: Fitting population dynamic models to time-series by gradient matching. Ecology **83**, 2256–2270 (2002)
Feder, M.: Time series analysis of repeated surveys: the state-space approach. Stat. Neerl. **55**, 182–199 (2001)
Harvey, A.C.: Forecasting, Structural Time Series Models, and the Kalman Filter. Cambridge University Press, Cambridge (1990)
Fuller, W.: Time Series Analysis. Wiley, New York (1996)
Hovestadt, T., Nowicki, P.: Process and measurement errors of population size, their mutual effects on precision and bias of estimates for demographic parameters. Biodivers. Conserv. **17**, 3417–3429 (2008)
Ives, A.R., Dennis, B., Cottingham, K.L., Carpenter, S.R.: Estimating community stability and ecological interactions from time-series data. Ecol. Monogr. **73**, 301–330 (2003)
Ives, A.R, Abbott, K., Ziebarth, N.: Analysis of ecological time series with ARMA(p,q) models. Ecololgy **91**, 858–871 (2010)
Jungbacker, B., Koopman, S.J.: Monte Carlo estimation for nonlinear non-Gaussian state space models. Biometrika **94**, 827–839 (2007)
Knape, J.: Estimability of density dependence in models of time series data. Ecology **89**, 2994–3000 (2008)
Knape, J., Jonzén, N., Skold, M.: Observation distributions for state space models of population survey data. J. Anim. Ecol. **80**, 1269–1277 (2011)
Koons, B.K., Foutz, R.V.: Estimating moving average parameters in the presence of measurement error. Comm. Stat. **19**, 3179–3187 (1990)
Lee, J.H., Shin, D.W.: Maximum likelihood estimation for ARMA models in the presence of ARMA errors. Comm. Stat. Theor. Meth. **26**, 1057–1072 (1997)
Lele, S.R.: Sampling variability and estimates of density dependence, a composite-likelihood approach. Ecology **87**, 189–202 (2006)
Lele, S.R., Dennis, B., Lutscher, F.: Data cloning, easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. **10**, 551–563 (2007)
Lillegard, M., Engen, S., Saether, B.E., Grotan, V., Drever, M.: Estimation of population parameters from aerial counts of North American mallards: a cautionary tale. Ecol. Appl. **18**, 197–207 (2008)
Mallick, T., Sutradhar, B.: GQL versus conditional GQL inferences for non-stationary time series of counts with overdispersion. J. Time Anal. **29**, 402–420 (2008)
Miazaki, E.S., Dorea, C.C.Y.: Estimation of the parameters of a time series subject to the error of rotation sampling. Commun. Stat. A - Theor. Meth. **22**, 805–825 (1993)
McCulloch, C., Searle, S., Neuhaus, J. : Generalized, Linear, and Mixed Models, 2nd edn. Wiley, New York (2008)
Morris, W.F., Doak, D.F.: Quantitative Conservation Biology: Theory and Practice of Population Variability Analysis. Sinauer Associates, Sunderland, MA (2002)
Pagano, M.: Estimation of models of autoregressive signal plus white noise. Ann. Stat. **2**, 99–108 (1974)
Parke, W.: Pseudo maximum likelihood estimation: the asymptotic distribution. Ann. Stat. **14**, 355–357 (1986)

Pfeffermann, D., Feder, M., Signorelli, D.: Estimation of autocorrelations of survey errors with application to trend estimation in small areas. J. Bus. Econ. Stat. **16**, 339–348 (1998)
Ponciano, J., Taper, M., Dennis, B., Lele, S.: Hierarchical models in ecology, confidence intervals hypothesis testing, and model selection using data cloning. Ecology **90**, 356–362 (2009)
Resendes, D.: Statistical methods for nonlinear dynamic models with measurement error using the Ricker model. Ph.D. thesis, University of Massachusetts, Amherst (2011)
Sakai, H., Soeda, T., Hidekatsu, T.: On the relation between fitting autoregression and periodogram with applications. Ann. Stat. **7**, 96–107 (1979)
Saether, B., Lilligard, M., Grotan, V., Drever, M., Engen, S., Nudds, T., Podruzny, K.: Geographical gradients in the population dynamics of North American prairie ducks. J. Anim. Ecol. **77**, 869–882 (2008)
Schmid, C.H., Segal, M.R., Rosner, B.: Incorporating measurement error in the estimation of autoregressive models for longitudinal data. J. Stat. Plann. Infer. **42**, 1–18 (1994)
Solow, A.R.: On fitting a population model in the presence of observation error. Ecology **79**, 1463–1466 (1998)
Solow, A.R.: Observation error and the detection of delayed density dependence. Ecology **82**, 3263–3264 (2001)
Staudenmayer, J., Buonaccorsi, J.P.: Measurement error in linear autoregressive models. J. Am. Stat. Assoc. **100**, 841–852 (2005)
Stefanski, L.: The effects of measurement error on parameter estimation. Biometrika **72**, 583–592 (1985)
Stefanski, L., Cook, J.: Simulation-extrapolation: the measurement error jackknife. J. Am. Stat. Assoc. **90**, 1247–1256 (1995)
Stenseth, N.C., Viljugrein, H., Saitoh, T., Hansen, T.F., Kittilsen, M.O., Bolviken, E., Glockner, F.: Seasonality, density dependence, and population cycles in Hokkaido voles. Proc. Natl. Acad. Sci. USA. **100**, 11478–11483 (2003)
Tripodis, Y., Buonaccorsi, J.P.: Prediction and forecasting in linear models with measurement error and unknown parameters. J. Stat. Plann. Infer. **139**, 4039–4050 (2009)
Viljugrein, H., Stenseth, N.C., Smith, G.W., Steinbakk, G.H.: Density dependence in North American ducks. Ecology **86**, 245–254 (2005)
Walker, A.M.: Some consequences of superimposed error in time series analysis. Biometrika **47**, 33–43 (1960)
Wang, G.: On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. Ecol. Model. **20**, 521–528 (2007)
Wang, G., Hobbs, N.T., Boone, R.B., Illius, A.W., Gordon, I.J., Gross, J.E., Hamlin, K.L.: Spatial and temporal variability modify density dependence in populations of large herbivores. Ecology **87**(1), 95–102 (2006)
Williams, C.K., Ives, A.R., Applegate, R.D.: Population dynamics across geographical ranges, time-series analyses of three small game species. Ecology **84**, 2654–2667 (2003)
Wong, W-K., Miller, R.B.: Repeated time series analysis of ARIMA-noise models. J. Bus. Econ. Stat. **8**, 243–250 (1990)
Wong, W-K., Miller, R.B., Shrestha, K.: Maximum likelihood estimation of ARMA models with error processes for replicated observations. J. Appl. Stat. Sci. **10**, 287–297 (2001)