FUNDAMENTAL LIMITATIONS OF AD HOC LINEAR AND QUADRATIC MULTI-LEVEL REGRESSION MODELS FOR PHYSICAL SYSTEMS

ANDREW J. MAJDA

Department of Mathematics and Center for Atmosphere and Ocean Science
Courant Institute for Mathematical Sciences, New York University
New York, NY 10012-1110, USA

YUAN YUAN

Courant Institute for Mathematical Sciences, New York University
New York, NY 10012-1110, USA

Abstract. A central issue in contemporary applied mathematics is the development of simpler dynamical models for a reduced subset of variables in complex high dimensional dynamical systems with many spatio-temporal scales. Recently, ad hoc quadratic multi-level regression models have been proposed to provide suitable reduced nonlinear models directly from data. The main results developed here are rigorous theorems demonstrating the non-physical finite time blow-up and large time instability in statistical solutions of general scalar multi-level quadratic regression models with corresponding unphysical features of the invariant measure. Surprising intrinsic model errors due to discrete sampling errors are also shown to occur rigorously even for linear multi-level regression dynamic models. All of these theoretical results are corroborated by numerical experiments with simple models. Single level nonlinear regression strategies with physical cubic damping are shown to have significant skill on the same test problems.

1. Introduction. A central issue in contemporary applied mathematics is the development of simpler dynamical models for a reduced subset of variables in complex high dimensional dynamical systems with many spatio-temporal scales. This is an important issue in systems ranging from biomolecular dynamics to climate atmosphere ocean science to engineering turbulence. The approaches to address these issues range from analytic stochastic mode reduction strategies [1, 2, 3, 4] to purely data driven strategies [5, 6, 7, 23] to methods which use physical analytic properties to constrain data driven methods [8, 9, 10, 21]. The references listed above extensively describe other related earlier work on these issues.

Another straightforward approach relies on adapting multi-level regression techniques for nonlinear time series to fit the data output of some variables of a given nonlinear system without regard to any physical constraints. Single level linear regression models of this type for a reduced subset of variables have some skill for...
low frequency variability ([11] and references therein) and suitable improved
versions have very high skill for active data assimilation of turbulent signals ([12]
and references therein) as well as some skill for the low frequency response to changes
in external forcing [13, 14]. Recently, quadratically nonlinear multi-level regression
models have been utilized to determine suitable reduced nonlinear models from data
[15, 16, 17] and significant skill for these models for fitting the time series as well
as for prediction has been reported in that work. All of these ad hoc regression
strategies suffer from severe model error; in other words, they do not model the
exact nonlinear dynamics of the physical system on the reduced subset of variables
[14, 22].

The goal of the present work is to develop a series of rigorous mathematical the-
orems which illustrate unambiguously the fundamental limitations of such ad hoc
linear and quadratic multi-level regression models and also to demonstrate these
effects in numerical experiments with simple models. These limitations include
non-physical finite time blow-up and large time instability in statistical solutions
of general scalar multi-level quadratic regression models with corresponding un-
physical features of the invariant measure. These results are presented in section
3 below while the preliminary section 2 contains a brief introduction to physics
constrained stochastic mode reduction models and ad hoc multi-level linear and
quadratic regression models. With the background from section 2, simple theorems
are developed in section 4 which point out some remarkable fundamental limita-
tions in multi-level regression modeling for reduced sets of variables even for linear
models due to model error involving discrete sampling of the time series. Other
interesting recent work on this topic involves discrete sampling of systems with
widely separated time scales [18, 19, 20]; the examples and theory here apply to
situations without such a wide separation of time scale and highlight the role of
non-Markovian features. Section 5 is a brief concluding discussion.

2. Physical stochastic mode reduction and Ad hoc multi-level nonlinear
regression models. Here we introduce physical stochastic mode reduction and
both linear and quadratic multi-level nonlinear regression models in the context of
an observed scalar variable which arises from a true or perfect model with more
degrees of freedom and with nonlinear dynamics which is not known in detail, in
other words, there is model error. The emphasis here is on simple models which
illustrate the typical features of vastly more complicated systems.

2.1. Physical stochastic mode reduction: A prototype model. The three-
mode dyad model discussed next is a simple nonlinear stochastic model in the family
of prototype models [3, 4, 8] which mimic the structure of dynamical cores in large-
scale climate models. One of the modes in the model, $x_1$, is intended to represent a
slowly-evolving variable accessible to observation, whereas the remaining modes, $x_2,
x_3$, act as surrogate variables for the unresolved (subgrid scale) degrees of freedom
in the climate model. The unresolved modes are coupled to $x_1$ linearly and via a
dyad interaction between $x_1$ and $x_2$, and the model also allows for external forcing
to act on mode $x_1$. Specifically, the governing stochastic differential equations are

$$
\begin{align*}
    dx_1 &= (I x_1 x_2 + L_{12} x_2 + L_{13} x_3 + F + L_{11} x_1) dt, \\
    dx_2 &= (-I x_1^2 - L_{12} x_1 - \gamma_2 \varepsilon^{-1} x_2) dt + \sigma_2 \varepsilon^{-\frac{1}{2}} dW_2, \\
    dx_3 &= (-L_{13} x_1 - \gamma_2 \varepsilon^{-1} x_3) dt + \sigma_3 \varepsilon^{-\frac{1}{2}} dW_3,
\end{align*}
$$

(2.1a) (2.1b) (2.1c)
where \( \{W_1, W_2\} \) are independent Wiener processes, and the parameters \( I, \{L_{11}, L_{12}, L_{13}\} \), and \( F \), respectively measure the dyad interaction, the linear couplings, and the external forcing. Model (2.1), as well as the associated reduced scalar model below in equation (2.2), has been used as a prototype model to develop methods based on the fluctuation-dissipation theorem (FDT) for assessing the low-frequency climate response to external perturbations and model error [13, 14, 22].

The parameters of the three-mode model provide a physical correspondence with a number of important dynamical processes in the cores of comprehensive climate models ([4], chap. 3). Specifically, \( F \) represents a large-scale external forcing (e.g., solar heating), while \( L_{11} \) describes linear dissipative effects, such as viscosity and surface drag (for this reason, \( L_{11} \) would generally be viewed as a component of a negative-definite operator). The combined effect of \( F \) and \( L_{11} \) is referred to as the bare truncation, as it would still be part of the equation of motion for \( x_1 \) if one were to completely ignore its interaction with the unresolved modes. The coefficients \( L_{ij} \) are components of a skew-symmetric operator representing, e.g., the \( \beta \) effect associated with the Coriolis force, and the parameter \( I \) measures of interaction between modes \( x_1 \) and \( x_2 \). Notice that this nonlinear dyad interaction has the physical property of conservation of energy and a physics based reduced model for \( x_1 \) alone should reflect this important physical feature [3, 8, 21]. In geophysical flows, this type of interaction is expected to arise whenever the state variables are expansion coefficients in general bases of empirical orthogonal functions (EOFs) for the core dynamical fields [8].

Aside from the above deterministic components, a key ingredient of model (2.1) is stochastic forcing in equations (2.1b) and (2.1c), which, together with the linear damping parameterized by \( \gamma_i \), acts as a surrogate for the nonlinear interaction between the unresolved modes. The parameter \( \varepsilon \) controls the time-scale separation of the dynamics of the slow and fast modes, with the fast modes evolving infinitely fast relative to the climate-variable time scale in the limit \( \varepsilon \to 0 \).

We now examine the scalar stochastic model associated with the three-mode model in the limit \( \varepsilon \to 0 \). This reduced version of the model is particularly useful in exposing in a transparent manner the influence of the unresolved modes, \( (x_2, x_3) \) on the resolved mode, \( x_1 \), when there exists a clear separation of time scales in their respective dynamics (i.e., when \( \varepsilon \) is small). As follows by applying the MTV mode reduction procedure [1, 2, 3, 8, 13] to the coupled system (2.1), the resolved mode, \( x_1 \), in the reduced model is governed by the nonlinear stochastic differential equation

\[
\begin{align*}
\frac{dx_1}{dt} &= (F + L_{11}x_1)dt + \varepsilon \left( \frac{\sigma^2 I L_{12}}{2\gamma_2^2} + \frac{\sigma^2 I^2}{2\gamma_2} - \left( \frac{L_{12}^2}{\gamma_2} + \frac{L_{13}^2}{\gamma_3} \right) \right) x_1 - \frac{2IL_{12}}{\gamma_2} x_1^2 - \frac{I^2}{\gamma_2} x_1^3 dt + \frac{\sigma}{\gamma_2} L_{12} dW_2 \\
&+ \varepsilon \frac{\sigma^2}{\gamma_2} (Ix_1 + L_{12})dW_2 \\
&+ \varepsilon \frac{\sigma^3}{\gamma_3} L_{13}dW_3,
\end{align*}
\]

which may also be expressed in the form

\[
\frac{dx_1}{dt} = (\tilde{F} + ax_1 + bx_1^2 - cx_1^3)dt + (A - Bx_1)dW_2 + \sigma dW_3
\]
with the parameter values
\[
\tilde{F} = F + \varepsilon \frac{\sigma_2^2 L_{12}}{2\gamma_2}, \quad (2.4a)
\]
\[
a = L_{11} + \varepsilon \left( \frac{\sigma_2^2 I_2^2}{2\gamma_2} - \frac{L_{12}^2 + L_{13}^2}{\gamma_2} \right), \quad b = -\varepsilon \frac{2IL_{12}}{\gamma_2}, \quad c = \varepsilon \frac{I^2}{\gamma_2}, \quad (2.4b)
\]
\[
A = \varepsilon \frac{\sigma_2 L_{12}}{\gamma_2}, \quad B = -\varepsilon \frac{\sigma_2 I}{\gamma_2}, \quad \sigma = \varepsilon \frac{\sigma_2 L_{13}}{\gamma_2}. \quad (2.4c)
\]

Among the terms in the right-hand side of equation (2.2) we identify (i) the bare truncation (2.2a); (ii) a nonlinear deterministic driving (2.2b) of the climate mode mediated by the linear and dyad interactions with the unresolved modes; (iii) correlated additive multiplicative noise (2.2c); and (iv) additive noise (2.2d). Moreover, we note the parameter interdependence \( B = 2cB \equiv 0 \), which is a manifestation of the fact that in scalar models of the form (2.3), whose origin lies in multivariate models with multiplicative energy conserving dyad interactions [e.g. the three-mode model (2.1)], a non-zero multiplicative-noise parameter, \( B \), is accompanied by a non-zero cubic damping, \( c \). As noted in [8] and proved rigorously in [21], cubic damping has the important role of suppressing power-law tails of the probability density function (PDF) which are not compatible with climate data [9, 24]. The explicit derivation above can be generalized substantially using general averaging theorems [25, 26].

2.1.1. Physics based single level Regression Strategies. The above stochastic mode reduction strategy and the associated normal form in (2.3) yield a physics-based parametric regression strategy where the coefficients \( \tilde{F}, a, b, c, A, B, \sigma \) in (2.3) are all estimated from data in a time series for \( x_1 \) even though the dynamics of the true system being modeled is vastly more complicated but with the general physical structural features of (2.1) and the parameter, \( \varepsilon \) is not necessarily small. Physical examples of this procedure are presented in primitive form in [10] and in [8], using the normal form in (2.3) explicitly.

For simplicity in exposition here, we set \( A = B = 0 \) in the regression models in (2.3) and utilize Euler’s method for (2.3) resulting in the discrete model
\[
\frac{x_{n+1} - x_n}{\Delta t} = F + ax_n + bx_n^2 + cx_n^3 + \frac{\sigma}{\sqrt{\Delta t}} \varepsilon_n. \quad (2.5)
\]

Since the coefficients \( F, a, b, c \) all enter linearly in (2.5) given the input time series, \( x_n, \quad n = 0, 1, 2... \), we follow [15, 16, 17] and use linear regression least squares estimates for the coefficients \( F, a, b, c \).

The resulting scalar cubic regression model (2.5) has parameter estimators
\[
\beta = (X^TX)^{-1}(X^TY) \quad (2.6)
\]
where
\[
\beta = [F \ a \ b \ c]^T; \quad (2.7a)
\]
\[
Y_n = \frac{x_{n+1} - x_n}{\Delta t} \quad (2.7b)
\]
\[
X_n = [1 \ x_n \ x_n^2 \ x_n^3] \quad (2.7c)
\]
So, we have that
\[
\beta = [F \ a \ b \ c]^T = \left( \begin{array}{cccc} 1 & \mu_1 & \mu_2 & \mu_3 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ \mu_2 & \mu_3 & \mu_4 & \mu_5 \\ \mu_3 & \mu_4 & \mu_5 & \mu_6 \end{array} \right)^{-1} (X^T Y) \tag{2.8}
\]
where \(\mu_k\) is the \(kth\) moment of \(x_n\), \(\mu_k = \bar{x}_k\). Therefore, we need a long enough time series (compared to the decorrelation time) to have accurate 6th moment estimation.

Since polynomial regression problem has such strong stiffness, we always replace the input time series by the centered and normalized one
\[
\tilde{x} = \frac{x - \bar{x}}{\text{std}(x)}
\]
with \(\bar{x}\), the numerical mean and \(\text{std}(x)\), the standard deviation. This centering and normalizing reduces the conditioning number of this least square problem and gives better estimation of the parameters.

The physical considerations developed earlier require that \(c\) satisfy \(c > 0\) in a time series analysis. The limitation of the above physics-based regression models is that they involve only a single level of the time series and cannot include memory effects.

2.2. Ad hoc multi-level quadratic regression models. The ad hoc quadratic multi-level quadratic regression models proposed and used in [15, 16, 17] for the scalar variables \(x\) have the form,
\[
\begin{align*}
dx &= (F + ax + bx^2)dt + r^1 dt \\
dr^i &= \left( q_i x + \sum_{j=1}^{i} a_{ij} r^j + r^{i+1} \right) dt, \quad i = 1, 2, ..., N - 1 \\
dr^N &= \left( q_N x + \sum_{j=1}^{N} a_{ij} r^j \right) dt + \sigma dW
\end{align*}
\tag{2.9a}
\]
which can be written in compact form as the following
\[
\begin{align*}
dx &= (F + ax + bx^2)dt + r^1 dt \\
dr &= (q x + A r) dt + \Sigma dW
\end{align*}
\tag{2.10a}
\]
where
\[
r = \begin{pmatrix} r^1 \\ \vdots \\ r^N \end{pmatrix}, \quad q = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^N; \tag{2.11a}
\]
\[
A = \begin{pmatrix} a_{11} & 1 & 0 & \ldots & 0 \\ a_{21} & a_{22} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} & 0 \end{pmatrix} \in \mathbb{R}^{N \times N}. \tag{2.11b}
\]

The motivation for these models which have \(N\)-levels is that they incorporate nonlocal linear memory effects in the time history of \(x(t)\) through the \(N\) memory levels in (2.9) as follows by successively integrating the equations for \(r^N, r^{N-1}, ..., r^1\)
and substituting the formula for $r^1$ into the equation for $x$. This is a potential advantage of these methods beyond the Markovian methods in section 2.1.1. All the coefficients $F, a, b, q, A, \Sigma$ are estimated by a similar linear regression procedure from the time series of successive time differences of $x_n, n = 0, 1, \ldots$. One disadvantage of the approach is that a huge number of non-unique coefficients need to be measured by the least squares procedure, of order $10^3$ for fifteen variables with three levels [15].

Another disadvantage of the quadratic regression model in (2.9) is that the quadratic term with $b \neq 0$ is introduced in an ad hoc fashion and has no direct physical meaning in contrast to the physical quadratic terms in (2.1), for example, that arise from conservation of energy in the nonlinear dynamics; these physical terms resulted in the normal form in (2.3) with the corresponding physics based regression model discussed in section 2.1.1 above. Even more troubling is that the quadratic coefficient, $b \neq 0$, in (2.9) might create non-physical blow-up and instability of statistical solutions of the multi-level regression model in contrast to the correct physical behavior of the underlying system [8, 9, 21, 24]. This can dramatically alter the predictive skill of the models in (2.9). Such unphysical features are established rigorously in section 3. Such possibilities were noted in [15, 16, 17] but discussed as being insignificant. Next we set $b = 0$ and $N = 1$ so that we introduce two-level linear regression models where the issue of model error and non-uniqueness in regression fitting strategies are transparent yet non-trivial.

2.3. Instructive test models for two-level linear regression. Consider a true signal $x(t)$ to be modeled which is the first component of the solution of the $2 \times 2$ system of the linear stochastic equations

\begin{align}
\text{dx} &= (\bar{a}x + y)dt \\
\text{dy} &= (\bar{q}x + \bar{A}y)dt + \bar{\sigma}dW(t)
\end{align}

The dynamics in (2.12) is linear and has a Gaussian invariant measure provided the stability conditions

\begin{align}
\bar{a} + \bar{A} < 0, \quad \bar{a}\bar{A} - \bar{q} > 0
\end{align}

are satisfied. Furthermore, under these circumstances, the zero-mean statistically stationary Gaussian climatological time series for the variable $x(t)$ is complete determined by the correlation function $C(\tau) = \langle x(t)x(t + \tau) \rangle$.

Now, suppose we observe only $x(t)$ and don’t know the detailed model in (2.12). We follow the ad hoc strategy in section 2.2 above and consider the family of two-level linear regression models as a special case of (2.9) with $b = 0$ with the form

\begin{align}
\text{dx} &= (ax + y)dt \\
\text{dy} &= (qx + Ay)dt + \sigma dW(t)
\end{align}

The models in (2.14) are candidate two-level regression models for the unknown dynamics in (2.12) which arise from fitting a long time series of $x(t)$ alone. Which of the models in (2.14) have equilibrium marginal dynamics for $x(t)$ which exactly recovers the equilibrium statistics, $\langle x(t)x(t + \tau) \rangle = C(\tau)$? This is easy to answer through the following result.

**Proposition 2.1.** Consider the two-level linear model

\begin{align}
\text{d} \begin{pmatrix} x \\ y \end{pmatrix} = L \begin{pmatrix} x \\ y \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma dW. \end{pmatrix}
\end{align}
where $\sigma \neq 0$, and the stability coefficients, $L = \begin{pmatrix} a & 1 \\ q & A \end{pmatrix}$, satisfy the realizability conditions

$$a + A < 0, \quad aA - q > 0. \tag{2.16}$$

Assume $L$ has eigenvalues $\lambda_{1,2}$, then, the equilibrium marginal dynamics, $x(t)$, are determined by $\{\lambda_{1,2}, \sigma\}$, equivalently, $\{a + A, aA - q, \sigma\}$.

Thus, given the observed climatological distribution in (2.12), any two-level regression model in (2.14) satisfying

$$\tilde{a} + \tilde{A} = a + A, \quad \tilde{a}\tilde{A} - \tilde{q} = aA - q, \quad \tilde{\sigma} = \sigma \tag{2.17}$$

automatically, has same equilibrium marginal dynamics for $x(t)$. This is intrinsic non-uniqueness in the two-level regression model in a transparent fashion. The proof of Proposition 2.1 is a direct calculation as follows.

The realizability conditions make sure that the real parts of the eigenvalues of $L$, $\text{Re}(\lambda_{1,2})$, are negative, which implies that the two-level model has a stable equilibrium measure.

Under the realizability conditions, this two-level model determines a 2D Gaussian process with zero mean $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, positive definite equilibrium covariance matrix $\Sigma = \begin{pmatrix} 1 & -a \\ -a & aA - q + a^2 \end{pmatrix}$, and autocovariance function

$$C(\tau) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \begin{pmatrix} x(t + \tau) \\ y(t + \tau) \end{pmatrix}^T = \Sigma e^{L\tau} \tag{2.19}$$

The equilibrium marginal dynamics, $x(t)$, is also Gaussian with zero mean $\bar{x} = 0$ and autocovariance function $C_{11}(\tau) = (\Sigma e^{L\tau})_{11}$. It is known that $x(t)$ is uniquely determined by the mean and the covariance function. We calculate the autocovariance function $C_{11}(\tau)$ in two cases.

Case 1: Stability coefficient matrix $L$ has two different eigenvalues $\lambda_1 \neq \lambda_2$ which requires

$$\frac{(A - a)^2}{4} + q \neq 0.$$

The autocovariance of the equilibrium marginal dynamics, $x(t)$, is the first entry of $C(\tau)$,

$$C_{11}(\tau) = (C(\tau))_{11} = \frac{\sigma^2(\lambda_1 e^{\lambda_2 \tau} - \lambda_2 e^{\lambda_1 \tau})}{(\lambda_2^2 - \lambda_1^2)\lambda_1 \lambda_2}.$$

Case 2: Stability coefficient matrix $L$ has the same eigenvalues $\lambda_1 = \lambda_2$ which requires

$$\frac{(A - a)^2}{4} + q = 0.$$

The autocovariance of the equilibrium marginal dynamics, $x(t)$,

$$C_{11}(\tau) = (C(\tau))_{11} = \frac{\sigma^2(\lambda \tau - 1)e^{\lambda \tau}}{4\lambda^3}.$$

Therefore, in order to reproduce the autocovariance function of $x(t)$, we need to keep the values of $\{\lambda, \sigma\}$, equivalently, $\{a + A, \sigma\}$. 
2.3.1. Two-Level Regression Algorithms. For completeness, here we introduce the Two-Level Regression Procedure suggested in [15].

For the linear case, the Two-Level Regression model is given by

\[ d \left( \begin{array}{c} x \\ y \end{array} \right) = \left( \begin{array}{cc} \tilde{a} \\ \tilde{q} \\ \tilde{A} \end{array} \right) \left( \begin{array}{c} x \\ y \end{array} \right) dt + \left( \begin{array}{c} 0 \\ 0 \end{array} \right). \] (2.20)

According to the multi-level regression procedure [15], in practice the increments \(dx, dy\) can be approximated by

\[ dx_i = x_{i+1} - x_i, \quad dy_i = y_{i+1} - y_i, \quad 0 \leq i \leq \tilde{N} - 1 \]

where \(x_i = x(t_i), t_i = i\Delta t\), \(\Delta t\) is assumed to be equal to the data set’s sampling interval. We also assume that we have the white noise residual \(dr = \sigma dW\) which can be approximated by

\[ dr_i = \tilde{\sigma} \xi_i \sqrt{\Delta t}, \xi_i \sim N(0, 1). \]

So, the Two-Level Regression model becomes

\[ \left( \begin{array}{c} x_{i+1} - x_i \\ y_{i+1} - y_i \end{array} \right) = \left( \begin{array}{cc} \tilde{a} \\ \tilde{q} \\ \tilde{A} \end{array} \right) \left( \begin{array}{c} x_i \\ y_i \end{array} \right) \Delta t + \left( \begin{array}{c} 0 \\ \tilde{\sigma} \xi_i \sqrt{\Delta t} \end{array} \right). \] (2.21)

Note that this procedure makes model errors compared to the original linear two level model (2.20). In section 4 below, we will show that this kind of model error will lead to an intrinsic error when we use this regression procedure to recover the marginal dynamics of the mode \(x\) in (2.20).

At the first level, we need to solve a least squares problem for

\[ \frac{x_{i+1} - x_i}{\Delta t} = \tilde{a} x_i + y_i, \quad 0 \leq i \leq \tilde{N} - 1 \]

where \(\tilde{a}\) is the unknown parameter, \(\{y_i\}\) is the residual. Let’s denote

\[ Z = \left\{ \frac{x_{i+1} - x_i}{\Delta t} \right\}, \quad X = \{x_i\}, \quad Y = \{y_i\}, \]

we get the compact form of the least square problem for estimating unknown parameter \(\tilde{a}\)

\[ Z = \tilde{a} X + Y. \]

So, by the standard least squares method, we have the parameter estimate

\[ \tilde{a} = \frac{X^T Z}{X^T X} = \frac{\sum_{i=0}^{\tilde{N}-1} (x_{i+1} - x_i)x_i}{\Delta t \sum_{i=0}^{\tilde{N}-1} x_i^2} \] (2.22)

and the residual

\[ Y = Z - \tilde{a} X, \quad y_i = \frac{1}{\Delta t} (x_{i+1} - (1 + \tilde{a} \Delta t)x_i) \]

with the following relationship

\[ X^T Y = X^T Z - \tilde{a} X^T X = 0 \] (2.23)

In the second level regression, we need to solve a least square problem for

\[ \frac{y_{i+1} - y_i}{\Delta t} = \tilde{q} x_i + \tilde{A} y_i + \eta_i, \quad 0 \leq i \leq \tilde{N} - 2 \]
where $\tilde{q}, \tilde{A}$ are the unknown parameters and $\{\eta_i\}$ is the residual. Here, the residual $\{\eta_i\}$ is assumed to have the form $\tilde{\sigma} \xi_i \sqrt{\Delta t}$, $\xi_i \sim N(0, 1)$. Let’s denote $H_i = \frac{y_i + 1 - y_i}{\Delta t}$, $\eta = \{\eta_i\}$, so we have the compact form for the second level least square problem

$$H = \tilde{q}X + \tilde{A}Y + \eta.$$ 

We get the parameter estimates

$$\tilde{q} = \frac{X^T H}{X^T X} = \sum_{i=0}^{N-2} (x_{i+2} - (2 + \tilde{a} \Delta t)x_{i+1} + (1 + \tilde{a} \Delta t)x_i)x_i$$

$$\Delta t \sum_{i=0}^{N-2} x_i^2 \sum_{i=0}^{N-2} x_i^2$$

$$= \frac{1}{\Delta t^2} \sum_{i=0}^{N-2} (x_{i+2} - (2 + \tilde{a} \Delta t)x_{i+1} + (1 + \tilde{a} \Delta t)x_i)^2$$

$$= \frac{\text{Var}(\eta) \Delta t}{\Delta t^2} \sum_{i=0}^{N-2} (x_{i+2} - (2 + \tilde{a} \Delta t)x_{i+1} + (1 + \tilde{a} \Delta t)x_i)^2$$

Similarly, the discrete version of the two-level quadratic regression model from (2.9), (2.10) becomes

$$\frac{x_{n+1} - x_n}{\Delta t} = (F + ax + bx^2) + r_n$$

$$\frac{r_{n+1} - r_n}{\Delta t} = (qx_n + Ar_n) + \frac{\sigma}{\sqrt{\Delta t}} \tilde{\xi}_n$$

Here, we apply linear regression procedure in each level. The residual forcing $r_i$ are determined by ordinary least-squares as described above in detail for the linear model (see [15]).

### 3. Finite time blow up, unstable statistical solutions, and pathological invariant measures for Ad hoc quadratic multi-level regression models.

With the background in section 2.1, it is natural question to ask if the ad hoc multi-level quadratic regression models from [15] introduced in section 2.2 exhibit similar properties as are known to occur both theoretically [21] and in time series sampled from low frequency data of comprehensive models [9, 24]; In these situations there is a smooth unique ergodic invariant measure with at most Gaussian tails [9, 21, 24] for the large fluctuations. Here in contrast, we demonstrate that the ad hoc quadratic multi-level regression models in (2.9), (2.10) have statistical solutions which exhibit pathological finite time blow-up, long-time instability of the mean under rather general hypothesis, and corresponding pathological invariant measures (if they exist) often with an infinite mean and at the least fatter than exponential tails in general. Numerical experiments confirm this pathological behavior of the two-level regression model as well as the skill of the physics based scalar regression model described in section 2.1.1, which has damping cubic terms.

Below, we study statistical solutions of the general multi-level regression models in (2.9)-(2.11), so it is natural to consider the Fokker-Planck equation for the
corresponding probability density, \( p(x, r, t) \) given by

\[
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}((F + ax + bx^2 + r^1)p) - \text{div}_r((qx + Ar)p) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial (r^N)^2}p
\]

\[ p(x, r, t)|_{t=0} = p_0(x, r) \]  

(3.1)

Without loss of generality, we assume that \( b > 0 \) in (3.1).

The average of any nice functional \( f(x, r) \) is defined by

\[
\langle f \rangle(t) = \int f(x, r)p(x, r, t)dxdr.
\]

(3.2)

Since the scalar ODE

\[
\frac{dx}{dt} = F + ax + bx^2 \quad x(0) = x_0
\]

(3.3)

has solutions with \( x_0 > \varepsilon_0 > 0 \) that blow-up in finite time provided \( b > 0 \), we anticipate a similar behavior for the statistical solutions in (3.1) of the multi-level quadratic regression provided we can show that the linear multi-level stochastic coupling terms are negligible in a suitable sense. To achieve such a comparison of (3.1) with (3.3) it is natural to compute equations satisfied by the mean statistic, \( \langle x \rangle(t), \langle r \rangle(t) \). It follows easily from (3.1) and (3.2) that these averages satisfy the statistical mean equations

\[
\frac{\partial}{\partial t} \langle x \rangle = F + a\langle x \rangle + b\langle x^2 \rangle + \langle r^1 \rangle
\]

\[
\frac{\partial}{\partial t} \langle r \rangle = q\langle x \rangle + A\langle r \rangle.
\]

(3.4)

The equations in (3.4) are not closed equations since the equation for \( \langle x \rangle(t) \) involves the higher moment, \( \langle x^2 \rangle(t) \); however, since \( \langle x^2 \rangle(t) \geq \langle x \rangle^2(t) \), we always have the differential inequality,

\[
\frac{\partial}{\partial t} \langle x \rangle \geq F + a\langle x \rangle + b\langle x^2 \rangle + \langle r^1 \rangle
\]

\[
\frac{\partial}{\partial t} \langle r \rangle = q\langle x \rangle + A\langle r \rangle.
\]

(3.5)

With these preliminaries we state two of the main results proved below.

**Theorem 3.1. Instability of the Mean for General Statistical Solutions with the forcing \( F \), suitable large:** Assume that \( A \) in (2.11) is invertible and let \( \vec{\alpha} \) be the unique solution of

\[
\vec{\alpha}A = -\vec{e}_1, \quad \vec{e}_1 = (1, 0, \ldots, 0).
\]

(3.6)

Consider the linear function, \( h(x, r) \) defined by

\[
h(x, r) = x + \vec{\alpha} \cdot \vec{r}
\]

(3.7)

then for all statistical solutions with nice initial data,

\[
\langle h \rangle(t) \geq \langle h \rangle(0) + \lambda t, \quad \lambda > 0
\]

(3.8)

provided that \( F \) is large enough so that

\[
\lambda \equiv F - \frac{(a + \vec{\alpha} \cdot \vec{q})^2}{b} > 0.
\]

(3.9)

This has the immediate consequence.
Corollary 3.1. With the assumptions in (3.6) and (3.9), if the equilibrium measure for (3.1) exists, the equilibrium mean, $\langle x \rangle_{eq}$, cannot be finite.

A natural requirement for the regression model in (2.9)-(2.11) is that the regression process for $\vec{r}$ alone is stable without coupling to $x$, i.e. the matrix $A$ in (2.11) satisfies the stability condition

$$\text{all eigenvalues of } A \text{ have negative real parts.} \quad (3.10)$$

With this hypothesis alone in (3.10), necessarily there are families of smooth statistical solutions which blow-up in finite time as stated next.

Theorem 3.2. Finite-time Blow-up of Statistical Solutions for stable multi-Level regression models for $\vec{r}$ alone: Assume that the stability condition in (3.10) for $r$ alone is satisfied. Assume that the mean of the initial data satisfies

$$\langle x \rangle(0) > M_1 > 0, \ M_1 \text{ sufficiently large}$$
$$|\langle r \rangle(0)| < M_2, \text{ fixed.} \quad (3.11)$$

Then the statistical solution from (3.1) blows up in finite time and in fact the mean satisfies

$$\langle x \rangle(t) \geq \varphi(t) \quad (3.12)$$

where $\varphi(t)$ is the solution of the ODE with finite time blow up for $b > 0$,

$$\varphi'(t) = F - C_1 + (a - C_2)\varphi + b\varphi^2$$
$$\varphi'(0) = \langle x \rangle(0) > M_1 > 0 \quad (3.13)$$

so the mean of $x(t)$, $\langle x \rangle(t)$, blows up in finite time. The constants $C_1$, $C_2$ in (3.13) depend on the stability matrix $A$ in (3.10) with $M_1$ depending accordingly on $C_1$, $C_2$ and $M_2$.

As the proof below shows, it is straightforward to track the explicit dependence of $C_1$, $C_2$ on the requirement in (3.10) and thus use the exact solution of (3.13) to obtain an explicit finite time blow up bound. Next we prove all the above results.

The proof of Theorem 3.1 is extremely simple. With (3.5) and (3.6), $\langle h \rangle(t) = \langle x \rangle + \vec{\alpha} \cdot \langle r \rangle$ satisfies the equation,

$$\frac{d}{dt}(h(t)) = F + (a + \vec{\alpha} \cdot \vec{q})\langle x \rangle + b\langle x^2 \rangle$$
$$\geq b \left( \langle x \rangle + \frac{a + \vec{\alpha} \cdot \vec{q}}{2b} \right)^2 + F - \frac{(a + \vec{\alpha} \cdot \vec{q})^2}{4b}$$
$$\geq \lambda \quad (3.14)$$

with $\lambda$ given in (3.9) and this completes the proof.

The proof of Corollary 3.1 is immediate since at equilibrium, the second equation in (3.4) implies

$$\langle r \rangle_{eq} = -A^{-1}q(x)_{eq}$$

and applying Theorem 3.1 with initial data, $p_{eq}(x, r)$, establishes that $\langle x \rangle_{eq} + \vec{\alpha} \cdot \langle r \rangle_{eq}$ is infinite so necessarily, $\langle x \rangle_{eq}$ is infinite.
The proof of Theorem 3.2 is slightly more complicated. First, the stability condition in (3.10) immediately guarantees the uniform bounds
\begin{align}
&\max_{0 \leq t \leq +\infty} |\bar{e}_1 \cdot e^{At}(0)| \leq C_1 \tag{3.15a} \\
&\max_{0 \leq t \leq +\infty} \int_0^t |\bar{e}_1 \cdot e^{A(t-s)}q| ds \leq C_2. \tag{3.15b}
\end{align}
Next, we use Duhamel’s formula to integrate the second equation in (3.5) and substitute the result into the differential inequality for \( \langle x \rangle(t) \) in (3.5) to obtain in general,
\begin{align}
\frac{\partial \langle x \rangle}{\partial t} &\geq (F + \bar{e}_1 e^{At}(0)) \\
&\quad + a(x) + b(x)^2 + \int_0^t \bar{e}_1 e^{A(t-s)}q\langle x \rangle(s)ds \tag{3.16}
\end{align}
Now, on any interval \([0, T]\) where \( \langle x \rangle(t) \) is increasing for \( 0 \leq t \leq T \), with (3.15) and (3.16), we have the local differential inequality,
\begin{align}
\frac{\partial \langle x \rangle}{\partial t} &\geq (F - C_1) + a\langle x \rangle + b(x)^2 - C_2 \max_{0 \leq s \leq t} \langle x(s) \rangle \\
&\geq (F - C_1) + (a - C_2)\langle x \rangle + b(x)^2. \tag{3.17}
\end{align}
Thus, on any interval \([0, T]\) where \( \langle x \rangle(t) \) is increasing for \( 0 \leq t \leq T \), \( \langle x \rangle(t) \geq \varphi(t) \) where \( \varphi(t) \) satisfies the explicit finite time blow-up differential equality in (3.13). Thus, it remains for us to show that a priori, \( \langle x \rangle(t) \) is increasing for the entire lifetime of the statistical solution provided that the initial data \( \langle x \rangle(0) \) is chosen suitably. With \( C_1, C_2 \) defined in (3.15) and \( b > 0 \), choose \( M_2 \) large enough so that
\begin{align}
by^2 + (a - C_2)y + F - C_1 \geq \varepsilon_0 > 0 \quad \text{for} \quad y \geq M_2. \tag{3.18}
\end{align}
and consider initial data with \( \langle x \rangle(0) \geq M_2 \). Then (3.17),(3.18) guarantee that initially, \( \frac{\partial \langle x \rangle}{\partial t}|_{t=0} \geq \varepsilon_0 > 0 \) and a priori \( \frac{\partial \langle x \rangle}{\partial t} \geq \varepsilon_0 > 0 \) for all later time in the life span, in other words, the set \( y \geq M_2 \) in (3.18) is invariant for the statistical mean under the dynamics and automatically solutions with mean in this set have a monotone increasing mean in time. This concludes the proof of Theorem (3.2).

Note that in the proofs of Theorem (3.1), (3.2), we have made the tacit assumptions that the mean of local statistical solutions of (3.1) are continuously differentiable in time with finite second moment, \( \langle x^2 \rangle(t) \); of course if these mild tacit conditions are already violated by the regression strategy, it already has unphysical pathological behavior. Theorem (3.1) and (3.2) show unphysical pathological behavior in the multi-level quadratic regression strategy but impose additional requirements like (3.9), (3.10) on the values of the regression parameters. The final theoretical result stated next does not restrict the structure of the regression coefficients in any fashion yet leads to pathological dynamics.

**Theorem 3.3. Instability of Statistical Solutions for General Ad hoc Quadratic Multi-level Regression:** Assume the mild conditions that there is non-zero random noise in (2.9), i.e. \( \sigma \neq 0 \) and that \( A \) in (2.11) is invertible as well as \( b > 0 \). Choose \( \tilde{\alpha} \) as in (3.6) and consider the functional
\begin{align}
h(x, r) = e^{\tilde{\alpha}(x + \tilde{\sigma}r)}. \tag{3.19}
\end{align}
Then for $\alpha$ large enough so that
\[ -\alpha(F - \frac{(a + \vec{\alpha} \cdot \vec{q})^2}{4b}) + \frac{1}{2} \alpha^2 \alpha_N^2 \sigma^2 = \lambda > 0 \] (3.20)
there is general statistical instability,
\[ \langle h(t) \rangle \geq e^{\lambda t} \langle h(0) \rangle. \] (3.21)

In particular, if the equilibrium measure, $p_{eq}$, exists necessarily $\langle h \rangle_{eq}$ is infinite as a consequence of (3.20) since $h$ is positive.

Note that the equilibrium distribution of the ad hoc quadratic multi-level regression strategy necessarily has pathological fat tails in contrast to the physical behavior of actual low frequency dynamics [9, 24, 8, 21] as discussed earlier.

Here is the proof of Theorem 3.3. We use the stochastic equations in (2.9). By Ito’s lemma, we have
\[ dh(x, r) = de^{\alpha(x+\vec{\alpha} \cdot \vec{r})} \]
\[ = \alpha h(F + ax + bx^2 + r^2)dt \]
\[ + \alpha h((\vec{\alpha} \cdot A r + (\vec{\alpha} \cdot \vec{q})x)dt \]
\[ + \frac{1}{2} \alpha^2 \alpha_N^2 \sigma^2 hdt + \alpha \alpha_N \sigma h dW. \]
\[ = \alpha h(F + (a + \vec{\alpha} \cdot \vec{q})x + bx^2)dt \]
\[ + \frac{1}{2} \alpha^2 \alpha_N^2 \sigma^2 hdt + \alpha \alpha_N \sigma h dW. \] (3.22)

where (3.6) has been utilized in the last equality.

Simplifying (3.22) further, we get
\[ dh(x, r) = \alpha h(F + (a + \vec{\alpha} \cdot \vec{q})x + bx^2)dt \]
\[ + \frac{1}{2} \alpha^2 \alpha_N^2 \sigma^2 hdt + \alpha \alpha_N \sigma h dW. \]
\[ = \alpha h \left( x + \frac{a + \vec{\alpha} \cdot \vec{q}}{2b} \right)^2 dt \]
\[ + \alpha h \left( F - \frac{(a + \vec{\alpha} \cdot \vec{q})^2}{4b} \right) dt + \frac{1}{2} \alpha^2 \alpha_N^2 \sigma^2 hdt \]
\[ + \alpha \alpha_N \sigma h dW. \] (3.23)

Now choose $\alpha > 0$ large enough such that
\[ \alpha \left( F - \frac{(a + \vec{\alpha} \cdot \vec{q})^2}{4b} \right) + \frac{1}{2} \alpha^2 \alpha_N^2 \sigma^2 = \lambda > 0. \]

and note that the form of the matrix $A$ in (2.11) guarantees automatically, $\alpha_N > 0$.

We get the equation for $\langle h(x, r) \rangle$ by integrating (3.23) yielding
\[ dh(x, r) \geq \lambda h(x, r)dt + \alpha \alpha_N \sigma h dW \Rightarrow \]
\[ d\langle h(x, r) \rangle \geq \lambda \langle h(x, r) \rangle dt \]
So, we get that $\langle h(x, r) \rangle$ grows exponentially as in (3.21) to finish the proof.
3.1. Pathology in Ad hoc quadratic multi-level regression: A case study. Here we illustrate the pathological behavior in Theorem 3.2 as it arises naturally in an instructive case study of a time series, \( x(t) \), associated with a physics motivated problem [8, 14, 22]. In [8], the one-dimensional normal form in (2.3) was fit to low frequency data from a comprehensive atmospheric model involving the leading order principal component, PC-1, which has features of the Arctic Oscillation of the zonal winds. The resulting stochastic model in (2.3) is nearly a linear model with subtle non-Gaussian features [14, 22] reflected by the coefficient values for (2.3)

\[
\tilde{F} = -0.005, \quad a = -0.018, \quad b = 0.006 \\
c = 0.003, \quad \sigma = 0.226 \quad (3.24)
\]

with \( A = B = 0 \). The resulting stochastic equation generated by this model has a decorrelation time of \( O(100) \) which serves as an important large scale reference time. A time series, \( x(t) \), was created by solving (2.3) with Euler’s method with a time step, \( \Delta t = 0.005 \), and an initial time series length \( \tilde{N} = 2 \times 10^7 \), which is roughly \( 10^3 \) decorrelation times. Both the ad hoc quadratic two-level regression strategy from [15] described in (2.20) of section 2.3.1 and the physics based single level regression strategy of section 2.1.1 with cubic damping terms were applied to this data set.

The values of the coefficients of the two-level regression model together with their 95% confidence intervals for \( \tilde{N} = 2 \times 10^7 \) are given by

\[
F = -0.004 \pm 0.00185, \quad a = 0.0265 \pm 0.0015 \\
b = 0.004 \pm 0.001, \quad q = -5.1 \pm 0.25 \\
A = 200.0 \pm 0.1, \quad \sigma = 45.119 \quad (3.25)
\]

The time series length, \( \tilde{N} \), was increased systematically from \( \tilde{N} = 10^9 \) to \( \tilde{N} = 7 \times 10^9 \) and Figure 1 shows the corresponding statistical convergence of the parameters, \( F, a \) and \( b \); we observe that there is statistical significance of the quadratic interaction parameter \( b > 0 \) in Figure 1. Note that the regression parameters in (3.25) are consistent with the hypothesis of Theorem 3.2 since (3.10) is satisfied with \( b > 0 \); thus, we expect finite time blow-up of some statistical solutions.

Figure 2 is a plot of statistical convergence of the coefficients of the physics based single level regression strategy of section 2.1.1 with cubic damping terms and shows the statistical significance of the coefficients as the time series length \( \tilde{N} \) increases. For the time series of shorter length, \( \tilde{N} = 2 \times 10^7 \), we have the estimates \( \tilde{F} = -0.0045, \ a = -0.0185, \ b = 0.00615, \ c = 0.0032, \ \sigma = 0.226 \) which crudely approximate those in (3.24). Thus, for this physics based single level regression model, there is a statistically significant positive quadratic term but also statistically significant cubic damping.

Individual trajectories of the ad hoc two-level quadratic regression model and the single level regression strategy are depicted in Figure 3 starting from the same initial value, \( x_0 = 10 \), at the top and \( x_0 = 12 \) at the bottom of the figure. The trajectories of the two level regression model are stable for \( x_0 = 10 \) but blow up in finite time for \( x_0 = 12 \) as predicted by Theorem 3.2. On the other hand, the trajectories of the one level regression model with cubic damping are always stable; furthermore, as shown in Figure 4, the trajectory and equilibrium probability density function of the original model are well-captured by those from the one-level regression model. These numerical results are robust and apply to all five of the examples described in
FUNDAMENTAL LIMITATIONS OF MULTI-LEVEL REGRESSION MODELS

Figure 1. Statistical convergence of parameter estimations for the two level quadratic regression model (2.25) from centered and normalized time series, PC-1, with the integration time step $\Delta t = 0.005$. The circles are parameter estimations and the stars are ends of 95% confidence interval. We observe statistical significance of the quadratic nonlinearity $b$.

detail in [14] and [22] with explicit finite time blow up of the quadratic multi-level regression model although further detailed discussion is omitted here.

4. Discrete sampling model errors in linear multi-level regression strategies. In section 2.3, we introduced elementary test models for two level regression strategies that exhibit transparent non-uniqueness with the same marginal equilibrium dynamics. We also introduced the algorithm from [15] for two-level regression in section 2.3.1. The question we address here is the following one: Can the two level regression algorithm of section 2.3.1 recover the original marginal equilibrium
Figure 2. Statistical convergence of parameter estimations for the physical based single level regression model (2.5) from centered and normalized time series, PC-1, with the integration time step $\Delta t = 0.005$. The crosses are the exact value, the circles are the parameter estimations and the stars are ends of 95% confidence interval. We observe statistical significance of the quadratic term $b$ as well as statistical significant cubic damping.

Dynamics for an infinitely long discrete time series and small sampling time $\Delta t \ll 1$? Surprisingly, the answer to this question is no as shown by the following.

**Theorem 4.1.** Assume the Gaussian random process $x(t)$ is the marginal dynamics of the stable two-level linear hidden model

$$d\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & 1 \\ q & A \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma dW \end{pmatrix}.$$  

(4.1)
Assume we have infinitely many observations \( \{ x_i = x(t_i), i = 0, 1, 2, ..., t_i = i\Delta t \} \). Then, the two-level regression model for \( x(t) \) for \( \Delta t \ll 1 \) is given by

\[
\begin{pmatrix}
    x_{i+1} - x_i \\
    y_{i+1} - y_i
\end{pmatrix} = \begin{pmatrix}
    0 & 1 \\
    q - aA & \frac{1}{3}(a + A)
\end{pmatrix} \begin{pmatrix}
    x_i \\
    y_i
\end{pmatrix} \Delta t + \begin{pmatrix}
    0 \\
    \sqrt{\frac{2}{3}} \sigma \xi \sqrt{\Delta t}
\end{pmatrix}. \tag{4.2}
\]

Thus, the regression model can capture the variance of the marginal dynamics of \( x(t) \)

\[
C(0) = \frac{\sigma^2}{(a + A)(q - aA)}.
\]
Figure 4. The trajectory and equilibrium probability density function of the original model are well-captured by those from the physical based single level regression model (2.5). Here, we choose parameter set PC-1, length of the time series $N = 2 \times 10^7$ and the integration time step $\Delta t = 0.005$.

However, it can not reproduce the autocovariance of $x(t)$, since the eigenvalues of the stability coefficients of the regression model are different from those of the true model as required in Proposition 2.1.

When we use the multilevel regression procedure, we are making model errors. There are intrinsic errors when we estimate the variance of the local change of the
trjectories
\[ dx_i = x_{i+1} - x_i, \quad dy_i = y_{i+1} - y_i = \frac{x_{i+1} - 2y_{i+1} + x_i}{\Delta t}. \]

A theoretical way to correct these model errors is not to use the time series of the hidden model directly but instead to apply Euler’s method to the hidden model and generate a new auxiliary discrete time series which is not the time series of the perfect model. This procedure is confirmed by the following.

**Theorem 4.2.** Assume the Gaussian random process \( x_i \) is the marginal dynamics of Euler scheme for the two-level linear hidden model

\[
\begin{pmatrix}
    x_{i+1} - x_i \\
    y_{i+1} - y_i
\end{pmatrix}
= \begin{pmatrix}
    a & 1 \\
    q & A
\end{pmatrix}
\begin{pmatrix}
    x_i \\
    y_i
\end{pmatrix}
\Delta t
+ \begin{pmatrix}
    0 \\
    \sigma \varepsilon_i \sqrt{\Delta t}
\end{pmatrix}
\tag{4.3}
\]

where \( \varepsilon_i \) is a Gaussian random number \( N(0,1) \). Assume we have infinite many observations \( \{x_i = x(t_i), t = 0,1,2,..., t_i = i\Delta t\} \). Then, as \( \Delta t \to 0 \) the two-level regression model for \( x(t) \) is given by

\[
\begin{pmatrix}
    x_{i+1} - x_i \\
    y_{i+1} - y_i
\end{pmatrix}
= \begin{pmatrix}
    0 & 1 \\
    q - aA & a + A
\end{pmatrix}
\begin{pmatrix}
    x_i \\
    y_i
\end{pmatrix}
\Delta t
+ \begin{pmatrix}
    0 \\
    \sigma \zeta_i \sqrt{\Delta t}
\end{pmatrix}.
\tag{4.4}
\]

Therefore, the regression model can reproduce the marginal dynamics of the Euler scheme for the two level linear hidden model, since according to Proposition 2.1 (4.4) has the same eigenvalues as the original hidden model. However, clearly the model in (4.4) has different coefficients from the hidden model.

The proofs of Theorem 4.1 and 4.2 are tedious exact calculations using ergodicity and time averaging and are relegated to the appendix. While Theorem 4.2 is interesting in identifying the model error in two-level linear regression strategies, this result is not very useful from a practical view point since we do not know the actual hidden model generating the dynamics for \( x(t) \) very well. So the auxiliary time series using Euler’s method is not available. The predictability properties for ensembles of \( x(t) \) which are not the marginal equilibrium dynamics can be very different for the two level regression model compared with the original model [22]; this is an important comment to keep in mind for [16, 17] and is readily quantified in the present simplified setting but lack of space prevents a detailed discussion here.

### 4.1. Simple numerical experiments for two level linear regression.

In Theorem 4.1 and 4.2, we have assumed an infinitely long time discrete series for \( x(t) \) and then taken the limit as \( \Delta t \to 0 \) (see the appendix for the proofs which do this explicitly). In practice we have a finite length time series, \( N \), and a finite sampling interval, \( \Delta t \), so it’s interesting to see if the behavior predicted by Theorem 4.1,4.2 actually arises in a numerical example. This is done next.

In the numerical example presented below, the hidden model in (4.1) has coefficients \( (a,q,A,\sigma^2) = (-2,1,-1,0.25) \) so the eigenvalues of the perfect stability matrix are

\[ \lambda_1 = -0.382, \quad \lambda_2 = -2.618. \tag{4.5} \]

The eigenvalues of the stability matrix for the two level regression model predicted by Theorem 4.1 by using the true dynamics of the hidden model discretely sampled in time are

\[ \lambda_1 = \lambda_2 = -1. \tag{4.6} \]

Below we will produce auxiliary time series for the dynamics of \( x(t) \) from two different procedures:
1. True dynamics of the 2D linear model

\[
\begin{pmatrix}
    x_{i+1} \\
y_{i+1}
\end{pmatrix} = Q(\Delta t) \begin{pmatrix}
    x_i \\
y_i
\end{pmatrix} + Q(\Delta t) \int_0^{\Delta t} Q(-s) \begin{pmatrix}
    0 & 0 \\
0 & \sigma
\end{pmatrix} dW
\]

(4.7a)

where

\[
Q(t) = e^{Lt} = \begin{pmatrix}
    Q_{11}(t) & Q_{12}(t) \\
Q_{21}(t) & Q_{22}(t)
\end{pmatrix}, \quad L = \begin{pmatrix}
a & 1 \\
q & A
\end{pmatrix};
\]

(4.7b)

2. Numerical integration by the Euler method

\[
\begin{pmatrix}
    x_{i+1} \\
y_{i+1}
\end{pmatrix} = (L\Delta t + I) \begin{pmatrix}
    x_i \\
y_i
\end{pmatrix} + \begin{pmatrix}
    0 \\
\sigma \sqrt{\Delta t} \varepsilon_i
\end{pmatrix}
\]

(4.7c)

where \( \varepsilon_i \) is an i.i.d Gaussian random process \( N(0, 1) \).

We will vary the observation time for the time series \( x(t), \Delta T = m\Delta t \) so that with \( m = 1 \) and the method 1 in (4.7) we have the true dynamics for Theorem 4.1 while with \( m = 1 \) and method 2 in (4.7) we have the Euler Dynamics for Theorem 4.2. By increasing the observation time, \( \Delta T \), for the time series, we briefly explore the role of subsampling in this context [19, 20]; this slight abuse of notation using \( \Delta T \) should not confuse the reader.

In Figure 5 we present the computed autocovariance function from the Euler scheme and perfect model two level regression strategies compared with the autocovariance for the marginal equilibrium dynamics of the perfect hidden model. In the upper panel of Figure 5 we have the small time step, \( \Delta t = 0.0005 \) and \( \Delta T = \Delta t \) with the long time series interval \( T = 5000 \) and the results confirm the prediction of Theorem 4.1, 4.2 very nicely. In the lower panel of Figure 5, we have a curious result; we used the large time step \( \Delta t = 0.7 \) which is unstable for Euler’s method for long time integration and a long time series interval \( T = 70000 \) with \( \Delta T = \Delta t \) and \( N = \frac{T}{\Delta T} = 10^5 \). In this case, all the parameter estimates have large departure from the predicted value for the Euler method but nevertheless, a very accurate marginal covariance for \( x(t) \) has been calculated and it is more accurate than using the perfect dynamics as the time series input!

In Figure 6, the integration time step is fixed at \( \Delta t = 0.005 \) and the length of the time series, \( N = \frac{T}{\Delta T} \), and the subsampling window, \( \Delta T = m\Delta t \), \( m = 1, 10, 100, 1000 \) are varied. The first column of Figure 6 shows that the Euler method recovers the exact covariance with model errors with the perfect hidden model like in Theorem 4.1, 4.2 even with undersampling \( m = 10, 100 \) for a long time series. The second and third columns show comparable behavior of the two methods with shorter time series and sampling with \( m = 1, 10, 100 \) and significant model errors in each case. The fourth column shows the complete inaccuracy of both methods with a very short time series while the bottom row demonstrates the superiority of generating time series through the perfect model for strongly undersampled situations with \( m = 1000 \).

5. Concluding discussion. The authors have demonstrated rigorously in section 3 of this paper that recently proposed [15, 16, 17] ad hoc quadratic multi-level regression strategies can have unphysical artifacts such as finite time blowup and instability of statistical solutions with corresponding pathological behavior of the attractor of the regression model compared to the physical time series actually being modeled [8, 9, 21, 24]. Section 3 also contains a series of numerical experiments confirming this predicted pathological behavior. It is also demonstrated in section
Figure 5. Compare AutoCovariance and Variance to the exact values. The upper panel: the observation time lag $\Delta T = \Delta t = 0.0005$, the interval length $T = 5000$. The lower panel: the observation time lag $\Delta T = \Delta t = 0.7$, the interval length $T = 70000$. The dash line is the true AutoCovariance, the solid line is from the time series generated by Euler Scheme, the solid line with dots is from the true dynamics.

3 that a physics based single level regression model described in section 2.1 with cubic nonlinear damping has significant skill as a regression model [8, 21]. In section 4, the surprising effects of discrete sampling in creating model error [22] in simple two level regression models were elucidated through both rigorous theorems and simple numerical experiments.

One important research direction is to find stable multi-level nonlinear regression models that can incorporate memory effects and nonlinearity in time in a simple fashion. One possibility is to use the three equation models as in (2.1) as simple multi-level regression models for data arising from large dimensional dynamical systems with a similar physical structure. Another interesting problem is to generalize the results of this paper to the case when $x(t)$ is a vector time-series.
Figure 6. Compare AutoCovariance and Variance to the exact values with fixed time step $\Delta t = 0.005$, varied length of the time series, $N = \frac{T}{\Delta T}$, and varied subsampling window, $\Delta T = m\Delta t$, $m = 1, 10, 100, 1000$. The dash line is the true AutoCovariance, the solid line is from the time series generated by Euler Scheme, the solid line with dots is from the true dynamics.

Acknowledgments: This research of Andrew J. Majda is partially supported by grants ONR-No0014-05-1-0164 and NSF-DMS-0456713. Yuan Yuan was partially supported as a graduate research assistant under the last grant.

Appendix A. The Proof of Theorem 4.1 and 4.2.

A.1. Proof of Theorem 4.1. Gaussian random process $x(t)$ is the marginal dynamics of the two-level linear hidden model

\[
\frac{d}{dt} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & 1 \\ q & A \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma dW \end{pmatrix}.
\]

Assume we have infinite many observations \( \{ x_i = x(t_i), i = 0, 1, 2, ..., t_i = i\Delta t \} \).
The two-level regression model has the form

\[
\begin{pmatrix}
  x_{i+1} - x_i \\
y_{i+1} - y_i
\end{pmatrix} = \begin{pmatrix}
  \tilde{a} & 1 \\
\tilde{q} & \tilde{A}
\end{pmatrix} \begin{pmatrix}
x_i \\
y_i
\end{pmatrix} \Delta t + \begin{pmatrix}
  0 \\
\tilde{\sigma} \xi_i \sqrt{\Delta t}
\end{pmatrix}.
\]  

(A.1)

Here, we will apply the two-level regression procedure we discussed in 2.3.1 to the true marginal dynamics of the two-level linear hidden model, \( x(t) \).

**Step 1:** Solve the true dynamics, we get that

\[
\begin{pmatrix}
x_{i+1} \\
y_{i+1}
\end{pmatrix} = Q(\Delta t) \begin{pmatrix}
x_i \\
y_i
\end{pmatrix} + Q(\Delta t) \begin{pmatrix}
\varepsilon_1^i \\
\varepsilon_2^i
\end{pmatrix},
\]  

(A.2)

where

\[
Q(t) = e^{Lt} = \begin{pmatrix}
Q_{11}(t) & Q_{12}(t) \\
Q_{21}(t) & Q_{22}(t)
\end{pmatrix}, \quad L = \begin{pmatrix}
a & 1 \\
q & A
\end{pmatrix},
\]

\[
\begin{pmatrix}
\varepsilon_1^i \\
\varepsilon_2^i
\end{pmatrix} \sim N(0_{2\times1}, \Sigma(\Delta t)).
\]  

(A.3)

We have the approximations

\[
Q(\Delta t) = I_{2\times2} + L \Delta t + O(\Delta t^2); \quad \Sigma(\Delta t) = \int_0^{\Delta t} Q(-s) \begin{pmatrix}
0 & 0 \\
0 & \sigma^2
\end{pmatrix} Q'(s) ds.
\]  

(A.4)

**Step 2:** Statistics of the true dynamics.

By section 2.3, we’ve learned the equilibrium mean is 0 and the covariance matrix \( \Sigma \)

\[
\Sigma = \begin{pmatrix}
1 & -a \\
-a & aA - q + a^2
\end{pmatrix}, \quad -2(a + A)(aA - q) > 0.
\]  

(A.6)

By ergodicity, we have

\[
\sum_{i=0}^{N-1} f(x_i, y_i, \varepsilon_i) \xrightarrow{N \to \infty} (f(x, y, \varepsilon)).
\]

So, we have the following relationships

\[
\sum_{i=0}^{N-1} x_i y_i \xrightarrow{N \to \infty} \langle xy \rangle = -a,
\]

\[
\sum_{i=0}^{N-1} x_i^2 \xrightarrow{N \to \infty} \langle x^2 \rangle = -aA - q + a^2
\]

\[
\sum_{i=0}^{N-1} g(x_i, y_i) \xrightarrow{N \to \infty} \langle \varepsilon g(x, y) \rangle = 0
\]  

(A.7)

where \( \varepsilon(t) \) can be any stationary and ergodic random process with mean 0 and independent with random process \( x(t) \) and \( y(t) \).

**Step 3:** Calculate the parameter estimators \( \tilde{a}, \tilde{q} \) under the true dynamics.

By the two-level regression procedure as discussed in Section 2.3.1, we’ve learned the expressions of the parameter estimates in (2.22, 2.24). So, we need the following
expressions

\[ x_{i+1} - x_i = (Q_{11}(\Delta t) - 1)x_i + Q_{12}(\Delta t)y_i + Q_{11}(\Delta t)\varepsilon_i^1 + Q_{12}(\Delta t)\varepsilon_i^2 \]
\[ x_{i+2} - 2x_{i+1} + x_i = [(Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)Q_{21}(\Delta t)]x_i \]
\[ + (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t)y_i \]
\[ + (Q_{11}(\Delta t)^2 - 2Q_{11}(\Delta t) + Q_{12}(\Delta t)Q_{21}(\Delta t))\varepsilon_i^1 \]
\[ + (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t)\varepsilon_i^2 \]
\[ + Q_{11}(\Delta t)\varepsilon_{i+1}^1 + Q_{12}(\Delta t)e_i^{2i+1} \]  

(A.8)

Let’s plug the true dynamics (A.8) into (2.22, 2.24) and consider the limiting behavior, take \( N \to \infty \) first, then take \( \Delta t \to 0 \).

\[ \bar{a} = \frac{X^TZ}{X^TX} = \frac{\sum(x_{i+1} - x_i)x_i}{\Delta t \sum x_i^2} \]
\[ = \frac{(Q_{11}(\Delta t) - 1)\sum x_i^2 + Q_{12}(\Delta t)\sum x_i y_i + Q_{11}(\Delta t)\sum x_i \varepsilon_i^1 + Q_{12}(\Delta t)\sum x_i \varepsilon_i^2}{\Delta t \sum x_i^2} \]
\[ \xrightarrow{N \to \infty} \frac{(Q_{11}(\Delta t) - 1)}{\Delta t} + \frac{Q_{12}(\Delta t)}{\Delta t} \sum x_i y_i + \frac{Q_{11}(\Delta t)}{\Delta t} \sum x_i \varepsilon_i^1 + \frac{Q_{12}(\Delta t)}{\Delta t} \sum x_i \varepsilon_i^2 \]
\[ \xrightarrow{\Delta t \to 0} 0 \]

(A.9)

\[ \bar{q} = \frac{X^TH}{X^TX} = \frac{\sum(x_{i+2} - 2x_{i+1} + x_i)x_i}{\Delta t^2 \sum x_i^2} \]
\[ = \frac{\sum(x_{i+2} - 2x_{i+1} + x_i)x_i}{\Delta t^2 \sum x_i^2} - \bar{a}^2 \]
\[ = [(Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)Q_{21}(\Delta t)] \frac{1}{\Delta t^2} \]
\[ + (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t) \frac{1}{\Delta t^2} \sum x_i y_i \]
\[ + (Q_{11}(\Delta t)^2 - 2Q_{11}(\Delta t) + Q_{12}(\Delta t)Q_{21}(\Delta t)) \frac{1}{\Delta t^2} \sum x_i \varepsilon_i^1 \]
\[ + (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t) \frac{1}{\Delta t^2} \sum x_i \varepsilon_i^2 \]
\[ + Q_{11}(\Delta t) \sum x_i \varepsilon_{i+1}^1 + Q_{12}(\Delta t) \sum x_i \varepsilon_{i+1}^2 - \bar{a}^2 \]
\[
\lim_{N \to \infty} \left[ (Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)Q_{21}(\Delta t) \right] \frac{1}{\Delta t^2} \\
+ (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t) \frac{1}{\Delta t^2} (-a) \\
- \frac{(Q_{11}(\Delta t) - 1)}{\Delta t} + \frac{Q_{12}(\Delta t)}{\Delta t} (-a)^2 \\
\lim_{\Delta t \to 0} q - aA
\] (A.10)

Step 4: Calculate the parameter estimators \( \tilde{A} \) under the true dynamics.

\[
\tilde{A} = \frac{Y^T H}{Y^T Y} = \frac{\sum (x_{i+1} - (1 + \Delta \hat{a})x_i)(x_{i+2} - (2 + \hat{a}\Delta t)x_{i+1} + (1 + \Delta \hat{a})x_i)}{\Delta t \sum (x_{i+1} - (1 + \Delta \hat{a})x_i)^2} \\
= \frac{\sum (x_{i+1} - x_i)(x_{i+2} - 2x_{i+1} + x_i) - \Delta \hat{a} \sum (x_{i+1} - x_i)^2 - \Delta t^2 \hat{a} \hat{b} \sum x_i^2}{\Delta t \sum (x_{i+1} - x_i)^2 - \Delta t^2 \hat{a}^2 \sum x_i^2}
\] (A.11)

Let’s discuss each term.

By the expansion of the covariance matrix (A.5) and the expansion of the matrix \( Q(\Delta t) \) (A.4), we have that

\[
\frac{1}{N} \sum_{i=0}^{N-1} [Q_{11}(\Delta t)\varepsilon_i^1 + Q_{12}(\Delta t)\varepsilon_i^2]^2 \\
\lim_{N \to \infty} \begin{pmatrix} Q_{11}(\Delta t) & Q_{12}(\Delta t) \end{pmatrix} \text{Cov} \begin{pmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \end{pmatrix} \begin{pmatrix} Q_{11}(\Delta t) \\ Q_{12}(\Delta t) \end{pmatrix} \\
= \frac{1}{3}\sigma^2 \Delta t^3 + O(\Delta t^4)
\] (A.12)
\[
\frac{1}{N} \sum_{i=0}^{N-1} (x_{i+1} - x_i)^2 = \frac{1}{N} \sum_{i=0}^{N-1} [(Q_{11}(\Delta t) - 1)x_i + Q_{12}(\Delta t)y_i]^2
\]
\[+ \frac{1}{N} \sum_{i=0}^{N-1} [Q_{11}(\Delta t)e^1_i + Q_{12}(\Delta t)e^2_i]^2 \]
\[+ (\text{Terms of independent interactions (equal to zero for } N \text{ large)})\]
\[= (Q_{11}(\Delta t) - 1)^2 \frac{1}{N} \sum_{i=0}^{N-1} x_i^2 + Q_{12}(\Delta t)^2 \frac{1}{N} \sum_{i=0}^{N-1} y_i^2 \]
\[+ 2(Q_{11}(\Delta t) - 1)Q_{12}(\Delta t) \frac{1}{N} \sum_{i=0}^{N-1} x_i y_i \]
\[+ \frac{1}{N} \sum_{i=0}^{N-1} [Q_{11}(\Delta t)e^1_i + Q_{12}(\Delta t)e^2_i]^2 \]
\[+ (\text{Terms of independent interactions (equal to zero for } N \text{ large)})\]
\[\frac{N \to \infty}{\text{by (A.7)}} \Rightarrow ((Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)^2(aA - q + a^2)) \]
\[+ 2(Q_{11}(\Delta t) - 1)Q_{12}(\Delta t)(-a)) \frac{\sigma^2}{2(a + A)(aA - q)} \]
\[+ \frac{1}{3} \sigma^2 \Delta t^3 + O(\Delta t^4)\]
\[= \frac{\sigma^2 \Delta t^2}{2(a + A)} + O(\Delta t^3) \quad (A.14)\]

\[
\frac{1}{N} \sum_{i=0}^{N-1} (x_{i+2} - 2x_{i+1} + x_i)(x_{i+1} - x_i)
\]
\[= \frac{1}{N} \sum_{i=0}^{N-1} [(Q_{11}(\Delta t) - 1)x_i + Q_{12}(\Delta t)y_i + Q_{11}(\Delta t)e^1_i + Q_{12}(\Delta t)e^2_i]
\cdot \{(Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)Q_{21}(\Delta t)]x_i
+ (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t)y_i
+ (Q_{11}(\Delta t)^2 - 2Q_{11}(\Delta t) + Q_{12}(\Delta t)Q_{21}(\Delta t)e^1_i
+ (Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t)e^2_i
+ Q_{11}(\Delta t)e^1_i + Q_{12}(\Delta t)e^2_{i+1}\}
\]
\[= (Q_{11}(\Delta t) - 1)[(Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)Q_{21}(\Delta t)] \frac{1}{N} \sum_{i=0}^{N-1} x_i^2 \]
\[+ Q_{12}(\Delta t)^2(Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2) \frac{1}{N} \sum_{i=0}^{N-1} y_i^2 \]
\[+ ((Q_{11}(\Delta t) - 1)(Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2)Q_{12}(\Delta t)\]
Let’s expend the variable $\eta$.

**Step 5:** Calculate the parameter estimators $\tilde{\theta}$ under the true dynamics.

$$\hat{\sigma}^2 = \text{Var}(\eta) \Delta t$$

Let’s expend the variable $\eta$ first.

$$\eta_i = \frac{1}{\Delta t^2} (x_{i+2} - (2 + \tilde{a} \Delta t + \Delta t \tilde{\theta})x_{i+1} + [((1 + \tilde{a} \Delta t)(1 + \tilde{A} \Delta t) - \Delta t^2 \tilde{q}]x_i$$

$$= \frac{1}{\Delta t^2} (x_{i+2} - 2x_{i+1} + x_i) - \frac{1}{\Delta t} (\tilde{a} + \tilde{A})(x_{i+1} - x_i) + (\tilde{a} \tilde{\theta} - \tilde{q})x_i$$

$$= A_1(\Delta t)x_i + A_2(\Delta t)y_i$$

$$+ A_3(\Delta t)\varepsilon_{i+1}^1 + A_4(\Delta t)\varepsilon_{i+1}^2$$

$$+ A_5(\Delta t)\varepsilon_{i+1}^1 + A_6(\Delta t)\varepsilon_{i+1}^2$$
where the coefficients satisfy

\[
A_1(\Delta t) = \frac{1}{\Delta t^2}[(Q_{11}(\Delta t) - 1)^2 + Q_{12}(\Delta t)Q_{21}(\Delta t)] - \frac{1}{\Delta t}(\bar{a} + \bar{A})(Q_{11}(\Delta t) - 1) + (\bar{a}\bar{A} - \bar{q}) = (a + A)a + O(\Delta t)
\]

\[
A_2(\Delta t) = \frac{1}{\Delta t^2}(Q_{11}(\Delta t) + Q_{22}(\Delta t) - 2Q_{12}(\Delta t) - \frac{1}{\Delta t}(\bar{a} + \tilde{A})Q_{12}(\Delta t) = \frac{a + A}{3} + O(\Delta t)
\]

\[
A_3(\Delta t) = \frac{1}{\Delta t^2}(Q_{11}(\Delta t)^2 - 2Q_{11}(\Delta t) + Q_{12}(\Delta t)Q_{21}(\Delta t)) - \frac{1}{\Delta t}(\bar{a} + \tilde{A})Q_{11}(\Delta t) = -\frac{1}{\Delta t^2} + \frac{-2(a + A)}{3\Delta t} + O(1)
\]

\[
A_4(\Delta t) = A_2(\Delta t)
\]

\[
A_5(\Delta t) = \frac{1}{\Delta t^2}Q_{11}(\Delta t) = \frac{1}{\Delta t^2} + \frac{a}{\Delta t} + O(1)
\]

\[
A_6(\Delta t) = \frac{1}{\Delta t^2}Q_{12}(\Delta t) = \frac{1}{\Delta t} + O(1)
\]

Therefore, we can estimate \( \tilde{\sigma} \) by three parts by independence between \( x_i, y_i \) and \( \varepsilon_i, \varepsilon_{i+1} \)

\[
\tilde{\sigma}^2 = \begin{pmatrix} A_1(\Delta t) & A_2(\Delta t) \end{pmatrix} \text{Cov} \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} A_1(\Delta t) \\ A_2(\Delta t) \end{pmatrix} \Delta t + \begin{pmatrix} A_3(\Delta t) & A_4(\Delta t) \end{pmatrix} \text{Cov} \begin{pmatrix} \varepsilon_i \\ \varepsilon_{i+1} \end{pmatrix} \begin{pmatrix} A_3(\Delta t) \\ A_4(\Delta t) \end{pmatrix} \Delta t + \begin{pmatrix} A_5(\Delta t) & A_6(\Delta t) \end{pmatrix} \text{Cov} \begin{pmatrix} \varepsilon_i \\ \varepsilon_{i+1} \end{pmatrix} \begin{pmatrix} A_5(\Delta t) \\ A_6(\Delta t) \end{pmatrix} \Delta t = O(\Delta t) + \frac{1}{3}\sigma^2 + O(\Delta t) + \frac{1}{3}\sigma^2 + O(\Delta t) = \frac{2}{3}\sigma^2 + O(\Delta t)
\]

A.2. Proof of Theorem 4.2. Step 1: Gaussian random process \( x_i \) is the marginal dynamics of Euler scheme for the two-level linear hidden model

\[
\begin{pmatrix} x_{i+1} - x_i \\ y_{i+1} - y_i \end{pmatrix} = \begin{pmatrix} a & 1 \\ q & A \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ \sigma \varepsilon_i \sqrt{\Delta t} \end{pmatrix}
\]

where \( \varepsilon_i \) is a Gaussian random number \( N(0, 1) \). The equilibrium covariance matrix of this model has the following form

\[
\begin{pmatrix} 1 & -a \\ -a & aA - q + a^2 \end{pmatrix} \frac{\sigma^2}{-2(a + A)(aA - q)} + O(\Delta t).
\]
The two-level regression model has the form

\[
\begin{pmatrix}
(x_{i+1} - x_i) \\
(y_{i+1} - y_i)
\end{pmatrix} = \begin{pmatrix}
\tilde{a} & 1 \\
\tilde{q} & \tilde{A}
\end{pmatrix} \begin{pmatrix}
x_i \\
y_i
\end{pmatrix} \Delta t + \begin{pmatrix}
0 \\
\tilde{a} \xi_i \sqrt{\Delta t}
\end{pmatrix}.
\]

Here, we will apply the two-level regression procedure we discussed in 2.3.1 to the true marginal dynamics of the two-level linear hidden model, \(x(t)\).

**Step 2:** Estimate parameter estimators \(\tilde{a}, \tilde{b}\)

\[
\tilde{a} = \frac{\sum_{i=0}^{N-1} (x_{i+1} - x_i)x_i}{\Delta t \sum_{i=0}^{N-1} x_i^2} = \frac{\sum_{i=0}^{N-1} (ax_i + y_i)x_i}{\sum_{i=0}^{N-1} x_i^2}
\]

\[
\xrightarrow{N \to \infty} a + \frac{\text{Cov}(x, y)}{\text{Var}(x)}
\]

\[
\xrightarrow{\Delta t \to 0} 0
\]

\[
\tilde{q} = \frac{\sum_{i=0}^{N-2} (x_{i+2} - 2x_{i+1} + x_i)x_i}{\Delta t^2 \sum_{i=0}^{N-2} x_i^2} - \tilde{a}^2
\]

\[
= \frac{\sum_{i=0}^{N-2} ((a^2 + q)x_i + (a + A)y_i + \sigma_x \xi_i / \sqrt{\Delta t})x_i}{\sum_{i=0}^{N-2} x_i^2} - \tilde{a}^2
\]

\[
\xrightarrow{N \to \infty} (a^2 + q) + (a + A) \frac{\text{Cov}(x, y)}{\text{Var}(x)} - \tilde{a}^2
\]

\[
\xrightarrow{\Delta t \to 0} (a^2 + q) + (a + A)(-a) = q - aA
\]

**Step 3:** Estimate parameter estimators \(\tilde{A}\).

\[
\tilde{A} = \frac{\sum_{i=0}^{N-2} (x_{i+1} - 1 + \Delta t \tilde{a})x_i(x_{i+2} - 2x_{i+1} + x_i) - \Delta t \tilde{a}^2 \frac{1}{N} \sum_{i=0}^{N-2} x_i^2}{\Delta t \sum_{i=0}^{N-2} (x_{i+1} - 1 + \Delta t \tilde{a})x_i^2}
\]

\[
= \frac{1}{N} \sum_{i=0}^{N-1} (x_{i+1} - x_i)(x_{i+2} - 2x_{i+1} + x_i) - \Delta t \tilde{a}^2 \frac{1}{N} \sum_{i=0}^{N-1} x_i^2}
\]

\[
\xrightarrow{\Delta t \to 0} \frac{1}{N} \Delta t^2 \sigma_x^2 - 2(a + A) + O(\Delta t^3)
\]

Let’s discuss each term here.
Therefore, we have the estimation of \( \hat{\sigma}^2 \):

\[
\hat{\sigma}^2 = \Delta t \text{Var}(\frac{1}{\Delta t^2} (x_{i+2} - (2 + \Delta \hat{\sigma} + \hat{\sigma} \Delta t) x_{i+1} + [(1 + \hat{\sigma} \Delta t)(1 + \Delta t \hat{\sigma}) - \Delta t^2 \hat{\sigma}^2] x_i))
\]

\[
\hat{\sigma}^2 = \Delta t \text{Var}(\{a^2 + q + a(\hat{\sigma} + \hat{\sigma}) + (\hat{\sigma} \hat{\sigma} - \hat{q})\} x_i + [(a + A) + (\hat{\sigma} + \hat{\sigma})] y_i + \frac{1}{\sqrt{\Delta t}} \varepsilon_i)
\]

\[
\hat{\sigma}^2 = \Delta t \{a^2 + q + a(\hat{\sigma} + \hat{\sigma}) + (\hat{\sigma} \hat{\sigma} - \hat{q})\} \text{Var}(x) + \Delta t [(a + A) + (\hat{\sigma} + \hat{\sigma})]^2 \text{Var}(y)
\]

\[
+2\Delta t [a^2 + q + a(\hat{\sigma} + \hat{\sigma}) + (\hat{\sigma} \hat{\sigma} - \hat{q})] \Delta t [(a + A) + (\hat{\sigma} + \hat{\sigma})] \text{Cov}(x, y)
\]

\[
+\sigma^2 \frac{1}{\Delta t} \text{Var}(\varepsilon)
\]

\[
\xrightarrow{\Delta \rightarrow 0} \sigma^2
\]

**Step 4:** Estimate parameter estimators \( \hat{\sigma} \).

\[
\hat{\sigma}^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_{i+1} - x_i)(x_{i+2} - 2x_{i+1} + x_i)
\]

\[
= \frac{1}{N} \sum_{i=0}^{N-1} (ax_i + y_i)((a^2 + q)x_i + (a + A)y_i + \sigma \varepsilon_i \frac{1}{\sqrt{\Delta t}}) \Delta t^3
\]

\[
\xrightarrow{N \rightarrow \infty} \{a(a^2 + q) \text{Var}(x) + (a + A) \text{Var}(y)
\]

\[
+[(a + A)a + (a^2 + q)] \text{Cov}(x, y)\} \Delta t^3
\]

\[
= \frac{\sigma^2}{-2(a + A)(aA - q)} \Delta t^3 + O(\Delta t^4)
\]

\[
\xrightarrow{\Delta t \rightarrow 0} \frac{-\frac{1}{2} \sigma^2}{-2(a + A)} = a + A
\]

**REFERENCES**


Received November 2010; revised January 2012.

E-mail address: jonjon@cims.nyu.edu
E-mail address: yuan@cims.nyu.edu