# Analyzing Experiments with Nonlinear Population Dynamics Data

Grace Chan[1,*], Kung-Sik Chan[1], Nils Chr. Stenseth[2], and Ole Christian Lingjærde[2,3]

[1] *Department of Statistics and Actuarial Science, University of Iowa, Iowa City, 52242-1409, U.S.A.*

[2] *Division of Zoology, University of Oslo, PO Box 1050, Blindern, N-0316 Oslo, NORWAY*

[3] *Department of Informatics, University of Oslo, PO Box 1080, Blindern, N-0316 Oslo, NORWAY*

[*] *email:* grchan@stat.uiowa.edu

## Summary

We propose a general approach for analyzing long-term experiments with panels of nonlinear time series data, in the framework of additive models. Among other things, our approach is useful for testing and estimating the (partial) common dynamic structure across treatment groups. We illustrated our approach with a detailed analysis of an ecotoxicological experiment on the effect of sublethal doses of a toxic substance (cadmium) on the long-run dynamic structure of the greenbottle blowfly (*Lucilia sericata*). The general model for the blowfly experiment is an additive model which is derived from a stage-structured ecological model. We discuss the relationship between the components of the additive model and the ecological parameters of the underlying stage-structured model. In particular, our proposed approach casts new insights on the effect of toxic diet on the population dynamic structure of the blowfly.

**Keywords:** Blowfly; Density-dependent; Density-independent; Ecology; Ecotoxicology; Additive Model; Nonlinear Time Series; Population Dynamics.

## 1. Introduction

Data on population dynamic structure often come in the form of a panel of counts of the different life stages of some populations (for example: larva, pupa and adult) taken at regular time intervals under different experimental conditions. A fundamental issue is to test whether or not the observed dynamic structure depends on the treatments, and whenever treatment effects do occur, to localize and estimate the changes in the dynamic structure. In the case of linear population dynamic structure, classical ANOVA techniques have been extended to time series data within the frameworks of time domain and frequency domain analyses; see Box and Tiao (1975) and Brillinger (1981). However, similar analyses of nonlinear time series data is fairly under-developed. Here, we propose a general approach to analyze such experimental data with nonlinear time series. We shall motivate and illustrate our approach using a panel of laboratory blowfly time series of counts from an ecotoxicology study. The analysis of the blowfly data set is also of great interest in itself and a main focus of this paper is to shed new insights on the effect of toxic diet on population dynamic structure.

The greenbottle blowfly (*Lucilia sericata*) data (Smith *et al*, 2000) was set up to study the effect of toxic diet on the blowfly population dynamic structure (i.e., changes in the populations abundance, also being referred to as density, over time). This data set consists of bi-daily counts of different life stages of 12 blowfly populations over a period of about 2 years. The experiment is a $2^2$ factorial design; the first factor is whether or not the blowfly population was fed a constant amount of sublethal doses of a toxic substance (cadmium), and the second factor is the initial population size. A stage-structured model, see Lingjærde *et al.* (2000) and Section 4, relating the counts of the blowfly at various life stages suggests a nonlinear relationship among these counts. However, the underlying biology is not sufficiently understood to specify the functional form of the nonlinearity. Thus, Lingjærde *et al.* (2000) adopted a nonparametric approach, specifically, the additive model, to estimate the underlying stage-structured model for the panel of time

series of counts; see Section 6 for test results supporting the additivity assumption for the blowfly data. Lingjærde *et al.* (2000) pointed out that some vital statistics including reproductive rates of the blowfly differ between the treatment group and the control group. Granted the differences in these vital statistics, it is of interest to determine whether or not the dynamic structure otherwise remains unchanged by the toxic diet. However, Lingjærde *et al.* (2000) did not carry out any formal testing procedure to determine how the density-dependent factors, for example, the cadmium diet, and other density-independent factors affect the blowfly population dynamic structure.

In the analysis of experiments with nonlinear time series data, several interesting issues may arise. First, it is of interest to check whether or not populations under identical experimental conditions share the same dynamic structure. Hence, several independent populations under identical experimental conditions should be included in the experiment and then one can examine population dynamic structure variation both within and between experimental conditions. Second, it is likely that the population dynamic structure will settle to its equilibrium (long-run) dynamic structure only after a lengthy period of transiency. For the blowfly data, half of the populations were started with low initial size, and the rest with high initial size. There were three independent populations under each experimental condition, hence altogether there were twelve independent populations in this study. Lingjærde *et al.* (2000) provided some evidence that the blowfly dynamic structure reached dynamical equilibrium after about one year. It is plausible that the treatments may affect the population dynamic structure differently over the transient phase than they do over the equilibrium phase. Third, the intercept terms in the governing equations of the population dynamic structure may be interpreted as vital rates such as some reproductive or mortality rates; see Section 4. A simple case occurs if the treatment effects are confined to possibly changing the intercept terms of the governing equations of the dynamic structure, but with otherwise unchanged dynamic structure; in this case, we say that the treatment group and the control group

share common dynamical structure. This is similar to the case of parallel regression models. Fourth, the treatments may affect both the intercept terms and the population dynamic structure. Even in this case, some components of the dynamic structure may be identical across the treatment groups and the control groups. For example, the transition mechanism between some life stages may be unaltered by the treatments. Thus, it is of interest to localize and estimate the changes in the dynamic structure in the case that treatment effect on the dynamic structure is deemed significant. The detection and pooling of (partial) common structures in experimental or observational data is of considerable scientific interest; see Stenseth *et al.* (1999) and Yao *et al.* (2000).

Here, we propose an approach, in the framework of additive models, to address some of the above issues. The blowfly data serves as an important case study for illustrating our approach and also for shedding insights on how the density-independent factor, cadmium diet interact with the density-dependent factors. Our analysis reported below suggests that adding cadmium in the diet increases the adult-to-adult recruitment rate, reduces the adult survival, and modifies the functional relationship between the number of larvae and the number of new pupae.

We now outline the content of this paper. In Section 2, we discuss how one can generalize the idea of expressing separate model equations under different experimental conditions into a single equation with the help of indicator functions. Detail description of the blowfly data is given in Section 3. We describe the model and the testing procedure in Sections 4 and 5. The main result is reported in Section 6, and we discuss future statistical and biological researches in Section 7.

## 2. A framework for Studying Common Structure

In classical simple linear regression, one is typically interested in determining a linear relationship between a response and an explanatory variable based on a given data set. Whenever the individuals in the data set are divided into groups/categories, different groups may have different relationships between the

response and explanatory variable. One way to model the relationship for all individuals is to treat the groupings as another explanatory variable, taking numerical values to indicate the grouping of each individual. Since this explanatory variable simply identifies the groups of a nominal variable and its values do not have any quantitative meaning, it is usually called an indicator/dummy/categorical variable. If there is only two groups, we often use the values "0" and "1", which may signify, for instance, the control and treatment groups; then this variable is called binary variable. There are a number of different ways to express models relating these variables. For example,

- separate equations, one for each group,
- single equation which is divided into cases, one for each group,
- single equation with indicator functions, one for each group, and
- single equation with one indicator function less than the number of groups.

In this section, we use $y$, $x_1$ and $x_2$ to denote the response, the original explanatory variable, and the categorical explanatory variable respectively. Hence, the data set is denoted by $\{y_t, x_{1t}, x_{2t}\}_{t=1}^{T}$. Assuming that $x_2$ is a binary variable, we will first consider the following expressions of the linear relationship between $y$ and $x_1$ for each group:

*Expression 1:* For $j = 0, 1$,

$$y_t \;=\; \alpha_j + \beta_j x_{1t} + \epsilon_{jt}, \quad x_{2t} = j,$$

where $\epsilon_{jt}$s are independent and identically distributed. Note the error distributions may not be identical between groups.

*Expression 2:*

$$y_t \;=\; \begin{cases} \alpha_0 + \beta_0 x_{1t} + \epsilon_{0t}, & \text{for } x_{2t} = 0, \\ \alpha_1 + \beta_1 x_{1t} + \epsilon_{1t}, & \text{for } x_{2t} = 1, \end{cases}$$

where the $\epsilon_{jt}$s are same as above.

*Expression 3:*

$$y_t \;=\; (\alpha_0 + \beta_0 x_{1t} + \epsilon_{0t}) I_0(x_{2t}) + (\alpha_1 + \beta_1 x_{1t} + \epsilon_{1t}) I_1(x_{2t}),$$

where the $\epsilon_{jt}$s are same as above, $I_0$ and $I_1$ are indicator functions on the binary variable $x_2$ defined as follows:

$$I_0(x_{2t}) = \begin{cases} 1 & \text{for } x_{2t} = 0, \\ 0 & \text{for } x_{2t} = 1, \end{cases} \quad \text{and} \quad I_1 = 1 - I_0.$$

*Expression 4:*

$$y_t = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 I(x_{2t}) + \gamma_3 x_{1t} I(x_{2t}) + \epsilon_t,$$

where $\epsilon_t$s are independent and identically distributed and $I = I_1$.

Expressions 1 – 3 define the same model, $M_1$. This model implies that the intercept, slope of the linear relationship between $y$ and $x_1$, and the error distribution may be different for the control and the treatment groups. Expression 4 defines a different model, $M_2$. It has an additional condition on the error distributions: different groups are assumed to have the same error distribution. Hence, these two models are equivalent if and only if the error distributions are identical across the different levels of $x_2$. Moreover, the hypothesis of a common (regression) structure between $y$ and $x_1$, (i.e., common structure across groups) can be expressed as

$$\alpha_0 = \alpha_1 \quad \text{and} \quad \beta_0 = \beta_1, \qquad \text{in } M_1,$$

and

$$\gamma_2 = \gamma_3 = 0, \qquad \text{in } M_2.$$

Next, let us generalize the above simple linear regression model to allow for nonlinear terms, (i.e., now the "constant$\times x_1$" term is replaced by "$f(x_1)$" where $f$ is a smooth function). Since the model is unchanged if we add a constant to $f$ and subtract the same constant from the intercept, we need to add a number of constraints to make sure that all equations are identifiable. For model $M_1$ (different error distributions for different groups), these nonlinear relationships are expressed as follows: for $j = 0, 1$,

$$y_t \;=\; \alpha_j + f_j(x_{1t}) + \epsilon_{jt}.$$

The common identifiability constraints for $M_1$ are to set the sums of the nonlinear terms $f_j(x_{1t})$ equal to zero for each group. Note that the sums are over individuals in the same group only. These constraints can be expressed as:

$$\sum_{x_{1t}:x_{2t}=j} f_j(x_{1t}) = 0, \qquad j = 0, 1. \tag{1}$$

For model $M_2$ (common error distribution), the expression is:

$$y_t \;=\; \beta_0 + g_0(x_{1t}) + \beta_1 I(x_{2t}) + g_1(x_{1t})I(x_{2t}) + \epsilon_t. \tag{2}$$

Now the common identifiability constraints are to set the sums of the two nonlinear terms $g_0(x_{1t})$ and $g_1(x_{1t})I(x_{2t})$ to zero over *all* individuals. Since $g_1(x_{1t})I(x_{2t})$ is equal to zero for all individuals in the control group, these constraints can be expressed as:

$$\sum_{\text{all } x_{1t}} g_0(x_{1t}) = \sum_{x_{1t}:x_{2t}=1} g_1(x_{1t}) = 0. \tag{3}$$

As before models $M_1$ and $M_2$ are equivalent if and only if the distributions of the error are identical across the different levels of $x_2$. This assumption on the error distributions is often satisfied in practice.

The common structure hypothesis can be expressed as

$$\alpha_0 = \alpha_1 \quad \text{and} \quad f_0 = f_1, \qquad \text{in } M_1$$

and

$$\beta_1 = 0 \quad \text{and} \quad g_1(\cdot) \equiv 0, \qquad \text{in } M_2.$$

Note that $M_2$ can also be expressed as follows:

$$y_t \;=\; \{\gamma_0 + h_0(x_{1t})\}I_0(x_{2t}) + \{\gamma_1 + h_1(x_{1t})\}I_1(x_{2t}) + \epsilon_t. \tag{4}$$

Now the identifiability constraints (3) and the common structure hypothesis become

$$\sum_{x_{1t}:x_{2t}=j} h_j(x_{1t}) = 0, \quad j = 0, 1, \tag{5}$$

$$\gamma_0 = \gamma_1 \quad \text{and} \quad h_0 = h_1,$$

respectively. Equation (2) enables us to fit models and carry out hypothesis testing at the same time whereas equation (4) leads to direct estimation of the regression function within each level of $x_2$. Henceforth, we shall employ the common error distribution model $M_2$ to study the common structure in nonlinear time series experimental data. Note that model $M_2$ is specified as an additive model (Hastie and Tibshirani, 1990) with the additive "main" effects of $x_1$ (nonlinear) and $x_2$, plus their interaction.

Model $M_2$ can be extended to apply to more complex experimental designs by including in (2) or (4) more continuous and categorical explanatory variables. In summary, this model makes it easy to adapt the indicator variable approach for analyzing linear experimental designs to our current case of assessing the extent of common structure with nonlinear time series experimental data.

**3. Blowfly Data**

In order to study the population dynamic structure of greenbottle blowfly (*Lucilia sericata*) and how it is affected by density-dependent and density-independent factors, an experiment was carried out at the University of Reading between 1989 and 1992. Detailed description of this data can be found in Daniels (1994) and Smith *et al.* (1999); see also Lingjærde *et al.* (2000).

Twelve laboratory populations of blowfly were kept in separate bottles. These populations were divided into four experimental categories, each consisting of three replicates. Six populations were given sublethal dosages of the toxic compound cadmium acetate through the larval diet; the other six, that were not given cadmium acetate, were considered as controls. Initial density was *low* with 30 pupae and 30 adults for three of the six populations in each group, whereas the remaining three populations were initialized with *high* density: 150 pupae and 150 adults. Throughout this paper, we refer to the populations either by an overall index (from 1 to 12) or by a triple index consisting of the terms *control* (CON), *cadmium* (CAD), with each subdivided into *low initial density* (L), *high initial density* (H), and then

followed by the replicate sub-index (from 1 to 3) (see Table 1 for detail).

Altogether there are 36 time series. All except one population (Population 8 — CAD(L,2)) were observed for approximately two years. Population 8 — CAD(L,2) died out on day 298. In Lingjærde *et al.* (2000), they excluded this population from their analysis. This population is included in our study. Also there are some missing counts during the beginning of the experiment.

< Insert Table 1 here >

There are 5 stages in the life cycle of blowfly: egg, larva, pupa, immature adult and mature adult. Hatching is completed within roughly one day. All other stages last for more than two days. The duration of life stages may be different between the control and the treatment groups. For each population, the number of

- larvae of *all* ages,
- "*new*" (less than 2 days old) and "*viable*" (survive to become adult) pupae, and
- adults of *all* ages,

were recorded every two days. Thus the entire data set consists of a panel of $3 \times 12$ time series.

## 4. A Biological Model for the Blowfly Data

### 4.1 Preamble

We recall in this Subsection a biological model developed by Lingjærde *et al.* (2000) for the blowfly data. They observed that the egg stage lasts less than a day, the larva stage lasts for approximately 8 days among populations in the control group, and 9 days among populations in the treatment group, the pupa stage lasts between 6 and 12 days, the immature adult stage lasts for approximately 5 days, and the mature adult stage lasts for approximately 12 days among populations in the control group, and 9 days among populations in the treatment group. Regarding the duration of each life stage, it is assumed that the egg stage lasts less than 2 days

(not included in the data); the larva stage lasts no more than 8 days; the pupa stage lasts no more than 10 days; the immature adult stage lasts no more than 4 days, but the mature adult stage has no specified upper limit on its duration. Regarding the demographic rates, we assume the followings. They are identical for all blowflies in a given age group and have identical density-dependent structure for all age groups within a life stage. The reproductive rate depend on the density of mature adults only. The survival probability of a larva from one age group to the next is assumed to be density-independent. The proportion of larvae pupating at any given time is modelled as a function of the total number of larvae eight days before. All pupae become adults (this data only include viable pupa), and no deaths occur in the immature adult stage.

Based on the above assumptions, a stage-structural model for the blowfly population dynamic structure (ignoring the stochastic components) is defined as:

$$L_{t+1}^1 = A_t^{\mathcal{M}} \exp\{\alpha_{L1} + f_L(\log A_t^{\mathcal{M}})\} \tag{6}$$

$$L_{t+1}^{i+1} = L_t^i \exp(\alpha_{L2}), \quad 1 \le i \le 3 \tag{7}$$

$$P_{t+1}^1 = L_t^4 \exp\{\alpha_P + f_P(\log L_{t-1})\} \tag{8}$$

$$P_{t+1}^{i+1} = P_t^i, \quad 1 \le i \le 4 \tag{9}$$

$$A_{t+1}^2 = A_t^1 \ = \ P_{t-1}^5 \tag{10}$$

$$A_{t+1}^{i+1} = A_t^i \exp\{\alpha_A + f_{A1}(\log A_t^{\mathcal{I}}) + f_{A2}(\log A_t^{\mathcal{M}})\}, \quad 2 \le i < \infty \tag{11}$$

where

$L_t^i$ = number of larvae $d \in [2i-2, 2i)$ days old at time $t$, $\quad i = 1, 2, 3, 4,$

$L_t$ = total number of larvae at time $t$,

$$= \sum_{j=1}^{4} L_t^j,$$

$P_t^i$ = number of pupae $d \in [2i-2, 2i)$ days old at time $t$, $\quad i = 1, 2, 3, 4, 5,$

$A_t^i$ = number of adults $d \in [2i-2, 2i)$ days old at time $t$, $\quad i = 1, 2, \ldots,$

$$
\begin{aligned}
A_t^{\mathcal{I}} &= \text{number of immature adults at time } t \\
&= \sum_{j=1}^{2} A_t^j \;=\; P_{t-5}^1 + P_{t-6}^1, \quad \text{and} \\
A_t^{\mathcal{M}} &= \text{number of mature adults at time } t \\
&= \sum_{j=3}^{\infty} A_t^j .
\end{aligned}
$$

The stochastic components are modelled as additive noise on the logarithmic scale, see below.

The raw blowfly data contains counts on

$$L_t, \quad P_t^1, \quad \text{and} \quad A_t = \sum_{j=1}^{\infty} A_t^j = A_t^{\mathcal{I}} + A_t^{\mathcal{M}}, \qquad t = 1, 2, \ldots 378.$$

The parameters in the model defined by (6) – (11) have the following biological interpretation:

$\alpha_{L1}$ denotes the mean log reproduction rate,

$\alpha_{L2}$ denotes the mean log larval survival rate,

$\alpha_P$ denotes the mean log larva-to-adult survival rate, and

$\alpha_A$ denotes the mean log adult survival rate.

Recall that the main objective of Lingjærde *et al.* (2000) was to put as little constraint as possible on the form of the functions entering (12) and (13) of the model, mainly due to inadequate theoretical understanding of the basic ecological theory necessary to deduce the functional forms from first principles. Hence, a nonparametric approach was adopted by Lingjærde *et al.* (2000). Through our analysis reported in this paper, we may as a side result, be able to specify some of the functions as simpler parametric functions. To be able to formulate models parametrically, rather than nonparametrically, is advantageous not least since biological reasoning becomes easier; it is, we believe, easier to think about the meaning of changing a parameter-value than to think about the meaning of changing the entire shape of the relationship.

We now rewrite equations (6) – (11) into another set of equations relating observed quantities as follows:

$$\log\left(\frac{P^1_{t+1}}{A^{\mathcal{M}}_{t-4}}\right) = \alpha + f_L(\log A^{\mathcal{M}}_{t-4}) + f_P(\log L_{t-1}) \tag{12}$$

$$\log\left(\frac{A^{\mathcal{M}}_{t+1}}{A^{\mathcal{M}}_t + P^1_{t-6}}\right) = \alpha_A + f_{A1}(\log A^{\mathcal{T}}_t) + f_{A2}(\log A^{\mathcal{M}}_t) \tag{13}$$

$$L_{t+1} = \exp(-\alpha_P)\sum_{s=0}^{3} P^1_{t+s+2}\exp\{-s\alpha_{L2} - f_P(\log L_{t+s})\} \tag{14}$$

where $\alpha = \alpha_{L1} + 3\alpha_{L2} + \alpha_P$. Equations (12) and (13) are derived based on the simplistic assumption of deterministic transitions between the life stages of blowfly. In practice, the transitions are stochastic in nature corresponding to inherent biological fluctuations. Hence, error terms have to be added to the right side of (12), (13) and (14), which shall be assumed henceforth in the paper. We will focus on the two additive model equations (12) and (13) in this paper. As the blowfly dynamic structure may be subject to transient effects at the beginning period of the experiment, we partition the data into two parts for analysis (the transient and the stationary part). Each part lasts for approximately one year. This way we can separately study the effect of cadmium on the population dynamic structure over the transient period and the stationary period.

### 4.2 The Full Model

Based on equations (12) and (13) from Subsection 4.1, we consider a model that has a common structure for all 12 populations, and at the same time allow for different constants and nonparametric functions which depend on (cadmium) treatment effect, (initial) density effect, interaction between treatment and density, and individual population effect.

Applying the methodology described in Section 2, we use the following notation to specify the full model:

**M**ain effect (common structure):

$$y = \begin{cases} \log\left(\frac{P^1_{t+1}}{A^{\mathcal{M}}_{t-4}}\right) & \text{equation (12)}, \\ \log\left(\frac{A^{\mathcal{M}}_{t+1}}{A^{\mathcal{M}}_{t}+P^1_{t-6}}\right) & \text{equation (13)}, \end{cases}$$

$$x_1 = \begin{cases} \log A^{\mathcal{M}}_{t-4} & \text{equation (12)}, \\ \log A^{\mathcal{I}}_{t}, & \text{equation (13)}, \end{cases}$$

$$x_2 = \begin{cases} \log L_{t-1} & \text{equation (12)}, \\ \log A^{\mathcal{M}}_{t} & \text{equation (13)}. \end{cases}$$

(Cadmium) **T**reatment effect:

$$x_T = \begin{cases} 1 & \text{cadmium included in the diet,} \\ 0 & \text{control, i.e. no cadmium in the diet.} \end{cases}$$

(Initial) **D**ensity effect:

$$x_D = \begin{cases} 1 & \text{initial density is high,} \\ 0 & \text{control, i.e. initial density is low.} \end{cases}$$

**I**nteraction:

$$x_I = \begin{cases} 1 & \text{interaction of cadmium and high initial density,} \\ 0 & \text{otherwise.} \end{cases}$$

(Individual) **P**opulation effect: for $p = 1, 2, 4, 5, 7, 8, 10, 11$

$$x_{Pp} = \begin{cases} 1 & \text{population } p, \\ 0 & \text{otherwise.} \end{cases}$$

For each experimental category, one of three replications (populations 3, 6, 9, and 12, respectively) are left out in order to make all regression parameters identifiable.

Now the full model — **MTDIP** can be presented as follows (below, the notation $E\{\cdot\}$ stands for the conditional expectation of the enclosed expression given the

covariates and their past lags):

$$\begin{aligned} E(y) = \quad & \{\alpha_0 + f_1(x_1) + f_2(x_2)\} \\ & + \{\alpha_T + f_{1T}(x_1) + f_{2T}(x_2)\} I(x_T) \\ & + \{\alpha_D + f_{1D}(x_1) + f_{2D}(x_2)\} I(x_D) \\ & + \{\alpha_I + f_{1I}(x_1) + f_{2I}(x_2)\} I(x_I) \\ & + \sum \{\alpha_j + f_{1,j}(x_1) + f_{2,j}(x_2)\} I(x_{Pj}), \end{aligned}$$

where the last sum sums over $j \in \{1, 2, 4, 5, 7, 8, 10, 11\}$. In the above equation, the first three terms in the first line denote the **M**ain common structure, the next three terms in the second line denote the **T**reatment effect due to cadmium, the following three terms in the third line denote the **D**ensity effect, the next three terms in the fourth line denote the **I**nteraction effect between treatment and initial density, and the rest of the terms denote **P**opulation effects.

**5. Model Selection Procedure**

There are two stages in this model selection procedure. In Stage I, we focus on the effects. In Stage II, we examine individual terms within an effect. In both stages, a number of model selection criteria were used:

1. Approximate F-Test,
2. $C_p$ Criterion,
3. $\chi^2$ Test, and
4. Akaike Information Criterion (AIC).

In Stage I, we start with the full model **MTDIP** and remove terms from it using a procedure similar to the standard backward elimination regression model selection procedure. Here we determine if we can remove all terms associated with a single effect at a time. The basic procedure follows these steps. *Step 1:* we test if we can remove the last 24 terms that associate with individual population effect from **MTDIP** and reduce it to **MTDI**. If not, skip the remaining steps. *Step 2:* we

test if we can remove the last 3 terms that associate with interaction effect between cadmium treatment and initial density from **MTDI** and reduce it to **MTD**. If not, skip the remaining steps. *Step 3:* we test if we can remove the last 3 terms that associate with initial density effect from **MTD** and reduce it to **MT**. If not, skip the remaining step. *Step 4:* we test if we can remove the last 3 terms that associate with cadmium treatment effect from **MT** and reduce it to **M**.

In Stage II, we start with the model obtained from Stage I and further simplify it by testing the significance (and the nonlinearity) of the remaining (nonlinear smooth) terms.

Suppose the model **MT**:

$$\begin{aligned} E(y) = \quad & \{\alpha_0 + f_1(x_1) + f_2(x_2)\} \\ & + \{\alpha_T + f_{1T}(x_1) + f_{2T}(x_2)\} I(x_T), \end{aligned}$$

is the model obtained from Stage I, to be denoted by **M.1**. The *first step* is to consider the following simpler models:

**M.2.1**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2) + \alpha_T I(x_T) + f_{1T}(x_1) I(x_T)$,

**M.2.2**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2) + \alpha_T I(x_T) + f_{2T}(x_2) I(x_T)$,

**M.3**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2) + \alpha_T I(x_T)$,

**M.4**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2)$,

**M.5.1**: $E(y) = \alpha_0 + f_1(x_1)$, and

**M.5.2**: $E(y) = \alpha_0 + f_2(x_2)$.

Testing **M.1** against **M.2.1** (**M.2.2**), we can determine whether cadmium affects the second (first) nonlinear term or not. If cadmium does not affect the nonlinear terms, we consider model **M.3** and test it against **M.4** (the common structure model) to determine whether cadmium affects the relationship by shifting the "curves" up and/or down. If model **M.4** is accepted, we further test the significance of the two nonlinear terms.

Suppose the model **M.2.2** (cadmium does not affect the first nonlinear term)

is obtained from the first step. The *second step* is to test whether some/all of the nonlinear terms can be simplified to linear terms. That is **M.2.2** is compared with the following models:

$$E(y) = \{\alpha_0 + f_1(x_1) + f_2(x_2)\} + \{\alpha_T + \beta_{2T}x_2\}I(x_T),$$

$$E(y) = \{\alpha_0 + f_1(x_1) + \beta_2 x_2\} + \{\alpha_T + \beta_{2T}x_2\}I(x_T),$$

$$E(y) = \{\alpha_0 + \beta_1 x_1 + f_2(x_2)\} + \{\alpha_T + f_{2T}(x_2)\}I(x_T), \quad \text{and}$$

$$E(y) = \{\alpha_0 + \beta_1 x_1 + \beta_2 x_2\} + \{\alpha_T + \beta_{2T}x_2\}I(x_T).$$

Suppose the second simpler model (all terms involving $x_2$ are linear) is chosen from the second step. The significance of these new linear terms are tested in the *third step*, and so on, until only significant terms in their simplest form appear in the final model.

## 6. Results and Biological Interpretations

### 6.1 Preamble

Recall that we have two different equations relating the population dynamic structure variables and we are also interested in studying the effect of cadmium separately before and after the transience. Hence, we analyze the following four cases separately:

Case 1: Equation (12) for the first year or part 1 of the times series data,

Case 2: Equation (12) for the second year or part 2 of the times series data,

Case 3: Equation (13) for the first year or part 1 of the times series data, and

Case 4: Equation (13) for the second year or part 2 of the times series data.

The `gam` function in Splus is used to fit these additive models with the `ns` function modeling the nonlinear functions parametrically with 4 degrees of freedom. The `ns` function generates a basis matrix for a natural cubic spline which consists of piecewise cubic polynomials separated by a sequence of internal knots with linear extension beyond the boundary knots such that it has two continuous derivatives and a step function with jumps at every internal knot as its third derivative. For

example, a natural cubic spline with 4 degrees of freedom has 3 internal knots positioned at the first quartile, the median and the third quartile, and 2 boundary knots at the minimum and the maximum. These sets of basis functions are used to approximate/estimate the nonlinear functions in (12) and (13). See de Boor (1978) for more detail discussion on splines.

### 6.2 Stage I

All four model selection criteria (see Section 5) were computed for all models considered in this stage for all four cases. These results suggest that **MT** is the best model for all four cases. Only the $C_p$ values are listed in Table 2. Choosing **MT** as the best model for all four cases implies that:

- Cadmium treatment has significant effect on both parts' population dynamic structure both in terms of mean demographic rates and nonlinear functional relationships.
- Initial density does not have significant effect on population dynamic structure.
- Interaction between cadmium treatment and initial density does not have significant effect on population dynamic structure.
- Individual population does not have significant effect on population dynamic structure.
- The model **MT**:

$$\begin{aligned} E(y) = \quad & \{\alpha_0 + f_1(x_1) + f_2(x_2)\} \\ & + \{\alpha_T + f_{1T}(x_1) + f_{2T}(x_2)\} I(x_T) \end{aligned}$$

  can be used to model the population dynamic structure. Also the empirical distributions of the error term of individual populations are compared with the empirical distributions of the error term in the common error distribution model using both quantile-quantile plot and Kolmogorov-Smirnov goodness-of-fit test; in particular, only a few populations show mild evidence that the identical error distribution assumption may not hold.

< Insert Table 2 here >

### 6.3 Stage II

In Stage II, all nonlinear terms are examined closely to see whether one can reduce the above models further by replacing some of the nonlinear functions with linear functions and remove other remaining terms. We follow four steps:

*Step 1*: Sequential F-tests and information from other model selection criteria such as $C_p$ are used to determine whether any terms can be removed from the model **MT**. Now the *full* model **MT** is denoted by **M.1** and it is tested against the following simpler models:

**M.2.1**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2) + \alpha_T I(x_T) + f_{1T}(x_1)I(x_T)$,

**M.2.2**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2) + \alpha_T I(x_T) + f_{2T}(x_2)I(x_T)$,

**M.3**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2) + \alpha_T I(x_T)$,

**M.4**: $E(y) = \alpha_0 + f_1(x_1) + f_2(x_2)$,

**M.5.1**: $E(y) = \alpha_0 + f_1(x_1)$, and

**M.5.2**: $E(y) = \alpha_0 + f_2(x_2)$.

Four sets of sequential F-tests are considered:

1. **M.1** vs **M.2.1** vs **M.3** vs **M.4** vs **M.5.1**
2. **M.1** vs **M.2.2** vs **M.3** vs **M.4** vs **M.5.1**
3. **M.1** vs **M.2.1** vs **M.3** vs **M.4** vs **M.5.2**
4. **M.1** vs **M.2.2** vs **M.3** vs **M.4** vs **M.5.2**

Tables 3 and 4 show the $C_p$ values for all models and p-values for all the above sequential F-tests. Based on these results, the most suitable models at this stage for the four cases are **M.1**, **M.2.2**, **M.2.2**, and **M.3**, respectively.

< Insert Tables 3 and 4 here >

*Step 2*: All nonlinear term(s) in each model is tested to determine whether or not they can be reduced to linear functions. Table 5 lists the $C_p$ values for all the models considered. Other model selection criteria all suggest that the models for Cases 1 and 2 should be unchanged whereas the nonlinear functions of $x_2$ for Cases 3 and 4 can be replaced by linear functions.

< Insert Table 5 here >

*Step 3*: All linear functions are tested for significance. (Note: Since there is no change in Cases 1 and 2, only Cases 3 and 4 are considered.) In Case 3, the linear function of $x_2$ is not significant and hence can be removed from the model. However, all terms in the current model for Case 4 are significant.

*Step 4*: All terms in the model are tested for the final time to remove any in-significance term. In Case 1, the first nonlinear function is not significantly different between the control and treatment groups. In Case 2, the difference of the second nonlinear function between the control and treatment groups can be modelled linearly.

The final set of models can be presented in forms of *single-equation* expressions as follows:

Equation (12):

Parts 1 and 2:

$$E\left\{\log\left(\frac{P^1_{t+1}}{A^{\mathcal{M}}_{t-4}}\right)\right\} = \alpha + f_L(\log A^{\mathcal{M}}_{t-4}) + f_P(\log L_{t-1}) + \alpha_T I(x_{Tt}) + f_{PT}(\log L_{t-1})I(x_{Tt}),$$

where, e.g., $f_{PT}(\cdot)$ represents the significant modification of $f_P(\cdot)$ due to the cadmium effect. In other words, $f_P(\cdot)$ becomes $f_P(\cdot) + f_{PT}(\cdot)$ for the treatment group, for the first part of the data. Note that the intercepts and the nonlinear functions differ over the two parts of the time series. In particular, $f_{PT}$ is a linear function in Part 2.

Equation (13):

Part 1:

$$E\left\{\log\left(\frac{A^{\mathcal{M}}_{t+1}}{A^{\mathcal{M}}_{t}+P^{1}_{t-6}}\right)\right\}=\alpha_A+f_{A1}(\log A^{\mathcal{I}}_{t})+\alpha_T I(x_{Tt}),$$

Part 2:

$$E\left\{\log\left(\frac{A^{\mathcal{M}}_{t+1}}{A^{\mathcal{M}}_{t}+P^{1}_{t-6}}\right)\right\}=\alpha_A+f_{A1}(\log A^{\mathcal{I}}_{t})+\beta_{A2}\log A^{\mathcal{M}}_{t}+\alpha_T I(x_{Tt}),$$

where $\beta_{A2}$ is a parameter such that

$$f_{A2}(\log A^{\mathcal{M}}_{t})=\beta_{A2}\log A^{\mathcal{M}}_{t}.$$

Expressing the above models in *separate-equations* forms, we have

Equation (12):

Parts 1 ($i=1$) and 2 ($i=2$):

$$E\left\{\log\left(\frac{P^{1}_{t+1}}{A^{\mathcal{M}}_{t-4}}\right)\right\}=\begin{cases}\alpha_{C,i}+f_{L,C,i}(\log A^{\mathcal{M}}_{t-4})+f_{P,C,i}(\log L_{t-1}) & \text{control},\\ \alpha_{T,i}+f_{L,T,i}(\log A^{\mathcal{M}}_{t-4})+f_{P,T,i}(\log L_{t-1}) & \text{treatment},\end{cases}$$

where $\alpha_{C,i}=\alpha$, $\alpha_{T,i}=\alpha+\alpha_T+c_i$, $f_{L,C,i}=f_{L,T,i}=f_{L,i}$, $f_{P,C,i}=f_{P,i}$, and $f_{P,T,i}=f_{P,i}+f_{PT,i}-c_i$; the generic constants $c_i$, which may differ from occurrence to occurrence, are chosen to satisfy the appropriate set of identifiability constraints.

Equation (13):

Part 1:

$$E\left\{\log\left(\frac{A^{\mathcal{M}}_{t+1}}{A^{\mathcal{M}}_{t}+P^{1}_{t-6}}\right)\right\}=\begin{cases}\alpha_{A,C,1}+f_{A1,C,1}(\log A^{\mathcal{I}}_{t}) & \text{control},\\ \alpha_{A,T,1}+f_{A1,T,1}(\log A^{\mathcal{I}}_{t}) & \text{treatment},\end{cases}$$

where $\alpha_{A,C,1}=\alpha_A$, $\alpha_{A,T,1}=\alpha_A+\alpha_T$, and $f_{A1,C,1}=f_{A1,T,1}=f_{A1,1}$.

Part 2:

$$E\left\{\log\left(\frac{A^{\mathcal{M}}_{t+1}}{A^{\mathcal{M}}_{t}+P^{1}_{t-6}}\right)\right\}=\begin{cases}\alpha_{A,C,2}+f_{A1,C,2}(\log A^{\mathcal{I}}_{t})+\beta_{A2,C,2}\log A^{\mathcal{M}}_{t} & \text{control},\\ \alpha_{A,T,2}+f_{A1,T,2}(\log A^{\mathcal{I}}_{t})+\beta_{A2,T,2}\log A^{\mathcal{M}}_{t} & \text{treatment},\end{cases}$$

where $\alpha_{A,C,2} = \alpha_A$, $\alpha_{A,T,2} = \alpha_A + \alpha_T$, $f_{A1,C,2} = f_{A1,T,2} = f_{A1,2}$, and $\beta_{A2,C,2} = \beta_{A2,T,2} = \beta_{A2,2}$.

As discussed in Section 2, there are two ways of specifying the identifiability constraints:

1. All nonlinear functions are scaled to satisfy condition (3), e.g.

$$\sum_{\text{all } \log L_{t-1}} f_P(\log L_{t-1}) = \sum_{\text{all } \log L_{t-1}: x_{Tt}=1} f_{PT}(\log L_{t-1}) = 0.$$

2. All nonlinear functions are scaled to satisfy condition (1) or (5), e.g. for $i = 1$, 2, and $j = 0$, 1,

$$\sum_{\log A^{\mathcal{M}}_{t-4}: x_{Tt}=j} f_{L,C,i}(\log A^{\mathcal{M}}_{t-4}) = \sum_{\log L_{t-1}: x_{Tt}=j} f_{P,C,i}(\log L_{t-1}) = 0.$$

   Below, we shall employ the second set of constraints. However, note that the two sets of constraints are identical if the fitted model includes an intercept term for each group.

We now present the final estimate of the population dynamic structure incorporating the cadmium treatment conditions: (C1: model for control group over the first part of data, T1: model for treatment group over the first part of data, C2: model for control group over the second part of data, T2: model for treatment group over the second part of data) with identifiability constraints (5):

$$\log\left(\frac{\widehat{P^1_{t+1}}}{A^{\mathcal{M}}_{t-4}}\right) = \begin{cases} -2.555 + \hat{f}_{L,1}(\log A^{\mathcal{M}}_{t-4}) + \hat{f}_{P,C,1}(\log L_{t-1}) & \text{C1} \\ -2.238 + \hat{f}_{L,1}(\log A^{\mathcal{M}}_{t-4}) + \hat{f}_{P,T,1}(\log L_{t-1}) & \text{T1} \\ -2.663 + \hat{f}_{L,2}(\log A^{\mathcal{M}}_{t-4}) + \hat{f}_{P,C,2}(\log L_{t-1}) & \text{C2} \\ -2.163 + \hat{f}_{L,2}(\log A^{\mathcal{M}}_{t-4}) + \hat{f}_{P,T,2}(\log L_{t-1}) & \text{T2} \end{cases}$$

$$\log\left(\frac{\widehat{A^{\mathcal{M}}_{t+1}}}{A^{\mathcal{M}}_t + P^1_{t-6}}\right) = \begin{cases} -0.122 + \hat{f}_{A1,1}(\log A^{\mathcal{I}}_t) & \text{C1} \\ -0.187 + \hat{f}_{A1,1}(\log A^{\mathcal{I}}_t) & \text{T1} \\ -0.085 + \hat{f}_{A1,2}(\log A^{\mathcal{I}}_t) - 0.063(\log A^{\mathcal{M}}_t - m_C) & \text{C2} \\ -0.141 + \hat{f}_{A1,2}(\log A^{\mathcal{I}}_t) - 0.063(\log A^{\mathcal{M}}_t - m_T) & \text{T2}, \end{cases}$$

where $m_C$ and $m_T$ are the sample means of $\log A_t^{\mathcal{M}}$ among the control and treatment groups respectively.

Figure 1 shows all significant smooth functions: $\hat{f}_L$'s, $\hat{f}_P$'s, $\hat{f}_{A1}$'s, and $\hat{f}_{A2}$'s are displayed in panels (a) – (d) respectively.

The fitted models reported above suggest that the blowfly dynamic structure over the first part of the data differs from that of the second part of the data, which may be attributed to transient effects.

The first part of the data suggests that for equation (12), cadmium affects the dynamic structure in two regards: it increases the intercept term (p-value = 0.0000) and modifies the functional form of $f_P(\cdot)$ (p-value = 0.0104) and for equation (13), cadmium affects the dynamic structure only in terms of the intercept which is decreased (p-value = 0.0000).

Assuming that the blowfly dynamic structure is in steady equilibrium over part 2 of the data, the fitted models suggest that cadmium affects the (steady-state blowfly) dynamic structure in three regards: it increases the intercept term (p-value = 0.0000), modifies the functional form of $f_P(\cdot)$ (p-value = 0.0000) in equation (12), and decreases the intercept term (p-value = 0.0000) in equation (13).

< Insert Figure 1 here >

The effect of cadmium on the function $f_P$ is shown in Figure 2. Here the estimated difference between the control and treatment group are shown in the same scale as in Figure 1. Panel (a) displays the estimated difference between $f_{P,T,1}$ and $f_{P,C,1}$ (p-value = 0.0104), and panel (b) displays the estimated difference between $f_{P,T,2}$ and $f_{P,C,2}$ (p-value = 0.0000).

< Insert Figure 2 here >

Finally, we carry out additivity tests on the above models. In particular, we

test to see whether any nonlinear or linear function of cross-product term is needed. This may be done via exploratory data analysis or formal test: Nonadditivity may be check graphically by generalizing the idea of added variable plot in Weisberg (1985). Alternatively, we can carry out a formal $F$-test. These two approaches are outlined below: For example, to test additivity for Equation (12) Part 1. The added variable plot method works as follows:

1. Save the residuals from the current final model.

2. Regress $\log(A^{\mathcal{M}}_{t-4}) \times \log(L_{t-1})$ on the same regressors as in 1, and save these residuals as well.

3. Study the relationship between these two sets of residuals. A strong relationship between the residuals indicate nonadditivity.

The approximate $F$-test method can be summarized as follows:

1. Enlarge the "final" model with an interaction term, say, $f_{LP}\{\log(A^{\mathcal{M}}_{t-4}) \times \log(L_{t-1})\}$.

2. Use ANOVA to test if this new model is significantly better than the "final" model..

For all of the above models, both methods conclude that there is no evidence of nonadditivity. Chen, Liu and Tsay (1995) proposed other methods to test for additivity with nonlinear time series data. Their approaches differ from ours in that they estimated the additive models nonparametrically using, e.g. ACE (Breiman and Friedman 1985) whereas we employ regression splines with pre-specified degrees of freedom to estimate the additive model. We should remark that the degrees of freedom of the regression splines ordinarily increase with the sample size.

To conclude this section, we list a number of biological interpretations of the final model. First, there is no significant difference among the three individual populations within each experimental condition. And the initial density does not affect the population dynamic structure. However, the dynamic structure during

the second year of the study (part 2) is different from that of the first year (part 1). In particular, during the first year, there is no inter-specific competition among the mature adults; (see (11)). On the other hand, in the second year $f_{A2}$ is estimated to be a linear function implying a significant inter-specific competition among the mature adults. Finally the effect of cadmium given through larval diet can be summarized as follows: it increases the mean adult-to-adult recruitment rate, it decreases the mean adult survival rate, and it changes the functional relationship $f_P$ between the number of larvae and the number of new pupae nonlinearly during the first year and linearly during the second year.

## 7. Conclusion

In this paper, we have developed a new approach to analyze experiments with nonlinear time series data, in the framework of additive model. In particular, the proposed approach is useful for detecting and pooling (partial) common dynamic structure in a panel of nonlinear time series data. We have used regression splines, specifically cubic splines with pre-specified knots, to parameterize the components of the additive model. With increasing degrees of freedom, equivalently increasing number of knots, the regression splines can provide increasingly accurate approximations to a smooth function. Hence, this approach is ordinarily satisfactory in practice. Nevertheless, it is of interest to adopt a fully nonparametric approach to estimate the additive models. Unfortunately, the gam function in Splus does not allow fully nonparametric specification of the regression functions in the case of ANOVA-type analysis needed in our approach. More research on such an implementation and related theoretical problems is clearly needed.

## Acknowledgments

We are grateful to Professor Robert Smith, Department of Biology, University of Leicester, for giving us access to the data.

## References

Box, G. E. P. and G. C. Tiao (1975). Intervention analysis with applications

to economic and environmental problems. *Journal of the American Statistical Association*, **70**, 70-79.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, **80**, 580-619.

Brillinger, D. R (1981). *Time series: Data analysis and theory.* Holt, Rinehart & Winston.

Chen, R., Liu, J. S., and Tsay, R. S. (1995). Additivity tests for nonlinear autoregression. *Biometrika*, **82**, 369-83.

Daniels, S. (1994). *Effects of Cadmium Toxicity on Population Dynamics of the Blowfly Lucilia sericata.* PhD thesis, University of Reading.

de Boor, C. (1978). *A Pratical Guide to Splines.* Springer-Verlag.

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall.

Lingjærde, O. C., Stenseth, N. C., Kristoffersen, A. B., Smith, R. H., Moe, S. J., Read, J. M., Daniels, S. and Simkiss, K. (2000) Exploring non-linearities in the stage-specific density-dependent structure of experimental blowfly populations using non-parametric additive modelling. *Ecology*, (in press).

Smith, R. H., Daniels, S., Simkiss, K., Bell, E. D., Ellner, S. P. and Forrest, B. (2000). Blowflies as a case study in non-linear population dynamics. In Perry, J. N., Smith, R. H., Woiwod, I. P. and Morse, D. (eds.), *Chaos in Real Data: The Analysis of Non-linear Dynamics in Short Ecological Time Series.* Kluwer, Netherlands.

Stenseth, N.C. Chan, K.-S., Tong, H., Boonstra, R., Boutin, R., Krebs, C.J., Post, E., O'Donoghue, M., Yoccoz, N.G., Forchhammer, M.C. and Hurrell, J.W. (1999). Common dynamic structure of Canada lynx populations within three climatic regions. *Science*, **285**, 1071-1073.

Weisberg, S. (1985). *Applied Linear Regression.* Second Edition. John Wiley & Sons.

Yao, Q., Tong, H., Finkenstäd, B. and Stenseth, N.C. (2000). Common structure in panels of short ecological time series. *Proceedings of Royal Society of London, B.* (in press).

**Table 1**

*Descriptions of the population labels.*

| Label | Description | | |
|---|---|---|---|
| | population | group | initial density |
| 1. CON(L,1) | first | control | low |
| 2. CON(L,2) | second | control | low |
| 3. CON(L,3) | third | control | low |
| 4. CON(H,1) | first | control | high |
| 5. CON(H,2) | second | control | high |
| 6. CON(H,3) | third | control | high |
| 7. CAD(L,1) | first | treatment | low |
| 8. CAD(L,2) | second | treatment | low |
| 9. CAD(L,3) | third | treatment | low |
| 10. CAD(H,1) | first | treatment | high |
| 11. CAD(H,2) | second | treatment | high |
| 12. CAD(H,3) | third | treatment | high |

**Table 2**

*Stage I Model Selection: $C_p$ values for five models in each case*

| Model | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|
| **MTDIP** | 2743.701 | 1950.639 | 285.466 | 59.327 |
| **MTDI** | 2734.115 | 1910.420 | 277.638 | 57.614 |
| **MTD** | 2736.740 | 1902.893 | 275.980 | 57.352 |
| **MT** | 2730.370 | 1900.782 | 275.919 | 57.224 |
| **M** | 2757.327 | 1910.828 | 278.117 | 59.322 |

**Table 3**

*Stage II, Step 1 Model Selection: $C_p$ values for seven models in each case*

| Model | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|
| **M.1** | 2732.020 | 1901.213 | 275.947 | 57.227 |
| **M.2.1** | 2742.612 | 1913.927 | 276.400 | 57.133 |
| **M.2.2** | 2734.431 | 1901.490 | 275.950 | 57.054 |
| **M.3** | 2741.122 | 1912.168 | 276.285 | 56.983 |
| **M.4** | 2758.152 | 1911.043 | 278.131 | 59.323 |
| **M.5.1** | 3016.608 | 1990.237 | 277.974 | 59.812 |
| **M.5.2** | 3951.442 | 2374.752 | 281.845 | 59.304 |

**Table 4**

*Stage II, Step 1 Model Selection: Hypothesis tests*

*— p-values for the two sets of sequential F-tests*

| $H_0$ | $H_A$ | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|---|
| **M.2.1** | **M.1** | 0.0027 | 0.0002 | 0.0208 | 0.3267 |
| **M.2.2** | **M.1** | 0.0425 | 0.0816 | 0.0914 | 0.7745 |
| **M.3** | **M.2.1** | 0.1457 | 0.1922 | 0.1320 | 0.6243 |
| **M.3** | **M.2.2** | 0.0102 | 0.0006 | 0.0308 | 0.2432 |
| **M.4** | **M.3** | 0.0001 | 0.3759 | 0.0000 | 0.0000 |
| **M.5.1** | **M.4** | 0.0000 | 0.0000 | 0.1503 | 0.0000 |
| **M.5.2** | **M.4** | 0.0000 | 0.0000 | 0.0000 | 0.1207 |

**Table 5**

*Stage II, Step 2 Model Selection: $C_p$ values for four models in each case*

| Model | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|
| all nonlinear | 2732.020 | 1901.545 | 275.956 | 56.983 |
| linear functions on $x_1$ | 2749.483 | 1916.833 | 279.406 | 57.073 |
| linear functions on $x_2$ | 2886.450 | 1980.445 | 275.937 | 56.916 |
| linear functions on both $x_1$ and $x_2$ | 2946.050 | 1999.744 | 279.798 | 57.034 |

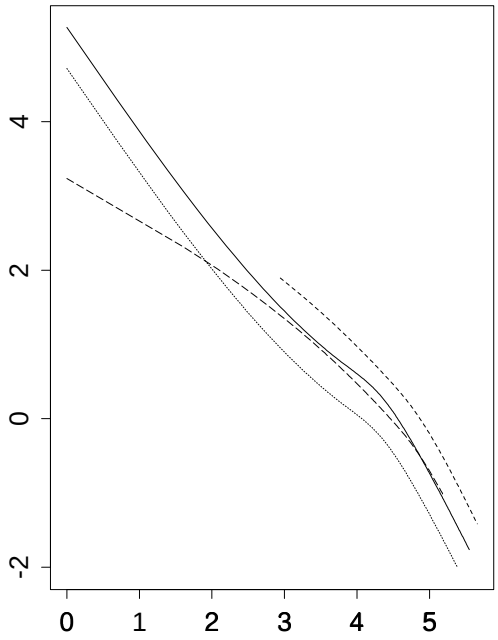
**Figure 1.** Presentation of the estimated smooth functions in the final set of equations satisfying identifiability constraints (3). Panel (a) displays $\hat{f}_L$'s, panel (b) displays $\hat{f}_P$'s, panel (c) displays $\hat{f}_{A1}$'s, and panel (d) displays linear $\hat{f}_{A2}$'s. In all four panels, solid line represents functions associated with part 1 control group, dotted line represents functions associated with part 1 treatment group, short-dash line represents functions associated with part 2 control group, and long-dash line represents functions associated with part 2 treatment group.

**Figure 2.** Estimated difference between functions due to cadmium treatment effect. Panel (a) displays the estimated difference between $f_{P,T,1}$ and $f_{P,C,1}$, and panel (b) displays the estimated difference between $f_{P,T,2}$ and $f_{P,C,2}$.

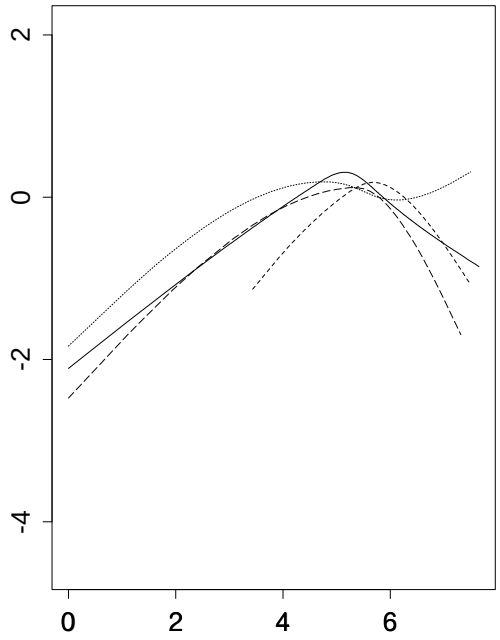
**Figure 1**

*Estimated smooth functions in the final set of equations satisfying identifiability constraints (5).*

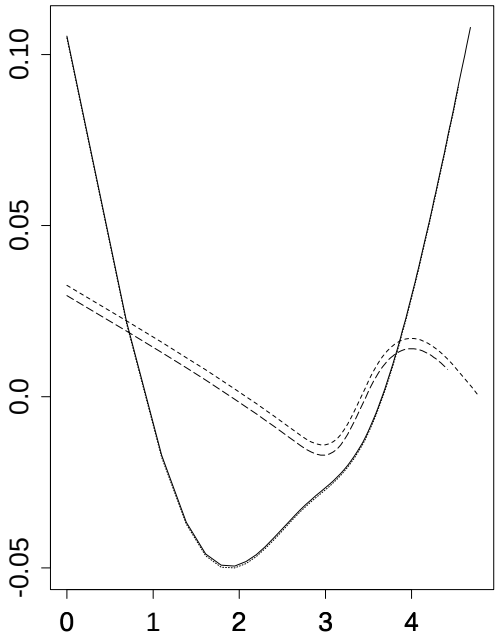
(a) $\hat{f}_L$ against $\log(A^M_{t-4})$




(b) $\hat{f}_P$ against $\log(L_{t-1})$




(c) $\hat{f}_{A1}$ against $\log(A^I_t)$




(d) $\hat{f}_{A2}$ against $\log(A^M_t)$

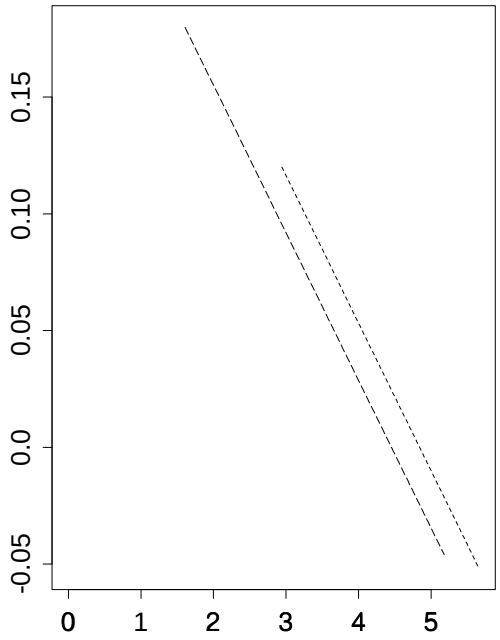
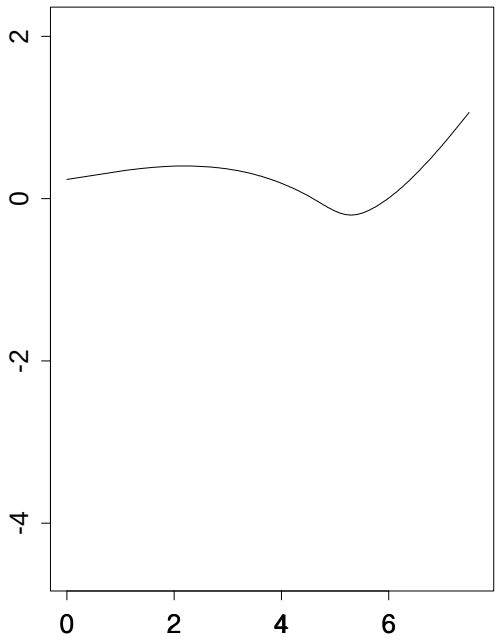


**Figure 2**

*Estimated difference between functions, $f_P$, due to cadmium treatment effect.*

$\hat{f}_{PT}$ against $\log(L_{t-1})$

(a) Part 1




(b) Part 2