

## **Chapter 1:**

# **A Study of the Statistical Evidence for Global Warming: Can the Normal Linear Regression Model Detect Global Warming?**

*“Greenhouse gases are accumulating in Earth’s atmosphere as a result of human activities, causing surface air temperatures and subsurface ocean temperatures to rise. Temperatures are, in fact, rising.”*

(United States Climate Action Report. 2002. pg. 251.)

## **Section 1: Introduction and Problem Statement**

From the Kyoto Protocol to domestic policies to restrict carbon emissions to proposed taxes on fuel and coal derivatives, national and international governments have embraced the call to combat global warming.<sup>1</sup> There remains little doubt in the minds of many in the scientific community that global warming is real and verifiable.

There are many ways to measure climate change; one factor that has received considerable attention by researchers and the public is surface air temperature.<sup>2</sup> The data available from monitoring stations and satellites around the world indicate the existence of a measurable upward trend in surface temperatures.<sup>3</sup> Because of such evidence and the increasingly evident link between the rate of temperature increases and human economic activity, governments have proposed numerous environmental and economic policies, often times at great social and financial cost, in order to slow down the rate of warming in the atmosphere.<sup>4</sup> The justification for these expenditures is that “temperatures are, in fact, rising,” as noted in the introductory quote.

### ***Problem Statement***

In the face of such costs, an important question to ask is how reliable are the empirical analyses upon which the evidence of global warming is founded. These analyses rely on data collected from several sources. The empirical analysis of global temperature data consists of the use of climate and statistical models (Angell, 2002;

---

<sup>1</sup> For information on United States policies: US Global Change Information Research Office (<http://www.gcrio.org>). For information on international policies: Intergovernmental Panel on Climate Change (<http://www.ipcc.ch/>).

<sup>2</sup> Easterling *et al.* (1999), US Climate Action Report (2002), Trenberth (2000).

<sup>3</sup> Trenberth (2000), IPCC (2001a, b), Lugina *et al.* (2001), Sterin (2001).

<sup>4</sup> US Climate Action Report (2002), “The Consumer Impact of Global Warming Policies: Some Sobering Facts.” (2002), IPCC (2001a), Taylor (1998).

Sterin, 2001; Folland *et. al.*, 2001; IPCC, 2001a; Trenberth, 2000). This paper focuses on the empirical analysis of temperature data using statistical models. Statistical analysis of trends in the temperature data involves the search for a statistically significant positive trend in the data—interpreted as evidence of global warming (Seidel and Lanzante, 2004; Angell, 2002; Sterin, 2001; Zheng and Basher, 1999; Folland *et. al.*, 2001; IPCC, 2001a; Trenberth, 2000).

The traditional method has been to fit a linear trend model to the data in order to assess the evidence for global warming (Seidel and Lanzante, 2004; Benestad, 2003; Zheng and Basher, 1999; Fomby and Vogelsang, 2000; Morris, 1999). However, there is increased recognition that the linear trend assumption is not appropriate for temperature series and that a non-linear function may better characterize trends in temperature patterns (Seidel and Lanzante, 2004; Benestad, 2003; Polsky *et. al.*, 2000; Harvey and Mills, 2000).

Polsky *et. al.* (2000) estimated temperature trends in the north-eastern United States using non-linear trend functions. Benestad (2003) modeled non-linear temperature trends in the Nordic region of Scandinavia using a cubic polynomial and found significant periods of warming and cooling. Seidel and Lanzante (2004) testes three alternative models that incorporate linear slopes and instantaneous step changes using surface temperature anomaly data and found that two of the three alternative models provide a better fit to the data than the model with a linear trend. Generalizing from a step transition to a smooth transition, Harvey and Mills (2000) test for trends in global temperature using logistic smooth transition functions and find that these functions provide a better fit to the temperature series than a linear trend.

In spite of the conclusions that the linear model does not properly characterize trends in temperature series, there is evidence that fitting a linear trend model to temperature series remains a common practice (Benestad, 2003; Zheng and Basher, 1999; Angell, 2002; Fomby and Vogelsang, 2000; Morris, 1999). This study provides formal evidence based on the results of a full set of misspecification tests, that the trend estimate from a linear model is not reliable. The study also contributes to the literature by estimating an alternative model including higher-order trends and lags that will be shown to be an appropriate model for the series based on misspecification test results.

Few studies that assess the statistical evidence for trends in global temperatures present the results from a full set of misspecification tests confirming the statistical adequacy of the statistical model. In other words, few studies provide evidence to confirm that the assumptions underlying the statistical model employed are adequate for the data. The search for a statistically adequate model is necessary because the validity of any inferences drawn from these models depends crucially on whether or not the model assumptions are valid for the data.

Benestad (2003) presents Q-Q plots of model residuals that indicate that the assumption of normality is adequate for the data. Seidel and Lazante (2004) present graphical plots of the residuals from the alternative models specified in the study as well as a table of root mean square errors from the models. However, to my knowledge, no one has reported the results of a comprehensive set of misspecification tests nor illustrated how to subsequently re-specify the model in the event that the model is misspecified. The complete set of misspecification tests provides formal tests of the assumptions relating to normality, parameter time invariance, heteroscedasticity, and independence (McGuirk *et. al.*, 1993). These assumptions are implicit in the formulation of a parametric time series model. Lacking formal misspecification tests, the reader cannot assess the overall reliability of the model.

## **Objective**

The objective of this study is to re-examine global surface temperature time series for trends in global temperature patterns paying careful attention to whether the assumptions underlying the statistical model formulated are appropriate for the data being analyzed.

## **Procedures**

Section 2 discusses certain pitfalls in the empirical analysis of temperature data, including the problems associated with fitting a linear trend model to the data. Section 3 lays out the formal statistical model that underlies the linear trend model, including model assumptions. An alternative model that includes more realistic assumptions is proposed and compared with the linear trend model. Section 4 presents the results from

an analysis of two temperature anomaly series based on the linear model and on an alternative model. The diagnostic test results show that the alternative model is an improvement over the linear model in terms of the fit to the data. The alternative model leads to different conclusions about temperature trends than those derived from the linear model.

This study does not address the important issue of measurement error, which is a real concern for environmental variables such as those used here.<sup>5</sup> Instead, the variables used to investigate warming trends in surface temperatures are assumed to be legitimate and the goal is to then specify a model that best captures the probabilistic features of these data.

## **Section 2: Pitfalls in Temperature Data Analysis**

A literature review revealed that climate change studies typically rely on annual series at a given spatial level (i.e. regional, country, or global) and to a lesser extent, on monthly series.<sup>6</sup> In most cases, daily records supplied by monitoring stations in the region of interest over a given period form the basis for these annual series.<sup>7</sup> Scientists first calculate the daily records as the average of the actual minimum and maximum temperature reading supplied by each station and then averaged over time in order to determine the monthly series for each station.<sup>8</sup>

If the researcher is interested in a regional monthly series, they will average the monthly series for each station within the region. If the researcher is interested in an annual series, the monthly series for each station undergo one more round of averaging over time at the station level in order to get the annual series for each station. They then average these annual series over the stations in order to determine the annual series for the region of interest.

---

<sup>5</sup> See Trenberth (2000) for a good discussion of the possible sources of bias in temperature data.

<sup>6</sup> IPCC (2001a), Trenberth (2000), Lugina *et. al.* (2001), Sterin (2001), Angell (2002) make reference to studies on global warming that base their findings on trends apparent in monthly or annual time series and/or use annual time series for their own analysis. IPCC (2001b) contains one of the few references to the analysis of trends in *daily* temperature data.

<sup>7</sup> USHCN, Hansen *et. al.* (2001), Angell (2002), Jones *et. al.* (2001), Easterling *et. al.* (1999), Lugina *et. al.* (2001).

<sup>8</sup> Hansen *et. al.* (2001) describes this method as the “standard way of calculating mean monthly temperature in the United States” (p. 3). Easterly *et. al.* also describe the method researchers commonly use to develop mean monthly and annual temperature time series using the mean daily temperature series.

There are two possible sources of distortion in monthly and annual series. Firstly, constructing the data necessarily introduces distortions because not all individual stations provide complete records for every month in every year. Therefore, not all stations provide complete records for the same months in the same year. For instance, in the case of a monthly series calculated using 150 stations in a given region of the United States, an individual monthly observation in the series may be calculated as the average of monthly observations from 45 stations while another monthly observation may be calculated as the average of monthly observations from 145 stations.

A monthly series calculated in this way will have few gaps in observations over time. However, it is not a series on mean monthly observations from 150 stations since not all individual observations are means over 150 stations. On the other hand, if the researcher wants to include only those months for which there are exactly 150 individual monthly station observations, there may be substantial gaps in the time series degrading the quality of the analysis.

The only way to address these problems is preemptively; by developing as complete a database as possible given the available technology and monitoring infrastructure. Implementing such improvements is beyond the reasonable reach of most applied researchers and this paper does not address such data problems.<sup>9</sup> However, this paper does address one major pitfall in data analysis that the researcher can reasonably avoid: fitting an incorrect statistical model to the data and drawing conclusions from the misspecified statistical model.

It is quite common to fit a simple model to the data in order to determine whether there is a statistically significant linear trend in the data (Seidel and Lanzante, 2004; Fomby and Vogelsang, 2000; Morris, 1999; Angell, 2002).<sup>10</sup> When analyzing air quality data, this simple model has been described as a way to “get a quick and useful overall estimate of the trend” (Schmidt, M.). The general model takes the form:

---

<sup>9</sup> Another potential problem with conducting analyses using station records is that the researcher is attempting to draw inferences about a process that manifests itself over hundreds of years using data of about 100 years whose overall quality is still debated (McKittrick, 2001; Trenberth, 2000). To some extent, this will always remain a drawback to the empirical analyses of temperature data. Meanwhile, researchers can continue to employ various tools of statistical analysis and diagnostic testing to learn as much as possible from the data that are available.

<sup>10</sup> For example, Angell (2002) employs a linear least squares regression to examine the existence of a trend in the data.

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where  $y_t$  is a time series on surface temperature anomalies measured over  $T$  time periods,  $t$  is the linear trend and  $\beta_1$  is the average change in temperature anomalies over the time period. Model (1) is most appropriate when dealing with data averaged over a large temporal and geographic scale, i.e. when dealing with annual data on global surface temperature (Schmidt, M.).

However, many researchers have realized that the model residuals from (1) tend to be serially correlated and that the standard t-statistic used to determine the significance of the trend parameter may thus be inflated (Seidel and Lanzante, 2004; Schmidt, M.; Fomby and Vogelsang, 2000; Fomby and Vogelsang, 2003). The presence of serial correlation might be indicative of departures from the assumption that the process is temporally dependent in which case the t-test is invalid. However, apart from independence, there are other statistical assumptions underlying the model in (1) and these must also be satisfied in order for the t-statistic to have statistical validity.

The Ordinary Least Squares (OLS) method of parameter estimation does not necessarily involve a distributional assumption (Spanos, 1999). However, statistical inference regarding the parameters estimated requires that a distributional assumption be made explicitly or indirectly via asymptotic arguments. The assumption typically made is that the data are drawn from a Normal distribution (Greene, 1997; Spanos, 1999.). If the assumption of a Normal distribution is not correct, any inference regarding the statistical significance of the parameter estimates will be invalid and likely misleading to a lesser or a greater degree.<sup>11</sup> The common assumption is that global surface temperature anomalies, having undergone several rounds of averaging over several hundreds or thousands of individual station observations, are normally distributed (Schmidt, M; Hansen *et. al.*, 2001).

Independence and normality are key assumptions underlying a model like (1), but these are not the only assumptions that must be satisfied. The model assumes a homoskedastic variance, and assumes not just linear independence but higher-order independence as well. If all of these assumptions do not hold, the t-statistics may be

---

<sup>11</sup> The inference may be reliable asymptotically. However, the annual temperature data analyzed in the literature typically contain around 100 observations and it is not clear that asymptotic properties will necessarily hold for a dataset of this size.

unreliable for statistical inference purposes. Most importantly, a misspecified statistical model cannot offer reliable conclusions regarding the significance and size of the trend estimate.

The following section illustrates why model (1) is invalid for global temperature series and presents an alternative model.

## **Section 3: Empirical Analysis of Temperature Anomaly Time Series**

### ***Model specification***

Model (1) may be inadequate for a time series because it is not a dynamic model. With time series data, observations may be dependent over time. In the case of temperature data, it is likely that today's temperature will be correlated with yesterday's temperature. If lagged dependent variables belong in the regression, omitting these lagged variables from the regression will lead to biased and inconsistent parameter estimates.<sup>12</sup>

Further, model (1) may not be adequate for the temperature time series if the data are trending nonlinearly rather than linearly. Simple graphical analyses could provide insight into the nature of this trend. Recent empirical research on trends in temperature series indicates that the linear trend term in model (1) does not adequately capture trends in temperature patterns over time.

Possible model misspecification can be easily avoided if the researcher estimates a model such as model (1) and then conducts a comprehensive set of diagnostic or misspecification tests (tests of the statistical assumptions underlying the model). If the model requires the inclusion of lagged dependent variables or nonlinear trend terms, misspecification tests will likely indicate that the model is misspecified.

The following table specifies a variant of model (1) which retains the distributional assumption of Normality but includes nonlinear trends and lags.<sup>13</sup> The reformulated model that includes nonlinear trends and lags is distinguished from model

---

<sup>12</sup> See Spanos (1987) for an insightful discussion of the omitted variable bias problem in model estimation.

<sup>13</sup> See Spanos (1999): 761-762 for the specification of the Normal autoregressive model with trend



(1) by referring to the latter as the static model and to the former as the dynamic model, even though the dynamic (that is, autoregressive) structure of the reformulated model is not its only distinguishing feature.

Table 1 compares and contrasts the dynamic model with the static model in terms of the probabilistic, statistical, and sampling assumptions that the statistical model imposes on the data.

**Table 1: Comparing the static and dynamic Normal regression models**

<i>Assumption</i>	Static Model	Autoregressive Model
<i>Statistical Generating Mechanism</i>	$y_t = \alpha_0 + \alpha_1 t + \varepsilon_t,$ $t \in \mathfrak{T}$	$y_t = \alpha_0 + \sum_{k=1}^P \gamma_k t^k + \sum_{i=1}^l \beta_i y_{t-i} + \varepsilon_t,$ $t \in \mathfrak{T}$
<i>Distribution</i>	$f(y_t; \phi)$ , is Normal	$f(y_t   Y_{t-1}^0; \theta)$ , is Normal
<i>Mean</i>	$E(y_t) = \alpha_0 + \alpha_1 t$ is linear in $t$	$E(y_t   \sigma(Y_{t-1}^0)) = \alpha_0 + \sum_{k=1}^P \gamma_k t^k + \sum_{i=1}^l \beta_i y_{t-i},$ is linear in $y_{t-l}$ *
<i>Variance</i>	$Var(y_t) = \sigma_y^2$ is the unconditional variance of $y$	$Var(y_t   \sigma(Y_{t-1}^0)) = \sigma_{y Y_{t-1}}^2$ , is the conditional variance of $y^*$
<i>Homogeneity</i>	$\phi := (\alpha_0, \alpha_1, \sigma_y^2)$ are $t$ - homogeneous $\forall t \in \mathfrak{T}$	$\theta := (\alpha_0, \gamma_1, \dots, \gamma_P, \beta_1, \dots, \beta_l, \sigma_{y Y_{t-1}}^2)$ are $t$ - homogeneous $\forall t \in \mathfrak{T}$ , where $P$ is the order of the trend, and $l$ is the order of the lag.
<i>Dependence</i>	$\{(\varepsilon_t), t \in \mathfrak{T}\}$ is a white noise process**	$\{(\varepsilon_t   Y_{t-1}^0), t \in \mathfrak{T}\}$ is an innovation process**
<i>Sampling</i>	$\{Y_t\}$ is a covariance stationary, independent and mean heterogeneous sample (due to a linear trend)	$\{Y_t\}$ is a covariance stationary, Markov( $l$ ) dependent, and mean heterogeneous sample (due to linear and higher order trends)

\*The conditioning set,  $\sigma(Y_{t-1}^0)$ , is the sigma-field of past values of the conditioning variable up till the value of “ $l$ .”

\*\* The properties of the white noise process (an unconditional mean equal to zero and a time-invariant unconditional variance) are defined in terms of the marginal distribution, while the properties of the innovation process (zero conditional mean, time-invariant and homoskedastic conditional variance, and zero covariance between the variables in the series) are defined in terms of the conditional distribution.

## **Data**

Two widely used and publicly available datasets will be used to conduct the analysis (Carbon Dioxide Information Analysis Center, 2003). P. Jones of the Climatic Research Unit, University of East Anglia, U. K., along with colleagues, compiles and continually updates the first dataset, hereafter referred to as the Jones dataset. The Jones dataset “has been used extensively in various Intergovernmental Panel on Climate Change (IPCC) reports” (Jones *et. al.*, 2001). Seidel and Lanzante (2004), Fomby and Vogelsang (2003), and Harvey and Mills (2000) are among several studies to make use of the Jones dataset. The dataset consists of global and hemispheric temperature anomaly time series, which incorporate land and marine data (Jones *et. al.*, 2001). The annual temperature anomalies cover the years 1856-2000 and are relative to the 1961-1990 reference period means.

The land data consists of surface temperature data collected from land-surface meteorological equipment and fixed-position weather ships. This data has undergone routine quality control adjustments, which correct for nonclimatic errors such as station shifts and/or instrument changes. Recent updates to the data have included the addition of up to 1000 more monitoring stations and a new reference period common to all stations (1961-1990; previously 1950-1979). The marine data consists of sea surface temperature data collected from ships and buoys. The marine data has undergone quality control adjustments and has been converted to anomalies relative to the 1961-1990 reference period means.

K. M. Lugina of St. Petersburg State University, St. Petersburg, Russia, along with colleagues, compiles and continually updates the second dataset, hereafter referred to as the Lugina dataset (Lugina *et. al.*, 2003). The data consists of mean annual values of surface air temperature anomalies covering the time period 1881-2002 and the anomalies are relative to the 1951-1975 reference period means. Lugina and colleagues compile the data from several published sources: *World Weather Records*, *Monthly Climatic Data for the World*, and *Meteorological Data for Individual Years over the Northern Hemisphere Excluding the USSR* (Lugina *et. al.*, 2003).

The researchers made quality control adjustments to the data by removing some stations and adding others, primarily to ensure homogeneity and unbiasedness in the

reported values from the individual stations. A station record is deemed to be homogeneous if it is “a record in which the temperature change is due only to local weather and climate” (Hansen *et. al.*, p. 3).

The quality control adjustments performed on the Jones dataset and on the Lugina dataset are assumed to control adequately for any systematic measurement error in the surface temperature series. Therefore, this study does not specifically address the issue of measurement error as a factor in determining the statistical adequacy of the linear regression model.

### ***Estimation Results***

The goal of this section is two-fold: 1) to reproduce published results about the existence of a linear trend in global temperatures and then show that the model that produced the reported trend estimates is misspecified; 2) re-specify the model that estimates trends in the temperature series and show that the re-specified model is statistically adequate for the data. Published trend estimates are available for the Lugina dataset but are only available for earlier versions of the Jones dataset.

According to Jones and Briffa (1992) (as referenced in Jones *et. al.*, 2001), an analysis of an earlier version of the Jones dataset indicates that “the average surface air temperature of the globe has warmed  $\sim 0.5^{\circ}\text{C}$  since the middle of the nineteenth century” (Jones *et. al.*, 2001). Using an earlier version of the Jones dataset than the one used in the present analysis, but a later version than that used by Jones and Briffa (1992), Fomby and Vogelsang (2003) estimate a linear trend parameter of  $0.45^{\circ}\text{C}$  for a series running from 1857-1998. The IPCC report, which uses the Jones dataset, reports an increase in surface air temperatures in the range of  $0.3$  to  $0.6^{\circ}\text{C}$ , stating that “additional data since 1990 and the reanalyses since then have not significantly changed this range of estimated increase” (IPCC, 2001).

The Lugina series, updated through 2002, shows that “the warming rate for the globe [...] is  $0.60^{\circ}\text{C}/100$  yrs” (Lugina *et. al.*, 2003). In what follows, the linear trend model is estimated and then the estimated trends from the model are compared with the published trend estimates. The subsection begins with a discussion of summary statistics.

Table 2 presents summary statistics for the annual temperature series from each dataset. The usefulness of these statistics depends crucially on the validity of the assumption of second-order stationarity (the mean and variance are constant and the covariances between any two observations in the series only depend upon the time interval between the two observations on temperature anomalies).

**Table 2: Summary statistics and p-values on global temperature anomalies**

<i>Variable</i>	<i>Jones dataset</i>	<i>Lugina dataset</i>
Mean	-0.1504828	-0.00503
Standard Error	0.01893003	0.02357
Median	-0.19000	0.01667
Standard Deviation	0.22794776	0.26038
Skewness	0.61787051	0.3465
Kurtosis	-0.025993	0.03487
Minimum	-0.5300000	-0.59667
Maximum	0.5900000	0.72917
T	145	122
Misspecification test	p-value	p-value
SW	0.0009	0.1977*
KS	<0.0100	>0.1500*
CvM	<0.0050	0.2356*
AD	<0.0050	0.1856*

SW: Shapiro-Wilk Normality test.

KS: Kolmogorov-Smirnov Normality test

CvM: Cramer-von Mises Normality test

AD: Andersen Darling Normality test

\* p-value > 0.05 implies that there is only weak evidence against the null hypothesis that data is drawn from a Normal distribution.

There are some differences between the two datasets that are worth noting. Significance tests of the mean of each time series indicate that the mean of the Jones data is significantly different from zero (p-value<0.0001) while the mean of the Lugina data is not significantly different from zero (p-value=0.8315). Therefore, assuming these summary statistics are valid, there is no deviation from the mean in the Lugina time series of temperature anomalies. The distributional tests also reveal differences between the two series. For the Lugina data, there is only weak evidence against the Normality assumption while in the case of the Jones series all four distributional tests indicate that there is strong evidence in the data against the assumption of Normality. Thus, assuming that the summary statistics are valid in the first instance, the Jones data do not appear to be Normally distributed. The next section presents the results from fitting model (1) to the data.

### ***Static Model***

The specification of the static model is duplicated here for convenience.

$$y_t = \alpha_0 + \alpha_1 t + u_t \quad (1)$$

Model (1) is estimated for each of the datasets, where  $y$  is the average annual surface temperature anomaly at time  $t$ . The results are presented in Table 3. The table presents parameter estimates, the associated standard errors, and p-values associated with the various misspecification tests.

**Table 3: Static Normal model estimation and misspecification testing results**

<i>Variables</i>	<i>Jones dataset</i>	<i>Lugina dataset</i>
$\alpha_0$	-0.4638** (0.0234)	-0.3750** (0.0275)
$\alpha_1$	0.0043** (0.0003)	0.0060** (0.0004)
$\sigma^2$	0.0196	0.0227
Adj. R <sup>2</sup>	0.6229	0.6649
Observations	144	121
Misspecification Tests	p-values	p-values
BJN	0.891*	0.618*
DPN	0.869*	0.621*
DSK	0.62*	0.329*
DKU	0.85*	0.964*
Results of Individual Tests	p-values	p-values
L(2)	2.05e-009	0.117*
AC(2)	1.67e-015	1.76e-008
ARCH(2)	0.00118	0.285*
H(2)	0.326*	0.0318
Results of Joint Unconditional Mean Tests	p-values	p-values
Overall	4.44e-016	0.05412
Trend(2)	0.95*	0.05544
L(2)	0.564*	0.00013
AC(3)	9.55e-010	0.00000
Results of Joint Unconditional Variance Tests	p-values	p-values
Overall	0.00399	0.085
Trend(2)	0.161	0.0453
H(2)	0.404*	0.681*
ARCH(3)	0.00228	0.413*

The *parameter estimates* with a double asterisk (\*\*) are significant at a 1 percent level if the model is correctly specified. The parameter estimates with a single asterisk (\*) are significant at a 5 percent level but not at a 1 percent level if the model is correctly specified. Standard errors are in parentheses.

The *p-values* with an asterisk (\*) indicate that there is only weak evidence against the original model assumptions vis-à-vis the assumptions built into the particular misspecification test.<sup>14</sup>

<sup>14</sup> Spanos (1999) has shown that p-values can be used as indicators of the evidence provided by the data against the null hypothesis under the assumption that the original model is the true model. I use the approximate cutoff of 0.1 to signify the level beyond which the data cease to provide strong evidence against the null hypothesis that the original model is the true model (see Spanos, (1999), p. 690 for general guidelines on interpreting p-values).



BJN:	Bera-Jarque test for Normality
DPN:	D'Agostino-Pearson test for Normality
DSK:	D'Agostino-Pearson test for Skewness=0
DKU:	D'Agostino-Pearson test for Kurtosis=3
H(x):	RESET test for static heteroskedasticity of order $x$
L(x):	RESET test for nonlinearity of order $x$
ARCH(x):	Test for dynamic conditional heteroskedasticity of order $x$
AC(x):	Test for autocorrelation of order $x$

As can be seen from the results in Table 3, the parameter estimates on the trend coefficient are positive and statistically significant, indicating that there is evidence of a warming trend in global surface air temperatures. Table 4 summarizes the published information and compares the published trend estimates with the reproduced trend estimates, with standard errors in parentheses.

**Table 4: Comparison between reported and estimated trend**

<b>Jones et. al.</b>			
<ul style="list-style-type: none"> <li>• 1856-2000 (temperature anomalies relative to 1961-1990)</li> <li>• used extensively in various IPCC reports</li> </ul>			
Latitude band	Area	Published trend estimate	<i>Reproduced trend estimate</i>
n/a	Global	0.45°C	0.43 °C(0.03)
<b>Lugina et. al.</b>			
<ul style="list-style-type: none"> <li>• 1881-2002 (temperature anomalies relative to 1951-1975)</li> </ul>			
Latitude band	Area	Published trend estimate	<i>Reproduced trend estimate</i>
0°-90°N	Northern hemisphere	0.65°C/century	0.65 °C/century(0.05)
0°-60°S	Southern hemisphere	0.54°C/century	0.52 °C/century(0.03)
90°N-60°S	Globe	0.60°C/century	0.60 °C/century(0.03)

The results in Table 4 indicate that model (1) reproduces the published results. While there is no published trend estimate using the most recent version of the Jones dataset, the reproduced trend estimate of 0.43°C is close to the trend estimate of 0.45°C that was reported using an earlier version of the Jones dataset. The reproduced trend estimate also lies within the range reported by the IPCC: “0.3 and 0.6°C since the late 19<sup>th</sup> century” (IPCC, 2001), a range which, according to the report, has not changed significantly even with updated data. In the case of the Lugina data where published trend estimates using the same series are available, the estimated trends using model (1) mirror the reported estimates. Using both the Jones and Lugina datasets, the estimation

of the model results in a trend estimate that is positive and highly statistically significant, judging from the standard errors. Thus, there appears to be statistical evidence of a linear increase in mean annual global temperatures based on these datasets.

However, before drawing conclusions based on hypothesis tests on the model coefficients, it is important to conduct misspecification tests in order to check whether the model satisfies the underlying assumptions of linearity, independence, homoskedasticity and homogeneity (t-invariance).

The basic logic behind misspecification tests is this: if the estimated mean function is a correct characterization of the mean function of the distribution (i.e. Normal, Student's t, etc.) from which we assume the data are drawn, then the mean function captures all the systematic information and the residuals will exhibit no systematic patterns such as dependence, nonlinearity, etc. Furthermore, the squared residuals will exhibit only the patterns specified by the underlying distribution (in the case of the Normal distribution, the squared residuals should exhibit no patterns since we assume that the variance is homoskedastic and t-invariant). If the residuals exhibit systematic patterns, then the mean function is not well-specified for the data, either because the distributional assumption is correct but the mean function is incomplete, or because the distributional assumption is incorrect and a different mean function is required altogether.

In many cases, the assumption of a Normal distribution will be adequate but the mean function will be incorrectly specified. The preliminary distributional tests of the time series indicated that only the Lugina data followed a Normal distribution. According to the misspecification tests in Table 3, both time series (actually, the detrended time series) appear to be normally distributed. However, the misspecification tests indicate that neither the mean nor the variance is correctly specified for either of the time series data. There is evidence against the independence assumption for both datasets based on the misspecification tests and on the t-plots in Figure 1.

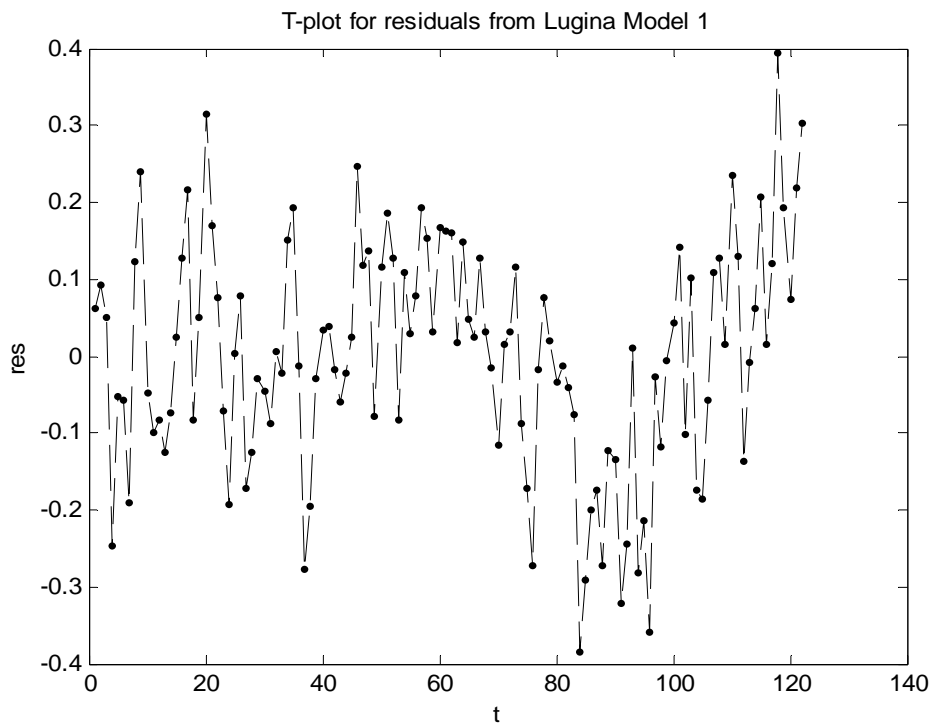
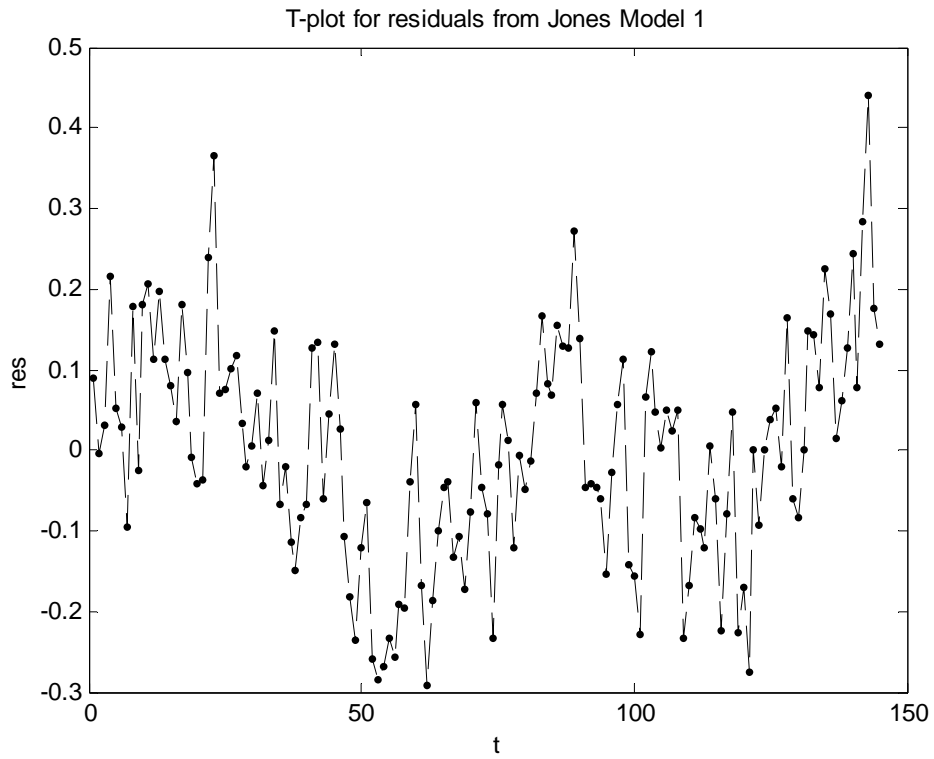


Figure 1: T-plots of residual from the static model using the Jones and Lugina datasets

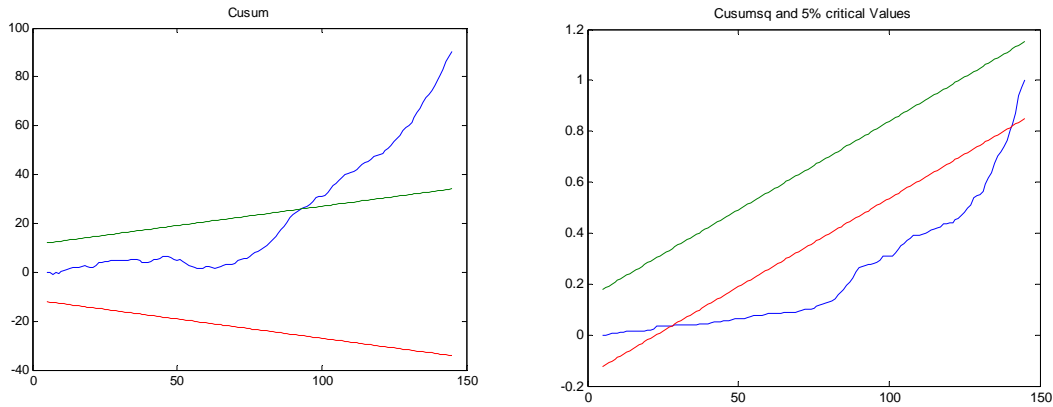
According to the misspecification tests, serial correlation in the residuals causes the joint mean tests to fail in the case of the Jones dataset. In the case of the Lugina dataset, the mean appears to be misspecified for various reasons: the residuals are trending, are correlated over time, and exhibit nonlinearity. The assumptions regarding the variance do not appear to be satisfied. Based on the individual tests and the overall joint variance test, the variance, which is assumed to be homoskedastic and  $t$ -invariant, is misspecified.

The assumption of parameter stability also needs to be assessed because the model assumes that the statistical parameters  $(\alpha_0, \alpha_1, \sigma^2)$  are  $t$ -invariant, or stable over time. The CUSUM and CUSUMSQ tests provide a test of whether the mean and variance parameters, respectively, are stable over time.<sup>15</sup>

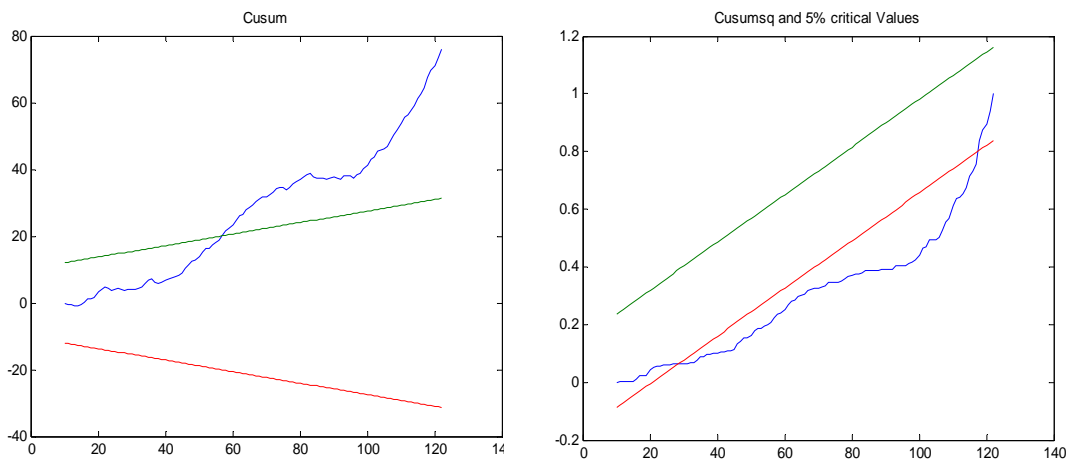
The CUSUM test assumes variance homogeneity and tests for mean homogeneity. The CUSUMSQ test assumes mean homogeneity and tests for variance homogeneity. Both statistics are based on standardized recursive residuals, which are residuals that are sequentially estimated by adding an observation each time, until the full sample is reached. If the plot of either statistic breaks out of the 5% confidence bounds that have been established for the mean of the statistic, there is indication of instability in the mean or the variance. Figures 2 and 3 present graphs of the CUSUM and CUSUMSQ tests for model (1) estimated using the Jones and Lugina series.

---

<sup>15</sup> Appendix A contains more recursive plots that check for parameter stability: plots of the recursive residuals, plots of the estimates of the conditional variance, and plots of the coefficient on the constant term for the Jones and Lugina datasets.



**Figure 2: CUSUM (a) and CUSUMSQ (b) graphs for the static model using the Jones dataset**



**Figure 3: CUSUM (a) and CUSUMSQ (b) graphs for the static model using the Lugina dataset**

The tests indicate that the parameters of the mean of the static model are unstable and that the instability increases over time. The plot of the CUSUM statistic breaks out of the confidence bounds and remains outside of the bounds for a significant portion of the graph for both data series. This result is not surprising given that the misspecification tests indicated problems with the mean function for the static model and that the mean function neither includes lagged dependent variables nor higher-order trends.

The CUSUMSQ test indicates that the variance is not stable. This result is not surprising given that the misspecification tests indicated problems with the variance. Thus, the data indicate that the variance is either heteroskedastic or more likely

heterogeneous (varies with  $t$ ), in violation of the underlying model assumptions of a constant variance.

The CUSUM test indicates that the mean function does not properly capture the systematic variations in the mean. If it did, the residuals would exhibit no systematic patterns over time, like correlation. Serial correlation in time series models is a major problem (Fomby and Vogelsang, 2000). One recognized consequence of serial correlation is that it inflates the size of the t-test (Fomby and Vogelsang, 2000; Fomby and Vogelsang, 2003). An increase in the size of the t-test means that the probability of rejecting the null hypothesis when it is true is now higher. In other words, an increase in the size of the test means that the researcher is more likely to reject the null hypothesis of a coefficient estimate equal to zero even when the coefficient estimate is equal to zero, in a probabilistic sense. That is, an inflated size means that the researcher is more likely to conclude that the trend coefficient is statistically significant when actually it may not be.

Thus, it is important to find ways to deal with serial correlation in the residuals. Fomby and Vogelsang (2000) develop a significance test that adjusts for the inflated size of the significance tests. Applying their methods to an earlier version of the Jones dataset used here, they find a statistically significant warming trend on a similar scale to the trend coefficient estimates in Table 4; their trend estimate for the Jones series is  $0.45^{\circ}\text{C}$  (Fomby and Vogelsang, 2003).

The approach taken here is to estimate a model that is statistically adequate for the data, dispensing with the need to adjust the significance test in order to account for serial correlation. The misspecification tests are crucial to the success of this approach in that they indicate when the statistical model is appropriate and satisfies the underlying probabilistic assumptions. These probabilistic assumptions in turn must be true in order for the significance test to yield useful and reliable conclusions about the significance of the trend coefficient estimate.

In summary, the misspecification tests indicate that the static model is not statistically adequate for the Jones and Lugina series in that the model assumptions are not satisfied. Furthermore, a graphical analysis of the recursive residuals indicates that

the conditional mean parameters and the conditional variance are unstable over time, despite model assumptions to the contrary.<sup>16</sup>

Given the inadequacy of these statistical models to capture the systematic information in the underlying data generating process, one cannot draw any reliable conclusions about a warming trend in the surface temperature based on the sign of the coefficient estimates of the static model. The task then is to find a well-specified model that leads to reliable conclusions about a warming trend using these data.

### ***Dynamic Model***

The misspecification tests indicate a need to capture information from the recent past and to include higher-order trends in the conditional mean therefore model (1) is modified accordingly. The problem with higher-order trends is that they are functionally dependent and can cause the data matrix to be of less than full rank. Orthogonal polynomials are, as their name implies, functionally independent. Hermite polynomials are orthogonal over the range of values represented by the real line and are also orthogonal with respect to the normal probability distribution.<sup>17</sup> Hermite polynomials will be used in the place of simple powers of  $t$  in the regression. The variable  $H(k)$  in the regression represents the Hermite polynomial of order  $k+1$ .<sup>18</sup>

The dynamic model specification that appears to be most appropriate for the Jones dataset is as follows:<sup>19</sup>

---

<sup>16</sup> See Appendix for graphs of the recursive plots.

<sup>17</sup> More information on Hermite polynomials and their relation to the normal probability distribution can be found at <http://mathworld.wolfram.com/HermitePolynomial.html> or at <http://encyclopedia.thefreedictionary.com/Hermite%20polynomials>

<sup>18</sup> The value of  $H(1)$  is 1.

<sup>19</sup> This specification in model (2) was arrived at through a combination of analysis of partial autocorrelations, sequential F-tests of lag lengths, and a literature review of studies of temperature anomalies using nonlinear trend terms. The graph of the partial autocorrelation function of order 10, including the 5 percent confidence intervals on the estimates, indicated that a lag length of 4 is appropriate for the Jones series. The lag length of 4 was tested against longer lag lengths (6 and 5). Longer lag lengths were strongly rejected in favor of a lag length of 4. A shorter lag length than 4 was not employed because the consequences of choosing the lag too small in terms of omitted variable bias were thought to be more severe than the consequences of overfitting. Concerning the order of the nonlinear trend, previous studies employing nonlinear trends favor using polynomials with an odd-numbered order (like the cubic function) over polynomials with even-numbered orders. A graphical analysis of the trending component of the regression model indicated that the 5<sup>th</sup>-order polynomial provided a better fit to the data than a cubic polynomial.

$$y_t = \alpha_0 + \sum_{k=1}^5 \alpha_k H(k) + \sum_{l=1}^4 \beta_l y_{t-l} + u_t \quad (2)$$

The specification that appears to be most appropriate for the Lugina series is as follows:<sup>20</sup>

$$y_t = \alpha_0 + \sum_{k=1}^4 \alpha_k H(k) + \sum_{l=1}^3 \beta_l y_{t-l} + u_t \quad (3)$$

The dynamic models in (2) and (3) are estimated using the Lugina and Jones datasets and Table 5 presents the estimation and misspecification test results. The table presents parameter estimates, the associated standard errors, and p-values associated with the various misspecification tests.

**Table 5: Dynamic Normal model estimation and misspecification testing results**

<i>Variables</i>	<i>Jones dataset</i>	<i>Lugina dataset</i>
$\alpha_0$	-0.2649* (0.0834)	-0.283* (0.08)
$\alpha_1$	0.0086 (0.0045)	-9.725e-4 (3.415e-3)
$\alpha_2$	-2.324e-4 (1.024e-4)	6.891e-5 (6.058e-5)
$\alpha_3$	2.363e-6* (9.442e-7)	-6.9e-7 (3.938e-7)
$\alpha_4$	-9.79e-9* (3.726e-9)	1.859e-9* (8.43e-10)
$\alpha_5$	1.433e-11* (5.28e-12)	---
$\beta_1$	0.4309* (0.0865)	0.385* (0.0948)
$\beta_2$	-0.1264 (0.0953)	-0.159 (0.1)
$\beta_3$	-0.0072 (0.0962)	0.0455 (0.9591)
$\beta_4$	0.1157 (0.0883)	---
$\sigma^2$	0.0107	0.0151
Adj. R <sup>2</sup>	0.7934	0.7758
Observations	145	122

<sup>20</sup> A similar approach to that used to specify (2) yielded the model specified in (3). Graphical analysis of the partial autocorrelation function of order 10 indicated that a lag length of 3 would be appropriate for the Lugina dataset. Longer lag lengths were strongly rejected and shorter lag lengths were avoided in order not to omit a relevant variable. Despite the preference in the literature for polynomials of an odd-numbered order, graphical analysis revealed that the 4<sup>th</sup>-order polynomial provided the best fit to the data.



Misspecification Tests		
BJN	0.429*	0.659*
DPN	0.305*	0.688*
DSK	0.905*	0.703*
DKU	0.125*	0.438*
Results of Individual Tests	p-values	p-values
L(2)	0.0162	0.2*
AC(2)	0.76*	0.432*
ARCH(2)	0.822*	0.691*
H(2)	0.121*	0.916*
Results of Joint Conditional Mean Tests	p-values	p-values
Overall	0.125*	0.282*
Trend(v) <sup>***</sup>	0.42*	0.801*
L(2)	0.136*	0.105*
AC(2)	0.717*	0.451*
Results of Joint Conditional Variance Tests	p-values	p-values
Overall	0.575*	0.814*
Trend	0.162*	0.364*
H(2)	0.432*	0.539*
ARCH(2)	0.848*	0.632*

The *parameter estimates* with an asterisk are significant at a 1 percent level. Standard errors are in parentheses.

The *p-values* with an asterisk indicate that there is only weak evidence against the original model assumption vis-à-vis the assumption built into the particular misspecification test.

Trend(v)<sup>\*\*\*</sup>: v=6 for the Jones dataset and v=5 for the Lugina dataset

BJN: Bera-Jarque test for Normality

DPN: D'Agostino-Pearson test for Normality

DSK: D'Agostino-Pearson test for Skewness=0

DKU: D'Agostino-Pearson test for Kurtosis=3

H(x): RESET test for static heteroskedasticity of order  $x$

L(x): RESET test for nonlinearity of order  $x$

ARCH(x): Test for dynamic conditional heteroskedasticity of order  $x$

AC(x): Test for autocorrelation of order  $x$

A superficial comparison between the estimation results for the static and dynamic models reveals several differences. The coefficient estimates on the linear trend are no longer significant in the dynamic model though they were highly significant in the static model. The estimate of  $R^2$  has increased from ~0.6 in the static model to ~0.8 in the dynamic model. The misspecification test results in Table 5 indicate that the dynamic model has captured the systematic information in the data, leaving no systematic information in the residuals.

Figure 4 presents a t-plot of the residuals from both datasets that shows no discernible pattern to the data. The t-plots of residuals from the dynamic model differ from the t-plots of residuals from the static model, where there was clear evidence of correlation in the data.

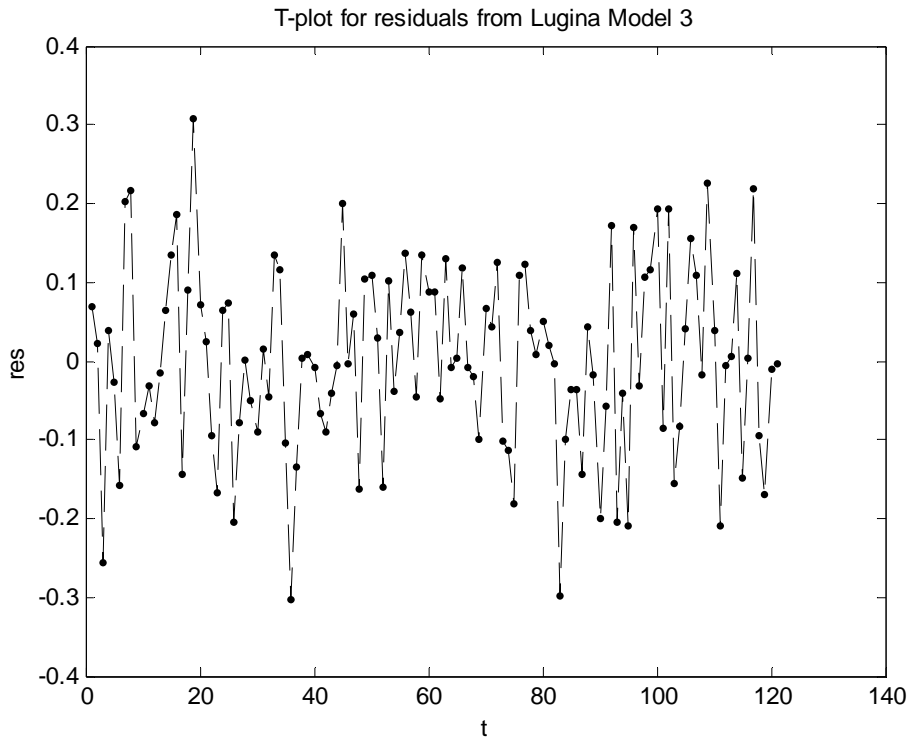
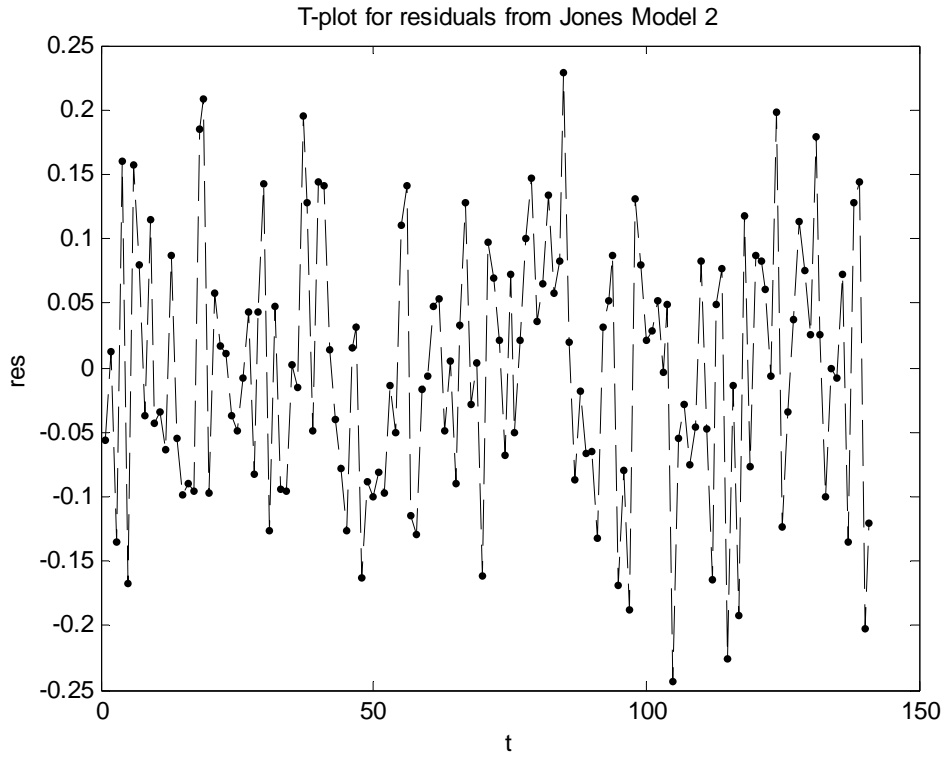
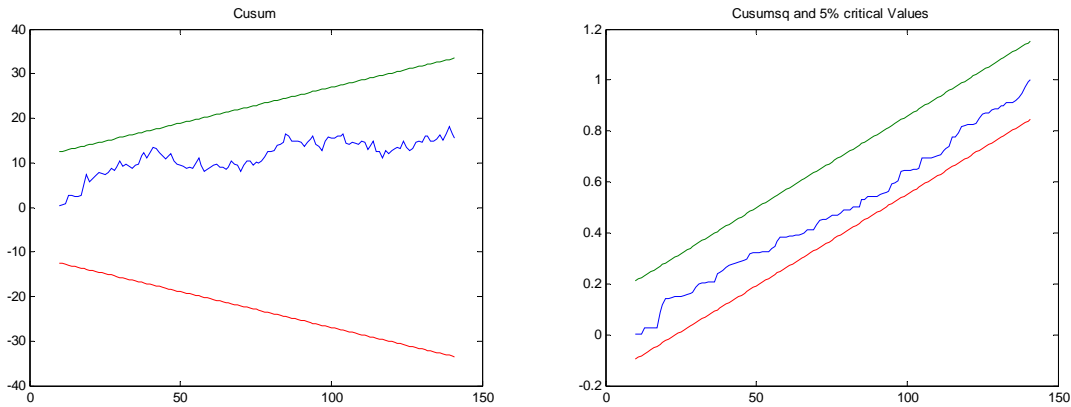
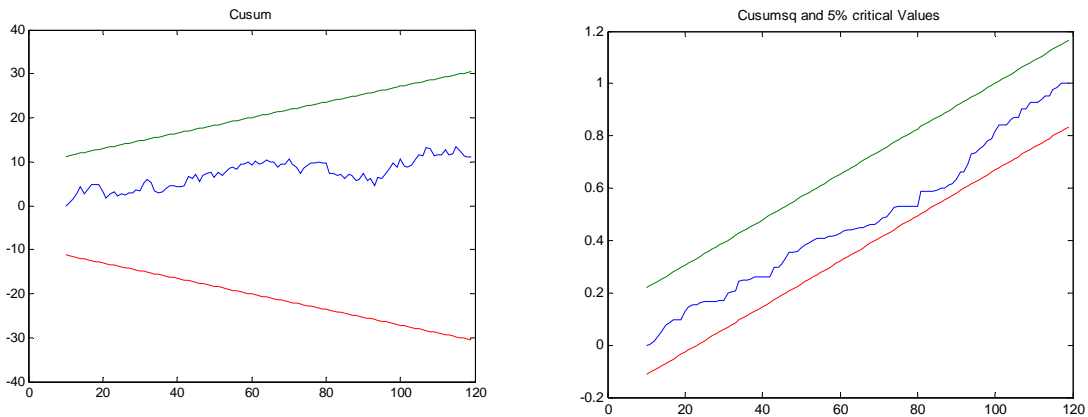


Figure 4: T-plot of residuals from the dynamic model using the Jones and Lugina datasets

Figures 5 and 6 present the graphs of the CUSUM and CUSUMSQ statistics, respectively, for each dataset.



**Figure 5: CUSUM and CUSUMSQ graphs for the dynamic model using the Jones dataset**



**Figure 6: CUSUM and CUSUMSQ graphs for the dynamic model using the Lugina dataset**

The graphs of the CUSUM and CUSUMSQ statistics indicate that the dynamic model does a better job of modeling the distributional moment parameters than does the static model. The values of the statistics stay within the confidence bounds, indicating that the parameters are stable over the time period in question. Furthermore, plots of the recursive values of the estimated variance, of the estimated coefficient on the constant

term, and of the recursive residuals present an overall picture of t-homogeneity in the model parameters, unlike the case with the static model.<sup>21</sup>

Since the model appears to be specified correctly, specification tests can now be conducted. From Table 5, the coefficient on the one-period lagged dependent variable in both models (2) and (3) is significant, indicating the presence of dependence in the data.

The coefficient estimates on the linear trend are not significant in either model (2) or model (3) at either the 1 percent or 5 percent level. The results present no statistical evidence of a linear increase in annual global temperature anomalies. This result is in contrast to the results from the static model where the coefficient estimate on the linear trend term is significant at the 1 percent level for both datasets. However, there are higher-order trends present in the temperature series. In the case of the Jones dataset, there is a 3<sup>rd</sup> and 5<sup>th</sup> order increase and a 4<sup>th</sup> order decrease in annual global temperature anomalies. In the case of the Lugina dataset, there is a 4<sup>th</sup> order increase in annual global temperature anomalies.

The following graphs plot the actual data series against the predicted trending component of the regression model estimated for each series.<sup>22</sup> For each data series, the first graph shows the predicted data series calculated from the linear trending component of the static model and the second graph shows the predicted data series calculated from the nonlinear trending component of the dynamic model.

The graphs in Figures 7 and 8 confirm that the trending component of the dynamic models provides a better fit to the data series and thus also provides a more reliable prediction of future trends. The graph also shows that the trend in the data series is highly nonlinear over the time period. Figures 7 and 8 show that the trend in temperature anomalies is not consistently positive, as a linear trend assumes. Rather, there are periods of cooling and warming.

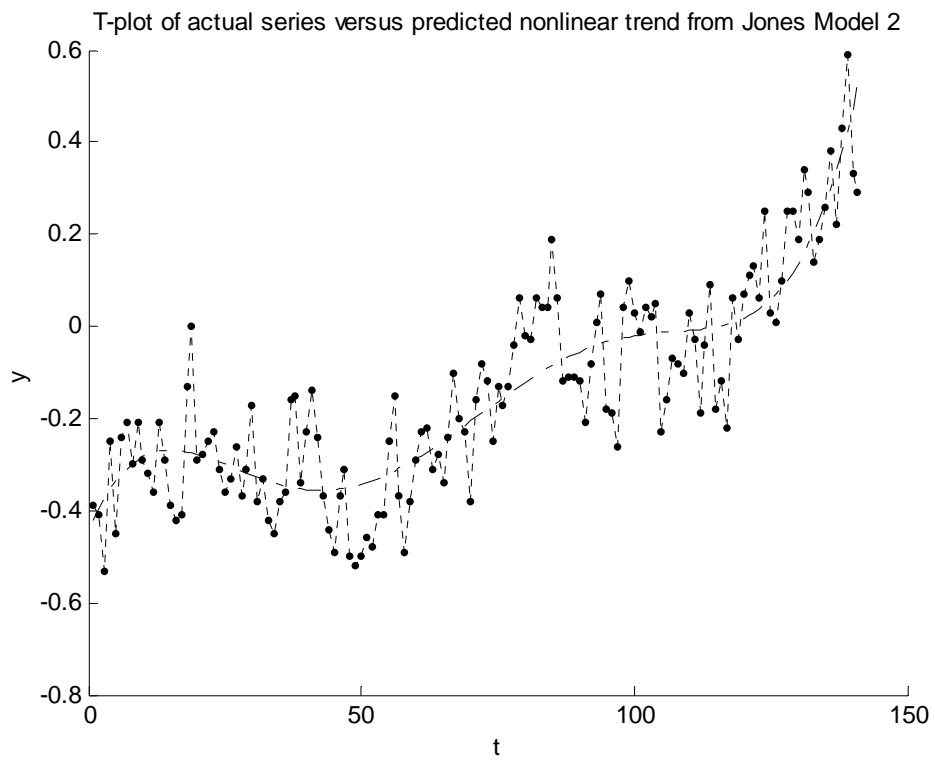
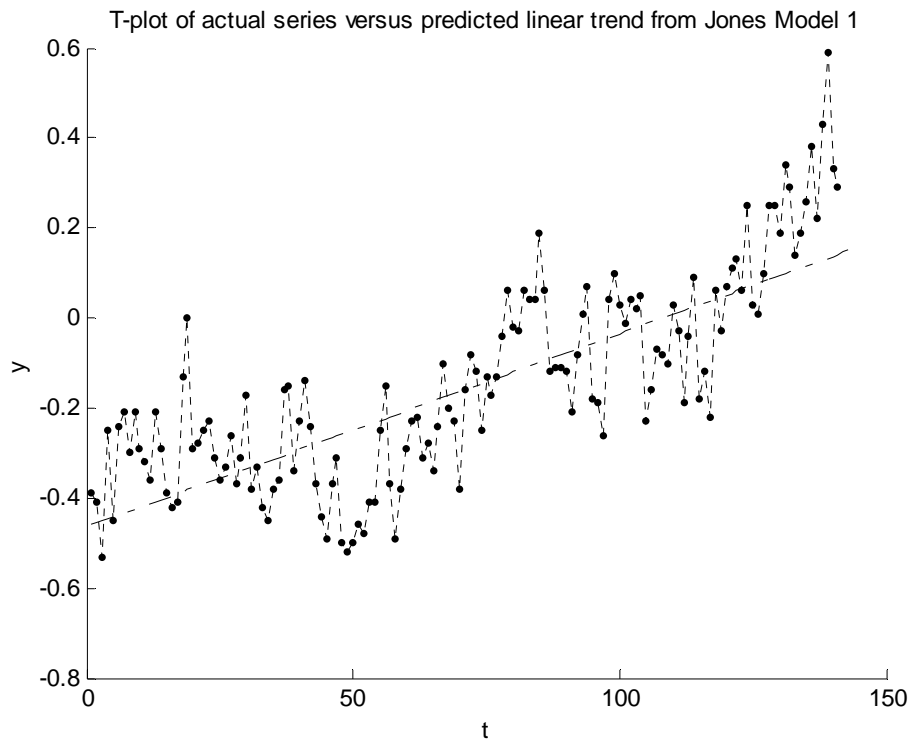
These plots of the linear (nonlinear) trend vs. the data series are similar to the graphs presented in Seidel and Lanzante (2004) in which they show that the linear trend is not adequate for the temperature anomaly series. Their study includes plots of the

---

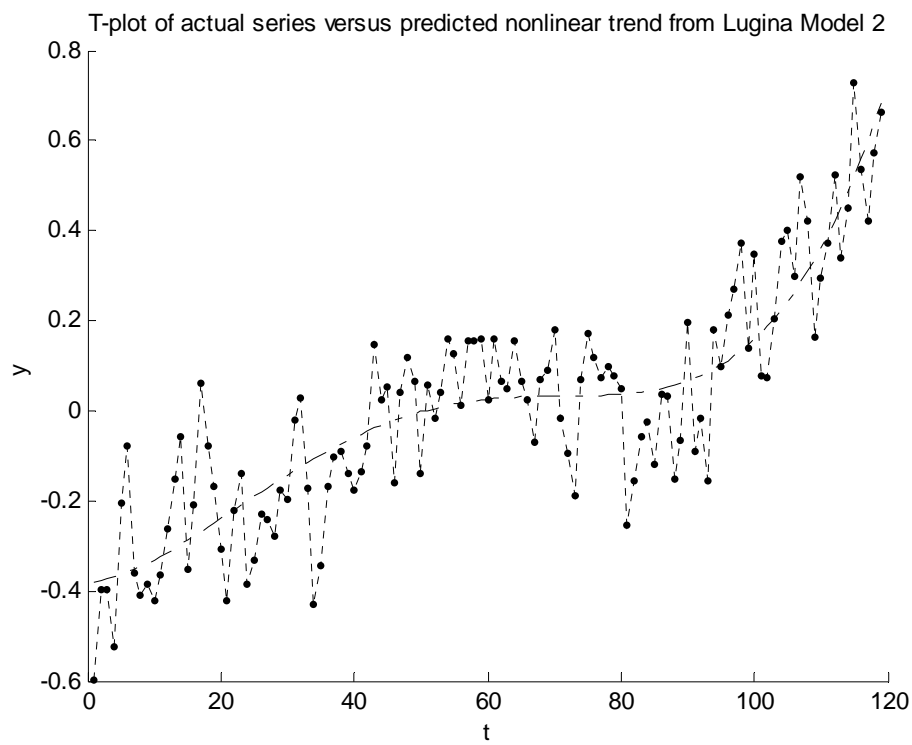
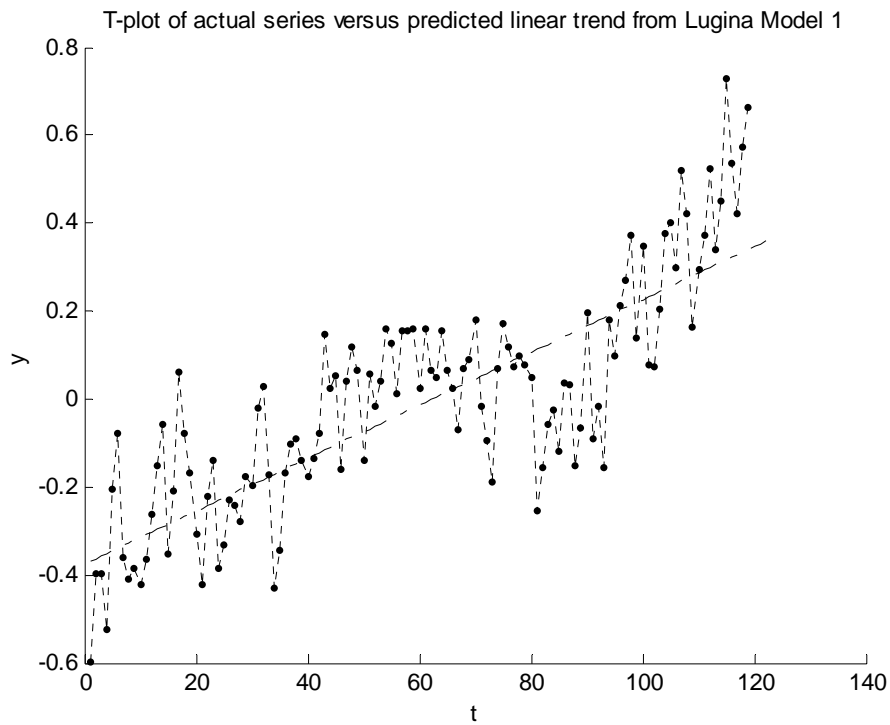
<sup>21</sup> These comparison plots can be found in Appendix A.

<sup>22</sup> Graphs of the nonlinear versus the linear trending component of the regression model using each of the Jones and Lugina series can be found in Appendix B.

linear trend against the time series and plots of the alternative piecewise linear trends against the time series, providing evidence that, in general, the alternatives are superior to the linear trend in terms of modeling temperature trends. Similarly, Harvey and Mills (2000) present graphical evidence of the superiority of smooth transition functions over the linear trend when modeling temperature trends.



**Figure 7: The actual (---●---) series vs. the predicted (-----) trend component of the regression model using the Jones series to estimate the static and dynamic models**



**Figure 8: The actual (---●---) series vs. the predicted (-----) trend component of the regression model using the Lugina series to estimate the static and dynamic models**



If global temperature anomalies started rising (or falling) at the beginning of the time period and maintained their rate of change, then the linear trend would be sufficient for modeling temperature patterns over time. However, if the rise (or fall) in temperature anomalies is delayed to a later time period and/or the rate of change in temperature anomalies varies over the time period, then a nonlinear trend will be more appropriate for modeling temperature patterns over time. The behavior of the nonlinear trend can be different depending on the order of the nonlinear trend and on the signs and magnitudes of the different coefficients.

The results from the empirical analysis support the growing evidence that trends in global temperature patterns are nonlinear in nature. Seidel and Lanzante (2004) employed a linear piecewise trend that was shown to be an improvement over the linear trend. This study provides evidence that continuous nonlinear functions may provide a better fit to the data. Unlike previous studies on the topic of trends in temperature data, this study provides evidence, based on formal diagnostic tests, that the linear model is misspecified for these commonly-used data and also provides evidence, based on formal diagnostic tests, that a nonlinear trend specification provides a better fit to the data. Graphical analyses of the trending component, similar to those provided in previous studies, confirm the results of the diagnostic tests.

## **Section 4: Conclusion**

To answer the question posed at the outset, the normal linear regression model can assess the evidence for a warming trend when the model is correctly specified. The dual goal of this paper was to illustrate why the static model may be invalid for some global temperature series and to propose an alternative formulation that is statistically adequate for the data.

Concerning the first goal, the results indicate that the static model is misspecified for two widely-used temperature series and is not reliable for drawing conclusions about trends in surface temperatures using these series. The model provides estimates of a statistically significant warming trend in temperature series that have been widely

reported in international policy arenas. The results indicate that these estimates may overestimate or mischaracterize the trend in temperature anomalies.

Concerning the second goal, the results indicate that the temperature series exhibit nonlinear trends rather than linear trends. A nonlinear trend with an overall positive slope provides a different picture of global warming compared to a linear trend with a positive slope. Given the recent sharp upward trending temperatures predicted by the dynamic model, some may claim that the better specified model provides more impressive and urgent evidence of global warming than the linear model. The model results neither support nor refute such a conclusion because, while the model provides statistical evidence of nonlinear trends in the data, it does not explain the cause of the trends nor indicate whether the trends are temporary or permanent.

In a sense, the trends in the statistical models are a measure of the researchers' ignorance about certain characteristics of the data generating process. It is up to climate scientists to explore alternative possible theories, as well as tests of these theories, that can explain the confirmed nonlinear patterns in the data.

## References

- Angell, J. K. "Global, hemispheric, and zonal temperature deviations from radiosonde records." In *Trends Online: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, Tennessee, U. S. A. (2002). Website: <http://cdiac.ornl.gov/trends/temp/angell/angell.html> Accessed 12.25.03.
- Benestad, R.E. "What can Present Climate Models Tell Us About Climate Change?" *Climate Change*. 59 (2003):311-331.
- Carbon Dioxide Information Analysis Center. *Trends Online: A Compendium of Data on Global Change*. Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, Tennessee. (2003). Website: <http://cdiac.ornl.gov/trends/abstract.htm> Accessed 2.24.04.
- "Consumer Impact of Global Warming Policies, The: Some Sobering Facts." Website: <http://www.globalwarming.org/consume.htm> Accessed 9.12.02.
- Easterling, D. R., T. R. Karl, J. H. Lawrimore, and S. A. Del Greco. *United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data for 1871-1997*. National Oceanic and Atmospheric Administration, National Climatic Data Center, Asheville, North Carolina, U. S. A. (1999). Website: <http://cdiac.esd.ornl.gov/ftp/ndp070/nd070.txt> Accessed 5.23.02.
- Folland, C.K., T. R. Karl, J. R. Christy, R.A. Clarke, G.V. Gruza, J. Jouzel, M.E. Mann, J. Oerlemans, M.J. Salinger, S.W. Wang. "Observed climate variability and change." In: Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, D. Xiaoxu (eds). *Chapter 2 of Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*. (2001). Cambridge University Press, Cambridge, UK. Pp. 99-181.
- Fomby, T. B. and T. J. Vogelsang. "The Application of Size Robust Trend Statistics to Global Warming Temperature Series." Working paper, Cornell University. (2000). Website: <http://people.cornell.edu/pages/tjv2/globwarm.pdf> Accessed 2.24.04.
- Fomby, T. B. and T. J. Vogelsang. "Tests of Common Deterministic Trend Slopes Applied to Quarterly Global Temperature Data." Working paper. (2003). Website: <http://faculty.smu.edu/TFomby/warming.pdf> Accessed 2.24.04.
- Greene, W. H. *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall. (1997).

- Hansen, J., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl. "A Closer look at United States and global surface temperature change." *Journal of Geophysical Research*. 106 (2001):23947.
- Harvey D. I. and T.C. Mills. "Detecting Climate Change Using Smooth Transitions." Loughborough University, Department of Economics. Economic Research Paper No. 00/21. (2000). Website: <http://www.lboro.ac.uk/departments/ec/Researchpapers/2000/00-21/erp00-21.pdf> Accessed: 8.15.04.
- IPCC (Intergovernmental Panel on Climate Change). "Summary for Policymakers." In *Climate Change 2001: The Scientific Basis*. (2001a). Website: <http://www.ipcc.ch/pub/spm22-01.pdf> Accessed 9.13.02
- IPCC (Intergovernmental Panel on Climate Change). "Technical Summary." In *Climate Change 2001: The Scientific Basis*. (2001b). Website: <http://www.ipcc.ch/pub/wg1TARtechsum.pdf> Accessed 9.13.02
- IPCC (Intergovernmental Panel on Climate Change). "Summary for Policy Makers." In *Climate Change 2001: The Science of Climate Change*. (2001c). Website: <http://www.ipcc.ch/pub/sarsum1.htm> Accessed 4.2.04
- Jones, P. D., D.E. Parker, T. J. Osborn, and K. R. Briffa. "Global and hemispheric temperature anomalies—land and marine instrumental records." In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, Tenn., U. S.A. (2001). Website: <http://cdiac.ornl.gov/trends/temp/jonescru/jones.html> Accessed 12.24.03.
- Lugina, K. M., P. Ya. Groisman, K. Ya. Vinnikov, V. V. Konaeva, and N. A. Speranskaya. "Monthly surface air temperature time series area-averaged over the 30-degree latitudinal belts of the globe, 1881-2002." In *Trends Online: A Compendium of Data on Global Change*. (2003). Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, Tenn., U. S. A. Website: <http://cdiac.ornl.gov/trends/temp/lugina/lugina.html> Accessed 8.25.02 and 12.24.03
- McGuirk, A., P. Driscoll, and J. Alwang. "Misspecification Testing: A Comprehensive Approach." *American Journal of Agricultural Economics*. 75 No.4 (1993): 1044-1055.
- McKittrick, R. "The Search for Warming in Global Temperatures: Data, Methods and Unresolved Questions." Working Paper, University of Guelph. (2001).

- Morris, M. "The Impact of Temperature Trends on Short-Term Energy Demand." Energy Information Administration. (September 1999). Website: <http://www.eia.doe.gov/emeu/steo/pub/special/weather/temptrnd.pdf> Accessed 3.24.04.
- National Oceanic and Atmospheric Administration. "The Top 10 El Nino Events of the 20<sup>th</sup> Century." Climate Perspectives Branch, Global Climate Laboratory, National Climatic Data Center, Asheville, NC. (1998). Website: <http://www.ncdc.noaa.gov/oa/climate/research/1998/enso/10elnino.html> Accessed 9.10.03.
- Polsky, C., J. Allard, N. Currit, R. Crane, and B. Yarnal. "The Mid-Atlantic Region and its Climate: Past, Present, and Future." *Climatic Research*. 14 (2000): 161-173.
- Schmidt, M. "Statistical Trend Detection and Analysis Methods." Appendix A. In *Recommendations for Analysis of PAMS*. by T. Stoeckenius. Website: <http://capita.wustl.edu/PMFine/Reports/TrendDetect/APXARAOP.pdf> Accessed 3.24.04.
- Seidel, D. J. and J. R. Lanzante. "An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes." *Journal of Geophysical Research*. 109 (2004). D14108.
- Spanos, A. *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge: Cambridge University Press. (1999).
- Sterin, A. M. "Tropospheric and Lower Stratospheric Temperature Anomalies Based on Global Radiosonde Network Data. In *Trends Online: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, Tennessee, U. S. A. (2001). Website: <http://cdiac.esd.ornl.gov/trends/temp/sterin/sterin.html> Accessed 8.25.02.
- Taylor, J. "Clouds over Kyoto: The Debate over Global Warming." *Regulation*. (Winter 1998): 57-63.
- Trenberth, Kevin E. "Climate Change and Temperatures." Testimony before the U. S. Senate Committee on Commerce, Science, and Transportation. (May 17, 2000). Website: <http://www.gcrio.org/OnLnDoc/pdf/pdf/trenberth000517.pdf> Accessed 9.14.02.
- United States Climate Action Report 2002. *Third National Communication of the United States under the United Nations Framework Convention on Climate Change*. Appendix D: Climate Change Science: An Analysis of Some Key Questions. (2002). Website: <http://www.gcrio.org/CAR2002/#toc> Accessed 9.14.02.

United States Historical Climatology Network (USHCN). Website:  
<http://lwf.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html> Accessed:  
5.23.02.

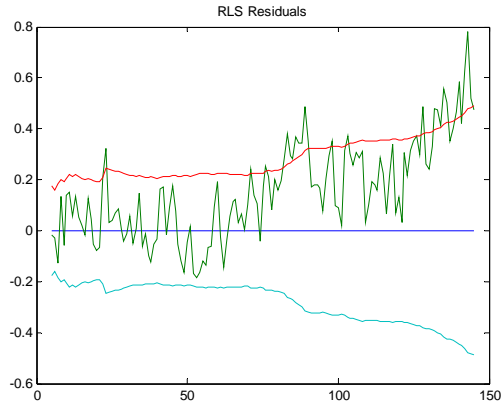
World Climate Report. "A Complete Temperature Record." July 5, 1999. As referenced  
in: Global Warming Information Page. (1999). Website:  
<http://www.globalwarming.org/sciup/sci7-7-99.html>. Accessed 9.12.02.

Zheng, X. and R. E. Basher. "Structural Time Series Models and Trend Detection in  
Global and Regional Temperature Series." *Journal of Climate*. 12 (1999): 2347-  
2358.

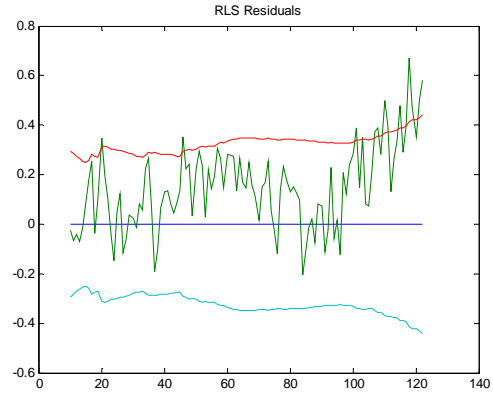
## Appendix A: More Recursive Plots

### Static model

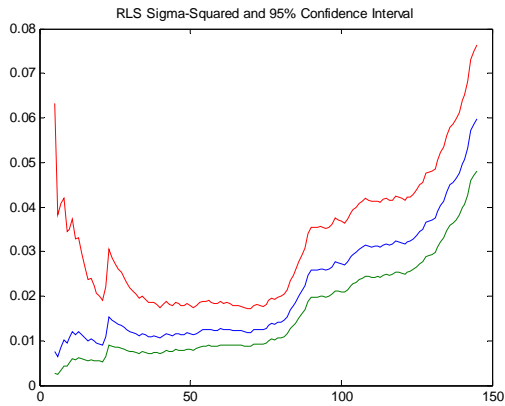
RLS residuals for Jones dataset



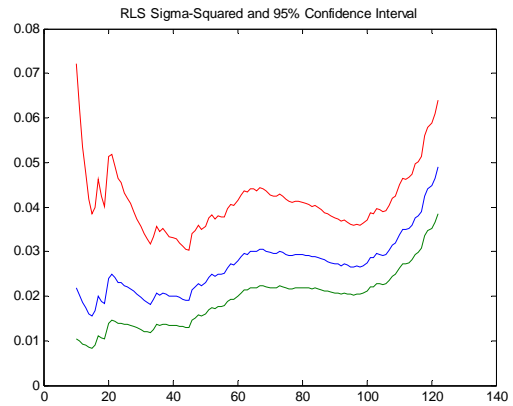
RLS residuals for Lugina dataset



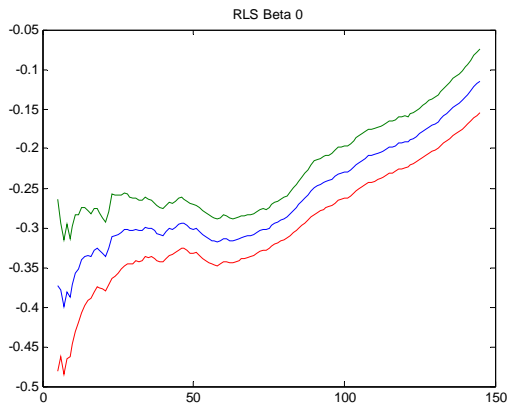
RLS sigma-squared for Jones dataset



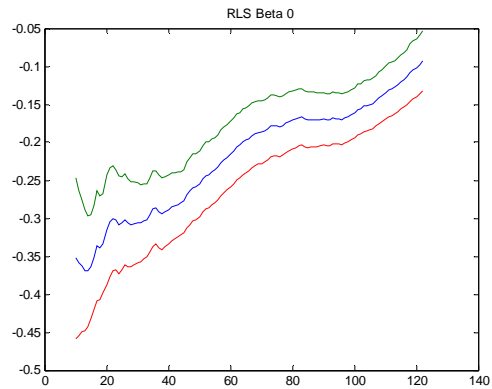
RLS sigma-squared for Lugina dataset



RLS Constant term for Jones dataset

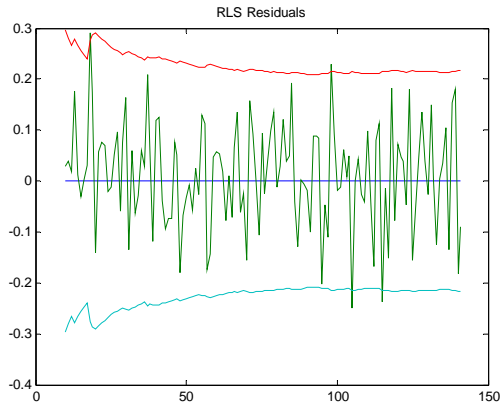


RLS Constant term for Lugina dataset

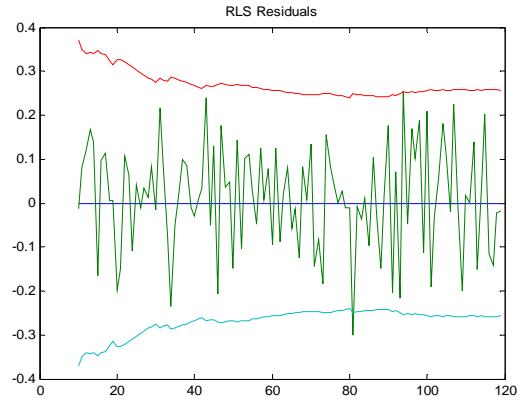


## Dynamic model

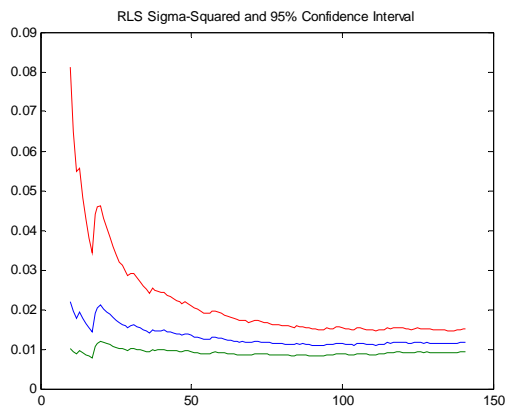
RLS of residuals for Jones dataset



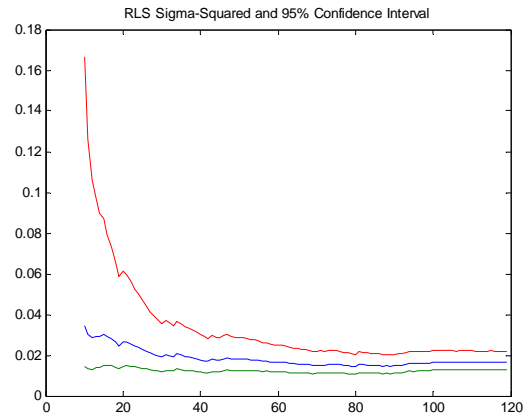
RLS of residuals for Lugina dataset



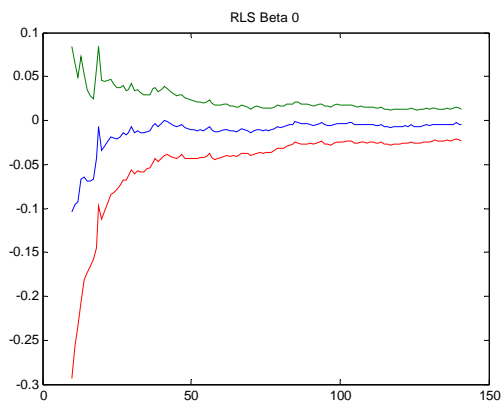
RLS of sigma-squared for Jones dataset



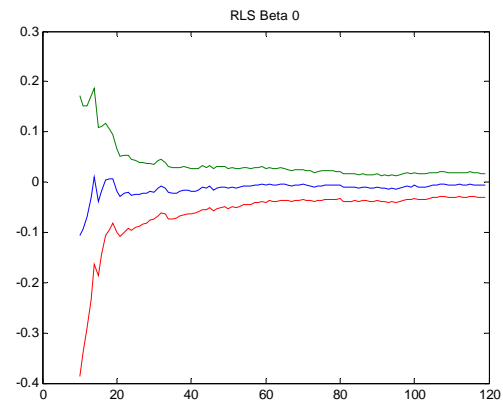
RLS of sigma-squared for Lugina dataset



RLS of Constant term for Jones dataset

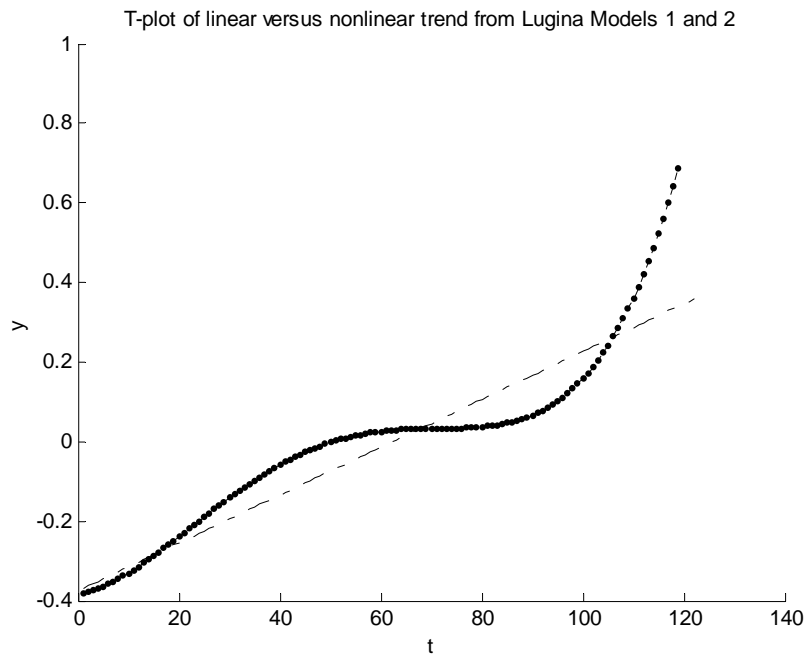
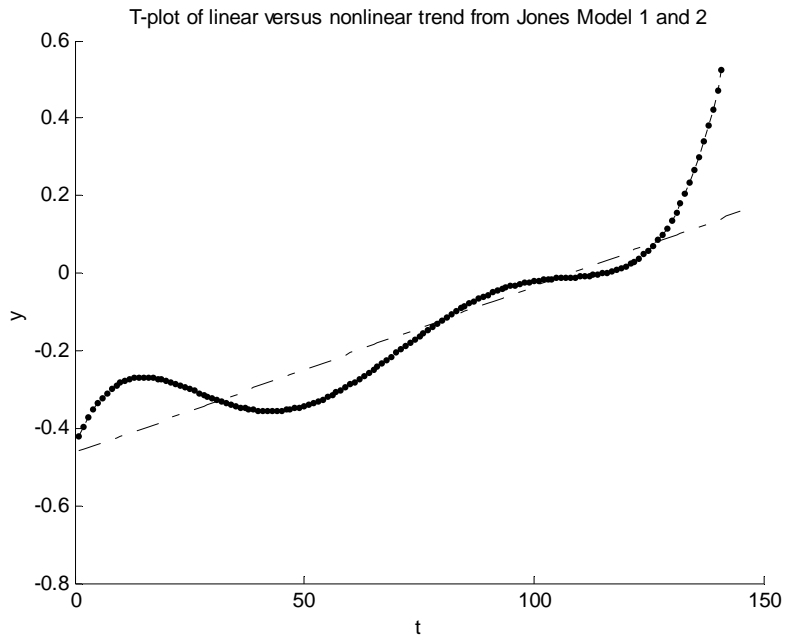


RLS of Constant term for Lugina dataset





## Appendix B: Linear versus nonlinear trending component from the static and dynamic models for the Jones and Lugina series



## **Chapter 2:**

# **An Analysis of Overcompliance with Effluent Standards among Wastewater Treatment Plants**

## Section 1: Problem Statement and Objectives

### *Introduction*

Regulatory programs to manage and protect environmental quality could be evaluated on a number of different criteria. Questions that may be asked of the program include: Have social environmental objectives related to the protection or restoration of natural resources been achieved? Are regulated parties complying with the regulatory requirements? Do regulated parties face economic incentives to comply with regulatory requirements, i.e. in terms of penalties for noncompliance and/or rewards for overcompliance? Does the program provide incentives to continuously seek and develop pollution prevention innovations?

The Clean Water Act (CWA) is the overarching statute that establishes national water quality goals and how those goals are achieved. The goal of the CWA is “to restore and maintain the chemical, physical, and biological integrity of the nation’s waters” (Federal Water Pollution Control Act, Sec. 101(a)/33 U.S. C. 1251). To achieve these broad objectives, the CWA identifies a list of conventional pollutants<sup>23</sup> to be regulated. Biochemical oxygen demand (BOD) is the most prominent of the conventional pollutants. BOD is an indirect measure of the amount of organic compounds in a unit of wastewater discharge.<sup>24</sup> Excessive loadings of organic

---

<sup>23</sup> The Clean Water Act classifies water pollutants into three groups: conventional, toxic, and a catchall group for any pollutant that is neither conventional nor toxic. There are 5 conventional pollutants and they are biochemical oxygen demand (BOD), total suspended solids (TSS), pH, fecal coliform, and oil and grease (Code of Federal Regulations, 40, Chapter 1, §401.16).

<sup>24</sup> These oxygen-demanding compounds are difficult to measure directly therefore the solution is to find a reliable indirect measure of the amount of organic matter in the wastewater (Stoddard *et. al.*, 2002). Given the direct relationship between the levels of oxygen demand and biodegradable organic matter, the best way to test for the presence of organic pollutants in the wastewater is to measure the amount of oxygen that is used during decomposition of the organic matter. The BOD test of wastewater effluent defines the change in oxygen demand as the change in organic matter in a sample of treated wastewater as it decomposes over a 5-day period at 20° Celsius in the dark (Liu and Liptak, 1999; Gray, 1999; Hammer and Hammer, Jr., 1996). Under controlled laboratory conditions, it is typical that only the readily oxidized carbonaceous organic matter will decompose in the first few days while the nitrogenous organic matter will start to decompose and exert a significant oxygen demand after 8-10 days (Gray, 1999). Thus, the BOD test measures carbonaceous oxidation. The BOD test commonly looks at the 5-day BOD and not, for instance, at the 20-day BOD for historical reasons. When sanitary engineers in England developed the BOD test, rivers in England took at most 5 days to reach the sea (Davis and Cornwell, 1991). Engineers

compounds may have an impact on the receiving water because there is a negative correlation between the level of organic compounds in the wastewater and the level of dissolved oxygen in the receiving water. Molecular dissolved oxygen is necessary for the decomposition of organic compounds and it is also necessary for supporting many higher forms of aquatic life in a river or lake system (Stoddard *et. al.*, 2002; Liu and Liptak, 1999; Davis and Cornwell, 1991).

Organic compounds in the wastewater, of which BOD is an indirect measure, can “have such a profound impact on almost all types of rivers that they deserve special emphasis. This is not to say that they are always the most significant pollutants in any one river, but rather that no other pollutant category has as much overall effect on our nation’s rivers” (Davis and Cornwell, p. 266).

The chief mechanism of the CWA for regulating the discharge of conventional pollutants like BOD is the National Pollutant Discharge Elimination System (NPDES). The CWA requires that all point sources must hold NPDES permits in order to legally discharge wastewater containing pollutants. A point source is any facility (i.e. industrial plant, wastewater treatment plant) that releases pollutants through a specific location or “point” such as an outfall pipe.<sup>25</sup> Wastewater treatment plants (WWTPs) are a significant class of point source dischargers (Liu and Liptak, 1999; Stoddard *et. al.*, 2002).

One of the primary conditions specified in the NPDES permits are average daily and monthly effluent concentration limits for conventional pollutants. The CWA specifically requires the Environmental Protection Agency (EPA) to establish effluent concentration limits for conventional pollutants for different classes of point source dischargers. EPA is instructed to establish concentration limits based on the best

---

were not concerned with the oxygen demand of organic matter in rivers at longer time spans since dilution with the sea would render the oxygen demand of the organic matter in the river irrelevant. No other time span is more logical than 5 days so the BOD<sub>5</sub> measure has become entrenched even though, unlike in England in times past, the BOD parameter may not achieve its maximum value in 5 days for a given sample. The BOD test is not useful for drawing conclusions about the actual oxygen demand of the *effluent* for three reasons: laboratory conditions do not mimic natural conditions in the receiving water; the varied biochemical reactions that determine the maximum level of oxygen demand due to biodegradable organic matter in wastewaters and natural waters are not completed within 5 days and; biological oxidation is powerless to break down some types of organic matter (Hammer and Hammer, Jr., 1996; Stoddard *et. al.*, 2002; Gray, 1999). Regardless of the shortcomings of the BOD<sub>5</sub> test, it “does indicate the potential possessed by a wastewater for deoxygenating a river or stream” (Gray, p.103).

<sup>25</sup> The Act defines a point source of water pollution as “any discernible, confined, and discrete conveyance... from which pollutants are or may be discharged” (33. U.S.C. Sec. 1362).

available secondary treatment technologies that have been identified by the Agency.<sup>26</sup> Consequently, concentration limits are sometimes called technology based performance standards. The activated sludge treatment process is the most common secondary treatment technology and therefore is used as the standard for establishing BOD technology-based performance standards for WWTPs (USEPA, 1996).<sup>27</sup> The 30-day average limit for BOD is 30 milligrams/liter and the 7-day average limit for BOD is 45 milligrams/liter (WEF, 1997). In addition to these effluent limits, the WWTP is to achieve at least 85 percent removal of BOD from the influent wastewater (Ibid.). Congress also instructs EPA to periodically review these standards and lower these technology-based performance standards as new technologies are developed (EPA, 1996).

Once effluent concentration standards are identified, either the EPA or an authorized state-level environmental protection agency is required to administer permits to individual point source dischargers. Currently, 44 out of 50 states are authorized to administer the permit program (State NPDES Program Authority). Permit granting agencies are required to issue NPDES permits to point sources every five years. In addition to specifying effluent concentration limits, the NPDES program administrator may also include technological and capital operational requirements as part of the permit (USEPA, 1996).

NPDES permits contain monitoring and enforcement provisions. The CWA authorizes federal or state regulators to conduct any inspections of permitted facilities that are required for developing NPDES permit conditions or determining the facilities' compliance with such conditions (WEF, 1997). As part of the requirements of their permits, regulated sources are also required to fill out monthly discharge monitoring reports (DMRs) in which they report their discharges of the permitted water pollutants.

---

<sup>26</sup> Secondary treatment is the minimum level of treatment required by law. It is a biological treatment method that is specifically designed to remove organic matter from wastewater by providing enabling conditions for the organic matter to solidify enough to settle out of the wastewater (Davis and Cornwell, 1991; Liu and Liptak, 1999).

<sup>27</sup> Lawmakers realized this inherent 'lock on a particular technology' and to counter this effect, established "equivalent-to-secondary treatment" standards for pollutants under the Clean Water Act (Stoddard *et. al.*, 2002; USEPA, 1996). They required EPA to "provide allowances for alternative biological treatment technologies, such as a trickling filter or waste stabilization pond" (USEPA 1996, p. 79). According to EPA's manual, "although this process has been used for industrial facilities, the concept [of alternative limits] has generally not been applied to municipal [i.e. WWTP] permits" (USEPA 1996, p. 79).

These DMRs are an essential tool for the regulator charged with monitoring point source pollution and determining compliance with environmental regulations.

Violations of the terms of the NPDES permit can result in judicial penalties and administrative penalties, with the latter being the more common of the two. Administrative penalties include fines that can be as high as \$11,500 per day for each day during which the violation continues (Civil Monetary Penalty Inflation Adjustment Rule; WEF, 1997). The regulator decides the severity of the violation after considering several factors such as the extent of the violation, the facility's ability to pay, the facility's compliance history, etc. (Ibid.)

The CWA requires states to develop ambient water quality standards for their surface waters and develop water quality-based effluent limits based on these designated uses (WEF, 1997). When the technology-based limits are not sufficient to meet ambient water quality standards in the receiving water, water-quality based limits may be imposed that replace the technology-based limits (Hammer and Hammer, Jr., 1996). Water quality-based effluent limits must be at least as stringent as the technology-based limits.

### ***Problem Statement***

The effectiveness of the NPDES program to manage point source effluent discharge could be evaluated on a number of criteria. A frequent measure of success used to evaluate the NPDES permitting program is the degree to which point sources comply with permit limits. Traditionally, observers have noted that the violation of permit limits is a persistent problem under the CWA (USGAO, 1996; USGAO, 1991; USGAO, 1988; Farber, 1999; Deegan, 1992; Clean Water Network; USEPA, 1998; Pianin, 2002).

Some observers do not find this surprising because the regulatory structure does not reward continuous compliance, in which plants lower their pollutant discharges over time. As the argument goes, plants respond to the lack of incentives by controlling pollution only to levels (i.e. at or just below the permit limit) that are sufficient to stay in compliance. If plants are polluting at or just below their permit limit, then they are more likely to violate their permit limit than plants that reduce their effluent discharge level further below the permit limit.

However, there is a growing body of recent evidence indicating that a significant number of regulated point source dischargers are consistently reducing discharge effluent below the level stipulated in their permits (Kagan *et. al.*, 2003; Bandyopadhyay and Horowitz, 2004; Houtsma, 2003; McClelland and Horowitz, 1999; Dasgupta *et. al.*, 2000; Brannlund and Lofgren, 1996). Defining overcompliance as the positive difference between the permitted effluent concentration levels and the reported effluent concentration levels, Bandyopadhyay and Horowitz (2002) found “substantial and unexpected” evidence of overcompliance with BOD permit limits amongst WWTPs from 1993-1997. The wastewater treatment plants in their study emitted on aggregate 32 percent of their permitted level of BOD.

In their study of pulp and paper plants, McClelland and Horowitz (1999) found that BOD effluent levels from pulp and paper mills in 1992 were no more than 50 percent of the level specified in the NPDES permit, despite the significant estimated marginal control costs. In a study of municipal WWTPs in Ontario, Canada, and privately operated WWTPs in the United States, Houtsma (2003) found that many plants were performing better than they were required to perform. Houtsma’s survey of 155 municipal plants in Ontario from 1992-1997 revealed that 84 percent of the plants were emitting at most 50 percent of their BOD permit limit and 43 percent of plants were emitting at most 25 percent of their BOD permit limit.

Houtsma’s survey of treatment plant superintendents and public sector officials suggests some possible reasons why WWTPs overcomply with their permit limits. In addition to the more obvious reasons related to the operator’s desire to keep his/her job or the operator’s desire to be a good public citizen, Houtsma’s survey identified process related reasons for overcompliance:

- WWTPs are expected to meet standards “all the time,” not just on a weekly or monthly basis;
- Plant operators do not exert enough control over the wastewater treatment process to allow for higher effluent levels. The implication is that if plants had better control over the level of BOD in the discharge, the treatment plant’s effluent would contain BOD concentrations that were closer to the level specified in the NPDES permit limit.

The desire to be a good public servant has been previously identified as a possible reason for overcompliance. McClelland and Horowitz (1999) find that those in the industry attribute the observed overcompliance to “their desire to be good neighbors and establish better relations with regulators” (p. 221).

Another reason not often mentioned is that BOD is one of several pollutants that the WWTP may have to control, and that the treatment plant’s control of two or more pollutants in the same process may improve its ability to provide BOD control. McClelland and Horowitz referred indirectly to this possibility when they noted that plants may overcomply with BOD limits when the control of another pollutant whose limit is binding also affects BOD control.<sup>28</sup> Using, as an example, the case of jointness between chlorine and BOD control, they discount this reason citing data on plants that control BOD showing that plants that do not release chlorine are more overcompliant than plants that release chlorine.

The recent and increasing evidence of significant and widespread overcompliance amongst WWTPs with their NPDES permit raises several questions. Does the overcompliance point to the existence of an incentive system established by the regulator to encourage overcompliant behavior on the part of the treatment plant? If so, then what is the nature of these incentives? If not, and there is no incentive structure in place that rewards pollution control, then what explains the observed overcompliance? Whether and to what degree regulated point sources are compliant and the degree to which they are overcompliant are important for determining how the CWA influences discharger behavior and whether reforms or changes to the CWA are warranted.

Both the qualitative and quantitative evidence suggests that there is significant and widespread overcompliance among regulated plants. There is a lot of discussion but much less consensus in the literature as to the reasons for the observed overcompliance. This paper contributes to the debate by exploring the evidence for overcompliance in

---

<sup>28</sup> They bring up this possibility within the context of the effect that a non-smooth technology would have on BOD overcompliance. They discount the possibility that a non-smooth control technology for BOD is causing the observed overcompliance by noting that several inputs “can be smoothly varied to affect BOD.” However, they consider that the non-smooth technology may still be a relevant factor when they discuss the effect on BOD overcompliance of jointness between BOD and another pollutant that has a non-smooth joint production relationship with BOD and whose permit limit is binding on the plant.



greater detail and offering plausible reasons for the observed overcompliance, based on qualitative and quantitative evidence.

## ***Objective***

The objective of this paper is to qualitatively and quantitatively evaluate explanations for the observed overcompliance with BOD permit limits. Based on a variety of data sources, the paper will identify plausible explanations for the observed degree of overcompliance with NPDES permit limits for BOD. Specific emphasis will be given to the role of the stochastic nature of BOD control and the technology-induced relationship between BOD control and the control of other pollutants in explaining BOD overcompliance.

## ***Procedures***

Section 2 provides a conventional conceptual behavioral model for understanding discharger compliance behavior under effluent performance standards. This behavioral model of wastewater treatment plant compliance yields plausible hypotheses about why plant operators might overcomply with their permit limits. Section 3 applies the conceptual framework to the special circumstances of BOD effluent control technology at WWTPs and further refines the possible explanations for overcompliance at WWTPs through an investigation of the technical processes involved in BOD control and the regulatory environment in which BOD permit limits are established. Section 4 presents the empirical evidence for the hypotheses developed in Sections 2 and 3 using effluent data from wastewater treatment plants in Connecticut and Maryland. Statistical methods are employed to test these hypotheses. Section 5 concludes with a discussion of the implications of the foregoing analysis for explaining the observed overcompliance at wastewater treatment plants.

## Section 2: Conceptual Framework

This section introduces the standard behavioral model of wastewater treatment plant compliance as it is presented in the economic literature. The model offers possible explanations for the observed overcompliance with NPDES permit limits for BOD. The monitoring and enforcement literature has studied the compliance decisions of regulated facilities under effluent standards from several dimensions.

More specifically, the literature has studied the firm's compliance decision with effluent performance standards given stochastic effluent discharges (Bandyopadhyay and Horowitz, 2004; Houtsma, 2003; Brannlund and Lofgren, 1996; Beavis and Walker, 1983a; Beavis and Walker, 1983b), uncertainty in monitoring and/or enforcement (Franckx, 2002; Malik, 1990; Malik, 1993; Harrington, 1988; Harford, 1991; Harford and Harrington, 1991; Fuller, 1987), inspections (Heyes and Doucet, 1999; Oljaca *et. al.*, 1998; Laplante and Rilstone, 1996; Liu, 1995; Magat and Viscusi, 1990;), variations in the characteristics of the firm's human capital (Kagan, *et. al.*, 2003; Dasgupta *et. al.*, 2000; Hartman *et. al.*, 1995; Doonan *et. al.*, 2002), and the existence of information sharing programs (Kagan *et. al.*, 2003; Foulon *et. al.*, 2002; Farber, 1999; Hipel *et. al.*, 1995).

Broadly, these studies conceptualize the operator's compliance decision as balancing the costs of reducing effluent discharges with the cost of possible penalties or fines. This standard behavioral model reflects the underlying behavioral assumptions presented in the monitoring and enforcement literature. The wastewater treatment plant seeks to minimize the present-value of its future stream of cost of meeting a specific effluent standard (regulatory limitation). The plant's costs are composed of abatement costs and expected penalties. In the current time period, the management of the plant has to balance its certain costs against its expected penalties.

The plant's abatement costs are related to the nature of the good that it produces. The plant incurs fixed and variable costs in order to acquire the fixed and variable inputs necessary for effluent control. Its fixed costs include the cost to acquire the land and capital equipment as well as the cost of constructing the plant. Its variable costs include its labor and energy costs. The labor costs are the remuneration paid per diem or on a salaried basis to the people at the treatment plant who clean, maintain, operate, and

upgrade plant machinery. Energy costs are the costs to provide electricity to run the machinery that moves the wastewater through the various stages of treatment. Energy costs are an essential component of variable costs since modern wastewater treatment has become increasingly mechanized and computerized (personal communication with S. Luckman, Maryland Department of the Environment).

The cost of providing an extra unit of effluent control, i.e. the marginal abatement cost, is generally thought to be an increasing function of pollution control or a decreasing function of pollution output. The marginal cost of pollution abatement at higher levels of pollution abatement is higher than the marginal cost of pollution abatement at lower levels of pollution abatement. The marginal cost of pollution abatement may also be nonlinear such that the rate of increase in the marginal cost of pollution abatement is higher at higher levels of pollution abatement.

The plant also faces a potential stream of future costs due to permit violations and the attendant penalties. That is, it has a penalty function that, for specified violations of its permit limits, determines the fine that the plant will incur with a given probability. There are several components of the penalty function: the actual monetary fine, the probability of exceeding the permit limit, the probability of inspections, the probability of detection, and the probability of being fined given that there was a violation.

The monitoring and enforcement literature has made a substantial contribution to identifying and explaining the determinants of the theoretical abatement cost and penalty functions. Some insights gained from the literature may be useful in explaining the individual plant's behavioral response and how this behavioral response may relate to the observed overcompliance among plants. The individual plant's behavioral response will be examined using a mathematical and graphical model of the plant's optimization problem.

Let  $Q$  represent the plant's effluent levels of a given pollutant into the receiving water. Expressed as a concentration standard,  $Q$  is often quantified as milligrams per liter (mg/l) per unit of time (such as a 30- or 7-day average). The maximum amount of  $Q$  that the plant discharges is specified in the NPDES permit. The plant produces effluent control so its economic activity involves *reductions* in  $Q$ . Therefore, the higher the amount of  $Q$  that is released from the plant, the lower will be the plant's total abatement

costs. Additionally, the higher the amount of  $Q$  that is released from the plant, the higher will be the probability of exceeding the permit limit, and the higher will be the plant's total expected penalties.

The plant seeks to minimize the present value (PV) of its future stream of costs. At the current time period, the plant management solves the following static cost minimization problem:

$$\text{Min}_Q H \equiv TAC(Q;T) + TEP(Q), \quad (1)$$

where  $H$  represents the objective function,  $TAC$  is the PV of total abatement costs,  $TEP$  is the PV of total expected penalty. The  $TEP$  function is defined over the probability distribution of BOD output, particularly the probability of exceeding the BOD permit limit. As such, moments of the BOD output distribution are parameters of the  $TEP$  function. Specifically, a change in the variance of the BOD distribution will shift the  $TEP$  function in a given direction.  $T$  is the particular technology in place at the plant that is used to treat the wastewater. Changes in the technology installed at the plant entail parametric shifts in the  $TAC$  function. Table 6 presents the curvature restrictions on the functions in (1) in terms of the derivatives of the cost and penalty functions with respect to  $Q$ .

**Table 6: Curvature restrictions**

<b>Function</b>	<b>Sign</b>	<b>Explanation</b>
$TAC_Q(\cdot)$	$<0$	Increasing the level of $Q$ by an additional unit reduces costs.
$TAC_{QQ}(\cdot)$	$<0$	The change in marginal abatement cost for a given change in $Q$ is higher, the lower the absolute level of $Q$ .
$TEP_Q(\cdot)$	$>0$	Increasing the level of $Q$ by an additional unit increases the probability of exceeding the permit limit, which increases the expected penalty.
$TEP_{QQ}(\cdot)$	$>0$	The change in the marginal expected penalty for a given change in $Q$ is higher, the higher the absolute level of $Q$ . That is, at higher levels of $Q$ , an increase in $Q$ will occasion a higher shift in the probabilities of exceeding the permit limit, being inspected, and/or being fined.

Equations (2) and (3) present the first- and second-order conditions associated with the optimization problem in (1).

$$0 = \frac{\partial H(.)}{\partial Q} \equiv \frac{\partial TAC(.)}{\partial Q} + \frac{\partial TEP(.)}{\partial Q} \quad (2)$$

$$0 = \frac{\partial TAC(.)}{\partial Q} + \frac{\partial TEP(.)}{\partial Q}$$

$$-\frac{\partial TAC(.)}{\partial Q} = \frac{\partial TEP(.)}{\partial Q} \quad (2')$$

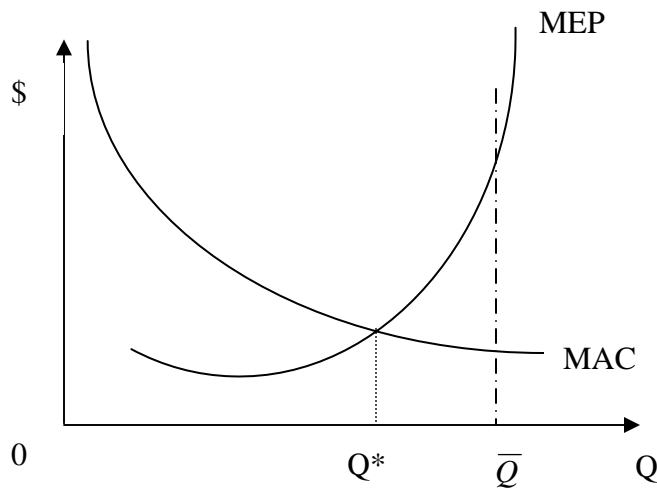
The negative sign on the marginal abatement cost in (2') indicates that the marginal abatement cost is decreasing in Q since Q is a 'bad,' that is, an undesirable product.

$$\frac{\partial^2 H(.)}{\partial Q^2} = \frac{\partial^2 TAC(.)}{\partial Q^2} + \frac{\partial^2 TEP(.)}{\partial Q^2} \quad (3)$$

Given the curvature restrictions on the cost and penalty functions, a necessary and sufficient condition, in addition to (2), for a minimum on  $H$  is that:

$$\frac{\partial^2 H(.)}{\partial Q^2} > 0 \Rightarrow \left| \frac{\partial^2 TEP(.)}{\partial Q^2} \right| > \left| \frac{\partial^2 TAC(.)}{\partial Q^2} \right| \quad (4)$$

Equation (4) states that at the optimum level of Q, the marginal expected penalty curve should be steeper than the marginal abatement cost curve. The result in (4) implies that the slope on the MAC cannot be 'too high.' Suppose that the MAC curve is nonlinear and increasing at an increasing rate as Q falls. Then the result in (4) implies that the optimum  $Q^*$  will be located at higher levels of Q, where the rate of change in the MAC is lower. Figure 9 presents a graphical representation of such an optimum.



**Figure 9: Observed choice of effluent level**

The facility chooses a level of pollution and consequently chooses a level of pollution abatement that is the positive difference between the concentration of BOD in the influent and  $Q^*$  and that equates its marginal abatement cost to its marginal expected penalty. If the facility finds itself to the right of  $Q^*$ , the penalty it expects to incur for any increase in  $Q$  is larger than the cost savings from increasing  $Q$  an additional unit therefore the facility would rather reduce  $Q$ . This moves it back towards  $Q^*$ . Similarly, if the facility finds itself to the left of  $Q^*$ , the additional abatement cost it incurs for an additional reduction in  $Q$  is larger than the money it saved in terms of avoided expected penalty therefore the plant would rather increase  $Q$ . This moves it back towards  $Q^*$ .

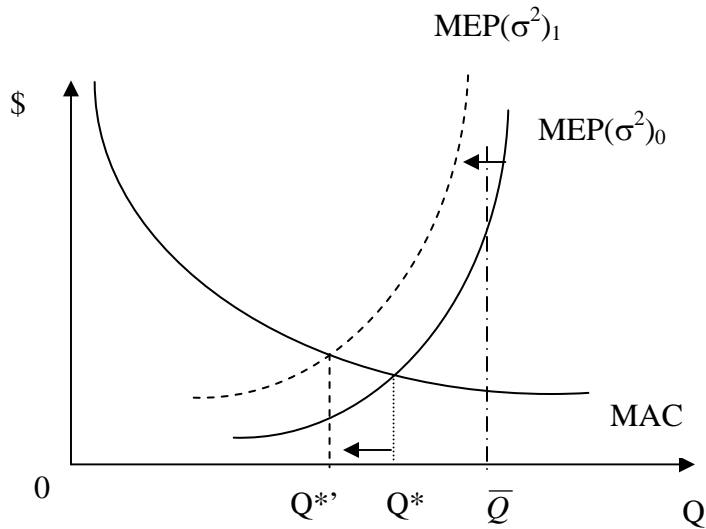
The graph in Figure 9 shows a plant with a permit limit at  $\bar{Q}$ . Note that  $Q^*$  will typically be less than the permitted level. If regulatory overcompliance is defined as the difference between reported discharge levels and effluent permit limits ( $\bar{Q}$ ) then overcompliance can be explained using the cost-minimization behavioral model above. No additional assumptions about the existence of pollution prevention incentives or altruistic motivations are necessary to explain why observed discharges might be less than permitted discharges. The model assumes a given distribution of effluent output. It

also assumes that the plant controls one pollutant. Changing the underlying assumptions of the model can affect the amount of observed overcompliance.

### ***Distribution of Effluent Discharge***

A plant with a high variance will exceed its permit limit more often than a plant with a lower variance, *ceteris paribus*. Under a cost-minimizing framework, the WWTP seeks to minimize the sum of control costs and expected penalties and therefore the plant with a high variance will reduce its average effluent discharge levels in order to compensate for the higher variance.

In terms of the behavioral model, the higher variance translates into a parametric shift upward of the MEP curve since the curve holds the distribution of effluent levels constant. In other words, a high variance means that the probability of exceedance and the expected penalty are higher for any level of  $Q$ . Figure 10 depicts this situation and its implications for overcompliance.



**Figure 10: Observed choice of effluent level with a shift in the MEP curve**

A higher variance increases the expected violation and the expected penalty for all levels of  $Q$ , shifting the MEP curve up and to the left from  $MEP(\sigma^2)_0$  to  $MEP(\sigma^2)_1$ . If the plant remains at  $Q^*$ , its MEP will exceed its MAC therefore it has an incentive to move to the left, increasing pollution abatement. Thus, the higher the variance of  $Q$ , the lower will be the mean level of  $Q$ .

In the case of BOD control, these results imply that plants with high variation in their BOD output will seek to reduce their mean BOD discharge levels such that  $Q^*$  will be further below  $\bar{Q}$ .

### ***Behavioral Model under Multiple Pollutants: Jointness in Production***

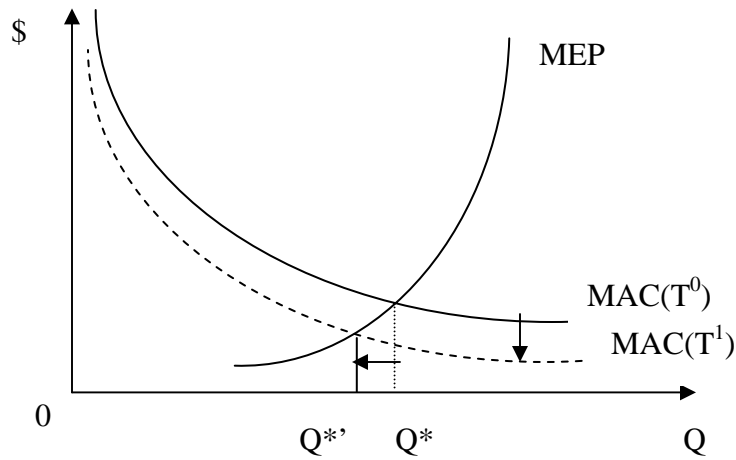
Jointness in pollution control may lower marginal abatement costs for all levels of  $Q$ . Jointness in technology occurs in multiproduct firms. It is the phenomenon whereby two or more outputs are technically interdependent either due to non-allocable inputs or due to interdependencies in the production process, or both (Blandford and Boisvert, 2001). The WWTP is a multiproduct firm because it typically controls several pollutants



in the wastewater such that  $Q$  is a vector, not a singleton. In this case, the WWTP faces a vector of permit limits, one limit for each of the pollutants it has to control. If there is jointness in the technology to control the pollutants, then either the plant cannot separate inputs in the control of each pollutant or there are characteristics of the treatment technology whereby the control of the pollutants are linked.

Suppose that a plant is permitted to discharge two pollutants (A and B) and that there is jointness in the technology to control A and B. Suppose also that the plant faces increasing restrictions on pollutant A due, for example, to concerns about the pollutant's environmental effects, resulting in more intense regulatory oversight of plant discharges of pollutant A. As a result of the tighter limits and/or increased regulatory oversight, the plant will seek to reduce its discharges of pollutant A through some combination of process changes and technology upgrades. The plant may even receive financial or other regulatory incentives to reduce its discharge of A. Regardless of the motivations for reducing discharges of pollutant A, the plant will see significant improvement in the removal of pollutant B irrespective of the current level of control of pollutant B because it has tightened its control of pollutant A.

Figure 11 illustrates this effect graphically. The jointness in pollution control between pollutants A and B lowers the marginal abatement costs for pollutant B control, analogous to the effect of a technological improvement in the control of pollutant B.



**Figure 11: Observed choice of effluent level with a shift in the MAC curve**

The original  $MAC(T^0)$  designates the plant's marginal abatement cost for controlling pollutant B when the plant only controls pollutant B. If the plant controls A in addition to B and their control is joint, its MAC curve for controlling pollutant B shifts to  $MAC(T^1)$ . If the firm remains at  $Q^*$ , its MEP exceeds its MAC for an additional unit of pollution output therefore it will reduce  $Q$  and move towards the new optimum level,  $Q^{**}$ .

In other words, the presence of jointness in pollution control between pollutants A and B allows the plant to achieve further reductions in pollutant B. A firm experiencing these additional reductions in pollutant B may overcomply with its permit limits for pollutant B through no additional effort, but due to parametric reductions in its marginal abatement costs. If the technology exhibits jointness such that the plant can achieve additional reductions in effluent control with zero additional cost, the shape of the MAC curve will change. There will be flats in the MAC over the applicable range of control.

To make the example more specific, there may be several pollutants that are related to BOD in the same way that pollutant A relates to pollutant B. For example, assuming that nitrogen is pollutant A and BOD is pollutant B, a plant that discharges

nitrogen to receiving waters and faces restrictions on its nitrogen output will have to upgrade the technology at the treatment plant so that it provides for treatment of nitrogen. Due to the jointness, the plant that has been upgraded to provide nitrogen control will see significant improvement in BOD removal, regardless of its current level of BOD control (Stoddard *et. al.*, 2002; Hammer and Hammer, Jr., 1996).

This jointness is due to the presence of non-allocable inputs like energy, oxygen, and aeration tanks, which are used in nitrogen control as well as in BOD control (Hammer and Hammer, 1996). The jointness is also due to the existence of a technical (i.e. biological) relationship between nitrogen and BOD (Gray, 1999).

### ***Stewardship Behavior***

A desire to be environmentally responsible can provide a noneconomic motivation for overcompliance with a performance standard. Studies have shown that management preferences for increased environmental protection can lead to reductions in effluent discharges and overcompliance (Doonan *et. al.*, 2002; Kagan *et. al.*, 2003). A survey of management staff at WWTPs suggests that some instances of overcompliance are due to plant managers' concern for water quality and for the environment in general (Houtsma, 2003).

The conventional behavioral model of the choice of pollution levels has not traditionally included stewardship as a relevant motivating factor. However, McClelland and Horowitz (1999) noted that the overcompliance amongst the plants in their dataset may be due to management's desire to be good neighbors, citing one mill manager's "good neighbor policy" which results in overcompliance at his plant. If present, stewardship motivations can lead the WWTP to emit levels of  $Q$  that are to the left of  $Q^*$ , resulting in both regulatory overcompliance and overcompliance from a cost-minimizing standpoint. However, stewardship does not fit into a strict cost-minimizing behavioral model and so will not be analyzed in the present paper.

### ***Identifying Explanatory Factors in Overcompliance***

The observed overcompliance in which  $\bar{Q}$  exceeds the reported level can be related to variation in effluent discharges or jointness. In theory, the relationships

between these factors and overcompliance can be investigated using cost and effluent discharge data. Cost data can be used to estimate  $Q^*$ . The estimated  $Q^*$  can then be compared to the reported  $Q$  and the established  $\bar{Q}$  and, with measures of variation and jointness, the relative contributions of each factor to the observed overcompliance can be established.

In most studies, including this one, data on the plant's costs are not observable or available so the level of  $Q^*$  cannot be estimated. Lacking these, a variety of information and analysis will be used in this study to investigate whether variation in BOD output and jointness are plausible reasons for overcompliance. The sources of information will include a review of the wastewater treatment engineering literature, interviews with federal and state regulatory officials involved in approving permits, enforcing permits, and designing WWTPs, and the statistical analysis of effluent data from WWTPs.

The next section discusses the conceptual model within the specific context of WWTPs and their control technology. This section reviews treatment processes for BOD control. Information is obtained from reviewing the engineering literature and interviewing regulatory permitting authorities.

This discussion is followed by a section that presents the results of an empirical investigation into the role that variation and jointness play in explaining overcompliance. The empirical section begins with a discussion of the available data that will be used to examine the relationships of interest. The data are from Maryland and Connecticut, two states with water quality concerns in which WWTPs have to monitor and continuously seek reductions in their nitrogen discharges. The presence of nitrogen control amongst the treatment plants allows for explicit consideration of the role that jointness between nitrogen control and BOD control may play in BOD overcompliance.

This section also presents a detailed picture of compliance (i.e. overcompliance and noncompliance) behavior at the plants in Maryland and Connecticut, evidence of variation in effluent discharge and jointness in pollution control, and a regression model that estimates the individual role of variation and jointness in explaining the observed overcompliance.

### **Section 3: Wastewater Treatment Technology**

This section presents an overview of wastewater treatment technology as it relates to BOD control. These discussions, based on a review of the literature and personal interviews, examine the extent to which a technological basis exists for two possible explanations for BOD overcompliance identified in the behavioral model: output variance and jointness in outputs.

#### ***Overview of Wastewater Treatment Technology***

The wastewater treatment process is a “sophisticated physical, chemical, and biological process” (Stoddard *et. al.*, p.23) that is on-call 24 hrs-a-day/365 days a year. By law, the WWTP provides primary and secondary treatment to wastewater. In some states, plants are encouraged to upgrade so that they can provide tertiary treatment. The discussion of the treatment technology is organized around the stages of wastewater treatment: primary, secondary, and tertiary.

There are several types of structures and machinery that possibly come together within a treatment plant including pumps, pipes, sedimentation tanks, grit chambers, aeration tanks, lagoons, and trickling filters (Liu and Liptak, 1999; Hammer and Hammer, Jr., 1996; Davis and Cornwell, 1991; Brain, M.; Water Environment Federation). Typically, treatment plants take up a sizable land area due to the space required for the construction of structures like lagoons, settling basins and aeration tanks that are necessary for treating wastewater.

These capital assets are necessary to ensure that the plant achieves at least secondary treatment of the influent, as required by the Clean Water Act. However, before applying secondary treatment, most plants apply primary treatment to the wastewater. Primary treatment is composed of pretreatment and clarification of the influent and is designed to apply a finer and finer treatment to the influent in order to divest it of settleable solids (Liu and Liptak, 1999; Davis and Cornwell, 1991; Hammer and Hammer, Jr., 1996). This stage of wastewater treatment is not designed to reduce BOD.

The major purpose of secondary treatment is to remove soluble BOD (Liu and Liptak, 1999; Davis and Cornwell, 1991). According to recent data on WWTPs, the

average BOD concentration in wastewater flowing to WWTPs is 205 mg/l (Ibid.). Primary treatment will reduce these concentrations to 143.5 on average (Ibid.). Secondary treatment will reduce these concentrations to 16.4 mg/l (Ibid).

During this crucial stage of wastewater treatment, organic matter and nutrients are converted to settleable microorganisms that can then drift to the bottom of the tank for disposal as sludge, leaving the water less polluted (USEPA, 1999). These biological processes mimic the natural processes that would go on in the receiving water if it had adequate capacity to assimilate the pollutants in the untreated wastewater (Ibid.). Thus, these biological processes “are designed to speed up natural processes so that the breakdown of the degradable organic pollutants can be achieved in relatively short time periods” (Davis, M. L. and D. A. Cornwell, p. 334).

An aerobic biological treatment method<sup>29</sup> must provide and possess certain characteristics to be effective at treating wastewater. These are: availability of microorganisms; good contact between microorganisms and the organic material to be decomposed; availability of oxygen; and a favorable environment (i.e. temperature of water and adequate time for organisms to decompose organics) (Davis and Cornwell, 1991). There are two major techniques for meeting these core requirements and each is composed of several specific secondary treatment methods. The two techniques are suspended growth (or fluid-film) processes and fixed-film processes (USEPA, 1999; Davis and Cornwall, 1991; Liu and Liptak, 1999).

Suspended growth processes keep the microorganisms suspended in the wastewater while providing optimal environmental conditions for the microorganisms to feed on organic matter and grow, until they can settle out of the wastewater as solids (Mancl, 1996). Examples of the suspended growth process include activated sludge, extended aeration, and lagoons (or ponds) (Liu and Liptak, 1999).

Fixed film processes grow microorganisms on rocks, sand, or plastic and pass the wastewater over the rocks to achieve biological treatment of the wastewater (Mancl,

---

<sup>29</sup> In addition to aerobic biological treatment methods, there are also anaerobic biological treatment methods (Liu and Liptak, 1999). These anaerobic methods enjoy some advantages over the more common aerobic methods but more research and application is needed before anaerobic methods become a viable alternative for wastewater treatment plants (Ibid). Another option, currently in use at some treatment plants, is to use pure commercial oxygen in the aeration tanks instead of air (Liu and Liptak, 1999.; Gray, 1999). This method enjoys some advantages over using air: shorter retention time and higher, more consistent BOD reduction (Ibid).

1996). Trickling filters and rotating biological contactors (RBCs) are examples of fixed film processes (Liu and Liptak, 1999; Gray, 1999). Of all these specific methods, the most common methods amongst municipal wastewater treatment plants are the activated sludge process, trickling filters and to a lesser extent, waste stabilization ponds (Stoddard *et. al.*, 2002; EPA, 1999; Liu and Liptak, 1999; Davis and Cornwell, 1991). Each method will be discussed in turn. Rotating biological contactors are also discussed as they are quite commonly used for wastewater treatment (Davis and Cornwell, 1991).

### ***Suspended Growth Processes***

The activated sludge process involves “the production of an activated (i.e. living) mass of microorganisms capable of aerobically stabilizing the organic content of a waste” (“Wastewater Treatment”). The process encourages the growth of microorganisms via the introduction of seed microorganisms and oxygen into the wastewater.

These microorganisms in turn feed on the pollutants in the wastewater, converting them into settleable microorganisms, or floc (USEPA, 1999; “Wastewater Treatment.”; WEF). As the microorganisms flocculate, they form an active mass of microbes, otherwise known as “activated sludge” (Davis and Cornwell, 1991). The activated sludge is continually circulated back to the aeration base to facilitate organic decomposition (Ibid.).

Some of the sludge is pumped away for treatment and disposal because the amount of sludge generated is typically more than is needed for recirculation (Ibid.). Once adequate flocculation has taken place (after 6-8hrs from when the wastewater first enters the treatment tank), the water then moves into settling tanks, or secondary (final) clarifiers, where the solids (sludge) are separated from the liquid. The sludge is then treated and disposed (Ibid.).

The second type of suspended growth process is a waste stabilization pond. This term is usually used to describe any type of wastewater treatment in which a pond or lagoon stores the wastewater while it undergoes biological treatment naturally. That is, the ponds treat organic waste by allowing the natural process of decomposition via natural aeration and microbial growth to run its course without mechanically aerating the wastewater or injecting seed microorganisms into the wastewater (Stoddard *et. al.*, 2002).

Natural aeration, instead of mechanical aeration, saves the treatment plant significant amounts of money given the high cost of mechanical aeration. In North America, there are \$600-800 million worth of aeration systems in operation and municipalities spend about \$600 million to power these aeration systems (Daigger and Butz, 1998).

Aerobic ponds, facultative ponds, anaerobic ponds, aerated lagoons, and polishing ponds are specific types of waste stabilization ponds (Davis and Cornwell, 1991). Of these, the most common and the type favored by smaller communities is the facultative pond. “Approximately 25 percent of the municipal wastewater treatment plants in this country are ponds and about 90 percent of these ponds are located in communities of 5,000 people or fewer” (Davis and Cornwell, p.367). These types of ponds are popular because the long retention times (about six months) “facilitate the management of large fluctuations in wastewater flow and strength with no significant effect on effluent quality. Also capital, operating, and maintenance costs are less than that of other biological systems that provide equivalent treatment” (Ibid.).

Facultative ponds are shallow ponds that treat wastewater in two zones that form in the pond: the aerobic zone and the anaerobic zone (Hammer and Hammer, Jr., 1996; Davis and Cornwell, 1991). In the former zone, oxygen from algal photosynthesis and aerobic bacteria facilitate organic decomposition and in the latter zone chemical species facilitate organic decomposition. Sludge collects at the bottom of the pond for later treatment and disposal.

### ***Fixed Film Processes***

The Lawrence (Massachusetts) Experiment Station developed the trickling filter treatment method in 1892 (Stoddard *et. al.*, 2002). This method involves the construction of a bed of coarse material (i.e. rocks, corrugated plastic) that is used to provide the medium for microbial growth necessary for organic decomposition (Liu and Liptak, 1999; Davis and Cornwell, 1991). The process begins by trickling wastewater, seeded with microorganisms, over the bed of rocks. Bacteria begin to grow on the rocks as the wastewater passes over the rocks.

The surface of the rocks provides contact between the organics and the microorganisms as well as sufficient aeration to the wastewater during the re-circulation



of the wastewater over the rocks (Ibid.). Contrary to its name, the trickling filter method does not actually filter the wastewater so once the wastewater has been sufficiently treated, it moves on to a secondary settler for solids to settle out of solution (Ibid.). The sludge is then treated and disposed.

The second type of fixed film process involves the use of rotating biological contactors (RBCs). The RBCs are a series of closely spaced plastic discs, mounted on a horizontal shaft and rotated while half of their surface area is immersed in the wastewater (Davis and Cornwell, 1991; Hammer and Hammer, Jr., 1996). The wastewater itself sits in a reservoir. As the discs slowly rotate in and out of the wastewater, they carry wastewater into the air where it trickles down the surface of the discs, absorbing oxygen and carrying it into the reservoir. Microbes in the wastewater adhere to the surface of the rotating discs and grow there until they cover the discs with a 1-3 mm layer of biological slime (Davis and Cornwell, 1991).

As the attached microbes pass through the reservoir, they absorb other organics for breakdown thus providing treatment to the wastewater. Any excess growth of microbes is sheared from the discs as they move through the reservoir and are prevented from settling in the reservoir by the rotating discs. These discs fulfill the requirements for secondary treatment methods by: providing media for microbial growth; bringing this growth into contact with the wastewater; and aerating the wastewater and suspended microbial growth in the reservoir (Ibid.). Once the water is fully treated, it flows from the reservoir into a basin for final settling. As in other treatment methods, the sludge is treated and disposed.

The RBC treatment method “can achieve secondary effluent quality or better. By placing several sets of discs in series, it is possible to achieve even higher degrees of treatment, including biological conversion of ammonia to nitrates” (Ibid., p. 370).

Regardless of the biological treatment method employed at a treatment plant, before the treated wastewater can be released to the receiving water it must be disinfected (Davis and Cornwell, 1991; Hammer and Hammer, Jr., 1996). To disinfect the treated wastewater, the common practice is to add chlorine gas or chlorine in some other form to the wastewater and leave it there for about 15mins (Davis and Cornwell, 1991). Another common practice is to expose the wastewater to ultraviolet light and avoid the use of

chlorine since it can be toxic to aquatic life at a high enough concentration (Mancl, 1996).

The processes just outlined are designed to directly reduce the biochemical oxygen demand of the wastewater before releasing it back into the environment. However, secondary treatment has a minimal impact on the nitrogen levels in the wastewater (Davis and Cornwell, 1991; Hammer and Hammer, Jr., 1996). This inability to reduce nutrient levels is relevant because some waterbodies fail to meet ambient water quality goals due to elevated nutrient levels.

### ***Biological Nitrogen Removal (BNR) and BOD control***

During the organic decomposition phase of the secondary treatment process, the microorganisms that degrade organic compounds in the wastewater do not oxidize the nitrogen in the compounds (Hammer and Hammer, Jr., 1996). Rather, the nitrogen is released into the water as ammonia. Industrial wastes and fertilizers are two other sources of ammonia in the wastewater (Davis and Cornwell, 1991). Even in low concentrations, nitrates and ammonia encourage excessive algal growth (Ibid.). These algae in turn consume dissolved oxygen when they die and decompose (Ibid.).

The ammonia remains in the wastewater unless a special group of nitrifying bacteria is present to oxidize the ammonia to nitrite and then to nitrate (Hammer and Hammer, Jr., 1996; Davis and Cornwell, 1991). These nitrifying bacteria are found in low concentrations in the influent but increase in concentration as the wastewater undergoes secondary treatment (Ibid.). Nitrate is less oxygen demanding and less toxic than ammonia so nitrification results in a chemical that is less detrimental to the water and to the fish that inhabit it (personal communication, Randall, C.). However, the nitrification process uses up large amounts of dissolved oxygen (Gray, 1999; Davis and Cornwell, 1991).

If the treated water spends enough time in the aeration tank for anoxic<sup>30</sup> conditions to develop, denitrification will occur, thereby reducing the amount of nitrogen in the water and by extension, the oxygen demand of the effluent (Davis and Cornwell, 1991). However, the amount of nitrogen removed here is not nearly as much as is

---

<sup>30</sup> Anoxic conditions in water mean that the water lacks enough dissolved oxygen and decomposition proceeds by using nitrogen instead of oxygen (Hammer and Hammer, Jr., 1996; Gray, 1999).

removed during a process that specifically targets nitrate-nitrogen, such as the biological nutrient removal treatment method.

Any type of treatment beyond conventional secondary treatment to reduce BOD is generally classified as advanced or tertiary treatment (Hammer and Hammer, Jr., 1996; Stoddard *et. al.*, 2002). Nutrient reduction is one type of tertiary treatment that operators can apply to wastewater. Nitrogen reduction has a chemical and biological form (Davis and Cornwell, 1991).

Ammonia stripping is the chemical method of nitrogen reduction (Ibid.). During this process, operators raise the pH of the treated effluent, converting the ammonium ion into ammonia, and then strip the ammonia by forcing large quantities of air through the water. Since the process cannot reduce nitrates, the secondary treatment process must operate at a short cell (i.e. microorganism) detention time so that nitrification cannot take place (Ibid.).

Environmental concerns about some of the chemicals used to reduce nitrogen have led to growing interest in biological methods of nitrogen removal. There are several methods of accomplishing biological nitrogen removal (BNR) and they basically rely on a two-step or simultaneous process of nitrification and denitrification.

One type of process forces the natural nitrification process to occur during the secondary treatment phase (i.e. during the activated sludge process) by increasing the cell detention time to 15 days or more (Davis and Cornwell, 1991). Then bacteria, feeding on a natural or synthetic source of organic matter, accomplish anoxic denitrification of the nitrates and convert them into nitrogen gas, carbon dioxide, and water (Ibid.). The conversion of nitrates to nitrogen gas releases the nitrogen from the water into the atmosphere. Other methods of biological nitrogen removal employ a two-stage method in which wastewater that has undergone conventional secondary treatment is successively circulated through aerobic and anaerobic conditions that each encourage nitrification and denitrification via an activated sludge process.

Biological nutrient removal (BNR) systems can significantly reduce the concentration of nutrients in the wastewater. A survey of municipal WWTPs indicated that incoming wastewater to the plant has nitrate concentrations (as ammonia) at 18 mg N/L (Stoddard *et. al.*, 2002). Advanced wastewater treatment is able to reduce the nitrate

concentrations to 2.0 mg N/L (Stoddard *et. al.*, 2002), a reduction of 89 percent. Conversely, secondary treatment, via incidental denitrification, can reduce nitrogen by 0 to 30 percent (USEPA, 1999b).

Efforts to reduce the concentration of nutrients also reduce the concentration of BOD in the plant's effluent (Stoddard *et. al.*, 2002; Hammer and Hammer, Jr., 1996; Gray, 1999). Given a concentration of 205 mg/l of BOD in the influent, secondary treatment will reduce this concentration of BOD to 16.4 mg/l. Advanced treatments that target nitrogen for removal will reduce the concentration of BOD to 4.1 mg/l (Ibid.).

There is a biological reason for the relationship between nitrogen control and BOD control. Denitrifying bacteria require anoxic conditions and a suitable organic carbon source in order to convert nitrates to nitrite and finally to nitrogen gas (Gray, 1999; Chesapeake Bay Program, 2003). The BOD test measures the reduction in carbonaceous organic matter therefore a process such as nitrification/denitrification that consumes organic carbon will also reduce the concentration of carbonaceous BOD. In activated sludge systems for example, "for each 1 g of nitrate N utilized 2.9 g of BOD is also removed" (Gray, 1999, p.426). Thus, there is a technical relationship between nitrogen control and BOD control.

In addition to highlighting the existence of jointness, this overview of wastewater treatment technology highlighted the uncertain nature of BOD control. Rather than target a predetermined level of control, the treatment plant can only facilitate optimal bacterial growth and organic decomposition by providing an enabling environment in terms of adequate food for the microorganisms, the right temperature, and adequate time in the tank. Any process that depends on bacterial growth, flocculation, water temperature, and other environmental variables over which the plant operator has little control will not be deterministic. The conceptual framework for analyzing plant-level decisionmaking described above is applicable to WWTPs, with some modifications.

### ***Implications for the Conceptual Framework of Overcompliance***

The graphs looking at the effect on Q of parametric shifts in the *MEP* and *MAC* curves can be viewed in the cross-section as modeling differences between plants where the differences are related to the degree of variation in BOD output and to whether or not

the plant controls nitrogen and thus can benefit from jointness. The cross-sectional perspective is consistent with the goal of the paper, which is to explain differences between firms in the degree of BOD overcompliance at any given point in time based on differences in the variance in their BOD output and in the degree of nitrogen control.

Two hypotheses will be analyzed within the conceptual framework of behavioral responses: the implications for variation in BOD output will be examined followed by a discussion about the implications of jointness.

### ***Implications of Stochastic Effluent for BOD Overcompliance***

The review of the economic, technology and engineering literature indicates that BOD output has an underlying distribution. This introduces the possibility that certain characteristics of the distribution of BOD output may make the plant more likely to overcomply. The literature has identified variation in BOD effluent discharges as a possible driver of observed overcompliance at wastewater treatment plants (Bandyopadhyay and Horowitz, 2002; McClelland and Horowitz, 1999; Brannlund and Lofgren, 1996).

While there has been little empirical research examining this relationship, some results are worth mentioning. Consistent with the predictions of the conceptual model, Bandyopadhyay and Horowitz (2002, 2004) found a statistically significant negative correlation between the variance in effluent discharges and the median effluent discharge for WWTPs in the U.S. from 1992-1999.

Looking at annual data from 1989-1990 on loadings of BOD, carbonaceous oxygen demand, and TSS, for 30 Swedish pulp plants, Brannlund and Lofgren (1996) observe that most plants are overcomplying with their permit limits on average.<sup>31</sup> They model firm behavior under a profit-maximizing framework given the stochastic pollutant output and the uncertain penalties for noncompliance. They find that firms faced with stochastic output, a probability of inspections, and a positive penalty for violating the permit limit “may find it profitable to pollute less, on average, than the allowed level in order to avoid a too-high expected penalty cost” (Brannlund and Lofgren, 1996). The

---

<sup>31</sup> Carbonaceous oxygen demand is the oxygen demanded by decomposing organic matter that release carbon during decomposition (Hammer and Hammer, 1996). Biochemical oxygen demand (BOD) is the oxygen demand of organic matter that releases carbon as well as nitrogen during decomposition.

reasoning is that the higher the variance of the distribution the higher the risk of a violation at any given mean discharge level and therefore, the lower the desired mean discharge level.

An implicit assumption made in this argument is that plants have sufficient control over aspects of the production process to adjust in real time for changes in the variation of BOD output. McClelland and Horowitz (1999) point out that there are several inputs that plants can vary to directly affect the level of BOD output. Brannlund and Lofgren (1996) assume that while a plant cannot control its BOD exactly, the plant has indirect control over BOD output because it has the ability to “improve the precision level” of BOD output through its control of a specific set of inputs (Ibid.).

Interviews with state regulators and plant operators confirm the negative association between mean output level and variance but some officials attribute the negative relationship to the decisions of dischargers and regulators during the construction of the plant when capital decisions are made rather than to the day-to-day decisions of plant operators. In the construction phase, regulators specify permit limits based on the results of water quality models and then engineers hired by the owner of the WWTP design the plant to meet the level of control specified in the permit limit, working with state regulators (and their engineers) until the design specifications are approved (personal communication with Ta-Shon Yu, Maryland Department of the Environment).

Decisions about what initial capital equipment to install, and by extension, what level of BOD control the plant can achieve via secondary treatment occur during the capital design phase. According to this viewpoint, any adjustment for high variance in BOD comes at the plant’s capital design phase, not at the operational phase. The motivation that is offered for the desire to adjust for variance in the capital design phase is similar to those identified by Brannlund and Lofgren (1996): the owners’ desire to avoid paying penalties for noncompliance and the recognition that BOD loadings will vary over time.

One state official described the engineers hired by the municipality as risk averse with respect to their preferences over the plant’s compliance with BOD permit limits given the natural variation in BOD output. Engineers working for the municipality would rather avoid client suits that may arise in the event that the plant cannot meet its NPDES

permit limit so they prefer to design the plant to control average BOD output to some fraction of the NPDES permit limit (personal communication with S. Luckman, Maryland Department of the Environment).

Another state official describes the engineers' decision making process in a bit more detail. Before building a plant, engineers use mathematical models to simulate flow and concentration to the plant in order to answer two questions: (1) Can the treatment plant handle the flow of water to the plant? (2) Can the treatment plant provide adequate biological treatment to the incoming wastewater? (personal communication with G. Johnson, Connecticut Department of Environmental Protection). The former is a hydraulic concern while the latter is an organic concern and the answers to the questions will be different for each plant since wastewater usage and composition is not the same for every municipal WWTP. With consideration for the volume of inflow to the plant, the variation in BOD output, and the plant's current BOD permit limit, engineers design treatment plants to produce BOD output that falls within a certain range and reduces the chance that the BOD output will exceed the permit limit to a predetermined, acceptable level.<sup>32</sup>

Given the variation in BOD output and the plant's desire to minimize expected penalties, a general rule of thumb is that plants are designed to operate (i.e. emit BOD output) at a level that is within half to two-thirds of the BOD permit limit (personal communication with S. Luckman and T-S. Yu, Maryland Department of the Environment). Presumably, in determining the extent of 'over-design', the owners of the treatment plant will be concerned with the actual modeled variation in BOD, the regulator's predisposition to conduct regular inspections at the plant, the plant owner's aversion to the risk of incurring penalties for noncompliance, and the regulator's predisposition to assessing penalties for noncompliance. In any case, the plant is built such that on average, the plant will overcomply with its permit limit.

In explaining the reason behind the negative relationship between the variance of BOD and the mean BOD output, the qualitative evidence focuses on capital decisions while the quantitative evidence assumes that the plant can smoothly vary several inputs

---

<sup>32</sup> From personal interviews with officials from the Connecticut Department of Environmental Protection and the Maryland Department of the Environment.

(not just capital) to affect either BOD or the variance of BOD. In either case, the negative relationship between the variation in BOD and the mean BOD level is established and is consistent with the conclusions derived from the conceptual model of BOD overcompliance in Section 2.

Concerning possible modifications to the form of the conceptual model, cost-sharing arrangements between the regulator and the discharger can affect the permit limit that the discharger finds acceptable in the first stage. States typically provide cost-share assistance to WWTP for capital projects in the form of low interest loans or outright grants (T-S Yu, Maryland Department of the Environment). Cost-share programs effectively lower abatement costs faced by the discharger and could result in discharger acceptance of capital plant designs with lower levels of discharge. This is a feature of the regulatory program that could be incorporated in to the conceptual model.

The conceptual model can also accommodate situations in which the plant is a multi-product plant. The next section considers a plant that controls nitrogen and BOD, and the implications for BOD overcompliance of jointness between nitrogen control and BOD control.

### ***Implications of Jointness for BOD Overcompliance***

Before discussing the relationship between nitrogen control and BOD control, it is useful to discuss the regulatory context for nitrogen control and how regulators provide plants regulated under the Clean Water Act with the incentives to control a pollutant that is not specifically regulated by the Clean Water Act.

In many parts of the country water pollution due to elevated levels of nitrogen is an increasing environmental concern. However, unlike the case of BOD discharges, there is no federal statutory requirement to control nitrogen discharges. Local water quality concerns have resulted in state-level initiatives to control nitrogen: requiring reductions via water-quality based limits and providing incentives to encourage voluntary reductions.

In the case of mandatory reductions, the limits must be specified in the plant's NPDES permit. The nitrogen limit is formally written into the permit per the requirements of a Total Maximum Daily Load requiring reductions in nitrogen



discharges, as is the case with WWTPs in Connecticut<sup>33</sup> (Connecticut Dept. of Environmental Protection, 2002). The permitted level may also be reduced over time, forcing additional nitrogen control (Ibid.).

Often, state regulators use voluntary financial inducements to limit nitrogen discharges from WWTPs to target levels (Chesapeake Bay Foundation, 2003a; Connecticut Department of Environmental Protection, 2002). In order to facilitate nitrogen reduction efforts, regulators can establish financial incentives via cost-share arrangements which reduce the WWTP's costs of upgrading the technology at the plant to provide nitrogen control (Maryland Department of the Environment(2); Chesapeake Bay Foundation, 2003a; Sarbanes, 2002).<sup>34</sup> Maryland, for example, encourages nitrogen reductions by providing financial incentives for plants seeking to upgrade their facilities via the provision of capital construction grants and low interest loans.

The Maryland Water Quality Infrastructure Program, through the Biological Nutrient Removal Program, provides capital grants funds on a 50/50 cost share basis to plants that treat at least 0.5 million gallons per day (Maryland Department of the Environment(2)). Engineering and pilot study, design and construction costs are eligible costs under this program (Ibid.). Once the BNR technology is installed, WWTPs

---

<sup>33</sup> The permit regulating discharges from wastewater treatment plants in Connecticut is a general permit rather than an individual, plant-level permit (Connecticut Department of Environmental Protection, 2002). A description of Connecticut's general permit reads: "The general permit for nitrogen discharges will be used to regulate Connecticut's nitrogen permitting needs on a state-wide cumulative basis rather than issuing individual permits to the 79 individual municipal facilities in the state regulated under the general permit a process that is more efficient and streamlined compared to issuing individual permits. Nitrogen discharges will be controlled on a statewide basis. Utilizing the general permit, reductions will be achieved by controlling the amount of discharge coming from individual watersheds throughout Connecticut that collectively feed into the larger Long Island Sound watershed. Over time, the nitrogen discharge levels will be ratcheted down as Connecticut moves towards the goal of reducing nitrogen to the Sound by more than 60% by 2010." Website: <http://www.dep.state.ct.us/whatshap/press/2002/mf0214.htm> There are three types of permits: individual permits, general permits, and "permit by rule." The individual permit is most tailored to the conditions at the plant in terms of its pollutants, operating conditions, and the quality of the receiving water. General permits are used when the covered plants are similar in the type of pollutants they discharge and in the technology employed. A "permit by rule" is one where permit requirements originate from a formal administrative rule. See: [http://www.michigan.gov/deq/0,1607,7-135-3313\\_3682\\_3713-10252--,00.html](http://www.michigan.gov/deq/0,1607,7-135-3313_3682_3713-10252--,00.html) or <http://www.epa.state.oh.us/dsw/permits/gpfact.html>

<sup>34</sup> As of May 2004, wastewater treatment plants in Maryland can apply for and receive financial assistance that is up to 100 percent of planning, design, and construction costs to upgrade the plant from BNR technology to ENR technology. This newer technology, providing Enhanced Nutrient Removal, allows plants to reduce nitrogen concentrations in the effluent to 3mg/l (Maryland Department of the Environment(1)).

participating in the cost share agreements are expected to achieve an effluent nitrogen concentration of 8mg/l (Chesapeake Bay Foundation, 2003a).<sup>35</sup>

Cost share programs frequently accompany the conventional regulatory permitting programs. Connecticut, for example, also administers a grant and low-interest loan program to facilitate the implementation of nitrogen control. Acceptance of the cost-share funds commits plants to a specific level of performance that may be more stringent than the level specified in their permit. Several plants in Connecticut have accepted cost-share funds so that even though they face a specified permit limit on nitrogen, the more relevant constraint will be the target level that they voluntarily agreed to achieve by accepting state funds. If the nitrogen limit is 10mg/l and the cost-share-induced target level is 6mg/l, then the cost-share target is the binding constraint, but only for plants that accepted the public cost share funds.

Whether plants are required or encouraged to control nitrogen discharges, the decision to install BNR capital equipment during the operational phase has implications for BOD control. The jointness that exists between nitrogen control and BOD control will mean that plants with BNR technology are more likely to have lower BOD discharges in the absolute sense and relative to the plant's BOD permit limit.

## **Summary**

A behavioral model of plant-level decision-making explains the possibility of overcompliance among WWTPs. A plant that faces a higher variation in BOD will control more BOD than an otherwise similar plant and a plant that installs nitrogen control technology will control more BOD than a similar plant without BNR technology. In the case of jointness, the reductions in BOD come at no additional marginal cost since they are due to the nature of the technology. The two hypotheses could be tested using cost and effluent data.

While there is no cost data available, there is effluent data available for regulated WWTPs in Maryland and Connecticut. The data will allow tests of the first component

---

<sup>35</sup> The Chesapeake Bay Foundation rates WWTPs in the Chesapeake Bay Watershed as Excellent, Good, Needs Improvement or Unacceptable depending on whether they are controlling nitrogen levels below 3mg/l, between 3mg/l and 5mg/l, between 5mg/l and 8mg/l, or above 8mg/l (Chesapeake Bay Foundation, 2003a).

of the marginal cost hypothesis: whether there is jointness in pollution control between nitrogen and BOD leading to incidental overcompliance. The data on BOD and nitrogen discharge contain discharge information from WWTPs located in two watersheds that have been targeted for nitrogen reduction. The states in the Chesapeake Bay and the Long Island Sound have well-developed programs in place to monitor and reduce nitrogen discharges to the estuary. The effluent data will be used to test whether and the extent to which the presence of nitrogen control translates into further BOD control than that achieved by secondary treatment.

The data can also be used to test whether higher variation in BOD output is associated with lower mean BOD output. If the relationship holds, then there will be a correlation between the variance of BOD and the mean BOD output. The expectation is that this correlation will exist and will be statistically significant since the qualitative and quantitative evidence point to a negative relationship between variation in BOD and mean BOD output. Using a multivariate regression framework allows one to test the relative contribution that variation in BOD output and joint production make in determining the observed degree of overcompliance.

Testing for the impact of jointness and variation in BOD on BOD overcompliance has not been done before in the literature. To the author's knowledge, there has been no study linking nitrogen control to BOD control in a regression framework. Only a few empirical studies examine the link between variation in BOD and mean BOD output and even fewer consider the relationship in a regression framework. Brannlund and Lofgren (1996) observe that a majority of plants in their dataset are in significant overcompliance with their BOD permit limits. McClelland and Horowitz (1999) come to the same conclusion based on a graphical analysis of the plants in their dataset.

To the author's knowledge, a recent working paper (Bandyopadhyay and Horowitz, 2002, 2004) is the only one to consider the relationship between variation in BOD and the degree of BOD overcompliance in a regression framework. Their paper finds that the effect of variation in BOD on the degree of BOD overcompliance was negative and statistically significant. Bandyopadhyay and Horowitz (2002, 2004) do not consider the role of jointness in pollution control on BOD overcompliance therefore their

paper cannot draw any conclusions about the relative contribution of variation in BOD to the observed degree of overcompliance.

The next section presents an empirical investigation of the relationship between certain explanatory factors and BOD overcompliance and begins with a general description of the wastewater treatment plants.

#### **Section 4: Empirical Analysis of Overcompliance**

This section presents the results from a quantitative analysis of the relationship between both jointness and overcompliance with permit limits for BOD, as well as the relationship between variation in BOD output and BOD overcompliance.

##### ***Data***

Since BOD is a conventional pollutant that is regulated under the Clean Water Act, the available effluent data on BOD discharges from WWTPs will cover plants in a wide geographic region. There is much less effluent data available on nitrogen discharges because comparatively fewer plants face federal monitoring requirements, let alone reduce their nitrogen discharges. However, in order to examine the implication of jointness on BOD overcompliance, the data needs to include plants that control nitrogen and not just BOD.

Since the typical motivation for initiating nitrogen monitoring or reduction programs is environmental degradation, it is reasonable to focus the search for nitrogen and BOD effluent data on areas with nutrient-enriched surface waters. Connecticut and Maryland are two states with a long-standing interest in controlling nutrient discharges to the surface water. Nutrient enrichment is a well-known problem in the Long Island Sound watershed in which Connecticut is located and in the Chesapeake Bay watershed in which practically the entire state of Maryland is located (Connecticut Department of Environmental Protection, 2001; Chesapeake Bay Foundation, 2003(b)).

Regulators in both states are committed to reducing nitrogen discharges to the surface waters (Connecticut Department of Environmental Protection, 2001; “Nutrient Pollution”; Maryland Department of the Environment (1); Connecticut Department of the Environment, 2002; Fahrenthold, 2004). Nitrogen discharge restrictions will thus be binding on plants in these watersheds with specific nitrogen limits or that have accepted

financial assistance to upgrade their facilities. There are two separate state-level datasets that will be used for the analysis. The Connecticut (CT) dataset was received from the Connecticut Department of Environmental Protection (CTDEP) and contains monthly measurements on nitrogen (N) effluent concentrations and monthly flow measurements for the time period 1997-2003 on CT wastewater treatment plants. Data were also collected from CTDEP on plant design flow capacity, and monthly BOD performance and permit limits. Data on BOD performance were limited to the time period 2002-2003. There are 79 WWTPs in the CT general permit during the 2002-2003 time periods.

The data also contain information that identifies the plants in Connecticut that have been retrofitted or completely overhauled in order to install Biological Nutrient Removal technology. Given the jointness in pollution control that exists between N and BOD, the effluent from a plant with BNR technology installed should have lower BOD concentrations than the effluent from a similar plant without BNR technology.

The data for Maryland are made available by the Maryland Department of the Environment (MDE). The dataset consists of monthly data on BOD, BOD permit limits, plant design flow, and monthly flow data on 171 wastewater treatment plants for the time period 2001-2003. The corresponding monthly nitrogen discharge data is from the Chesapeake Bay Program (CBP) office's Point Source database.

The CBP Point Source database is compiled by the Chesapeake Bay Program<sup>36</sup> and covers the time period 1984-2002. The dataset includes over 600 selected facilities that discharge treated wastewater to the Chesapeake Bay or to one of its tributaries. The data cover several states including Maryland, Delaware, Pennsylvania, Virginia, West Virginia, New York, and the District of Columbia.

The CBP Point Source database collects data on pollutant discharges from wastewater treatment facilities and industrial facilities. The dataset includes monthly flow, nitrate, nitrite, total kjeldahl nitrogen, total organic nitrogen, total organic phosphorus, total nitrogen (N), and total phosphorus (P) concentrations in treated effluent that is discharged to the Bay.

---

<sup>36</sup> The Chesapeake Bay Program is a regional partnership of individual states in the Chesapeake Bay, the Chesapeake Bay Commission, the Environmental Protection Agency, and advisory groups.

Maryland submits data only on municipal wastewater treatment plants discharging at least 1,000 gallons per day as an annual average to the Bay or its tributaries. The CBP database contains all the information in the MDE dataset except for the data on permit limits that is necessary for analyzing overcompliance therefore the CBP dataset must be merged with the MDE dataset. Merging the MDE dataset with the CBP dataset reduced the number of plants originally in the MDE dataset from 171 to 133. Attrition in the cross-section of the merged dataset resulted from the fact that many plants were missing observations on key pollutants like BOD and nitrogen. Formal statistical tests showed that the plants missing BOD and nitrogen data were smaller plants.

Attrition in the cross-section also resulted from the fact that the CBP dataset only includes plants discharging at 1,000 gallons per day while the MDE dataset did not have a similar exclusion criterion. Attrition over time in the merged dataset resulted from the fact that the time period in the MDE dataset (24 months) is shorter than the time period in the CBP dataset (228 months).

One of the benefits of having more than one source of data on the same variables is that it provides the opportunity to cross check the values in each dataset. The MDE dataset included data on BOD monthly concentrations, TSS monthly concentrations, design flow and actual flow for plants in Maryland for 2001 and 2002. The data from the Chesapeake Bay Program also contained similar information for the time period in question, confirming the consistency of the values across datasets.

The geographic scope is limited to these two states because attempts to acquire a more comprehensive database from the EPA were not fruitful. Initially, a data request was presented to the EPA's Office of Permit Compliance, which maintains the Permit Compliance System (PCS) database.<sup>37</sup> The PCS database contains data on NPDES-permitted facilities nationwide from 1974 to the present. The PCS data promises to be very comprehensive since it is a repository of information on monthly pollutant discharges and permit limits from all the wastewater treatment plants in the country.

This comprehensive nature of the database makes it a very attractive source of information for researchers interested in analyzing discharges to surface waters.

---

<sup>37</sup> The PCS data were acquired through a Freedom of Information Act (FOIA) request to EPA's FOIA office.

Bandyopadhyay and Horowitz (2002, 2004) make use of PCS data in their study of BOD overcompliance. PCS data of monthly performance (effluent concentration) and permit limit data from major treatment facilities were requested and obtained for the time period 1997-2003.

However, the PCS data are inadequate for the present investigation for the primary reason that the data on nitrogen (N) discharges are practically nonexistent over plants and over time. Even when plants are likely to face nutrient discharge monitoring requirements (i.e. plants in Maryland, Virginia, New York, and Connecticut), the PCS nutrient discharge data on these plants remain thin. Without nitrogen discharge information, the implications of jointness between nitrogen control and BOD control for BOD overcompliance cannot be investigated.

### ***Description of WWTPs in Connecticut and Maryland***

This section will provide a more detailed description of the WWTPs in Connecticut (CT) and Maryland (MD). It will discuss the regulatory environment in which they operate in terms of limits on BOD and N, the prevalence of overcompliance and/or noncompliance with BOD limits, the pattern of variation in BOD output over time, the number of plants with BNR technology, and the evidence for jointness.

### ***BOD Permit Limits in CT and MD***

The 30-day average technology-based performance standard for BOD is 30mg/l. Table 7 shows the number of plants with permitted BOD limits at, below, or above the technology-based standard of 30mg/l for plants in Connecticut and Maryland.

**Table 7: Wastewater treatment plants grouped by BOD permit limit**

<b>State</b>	<b>BOD limit&lt;30mg/l (% of total)</b>	<b>BOD limit=30mg/l (% of total)</b>	<b>BOD limit&gt;30mg/ (% of total)</b>
<b>Connecticut</b>	20 plants (25%)	59 plants (75%)	0 plants (0%)
<b>Maryland</b>	49 plants (36%)	83 plants (62%)	1 plant (1%)

There is very little variation in the BOD permit limit across the plants in Connecticut. Seventy-five percent of all Connecticut WWTP plants face a BOD limit of 30mg/l. Nine percent face a permit limit of 25mg/l; 13 percent face a permit limit of

20mg/l; and 4 percent face a permit limit of 10mg/l. Plants with limits below the technology-based standard of 30mg/l have had their limits lowered based on water-quality based standards established to address a water quality problem in the plant's receiving water.

Maryland WWTPs exhibit slightly more variation in BOD permit limits. Sixty-two percent of plants face a permit limit of 30 mg/l. Eighteen percent of plants are in the 20-25mg/l range, 9 percent of plants are in each of the ranges: 25-30mg/l and from 10-20mg/l. There is one plant at 45mg/l and there is one plant less than 10mg/l.

Many MD plants face seasonal BOD limits where the summer BOD limit is lower than the winter BOD limit.<sup>38</sup> Summer limits usually apply from June to October while winter limits apply for the remaining months. In Maryland, 42 plants (i.e. 32 percent of the plants) have seasonal limits where the winter limit is at 30mg/l and the summer limit ranges from 5mg/l to 26mg/l, with most plants at summer limits between 10mg/l and 20mg/l.

According to an MDE official, the summer limits are lower because despite the warmer temperatures, some surface waters experience higher BOD levels in the summer (personal communication with S. Luckman, Maryland Department of the Environment). The culprit may be the excess production of algal blooms in the water due to nitrogen over-enrichment, increased availability of sunlight, and warmer temperatures. When these blooms die, they deplete the dissolved oxygen in the water, increasing the oxygen demand.

Low flow conditions may also be a contributing factor. During low flow conditions brought on either by a drought or excessive demand for water, the stream flow is at its lowest levels, the water holds less oxygen since it is not as dense (i.e. low levels

---

<sup>38</sup> The existence of seasonal limits seems to imply that operators can tighten BOD control at certain times of the year, discharging less BOD in the summer and more BOD in the winter. In a personal interview, an official of the Municipal Permits Division at the Maryland Department of the Environment explained that seasonal limits in Maryland reflect the fact that water quality conditions generally deteriorate during the summer due to elevated levels of dissolved oxygen (DO). As such, the permit reflects the need to discharge less BOD in the summer than in the winter, in order to protect water quality. Independently of the critical summer conditions, wastewater treatment plants are better able to control BOD in the summer because the biological processes are more efficient in the summer due to the warmer temperatures. So, in the absence of seasonal limits, a wastewater treatment plant running according to its design will discharge less BOD in the summer, *ceteris paribus*. Thus, while the seasonal BOD limits imply that plants can increase BOD control in the summer, the increased level of BOD control may not necessarily result from a behavioral choice but may be a consequence of the technology.



of dissolved oxygen), and pollutant concentrations increase as a result of the reduced dissolution characteristic of a low flow environment.<sup>39</sup> Low flow conditions can occur at any time of the year but in Maryland they occur during the June-September period.<sup>40</sup> The regulatory response to low levels of dissolved oxygen has been to reduce BOD and nitrogen discharge to the receiving water (personal communication with S. Luckman, MDE).

In addition to BOD, the plants in Connecticut face a formal limit on nitrogen due to the establishment of a Total Maximum Daily Load that requires that a wasteload allocation be made to WWTPs discharging to Long Island Sound. Those who have accepted cost-share funds will also face a nitrogen target that will typically be lower than the limit specified in the permit. The plants that accepted cost-share funds to retrofit the plant for BNR technology face a target nitrogen discharge level of 7.4mg/l on average (Connecticut Department of Environmental Protection)<sup>41</sup>. Those that accepted cost-share funds to completely upgrade to BNR technology face a nitrogen discharge level of 4.4mg/l on average (Connecticut Department of Environmental Protection). Table 8 shows the distribution of nitrogen target levels across the plants in Connecticut that installed BNR technology.

---

<sup>39</sup> DNREC News. (2001). Vol. 21, No. 20. Delaware Department of Natural Resources and Environmental Control. <http://www.dnrec.state.de.us/DNREC2000/Admin/News/01BreakingNews/0122TMDL.htm> Accessed 10.18.04.

<sup>40</sup> From official website of the Maryland Department of Natural Resources: <http://www.dnr.state.md.us/streams/mbss/synopsis.html> Accessed 10.18.04.

<sup>41</sup> The nitrogen targets for plants in Connecticut are expressed as a range, usually with a size of 1-2mg/l from the minimum to the maximum nitrogen discharge level. In estimating the average target nitrogen discharge level, the midpoint of the range was taken as the target level for a given plant.

**Table 8: Nitrogen target levels amongst plants in Connecticut and Maryland**

Type of BNR upgrade	Number of plants	Target nitrogen range in mg/l
<b>Connecticut</b>		
<b>Retrofit</b>	1	3-4
<b>Retrofit</b>	1	5-6
<b>Retrofit</b>	10	6-8
<b>Retrofit</b>	6	8-10
<b>Full-scale</b>	5	3-4
<b>Full-scale</b>	2	4-5
<b>Full-scale</b>	5	5-6
<b>Maryland</b>		
---	38	8

Maryland does not impose a mandatory permit limit on nitrogen, but instead promotes nitrogen control by cost-sharing BNR upgrades at WWTPs. The acceptance of cost-share funds, however, carries an expectation (through cost share agreements) that specific levels of nitrogen control will be achieved. Maryland plants receiving financial assistance to install BNR technology agree to achieve a target nitrogen discharge level of 8mg/l.

***Evidence of Overcompliance and Noncompliance at WWTPs***

Overcompliant and noncompliant behavior exists amongst the CT and MD plants in the data. Twenty-seven percent of plants in Connecticut (21 out of 79) experienced at least one incidence of noncompliance in which the reported monthly BOD output level exceeded the permitted monthly level. Twenty percent of plants in Maryland (26 out of 133) experienced at least one incidence of noncompliance with a monthly BOD limit. Table 9 shows the number of exceedances over the 2-year time period and the number of plants with this exceedance value. Amongst those who are out of noncompliance, most plants will have no more than 3 exceedance events over the 2-year time period.

**Table 9: Incidence of noncompliance with monthly BOD permit limits in Maryland and Connecticut**

<b>Incidence of noncompliance</b>	<b>Number of plants (% in parentheses)</b>
<b>Connecticut</b>	
0	58 (73%)
1	14 (18%)
2	5 (6%)
4	1 (1%)
6	1 (1%)
<b>Maryland</b>	
0	107 (80%)
1	12 (9%)
2	6 (4%)
3	2 (2%)
4	2 (2%)
5	1 (1%)
8	1 (1%)
9	1 (1%)
23	1 (1%)

However, there is also strong evidence of overcompliance. All plants are overcompliant over some range of the time period in both datasets. To examine the amount of overcompliance, a compliance ratio is calculated. The compliance ratio is defined as the ratio of the reported monthly average BOD output to the monthly permitted level of BOD output. A compliance ratio less than 1 indicates overcompliance and a value greater than 1 indicates that the plant is in violation of its permit limit.

On average, plants in Connecticut are overcomplying with their NPDES BOD permit limits by a large margin. The median (mean) compliance ratio over time and over plants is 0.22 (0.29). So, on an average month, the typical CT WWTP is discharging only 22% of its permitted limit. Forty-four plants have a compliance ratio between 0 and 0.25; 24 have a compliance ratio between 0.25 and 0.5; 9 have a compliance ratio between 0.5 and 0.75 and 2 have a compliance ratio between 0.75 and 1. Fully 86 percent of plants are emitting on average less than half of their permitted levels of BOD.

Maryland WWTPs shows a similar pattern. The median (mean) value of the compliance ratio is 0.2 (0.3). A breakdown of plants by their median compliance ratios shows that of the 133 plants, 82 plants have an average compliance ratio less than or

equal to 0.25, 38 plants have a compliance ratio between 0.25 and 0.5, 8 plants have a compliance ratio between 0.5 and 0.75, 4 plants have a compliance ratio between 0.75 and 1, and 1 plant has a compliance ratio above 1. Fully 90 percent of plants are emitting on average less than half of their permitted levels of BOD. The data suggests that overcompliance is widespread and significant. This level of overcompliance is similar to that found in other studies (McClelland and Horowitz (1999), Brannlund and Lofgren (1996)).

### ***Stochastic Pollution***

One potential explanation for the observed overcompliance identified in section II is that plant operators compensate for variation in BOD output by reducing average BOD output. This explanation relates to the natural variation in BOD output over time. This natural variation may also explain the occasional noncompliance that is observed at these plants. The variation in BOD output can be investigated empirically because the Connecticut and Maryland datasets are in longitudinal form.

Tables 10 and 11 present information on the variation in BOD output over time for plants in Connecticut and Maryland, followed by graphs in Figures 12 and 13 of the mean (estimated across plants for each month) monthly BOD over the year for plants in both states.

**Table 10: Monthly variation in BOD output for Connecticut**

<b>State</b>	<b>Month</b>	<b>BOD Mean</b>	<b>BOD Std.Dev</b>
<b>Connecticut</b>			
	January	12.9	21.76
	February	10.5	7.52
	March	10.51	9.55
	April	8.66	6.76
	May	8.41	6.17
	June	8.93	9.11
	July	9.03	15.2
	August	7.41	6.23
	September	7.66	6.87
	October	7.85	8.02
	November	8.38	7.03
	December	11.05	21.8

**Table 11: Monthly variation in BOD output for Maryland**

<b>State</b>	<b>Month</b>	<b>BOD Mean</b>	<b>BOD Std.Dev</b>
<b>Maryland</b>			
	January	9.7	10.13
	February	9.51	9.28
	March	9.5	9.13
	April	8.99	8.44
	May	8.34	7.38
	June	7.52	7.03
	July	6.87	7.21
	August	6.5	7.19
	September	6.27	6.75
	October	6.71	7.22
	November	7.02	6.96
	December	7.53	7.1

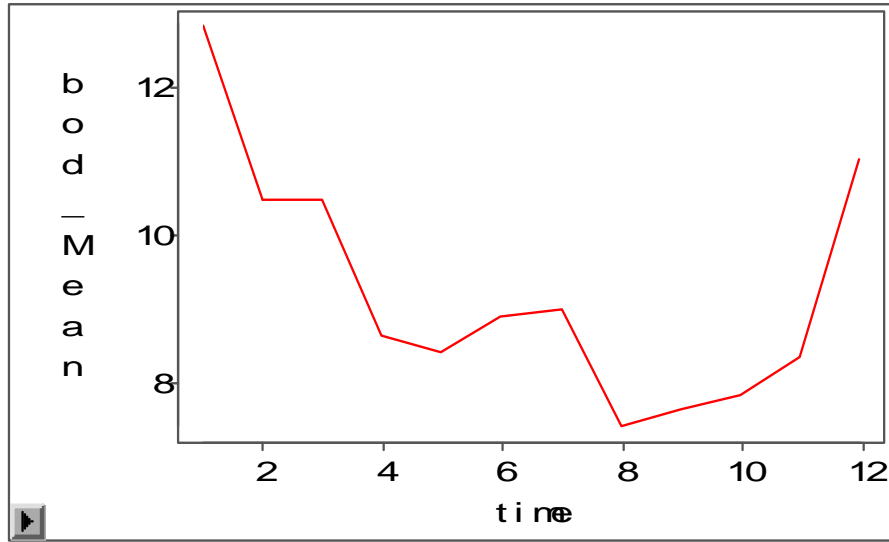


Figure 12: Plot of mean (over plants) monthly BOD output over the year for Connecticut

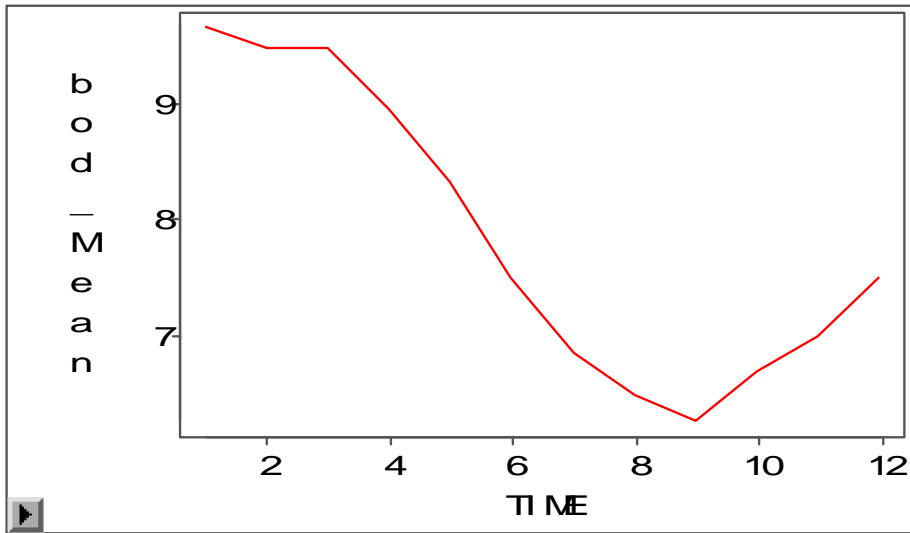


Figure 13: Plot of mean (over plants) monthly BOD output over the year for Maryland

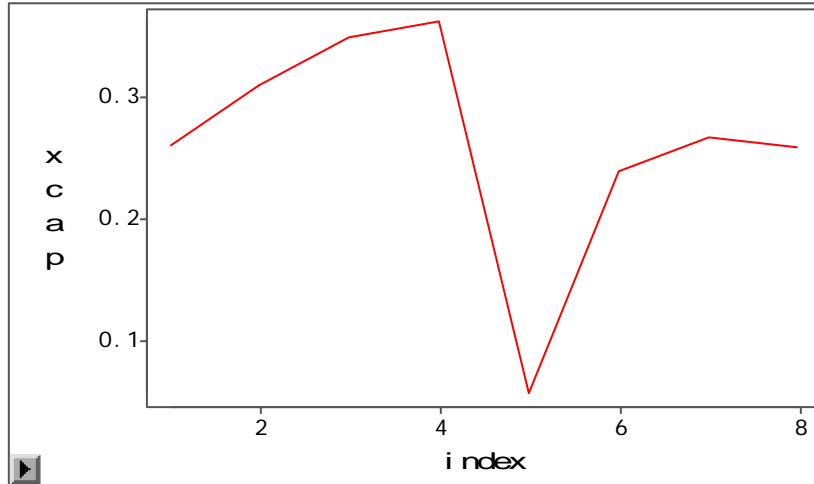
The graphs indicate that BOD output follows a slight U-shape over the course of the year: higher at the beginning and at the end of the year and lowest during the middle of the year, that is, during the summer months. The effect of temperature on BOD control may help explain the shape of the curve (Liu and Liptak, 1999; Hammer and Hammer, 1996). BOD control is easier when the water temperature is warmer because the biological growth processes needed for adequate flocculation and settling proceed faster in warmer than in colder temperatures (Liu and Liptak, 1999; Hammer and Hammer, 1996; personal communication with S. Luckman, Maryland Department of the Environment and G. Johnson, Connecticut Department of Environmental Protection). Based on this, one would expect to see lower BOD output in the summer.

Even so, plants may experience weather-related events that cause them to release unusually high levels of BOD despite the relatively warmer temperatures. Individual plots of each of the 79 Connecticut plants reveal that twenty plants saw pronounced spikes in their BOD output in the summer of 2003, that is, for the time period June-September, 2003.

The year 2003 was a particularly wet year on the east coast and some plants faced challenges with managing large inflows during this time. Large inflows to a WWTP can hamper the plant's ability to provide adequate treatment. In some cases, it is possible that excess inflow led the WWTP to release partially treated or untreated wastewater to the receiving water, causing unusually high BOD readings. In such situations, the plant's excess capacity, defined as the positive difference between the plant's maximum capacity to treat a volume of water and the volume of water it treats, will be near zero or will be negative, indicating that the plant is overloaded.

Therefore, changes in the excess capacity at these 20 plants can be used to indirectly test for large inflows to the plants since large inflows will reduce the plants' excess capacity. It is necessary to define a variable that measures the excess capacity at the plant. This variable can be defined as the ratio of two variables: 1) the difference between the plant's total capacity and the current monthly capacity usage and; 2) the plant's total capacity. This variable will be called *xcap*.

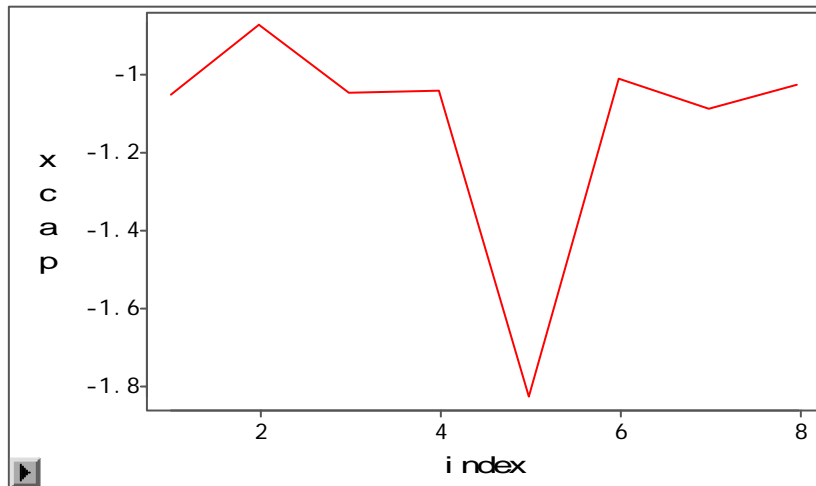
Graphs of  $xcap$  over each of the months in summer 2002 and 2003 for each of the twenty plants reveals pronounced dips during the middle to latter part of 2003 for this group of twenty plants. Figure 14 presents a typical graph for one of these twenty plants.



**Figure 14: Typical graph of excess capacity for the selected twenty plants in Connecticut**

Seven plants out of the 20 plants had at least one value of  $xcap$  in the negative range, indicating that the average monthly flow of wastewater to the plant exceeded its design capacity for that particular month. One plant, shown in Figure 15, had all values of  $xcap$  in the negative range but one:





**Figure 15: Graph of excess capacity for a selected plant in Connecticut**

Next, means tests for changes in the excess capacity, *xcap*, for these twenty plants between the summer 2002 month and the summer 2003 months are conducted. The test reveals that the plants' excess capacity in summer 2003 was significantly lower than in summer 2002.

Two Sample t-test for the Means of *xcap* within year

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
2002	79	0.327426	0.3529	0.0397
2003	80	0.194019	0.4088	0.0457

Hypothesis Test

Null hypothesis:	Mean 1 - Mean 2 = 0		
Alternative:	Mean 1 - Mean 2 $\neq$ 0		
If Variances Are	t statistic	Df	Pr > t
Equal	2.202	157	0.0292
Not Equal	2.204	154.25	0.0290

These results suggest that a few plants may have been overloaded in the summer of 2003 and that their performance regarding BOD output was adversely affected as a result. An interview with an official from the Connecticut Department of Environmental Protection revealed that plants will release high values of BOD during “severe washouts” and confirmed that rainfall amounts in the summer of 2003 in Connecticut were 20-25%

above normal (personal communication with G. Johnson, Connecticut Department of Environmental Protection). These events illustrate how BOD output can vary, sometimes widely, regardless of the operator decisions at the treatment plant.

An interview with a Connecticut state official about the operating conditions at WWTPs that cause concern about permit violations revealed that whatever increases the variation of BOD is of concern to WWTPs (personal communication with G. Johnson, Connecticut Department of Environmental Protection). According to the official, a wastewater treatment plant's worst nightmare in terms of remaining in compliance with its NPDES permit conditions is "a big snowmelt in March" (personal communication G. Johnson, CTDEP). This is because a higher flow impedes the plant's ability to provide treatment since the wastewater has to spend a certain amount of time in the settling tank in order to grow enough microorganisms. Furthermore, colder temperatures also impede the plants' ability to provide treatment since the microorganisms grow faster in warmer temperatures and the water is still very cold in Connecticut in March (personal communication G. Johnson, CTDEP).

Plots over time of the standard deviation of BOD estimated across plants for each month can provide information about the months in which WWTPs may be most concerned about violating their BOD permit limit. These plots are presented in Figures 16 and 17.

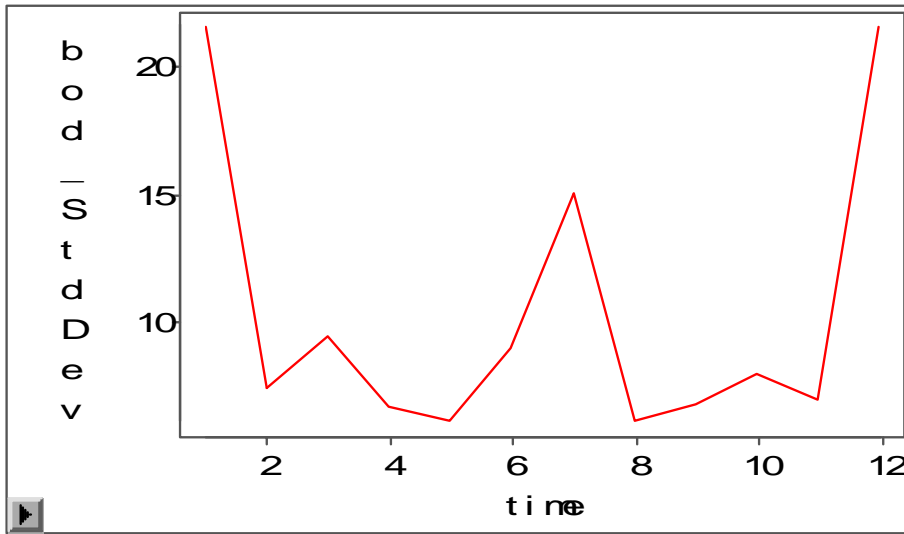


Figure 16: Plots of standard deviation (across plants) of monthly BOD output over the year in Connecticut

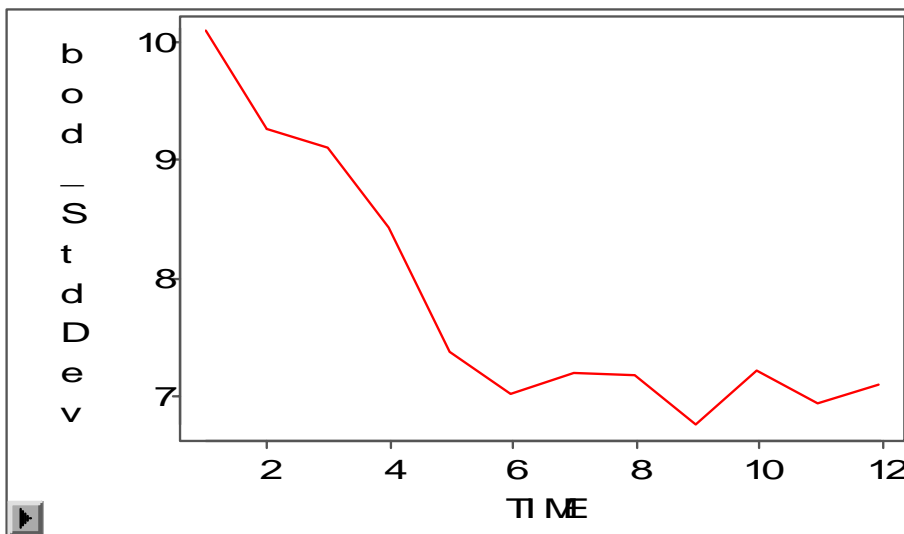


Figure 17: Plots of standard deviation (across plants) of monthly BOD output over the year in Maryland

According to Figures 16 and 17, BOD output is most variable in the winter. The high variability in the standard deviation of BOD over the year is especially evident in the plot of Connecticut plants. The graphs indicate that WWTPs, *ceteris paribus*, will be more concerned about violating their permit limits in the winter than in the summer.<sup>42</sup> This conclusion from the graphs corroborates the state official's remark about violations of permit limits: that BOD control is more difficult in colder temperatures and, in addition to the implications for noncompliance, a result of this relative lack of control is that operators will have even less flexibility to react to unplanned events like hydraulic overloads (i.e. "a big snowmelt in March").

To summarize, there is considerable variation over time and over plants in BOD output. This is not surprising since BOD output is a function of the temperature of the incoming wastewater, the amount of time the wastewater spends in the settling tanks, the food-to-microorganism ratio in the wastewater, and the degree of flocculation that occurs (Hammer and Hammer, Jr., 1996; Liu and Liptak, 1999; Davis and Cornwell, 1991; interviews with officials from Connecticut Department of Environmental Protection and Maryland Department of the Environment).

There is also variation over plants in the pattern of BOD outputs over the year. In a cross-section of plants, this leads to variation in the plant-level estimates of the variance of BOD output over time. Differences across plants in the variance of their BOD output levels over time may explain differences across plants in the degree of overcompliance.

### ***Jointness between Nitrogen Control and BOD Control***

Another potential source of observed BOD overcompliance is the WWTP's efforts to limit nitrogen output. If the significant degree of overcompliance that is observed in Connecticut and Maryland is driven by the presence of the operation of BNR technologies, then the overcompliance with BOD is incidental. The existence of jointness can be assessed descriptively using the data. Both datasets include plants that control nitrogen discharges and to this end have installed Biological Nitrogen Reduction

---

<sup>42</sup> Bandyopadhyay and Horowitz (2004) come to a similar conclusion based on their calculation of BOD overcompliance ratios using their data for each of the twelve months: the "bad season [for compliance] occurs in the winter and early spring" (p. 18).

technology. In Connecticut, 30 plants (38%) have BNR technology installed. In Maryland, 38 plants (29%) have BNR technology installed.

The Connecticut plants can be divided into three groups depending on whether they have not installed any BNR technology, whether they have retrofitted the plant for BNR, or whether they have completely upgraded the plant to perform BNR. The Maryland plants can be differentiated depending on whether they have some form of BNR or not. Plant performance regarding nitrogen control (and by extension BOD control) for plants with BNR that accepted public funds will be driven by any nitrogen targets that these plants have to meet as part of the cost-share agreement. Table 12 presents summary statistics on overall N output as well as statistics on BOD output and nitrogen output averaged over time and then over plants within the BNR groupings.

**Table 12: BOD and Nitrogen output summarized within BNR groupings**

	Obs	Effluent	Target (mg/l)	Mean (mg/l)	Std Dev	Minimum	Maximum
<b>Connecticut</b>							
<b>Total</b>	79	N	---	11.57	5.73	2.36	26.08
<b>No BNR</b>	49	BOD	---	9.32	5.86	1	28
		N	---	13.93	5.47	2.98	26.08
<b>BNR retrofit</b>	18	BOD	---	6.67	3.08	2.05	14.6
		N	3-10	8.95	3.53	4.08	16.89
<b>BNR upgrade</b>	12	BOD	---	6.01	5.32	2	21
		N	3-6	6.08	3.43	2.36	15.09
<b>Maryland</b>							
<b>Total</b>	133	N	---	14.25	8.15	0.78	60.59
<b>No BNR</b>	95	BOD	---	7.65	5.74	1	28
		N	---	17.01	7.66	0.78	60.59
<b>BNR technology</b>	38	BOD	---	5.42	7.04	1	43.5
		N	8	7.36	4.43	2.18	21.84

Not surprisingly, the aggregate nitrogen concentration in the effluent falls when plants install BNR technology. In Connecticut, plants with no BNR discharge about 14 mg/l, plants retrofitted for BNR discharge about 9 mg/l and plants that have been completely upgraded to BNR discharge 6 mg/l. In Maryland, the same trend is observed where plants with no BNR discharge approximately about 17mg/l while plants with BNR installed discharge about 7 mg/l.

Furthermore, BOD output is lower for those WWTPs with installed BNR technology. In Connecticut, plants with no BNR discharge 9 mg/l of BOD, plants retrofitted for BNR discharge about 6.5 mg/l of BOD and plants that have been completely upgraded to BNR discharge 6 mg/l of BOD. In Maryland, the same trend towards lower BOD output is observed where plants with no BNR discharge approximately 7.5mg/l of BOD while plants with BNR installed discharge about 5.4 mg/l of BOD. Thus, there appears to be some jointness in nitrogen and BOD control where the additional control of nitrogen also reduces BOD output.

## ***Empirical Model***

In order to formally test the effect of the various possible determinants of overcompliance, a statistical model is specified. The goal is to construct a statistically adequate model that controls for the varied determinants of BOD overcompliance in order to best isolate the impacts of jointness and discharge randomness on BOD overcompliance. The degree of overcompliance is defined as the ratio of observed BOD output to permitted BOD output. The degree of overcompliance can be estimated directly. The degree of overcompliance can also be estimated by first estimating a model of BOD output and then calculating the overcompliance ratio using the predicted BOD output level.

This latter method is preferred to directly modeling the overcompliance ratio because the hypotheses concerning the effects of jointness in pollution control and variation in BOD relate directly to the average BOD output and relate indirectly to the average BOD overcompliance ratio through their effect on the average BOD output. As a result, estimating the average BOD output as a function of nitrogen control and the variation in BOD output provides a more direct test of these hypotheses than estimating the average BOD overcompliance ratio as a function of nitrogen control and the variation in BOD.

The dependent variable will be the WWTP's BOD output level. Several factors appear *a priori* to influence the observed level of BOD output (as well as the degree of overcompliance at the plant) and each will be discussed in turn. The permit limit stipulated in the plant's NPDES permit may influence the observed BOD discharges. If plants seek to avoid penalties for noncompliance, the permit limit will be important to decisions about the level of BOD discharges, based on the conceptual model presented in Section 2.

In terms of the conceptual model presented in Section 2, if  $\bar{Q}$  in Figure 9 shifts to the left, the MEP curve will shift up, reflecting the plant's expectations that with a lower permit limit, permit exceedances and penalties will rise. The plant facing a lower permit limit will control more BOD because otherwise, its expected penalty exceeds its abatement costs savings on the margin. For example, a plant facing a permit limit of 20mg/l may differ in incentives (i.e. higher probability of exceedance leading to higher

expected penalties) for BOD control, technology, and/or capital inputs than a plant facing a permit limit of 30mg/l. Therefore, differences in BOD permit limits across plants may explain variation in BOD discharges across plants.

Differences in the plants' design flow may explain differences in observed BOD output in that, as a measure of plant size, design flow is a proxy for plant efficiency (i.e. marginal productivity of BOD control), which latter is likely to be positively correlated with plant size. Given two plants with otherwise equal characteristics, the more efficient plant (in terms of producing BOD control) will have lower BOD effluent levels. Furthermore, if larger plants are more efficient than smaller plants due to economies of scale, then plant size, as measured by the design flow, will be a factor in BOD output.

Difference in the excess flow capacity related to the flow volume and to the plant's design capacity may explain differences in observed BOD output across plants. The excess capacity is the excess of the plant's design flow over its actual flow. It is clear that when the actual flow exceeds the plant's design flow, the plant will be unable to provide complete secondary treatment and its BOD effluent levels may well exceed its permit limit. A similar situation has been described for a group of twenty plants in Connecticut. Likewise, it seems plausible that a plant that operates at fifty percent capacity will have a lower marginal control cost than a plant operating at ninety-five percent capacity.

In order to provide adequate secondary treatment, the wastewater must receive adequate aeration, must spend a minimum amount of time being recycled through the aeration tank and must spend a prescribed amount of time in the settling tank (Gray, 1999). A plant operating at ninety-five percent capacity will find it difficult to adequately aerate the wastewater in order to encourage flocculation. In order to achieve the *same* level of BOD control as a plant of similar size that is operating at 50 percent capacity, a plant that is operating at ninety-five percent capacity would have to intensify certain processes (i.e. increase the amount of time that the wastewater spends in the aeration and/or settling tank) and spend more energy resources recycling the wastewater and pumping away the sludge, all at added cost. Treatment of each additional unit of wastewater may be more costly to the plant in terms of energy (Liu and Liptak, 2000), labor, and chemical costs.



Conversely, a plant operating at 50 percent capacity is dealing with a smaller volume of wastewater. The extra capacity in the settling and aeration tanks makes it easier to provide complete aeration of the wastewater and encourage adequate flocculation of the bacteria. The plant operating at 50 percent capacity can provide more complete and intensive treatment without a significant increase in treatment costs. It is thus more likely to overcomply than a plant that has to treat a larger volume of wastewater with the same fixed capacity. The expectation is that the excess capacity will be negatively related to BOD output.

Most treatment plants will not be controlling BOD alone. As seen in Figure 11 from the conceptual model presented in Section 2, a plant that controls nitrogen may have a lower BOD output than a plant that does not control nitrogen but in all other respects, is similar to the other plant. The expectation is that the more nitrogen a plant controls, the lower will be its BOD output.

As seen from the conceptual model, variation in BOD output may help explain BOD overcompliance because plants seeking to avoid noncompliance penalties may adjust for high variation in BOD output by adjusting their average BOD output downward. This relationship has been confirmed empirically using effluent data; however, the reasons for the inverse relationship have not been conclusive.

The variance may not be the only distributional moment that is relevant in explaining overcompliance. The common assumption in the literature is that BOD follows a lognormal distribution and not a symmetric distribution (Bandyopadhyay and Horowitz, 2002; McClelland and Horowitz, 1999; Brannlund and Lofgren, 1996). The more skewed is the distribution of the plant's BOD effluent the more likely we are to observe lower concentrations of BOD in the plant's effluent, all other factors being equal. The plant will have a lower probability of exceedance since it has a lower mean output level compared to another plant with a less skewed distribution. The expectation is that the more skewed the distribution, the lower will be the absolute level of BOD control and the level of BOD relative to its permit limit.

The following table presents the variable naming conventions in order to facilitate the discussion of the empirical model and the presentation of the empirical results. The variables that enter into the regression are calculated by collapsing the panel over time to

form a cross-sectional dataset. In other words, the mean value over the time dimension, the median value over the time dimension, the variance over the time dimension, and the skewness over the time dimension are calculated by collapsing the panel over time. The prefixes on the variables, depending on the moment are listed in Table 13.

**Table 13: Prefixes and their meaning**

<b>Prefix</b>	<b>Meaning</b>
m	Mean
d	Median
v	Variance
s	Standard deviation
k	Skewness

Given the short time period ( $T=24$ ) over which these values are calculated, the median measure rather than the mean is used as a measure of location, because the mean is less precise than the median when the sample size is small (Greene, 1997). Thus, the level variables used in the statistical models will be prefixed by *d*. Table 14 presents definitions of variables used in the empirical analysis.

**Table 14: Variable Definitions used in Empirical Model**

<b>Variable</b>	<b>Definition</b>	<b>Units</b>
<b>bma</b>	The measured concentration of BOD in the wastewater effluent	Mg/l
<b>bla</b>	The specified upper limit on monthly average <i>bma</i> in the NPDES permit	Mg/l
<b>tn</b>	The measured concentration of N in the wastewater effluent	Mg/l
<b>mflow</b>	The average monthly flow volume of water through the wastewater treatment plant	Million gallons per day (mgd)
<b>dflow</b>	The wastewater treatment plant's design capacity	Million gallons per day (mgd)
<b>xcap</b>	The monthly average excess capacity at the plant defined as the ratio of the excess of <i>dflow</i> over <i>mflow</i> and <i>dflow</i> or $((dflow-mflow)/dflow)$	(no units)
<b>oc</b>	Overcompliance ratio equal to the ratio of <i>bma</i> to <i>bla</i>	(no units)
<b>dbma</b>	Median value over time of the natural log of the plant's average monthly effluent concentration of Biochemical Oxygen Demand = $median(\log(bma))$	Mg/l
<b>dbla</b>	Median value over time of the plant's monthly upper limit on the concentration of Biochemical Oxygen Demand in the effluent = $median(bla)$	Mg/l
<b>dtn</b>	Median value over time of the natural log of the average monthly concentration of total nitrogen in the effluent = $median(\log(tn))$	Mg/l
<b>dmflow</b>	Median value over time of the natural log of average monthly flow through the plant = $median(\log(mflow))$	Million gallons per day (mgd)
<b>dflow</b>	Median value over time of the plant's design (total) capacity. This value is unchanged over time.	Million gallons per day (mgd)
<b>dxcap</b>	Median value over time of the measure of excess capacity = $median(xcap)$ for each plant	(no units)
<b>vbma</b>	Variance over time of the natural log ( <i>bma</i> ) = $var(\log(bma))$	
<b>kbma</b>	Skewness over time of the natural log ( <i>bma</i> ) = $skew(\log(bma))$	
<b>ddt</b>	Dummy variable equal to 1 if the plant has either been retrofitted or overhauled to install BNR technology; equal to 0 otherwise	
<b>lower</b>	Dummy variable equal to 1 if plant has permit limits that are uniformly lower than the technology-based standard of 30mg/l; equal to 0 otherwise	
<b>seaslim</b>	Dummy variable equal to 1 if plant has seasonal limits;	

equal to 0 otherwise
----------------------

## **Model Specification**

The empirical model takes the following form:

$$BOD_t = \beta_0 + \beta_1 Nitrogen_t + \beta_2 BOD\ limit_t + \beta_3 FLOW_t + \beta_4 Excess\ capacity_t + \beta_5 Variance\ BOD_t + \beta_6 Skewness\ BOD_t + u_t, \quad t = 1 \dots T \quad (5)$$

In terms of the variables defined in Table 14, (5) can be rewritten as:

$$dbma_t = \beta_0 + \beta_1 dtn_t + \beta_2 dbla_t + \beta_3 dflow_t + \beta_4 dxcap_t + \beta_5 vbma_t + \beta_6 kbma_t + u_t, \quad t=1, \dots, T \quad (6)$$

There are probabilistic, statistical, and sampling assumptions implicit in (6) that must be satisfied in order for the model's estimated parameters to be reliable.<sup>43</sup> These assumptions are laid out in Table 15.

**Table 15: The Normal Linear Regression model**

<i>Statistical Generating Mechanism</i>	$y_t = \beta'x_t + \varepsilon_t, \beta \equiv (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6),$ $y \equiv dbma, x \equiv (1, dbla, dtn, dflow, dxcap, vbma, kbma),$ $t \in \mathfrak{T}$
<i>Distribution</i>	$f(y_t   X_t; \phi)$ , is Normal
<i>Mean</i>	$E(y_t   X_t) = \beta'x_t$ , the systematic component of the regression function, is linear in $x_t$
<i>Variance</i>	$Var(y_t   X_t = x_t) = \sigma_y^2$ is the conditional variance of $y$ and is free of $x_t$
<i>Homogeneity</i>	$\phi := (\beta, \sigma_y^2)$ are t-homogeneous $\forall t \in \mathfrak{T}$
<i>Dependence</i>	$\{(\varepsilon_t), t \in \mathfrak{T}\}$ is a white noise process

<sup>43</sup> The specification of the Normal linear regression model follows the discussion in Spanos (1986), Chapter 19.

<i>Sampling</i>	{ $Y_t$ } is an independent sample, sequentially drawn from $f(y_t   X_t ; \phi), t=1, 2, \dots, T$ .
-----------------	---

The assumption of a constant variance is particularly important when dealing with cross-sectional data, as the estimated variance can vary over the individual observations or over groups of individual observations. The distributional assumption must also be tested since it may not necessarily hold for small samples, and since it is important for the validity of the specification tests that the Normality assumption be valid for the data. If the distributional assumption and the homoskedasticity assumption do not hold, then the model is misspecified and the specification tests will not be valid. Misspecification tests will be performed on (6) before conducting specification (i.e. significance) tests.

## ***Estimation Results***

### ***Connecticut***

The data from Connecticut include all the wastewater treatment plants in the statewide nitrogen general permit and constitute a population of treatment plants. Table 16 presents summary statistics of the average performance over time and over plants in terms of BOD concentrations, N concentrations, degree of overcompliance, monthly flow, design flow, excess capacity, and BOD limit. The usefulness of these statistics depends on the validity of the assumptions in Table 15. If these assumptions are not correct, then the moment estimates are not reliable.

**Table 16: Summary Statistics for Connecticut data**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Obs.</b>	<b>Min</b>	<b>Max</b>	<b>Median</b>	<b>Skew.</b>	<b>Kurt.</b>
<b>bma</b>	9.48	12.72	1414	0	195	6.62	8.46	101.52
<b>bla</b>	27.53	4.9	1573	10	30	30	-2.13	4.03
<b>tn</b>	12.09	6.63	1569	0.9	42.68	10.74	0.72	0.15
<b>mflow</b>	4.87	7.9	1569	0.08	70.27	2.14	4.04	21.18
<b>dflow</b>	6.78	9.18	1573	.26	60	3.5	3.39	14.47
<b>xcap</b>	0.34	0.28	1569	-1.58	0.85	0.38	-2.23	9.74
<b>oc</b>	0.33	0.44	1493	0	6.8	0.23	8.33	101.06

In the literature, BOD effluent levels are assumed to follow a lognormal distribution (McClelland and Horowitz, 1999; Brannlund and Lofgren, 1996). Furthermore, histograms of BOD and other effluent variables over time showed that the distribution is skewed. Taking logs of the variables and then constructing a histogram resulted in a more symmetric distribution. Therefore, the variables that measure effluent concentrations, such as *bma* (*BOD output*) and *tn* (*nitrogen output*), undergo a log transformation before the estimation of equation (5). The variance and skewness measures of BOD are calculated from the logged value of BOD. Additionally, following Bandyopadhyay and Horowitz (2004), the BOD permit limit, *dbla*, undergoes a log transformation before estimating the models.

The model includes a dummy variable, *ddt*, that is equal to 1 if the plant has either been retrofitted or overhauled to install BNR technology.<sup>44</sup> Table 17 presents the estimation results and misspecification test results.

<sup>44</sup> Originally, the regression model included a separate dummy variable for plants that had been retrofitted and a dummy variable for plants that had been completely overhauled for BNR technology. However, the coefficient estimates were not significant. Additionally, due to the small sample size, the preference is for a parsimonious model. As a result, the model retains the present form.

Table 17: Estimation and Misspecification Test Results for Connecticut data

<b>Dependent variable: Observed BOD concentration in mg/l</b>		
<b>Variable</b>	<b>Estimate</b>	<b>Std. Error</b>
<b>Constant</b>	-1.13	1.02
<b>df1</b>	0.007	0.008
<b>dtn</b>	0.48 <sup>**</sup>	0.15
<b>vbma</b>	-0.09	0.29
<b>kbma</b>	-0.07	0.07
<b>dxcap</b>	0.06	0.28
<b>dbla</b>	0.57 <sup>*</sup>	0.3
<b>ddt</b>	0.08	0.17
<b>Number of observations</b>	79	
<b>R<sup>2</sup></b>	0.27	
<b>Adj. R<sup>2</sup></b>	0.2	
<b>Misspecification Tests</b>		
<b>Assumption</b>	<b>Test</b>	<b>P-value</b>
<b>Normality</b>	Bera-Jarque	0.57
<b>Normality</b>	D'Agostino Pearson	0.81
<b>Skewness=0</b>	D'Agostino Pearson Skewness	0.58
<b>Kurtosis=3</b>	D'Agostino Pearson Kurtosis	0.73
<b>Linearity(2)</b>	RESET	0.44
<b>Homoskedasticity</b>	White's	0.46
<b>Homoskedasticity(2)</b>	RESET	0.56

\* Parameter estimates significant at the 5 percent level

\*\* Parameter estimates significant at the 1 percent level.

Based on the results in Table 17, the model appears to be correctly specified. There is little evidence against Normality, homoskedasticity, and linearity. Residual plots indicate that the residuals appear to be random, exhibiting no systematic patterns.<sup>45</sup> Since the model appears to be correctly specified, specification tests on *dtn* and *vbma* can be conducted.

The coefficient on *dtn* is statistically significant at the 1 percent level providing compelling evidence of jointness in production between nitrogen (N) control and BOD control. Since the data is cross-sectional, this result means that plants that provide for more N control also provide for more BOD control. Since both the dependent variable

<sup>45</sup> Residual plots can be found in the appendix.

and  $dm$  are in logarithmic form, the coefficient on  $dm$  can be interpreted as an elasticity. The 'BOD elasticity of nitrogen control' is 0.48, meaning that a 1 percent reduction in nitrogen effluent discharge results in a 0.48 percent reduction in BOD effluent discharge. While the response is inelastic, the interesting finding is that there is a response at all since the intent is to determine whether there exists jointness in pollution control.

The coefficient on the variable  $vbma$  has the expected negative sign. The negative sign means that the higher the plant's variation in BOD, the lower is the plant's median BOD. However, while plants with higher variance of BOD emit less BOD than plants with a lower variance of BOD, there is no statistical evidence that higher variance is correlated with the observed BOD discharge levels, all else equal.

This result is unexpected given that the qualitative and quantitative evidence suggest a strong negative correlation between the variation in BOD and the average BOD output. The empirical evidence (i.e. Brannlund and Lofgren (1991); Bandyopadhyay and Horowitz (2004)) that found a (statistically significant) negative correlation between the variance of BOD and the mean BOD output did not simultaneously control for nitrogen discharges, as done here. It may be that the two effects on BOD have to be modeled separately.

The qualitative evidence may provide some insight into the lack of statistical significance on  $vbma$  and the need to model the two effects separately. The qualitative evidence suggests that there may be two stages to the decision making process in what concerns capital installation decisions that impact BOD output levels: a first stage decision in which allowance is made for the effect of variation in BOD on mean BOD output and a second stage decision to upgrade to nitrogen control technology.

If a two-stage process adequately describes the decision making process at most Connecticut plants, then the empirical model above which models the process at a given point in time, rather than in stages, may be unable to detect the independent impact that the adjustment for variance has on the mean BOD output levels. A model that includes a time dimension may resolve this issue by separating out first-stage relationships from second-stage relationships. The time dimension is not present in the current model because the time dimension was collapsed in order to derive a measure of variation in



BOD for use in the analysis of the relationship between measures of dispersion and of location.

In terms of other coefficient estimates, the coefficient on the BOD permit limit is statistically significant at the 5 percent level providing evidence that variation in BOD permit limits across plants explains some of the variation in BOD discharge levels across plants in Connecticut. The coefficient is positive indicating that plants with lower BOD permit limits also have lower BOD discharge levels. The response is inelastic: a 1 percent reduction BOD permit limits translates into a 0.6 percent reduction in BOD discharge levels. The response may be inelastic but the interesting fact is that there is a response at all.

The result is consistent with the conclusions from the conceptual model in that plants with lower permit limits have lower BOD output. However, since the coefficient estimate is an estimate of correlations, the direction of causality is not clear: whether reductions in permit limits induce a response from the plant to lower BOD output or whether regulators target plants with already low BOD output for lower BOD permit limits.

Furthermore, if there are stages in the decision making process, as implied by the qualitative evidence, it is not clear whether this correlation holds only in the first-stage, only in the second-stage, or during both stages of the process. If the result applies to the first stage, then it may be the case that the lower BOD output results from a lower BOD permit limit. In other words, if regulators, based on water quality modeling results, determine that a treatment plant should be permitted at a level that is below the technology-based standard, the treatment plant will be designed to comply with the lower permit limit and will control less BOD compared to a plant with a permit limit of 30mg/l.

On the other hand, if the permit limit is positively correlated with the BOD output level during the plant's operational phase, the question then arises whether the lower output level is the result of a reduction in the permit limit and if so, to what extent the plant operator has the flexibility to respond to reductions in its permit limit. The empirical model answers questions related to the existence of the relationship between the permit limit and the output level but raises other questions related to causality that cannot be answered without additional information.

In addition to specification tests, the regression equation can be used to calculate the estimated average BOD discharge. Since the dependent variable is in logarithmic form, the regression equation must be transformed in order to calculate the average discharge level on a raw scale rather than on the logarithmic scale (Manning and Mullahy, 2001). When the dependent variable is in logarithmic form, the formula for transforming the regression equation back into the raw scale value of the dependent variable is:

$$E[y | x] = \exp(\beta'x + 0.5\sigma_u^2) \quad (7)$$

Equation (7) states that the expectation of the conditional distribution of the dependent variable is equal to the exponential function evaluated at the sum of the systematic component of the regression function and one-half the variance of the regression. It applies as long as the conditional distribution is log normal (Manning and Mullahy, 2001). In practice, the mean of the dependent variable can be calculated by using the estimated parameter coefficients and the estimated variance of the regression in the place of the statistical parameters in (7).

Based on (7), the estimated mean BOD discharge level using the Connecticut data is 7.67 at the mean of the explanatory variables.<sup>46</sup> The model states that, on average, plants in Connecticut are discharging 7.67 mg/l of BOD. The predicted value of the dependent variable is quite sensitive to variation in the level of nitrogen control and somewhat sensitive to variation in the BOD permit limit. Table 18 displays the predicted mean BOD discharge level at various values of nitrogen discharge and the BOD permit limit, holding all other variables at their mean. The nitrogen discharge levels reflect the average discharge of plants without BNR technology (14mg/l), with a retrofit (9mg/l), and with a complete upgrade to BNR technology (6mg/l). The BOD values reflect the most common BOD permit limit (30mg/l) and the second most common permit limit (20mg/l) among Connecticut plants.

---

<sup>46</sup> This value is close to the mean of 6.89 that was calculated using the raw sample data.

**Table 18: Sensitivity analysis on mean BOD discharge using Connecticut data**

<b>Parameter</b>	<b>Value</b>	<b>Estimated average BOD discharge (mg/l)</b>
<b>Nitrogen output</b>	14mg/l	8.98
<b>Nitrogen output</b>	9mg/l	7.27
<b>Nitrogen output</b>	6mg/l	5.98
<b>BOD limit</b>	30mg/l	8.17
<b>BOD limit</b>	20mg/l	6.48

Based on Table 18, the model predicts that plants that have been completely overhauled for BNR technology will emit BOD concentrations that are 3mg/l less than the level discharged by plants without any BNR technology. This difference in BOD output is remarkable given that the only difference between the plants is in installation of technology to control another pollutant, nitrogen. The result provides further evidence of jointness between nitrogen and BOD.

In terms of sensitivity analysis on other parameters, the data indicates that BOD concentrations in effluent discharge will vary somewhat with variation in the permit limit. Three-quarters of plants in Connecticut face the technology-based permit limit of 30mg/l. After the standard of 30mg/l, the next most common value for the permit limit is 20mg/l. The model predicts that the effluent discharge from the plants facing a limit of 20mg/l will contain smaller concentrations of BOD than the effluent from plants facing the 30mg/l permit limit.

It is interesting to note, through all the permutations in Table 18, that the predicted BOD levels remain low relative to the BOD permit limit. In other words, the degree of overcompliance remains large. Ultimately, the regression model can be used to analyze the degree of overcompliance amongst plants in Connecticut. The predicted mean BOD discharge is 7.67mg/l and the average permitted level of BOD is 27.53 mg/l, therefore the predicted overcompliance ratio is 0.28.<sup>47</sup> In other words, the model states that, on average, plants in Connecticut are discharging 28 percent of their permitted concentration of BOD. The predicted overcompliance ratio is quite sensitive to variation in the level of nitrogen control and somewhat sensitive to variation in the permit limit.

---

<sup>47</sup> This value is close to the value of 0.25 that was calculated using the raw data.

Table 19 displays the mean overcompliance ratio at the values of average nitrogen discharge and the BOD permit limit specified in Table 18, holding all other variables at their mean.

**Table 19: Sensitivity analysis on mean overcompliance ratio using Connecticut data**

<b>Parameter</b>	<b>Value</b>	<b>Estimated average BOD overcompliance ratio</b>
<b>Nitrogen output</b>	14mg/l	0.33
<b>Nitrogen output</b>	9mg/l	0.26
<b>Nitrogen output</b>	6mg/l	0.22
<b>BOD limit</b>	30mg/l	0.3
<b>BOD limit</b>	20mg/l	0.24

As seen from Table 19, changes in nitrogen control have a double-digit effect on plant discharges as a fraction of the permitted level. The effect on the BOD overcompliance rate of varying only nitrogen levels between the average ‘no-BNR’ level and the ‘complete BNR’ level is that the average plant BOD discharge falls from 33 percent of the permitted level to 22 percent of the permitted level. Moving from the technology-based standard to the second most common permit limit, the average plant BOD discharge falls from 30 percent of the permitted level to 24 percent of the permitted level.

Thus, the regression model is useful in that it provides information on the technical relationships as well as yields an estimate of the overcompliance ratio. Most of the explanatory variables are not significant in the regression. This may mean that these variables do not play a role in explaining variation in BOD output. However, it is also possible that at 79 observations, there are simply too few data points available to isolate the individual influences on observed BOD discharges.

When using cross-sectional datasets, regressions are better able to properly isolate systematic effects from the noise in the data when the data contains 100 observations or more because the normality assumption becomes more acceptable as the sample size approaches and surpasses 100, as a general rule. Unfortunately, at 79 plants, the CT dataset constitutes a population for WWTPs so it is not possible to increase the sample size and test whether increasing the number of observations will change the model

results. The dataset from Maryland includes over 100 observations; it may yield better results in terms of achieving statistical significance.

### ***Maryland***

Table 20 presents summary statistics for the Maryland data. These summary statistics are over plants and over time. The usefulness of these statistics depends on the validity of the assumptions in Table 15. If these assumptions are not correct, then the moment estimates are not reliable.

**Table 20: Summary Statistics for Maryland data**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Obs.</b>	<b>Min</b>	<b>Max</b>	<b>Median</b>	<b>Skew.</b>	<b>Kurt.</b>
<b>bma</b>	7.88	7.98	3125	0.2	81.6	5	2.73	11.69
<b>bla</b>	26.92	6.84	3157	5	45	30	-1.73	2.09
<b>tn</b>	14.58	9.28	3186	0.47	106.11	16.49	3.7	34.15
<b>mflow</b>	2.56	9.82	3192	0	122	0.15	7.53	63.97
<b>dflow</b>	4.11	17.29	3099	.002	180	0.28	8.61	81.67
<b>xcap</b>	0.43	0.34	3099	-9.5	1	0.45	-8.29	223.77
<b>oc</b>	0.3	0.29	3125	.007	3.36	0.2	2.89	14.32

Before estimating the model, the variables measuring effluent concentrations undergo a log transformation. Additionally, the natural log of BOD is calculated before estimating the variance and skewness. Additionally, following Bandyopadhyay and Horowitz (2004), the BOD permit limit undergoes a log transformation before estimating the model. Table 21 presents the results.

Table 21: Estimation and Misspecification Test Results for Maryland data, Part I

<b>Dependent variable: Observed BOD concentration in mg/l</b>		
<b>Variable</b>	<b>Estimate</b>	<b>Std. Error</b>
<b>Constant</b>	-1.1	0.61
<b>dtm</b>	0.24**	0.08
<b>vbma</b>	-0.46	0.25
<b>kbma</b>	-0.22**	0.06
<b>dxcap</b>	-0.53*	0.24
<b>dbla</b>	0.78**	0.19
<b>dflow</b>	-0.00004	0.003
<b>Number of observations</b>	133	
<b>R<sup>2</sup></b>	0.31	
<b>Adj. R<sup>2</sup></b>	0.28	
<b>Misspecification Tests</b>		
<b>Assumption</b>	<b>Test</b>	<b>P-value</b>
<b>Normality</b>	Bera-Jarque	0.94
<b>Normality</b>	D'Agostino Pearson	0.96
<b>Skewness=0</b>	D'Agostino Pearson Skewness	0.78
<b>Kurtosis=3</b>	D'Agostino Pearson Kurtosis	0.97
<b>Linearity(2)</b>	RESET	0.29
<b>Homoskedasticity</b>	White's	0.31
<b>Homoskedasticity(2)</b>	RESET	0.42

\*Parameter estimates significant at the 5 percent level.

\*\*Parameter estimates significant at the 1 percent level.

Based on the results in Table 21, the misspecification tests indicate little evidence against Normality, linearity (2) and homoskedasticity (2). From the residual plots, the residuals appear to randomly distributed, with no systematic patterns.<sup>48</sup> The model appears to be correctly specified so specification tests can be considered.

The parameter estimate on *dtm* is statistically significant at the 1 percent level providing evidence of a technical relationship in pollution control between BOD and N. The parameter estimate on *dtm* is positive, indicating that a reduction in N effluent is correlated with a reduction in BOD effluent.

Since both the dependent variable and *dtm* are in logarithmic form, the coefficient on *dtm* can be interpreted as an elasticity. The 'BOD elasticity of nitrogen control' is 0.25

<sup>48</sup> Residual plots are located in the appendix.

meaning that a 1 percent reduction in nitrogen effluent discharge results in a 0.25 percent reduction in the BOD effluent discharge. As with the Connecticut plants, the response here is inelastic but statistically significant.

The coefficient on the variance term is negative as expected but is only statistically significant at the 10 percent level with a p-value of 0.07. Again, there is not compelling evidence of a relationship between the variance in BOD and the average BOD discharge levels. Most of the independent variables are statistically significant therefore the model is able to isolate individual effects. If the model cannot isolate the influence of variation in BOD on BOD output, either it is because there is no influence or it is because the model is not appropriate.

If it is true that the relationship between variation in BOD and BOD output must be modeled separately in time, then the current model will be unable to isolate separate effects because it has no time dimension. Collapsing the time dimension was necessary in order to be able to derive some measure of variation in BOD at the cross-sectional level. However, there may be a way to estimate a dynamic model of BOD output while including a measure of variation in BOD as an explanatory variable. The results from estimating such a model will indicate whether variation in a plant's BOD output plays a role in explaining its average level of BOD output and whether this role is independent of the role that jointness plays in explaining average BOD output.

Unlike the case with the regression using data from Connecticut, several variables in the model are statistically significant. The BOD permit limit is statistically significant at the 1 percent level, providing evidence that differences in BOD permit limits across plants in Maryland explain differences in BOD discharge levels across plants. The coefficient estimate of 0.78 is positive, indicating that plants with lower BOD permit limits discharge less BOD output than plants with higher BOD permit limits. The response is inelastic: a 1 percent reduction in the BOD permit limit translates into a 0.78 percent reduction in the BOD discharge level, on average.

The parameter estimates on *kbma* and *dxcap* have the expected sign and are statistically significant. The negative sign on the parameter estimate of *kbma* indicates that plants with a more skewed distribution of BOD effluent have lower BOD output levels. The parameter estimate on *dxcap* is negative and statistically significant at the 5



percent level. The negative sign on the parameter estimate is expected, indicating that plants with more excess capacity have lower BOD effluent levels.

Using the formula in (7), the predicted BOD discharge level can also be calculated for the Maryland data. Based on (7), the estimated BOD discharge using the Maryland data is 6.37 at the mean of the explanatory variables.<sup>49</sup> The model states that, on average, plants in Maryland are discharging wastewater containing 6.37mg/l of BOD. The mean of the dependent variable is quite sensitive to variation in the level of nitrogen control and is also sensitive to variation in the BOD permit limit. Table 22 displays the predicted mean BOD discharge level at various values of nitrogen discharge and the BOD permit limit, holding all other variables at their mean. The nitrogen discharge levels reflect the average discharge of plants without BNR technology (17mg/l) and plants with BNR technology (7mg/l). The BOD values reflect the most common BOD permit limit (30mg/l) and the approximated level of the second most common permit limit (20mg/l) among Maryland plants.

**Table 22: Sensitivity analysis on mean BOD discharge using Maryland data**

<b>Parameter</b>	<b>Value</b>	<b>Estimated average BOD discharge (mg/l)</b>
<b>Nitrogen output</b>	17mg/l	6.98
<b>Nitrogen output</b>	7mg/l	5.65
<b>BOD limit</b>	30mg/l	7.1
<b>BOD limit</b>	20mg/l	5.2

Based on Table 22, the model predicts that plants with BNR technology will emit BOD concentrations that are 1.5mg/l less than the level discharged by plants without any BNR technology. This difference in BOD output is not as large as the difference obtained based on the Connecticut data even though the range of nitrogen output is larger: 17-7mg/l=10mg/l for Maryland and 14-7mg/l=7mg/l for Connecticut. One reason for the smaller range of the estimated BOD discharge in the Maryland data compared to the Connecticut data may be that the identified BNR plants in Maryland include plants that have been retrofitted for BNR, thereby dampening the effect on BOD of BNR technology. The data summarized in Table 12 indicate that plants that have been

<sup>49</sup> This value is close to the mean of 5.26 that was calculated using the raw data.

completely overhauled for BNR perform better in terms of nitrogen control and BOD control than plants that were retrofitted for BNR. The Connecticut data include a variable that identified plants that have been retrofitted for BNR separately from plants that have been completely overhauled for BNR.

There is another difference between Tables 22 and 18. Concerning the higher end of the range of BOD discharge levels, the results indicate that Maryland plants without BNR technology emit lower BOD discharge levels (6.98mg/l) on average than Connecticut plants without BNR technology (8.98mg/l). In any case, the model predictions about BOD discharge levels for plants in Maryland with and without BNR indicate that, as in Connecticut, the existence of nitrogen control has an impact on BOD control in Maryland.

In terms of sensitivity analysis on other parameters, the data indicates that BOD concentrations in effluent discharge will vary somewhat with variation in the permit limit. Sixty-two percent of the plants in Maryland face the technology-based permit limit of 30mg/l. After the standard of 30mg/l, most plants are in the range of 20mg/l as an annual average. The model predicts that the effluent discharge from the plants facing an average limit of 20mg/l will contain smaller concentrations of BOD than the effluent from plants facing the 30mg/l permit limit.

Ultimately, the model can be used to calculate the degree of overcompliance amongst plants in Maryland. The predicted mean discharge is about 6.37mg/l and the permitted level is about 26mg/l, therefore the predicted overcompliance ratio is 0.25.<sup>50</sup> In other words, the model states that, on average, plants in Connecticut are discharging 25 percent of their permitted concentration of BOD. The predicted overcompliance ratio is quite sensitive to variation in the level of nitrogen control and somewhat sensitive to variation in the BOD permit limit. Table 23 displays the mean overcompliance ratio at the values of average nitrogen discharge and the BOD permit limit specified in Table 22, holding all other variables at their mean.

---

<sup>50</sup> This value is close to the mean of 0.2 that was calculated using the cross-sectional data.

**Table 23: Sensitivity analysis on the mean overcompliance ratio using Maryland data**

<b>Parameter</b>	<b>Value</b>	<b>Estimated average BOD overcompliance ratio</b>
<b>Nitrogen output</b>	17mg/l	0.27
<b>Nitrogen output</b>	7mg/l	0.22
<b>BOD limit</b>	30mg/l	0.27
<b>BOD limit</b>	20mg/l	0.2

As seen from Table 23, changes in nitrogen control have an effect on plant discharges as a fraction of the permitted level. The effect on the overcompliance rate of varying only nitrogen levels between the average ‘no-BNR’ level and the BNR level is that the average plant BOD discharge falls from 27 percent of the permitted level to 22 percent of the permitted level. This effect of BNR on BOD overcompliance is not large; the effect may be dampened because the BNR group includes plants that have been retrofitted, not just plants that have been completely overhauled for BNR technology, as with the third category of plants in Connecticut.

Moving from the technology-based standard to the second most common permit limit, the average plant BOD discharge falls from 30 percent of the permitted level to 24 percent of the permitted level. Thus, a lower permit limit is correlated with a lower BOD discharge level. Since the model estimates correlations and cannot be used to determine causation, these results do not indicate whether the lower BOD discharge follows the lower BOD permit limit (implying some degree of operational flexibility to manipulate BOD levels) or vice versa. More qualitative analysis and a dynamic regression model are needed to gain some insight as to the direction of causality.

The reason why the second most common permit limit is approximated rather than exact is because many plants in Maryland with an annual limit below 30mg/l have seasonal limits rather than uniformly lower limits and therefore there will not be many plants at a given level below 30mg/l. Rather, for plants with seasonal limits, their annual level will be the average of two limits: the winter limit and the lower summer limit. Given the correlation that exists between BOD permit limits and BOD discharge levels, and the fact that the current measure masks variation in the plants’ permit structure, it might be useful to know whether it is plants facing uniformly lower limits or plants with seasonal limits that are driving the observed correlation.

To test this possibility, the model in (5) is estimated by removing *dbla* and adding interaction terms between each of the dummy variables *lower* and *seaslim* and *dbla*. The dummy variable *lower* is equal to one for plants with uniformly lower permit limits than the technology-based standard of 30mg/l and equal to zero otherwise. The dummy variable *seaslim* is equal to one for plants with seasonal limits (lower limits in the summer than in the winter) and equal to zero otherwise. Table 24 presents the estimation and misspecification test results.

**Table 24: Estimation and Misspecification Test Results for Maryland data, Part II**

Variable	Estimate	Std. Error
<b>Constant</b>	1.55**	0.25
<b>dtn</b>	0.28**	0.08
<b>vbma</b>	-0.44	0.26
<b>kbma</b>	-0.21**	0.07
<b>dxcap</b>	-0.63**	0.24
<b>lower*dbla</b>	-0.17*	0.09
<b>seaslim*dbla</b>	-0.13**	0.04
<b>dflow</b>	-0.001	0.004
<b>Number of observations</b>	133	
<b>R<sup>2</sup></b>	0.28	
<b>Adj. R<sup>2</sup></b>	0.24	
<b>Misspecification Tests</b>		
<b>Assumption</b>	<b>Test</b>	<b>P-value</b>
<b>Normality</b>	Bera-Jarque	0.94
<b>Normality</b>	D'Agostino Pearson	0.99
<b>Skewness=0</b>	D'Agostino Pearson Skewness	0.91
<b>Kurtosis=3</b>	D'Agostino Pearson Kurtosis	0.92
<b>Linearity(2)</b>	RESET	0.25
<b>Homoskedasticity</b>	White's	0.24
<b>Homoskedasticity(2)</b>	RESET	0.68

\*Parameter estimates significant at the 5 percent level.

\*\*Parameter estimates significant at the 1 percent level.

The misspecification test results do not indicate that there are any problems with the model therefore the specification test results can be discussed. The results indicate that seasonal limits are driving the correlation between BOD limits and BOD discharge. Based on the significance levels, plants with seasonal limits have lower BOD discharge

levels while there is no compelling evidence that plants with uniformly lower permit limits have lower BOD discharge levels.<sup>51</sup>

Similarly, plants with seasonal limits discharge BOD at a level that is 0.46 percent lower than plants with no seasonal limits. Judging from the results, the positive correlation between BOD limits and BOD discharge is likely driven by the positive correlation between seasonal limits and BOD discharge levels. If this is true, then qualitative evidence may shed some light on the direction of causality.

An official with the Maryland Department of the Environment indicated that seasonal limits are set for plants when water quality conditions dictate the need for higher DO levels and less BOD (personal communication, S. Luckman, Maryland Department of the Environment). If this is the case, then permit limits are not lowered in response to lower BOD discharge and the conclusion is that plants with lower permit limits as a result of seasonal limits reduce their BOD discharges in response to the lower limit. For these plants, the correlation between permit limits and BOD discharge levels implies that plants are responding to the permit limits by lowering BOD discharge levels.

### ***Summary of Empirical Results***

To summarize, the regression model predicts that wastewater treatment plants in Connecticut and Maryland will emit BOD output that is about a quarter of their BOD permit limit and that variation across plants in the degree to which plants control nitrogen, in the level of the permit limit, and in the level of excess capacity explain variation across plants in BOD discharge levels and BOD overcompliance.

The data do not provide evidence to suggest that observed variation in BOD plays a role in explaining observed BOD discharge levels or BOD overcompliance in Connecticut and Maryland.<sup>52</sup> This result contrasts with previous work in this area. Bandyopadhyay and Horowitz (2004) are the first to estimate the effect of variation in BOD output on the observed BOD overcompliance. The coefficient on the variance of

---

<sup>51</sup> The coefficient estimate on *lower\*lmbl* has a p-value of 0.0556. While there are several plants with seasonal limits, there are not many plants with uniformly lower limits in Maryland: 43 plants with seasonal limits vs. 6 plants with uniformly lower limits.

<sup>52</sup> The coefficient on the variance of BOD output is negative using both datasets and marginally significant only for the Maryland data.

BOD in their static regression model, which does not simultaneously control for nitrogen output, is negative and statistically significant at the 1 percent level (Bandyopadhyay and Horowitz, 2004).<sup>53</sup>

Bandyopadhyay and Horowitz (2004) are open to the role that multiple pollutant control may play in explaining BOD overcompliance, acknowledging that variation in BOD may not fully explain the observed degree of BOD overcompliance. They find that “even when there’s no uncertainty, plants will pollute at just 60% of their permitted levels” (Bandyopadhyay and Horowitz, 2004, p. 17). The authors suggest that community pressure may also cause plants to overcomply but note that “subsequent joint analysis of multiple pollutants would also be valuable” (Ibid., p. 27).

The empirical results in the current paper provide evidence that multiple pollutant control provides some explanation for the observed BOD overcompliance. The results indicate that the regulation of nitrogen is an important determinant of BOD overcompliance amongst plants in Connecticut and Maryland. The parameter estimate on nitrogen is statistically significant at the 1 percent level in both regressions indicating that the effect of jointness on the observed level of BOD overcompliance is robust. The reductions in BOD due to reductions in N are not due to smooth variations in input levels at the plants but more likely result from discrete infrastructural changes in the plant’s capital due to the installation of BNR technology.

The BOD output response to changes in the permit limit is also robust. Plants with lower BOD permit limits also have lower BOD discharge levels. Since the regression determines correlations and not causation, the direction of causality is not clear: either plants are lowering their discharge levels in response to the lower permit limit or the regulators are lowering permit limits for plants that are discharging relatively low levels of BOD.

More research is needed in order to establish the direction of causality. In the case of Maryland plants, the direction of causality can be imputed using currently available evidence. The results in Part II and qualitative evidence from an interview with a Maryland state official indicate that plants with seasonal limits react to these limits by

---

<sup>53</sup> The model estimates the relationship between the overcompliance ratio and each of the following: a constant term, the variance, and a measure of plant size. Their model estimates that the average overcompliance ratio is 0.33. Based on their raw data, the observed average overcompliance ratio is 0.41.

lowering their BOD discharges. If there are stages to the decision making process, then the adjustment for the lower permit limit is either made during the plant's capital design phase and/or during the operational phase. Since the regression has no time dimension, it is difficult to determine whether the positive correlation results from earlier choices made in the capital design phase with a lower limit in mind or from later choices made in the operational phase in response to a lowering of the permit limit. Defining a dynamic model is a step towards resolving this difficulty and the estimation of a dynamic model will involve determining an alternative way to measure the variance of BOD while still preserving the time dimension.

Finally, these results are not complete because some cases are not considered. Not all plants that are overcomplying with their permit limit will also be controlling for nitrogen. More research is needed to document the number of plants that overcomply with BOD limits without providing tertiary treatment and more research is needed to document their degree of overcompliance relative to those that do provide tertiary treatment.

## **Section 5: Conclusion**

The objective of this paper was to identify plausible explanations for the observed degree of overcompliance with NPDES permit limits for BOD. Based on a variety of sources including reviews of the engineering literature on wastewater treatment technology, reviews of the economic literature on permit compliance, interviews with state and federal regulators of wastewater treatment plants, and an analysis of effluent data, the research indicates that certain characteristics of the treatment plant explain much of the variation in BOD overcompliance across treatment plants. In terms of plant characteristics, the type of technology installed at the plant and whether the plant controls multiple pollutants (i.e. whether there is BNR technology or not) can affect the plant's performance in terms of BOD control. Also, variation in BOD influent levels and in weather conditions can affect the plant's ability to treat the wastewater and ultimately its performance relative to the BOD permit limit.

There is recognition in the economic literature on permit compliance that wastewater treatment is a dynamic process and that BOD output levels will vary over time and across plants. The quantitative results using data from Connecticut and

Maryland corroborate these conclusions by showing that, independent of decisions at the plant:

- BOD output varies naturally over the course of a year, falling during the warmer months and rising during the colder months;
- Weather events can affect the plant's ability to treat the wastewater by causing severe washouts;
- There is some evidence of a negative relationship between the variation in BOD output and the plant's average level of BOD output. However, the relationship is only statistically significant at the 10 percent level for the Maryland plants.

In addition to providing additional confirmation of observations in previous studies concerning the stochastic nature of BOD output, the paper makes a contribution to the literature on permit compliance by providing additional explanations for BOD overcompliance. The results indicate that:

- The installation of nitrogen control technology at the plant leads to more BOD control and that this result is statistically significant, providing evidence of jointness in the control of nitrogen and BOD.
- For some plants, the more excess capacity at the plant, the better will be the plant's performance in terms of BOD control and overcompliance.
- Plants with limits below the technology-based standard of 30mg/l and plants with seasonal BOD limits have a lower level of BOD output than plants with BOD limits at 30mg/l.

The reasons for the positive relationship between BOD output and the BOD permit limit are not clear and warrant further research. The regression model is unable to shed some light on this issue because lacks a time dimension and therefore cannot distinguish between decisions taken in the capital design phase and those taken in the operational phase. On a related note, the paper raises questions about common assumptions regarding behavioral choice. The results (particularly concerning the relationship between variance in BOD and the average BOD output) suggest a need to distinguish between the capital design and operational phases when studying choices regarding the level of BOD control.



The economic literature has tended to characterize the management at the WWTP as actively involved in setting effluent discharge levels via its choice of variable inputs like labor and energy. However, interviews with regulators and plant operators suggest that the WWTP management does not exercise as much control over the compliance decision in the plant's operational phase as suggested in the literature.<sup>54</sup> Furthermore, a review of wastewater treatment technology confirms the lack of direct control over the treatment process. The technology review reveals that the BOD treatment is largely biological; that the ability of the WWTP to reduce BOD concentrations in the wastewater will be a function of the temperature of the water, the flow of the water, and the interactions of microbial organisms in the wastewater.

The lack of recognition in current behavioral models of the decision maker's inexact control over the pollution control process may be due to an error of extrapolation whereby certain characteristics of the typical technologies encountered in the literature are imposed on the pollution control process without a full investigation as to their relevance. For most types of production processes studied in economics, the technology is such that output can be varied in a given direction by adjusting the variable (as opposed to fixed) inputs in the right direction.<sup>55</sup> However, there are some unique situations in which changes in the variable inputs have no measurable impact on output levels, or there is a limited range over which variable inputs can have an impact on output levels, or there is a limited number of variable inputs that can have an impact on output levels at any given stage in a decision-making process.

In the case of a treatment plant, the technology is not as flexible as the technologies encountered in other settings. Rather, the plant is designed to operate at a certain level of pollutant control such that once the plant is operational and as long as the plant is being operated according to specifications, the level of BOD control cannot be increased or decreased via changes in economic inputs like labor and energy.

---

<sup>54</sup> Most of the officials interviewed confirmed that treatment plant operators have little direct control over the level of BOD output. However, one of the officials, by stating that plants in Maryland respond to seasonal limits by lowering summer BOD output levels, implied that there is some degree of operational flexibility. More research is needed to quantify the level of operational flexibility that exists in the treatment process and harmonize the evidence.

<sup>55</sup> This includes technologies that are smooth and lumpy, so called.

There are behavioral decisions that affect the plant's level of BOD control but it appears that these decisions are medium-to-long term capital decisions rather than day-to-day operational decisions. When engineers in charge of designing the treatment plant adjust the mean BOD output relative to the permit limit due to a high variance of BOD discharges, they do so at the capital design phase. Their decision has a direct bearing on the capital inputs used in construction and on the eventual level of BOD control at the plant. Another behavioral decision, also having to do with capital installation, that affects the level of BOD control at the plant is the decision whether or not to install nitrogen control technology. While this decision is not necessarily made at the design phase, it is nonetheless a capital decision like that of the engineers'. Thus, the evidence suggests that changes in capital inputs have a direct impact on the plant's BOD output levels and that these are the economic inputs that can affect the observed level of BOD overcompliance.

Additional insight might be gained from modeling effluent compliance decisions as a two stage process that models choice of effluent discharge levels during the capital construction phase separately in a first stage and models the choice of effluent discharge levels during the operational phase separately in a second stage. Modifying the behavioral model in this way opens up new lines of research, specifically:

- 1) With a two stage decision process, one can separate out the relative contributions, at the design and operational levels, to variation in average BOD output relative to the permit. For instance if one has time series that identify plants with more stringent BOD limits imposed after capital inputs were already in place separately from plants that were designed to achieve more stringent BOD limits, one could estimate the relative contribution that operational decisions play in explaining a plant's response to changes in BOD limits. This type of information would be important for regulators to know when faced with BOD related water quality problems because they can estimate both the discharger's response to changes in the BOD permit limit and whether the response will be sufficient to address the water quality need.

- 2) It is possible that objective functions will differ between the two stages. The first stage might be characterized as a negotiation between the regulator and the discharger as to the plant's design specifications. The outcome of this process establishes the BOD permit limit that guides the design and construction of the plant in the first stage and determines the plant's performance in terms of pollutant control. The regulator sets the permit limit based on water quality modeling and CWA requirements concerning technology-based standards (personal communication, T-S Yu, Maryland Department of the Environment.) The regulator's goal may be to maximize treatment by reducing permit limits given the plant's control costs.<sup>56</sup> If the municipality that owns the WWTP is receiving some financial assistance (i.e. low-interest loan or cost-share funds) to defray capital construction or technology upgrade costs, it may be more willing to accept a lower permit limit in exchange for the financial assistance. The discharger's goals may be to minimize treatment costs, including capital construction costs. However, the regulator controls the final decision as to the level of the permit limit (personal communication, T-S Yu, Maryland Department of the Environment). So the first stage objective might be to maximize pollutant control subject to pollutant control costs, reflecting the regulator's and the discharger's preferences. In the operational stage, however, the discharger is more concerned with remaining in compliance and operating the treatment plant according to its design specifications. There is less of a role for the regulator in the operational phase and the goal of expected cost-minimization may

---

<sup>56</sup> There is evidence to suggest that regulators will actively push for tighter effluent controls in capital decisions. The technology-based performance standards for BOD are known as "best control technology" (BCT) limits (WEF, 1997). BCT limits are based on the best performance of a given conventional treatment method. In the case of toxic pollutants, permit limits are based on the most stringent of technology-based limits, which are those based on the "best available technology economically achievable," which are "the very best control and treatment measures that have been or are capable of being achieved" (WEF, 1997). Thus, BCT standards seek the highest level of treatment that is possible at reasonable cost to the treatment plant. Costs are "reasonable" in the sense that the standards are based on the conventional treatment technology and not on the most recently developed technology, the use of which may come at a premium.

better reflect the objective function in this stage. The goal of this discussion is to illustrate how moving to a two stage decision framework can modify the objective functions governing behavioral choice.

- 3) More research on the capital design phase will increase knowledge, not only about the nature of the behavioral objective function but will also bring new insight into the role that variation in BOD output plays in the capital design phase and in the eventual design specifications related to BOD output. More specifically, research on the capital design phase will reveal how the engineers hired by the treatment plant owner account for variation in BOD output in their capital design specifications. Is there a ‘general rule of thumb’, such as described by an official of the Maryland Department of the Environment, like designing a plant for BOD output that is one-half to two-thirds of the permit limit? If so, how did this general rule of thumb come to be? If not, then what specific measures (i.e. formulae) are used to account for BOD variation when designing the plant?
- 4) If decisions affecting the level of BOD control largely occur at the capital design phase and the access to cost-share funds has implications for BOD control via its effect on the plant’s capital acquisition costs, then state-level variation in the availability of these construction assistance funds may explain variation in BOD control, and by extension, BOD overcompliance.

More research is needed into building appropriate theoretical and empirical models of the two-stage decision making process whereby plants arrive at their observed levels of BOD overcompliance. In terms of the theoretical implications of the shift from a one-stage to a two-stage framework, conceptualizing the observed BOD overcompliance levels as the result of a two-stage negotiation between plant management and regulators allows for the use of game theoretic and bargaining principles to explore the dynamics of the institutional decision making process. It also necessitates more research on the type of information that is available to each party at the bargaining table

in each stage as well as the costs and benefits of acquiring more information in each stage. Furthermore, since the bargaining process may likely occur at 5 year intervals once a permit expires, the impact of time and increased knowledge about the bargaining process on the bargaining outcome can also be explored using dynamic behavioral models.

The empirical implication of the two-stage behavioral model is that 2-stage econometric models may offer new lines of research compared to the one-stage models in terms of a better characterization of the data generating process. In the first stage, the researcher may model the relationship between the design BOD output levels and the variance in BOD output and other variables related to the design of the plant. In the second stage, the researcher may model the relationship between the observed BOD overcompliance decision as it relates to the plant's design standards, the installation of BNR (if it occurs in the second stage), and other variables related to the operation of the plant that may affect BOD overcompliance. Furthermore, if the negotiating parties change their behavior as a result of gaining more knowledge over time then a dynamic empirical model may be more appropriate in order to incorporate the effect of dynamic behavioral changes on the observed choice.

Finally, to the extent that regulator and management preferences, the available cost-share funds, the capital costs of construction and the need for technology upgrades differ from state to state, the dynamics of the first-stage bargaining process as well as the observed BOD overcompliance outcomes will differ from state to state. The availability of data that cover a wide geographic area will make it possible to test whether between-state variation in observed BOD overcompliance exists and what explains the variation.

In order to understand the determinants of the observed widespread and significant levels of BOD overcompliance, the theoretical and empirical models must properly characterize the technological, regulatory, and economic constraints faced by all parties in the bargaining process. Current models may offer occasional insights but must be improved upon in order to fully explain the evidence for overcompliance.

## References

- Bandyopadhyay, S. and J.K. Horowitz. "Variability and Overcompliance in Water Pollution Discharges: Evidence and Implications." Working Paper. (2004). University of Maryland. Website: <http://www.arec.umd.edu/jhorowitz/images/WebVersion.pdf> Accessed: 5.20.04.
- Bandyopadhyay, S. and J.K. Horowitz. "Randomness in Point Source Water Pollution: Evidence and Implications." Working Paper. (2002). Website: <http://www.crde.umontreal.ca/cree/textes/Horowitz.pdf> Accessed: 5.20.03
- Beavis, B. and M. Walker. "Random Wastes, Imperfect Monitoring and Environmental Quality Standards." *Journal of Public Economics* 21 No.3 (1983a):377-87.
- Beavis, B. and M. Walker. "Achieving Environmental Standards with Stochastic Discharges." *Journal of Environmental Economics and Management* 10 No. 2 (1983b):103-11.
- Blandford, D. and R. N. Boisvert. "Non-trade concerns and domestic/ international policy choice." 77<sup>th</sup> EAAE Seminar/ NJF Seminar No. 325. (2001). Website: <http://www.ptt.fi/eaee-njf/papers/blandford.pdf> Accessed: 5.15.04.
- Brannlund, R. and K-G Lofgren. "Emission Standards and Stochastic Waste Load." *Land Economics*. 72 No. 2 (1996): 218-30.
- Chesapeake Bay Foundation. "Sewage Treatment Plants: The Chesapeake Bay Watershed's Second Largest Source of Nitrogen Pollution." (October 2003a). Website: [http://www.cbf.org/site/DocServer/STP\\_final\\_no\\_watermark.pdf?docID=1283](http://www.cbf.org/site/DocServer/STP_final_no_watermark.pdf?docID=1283) Accessed: 8.30.04.
- Chesapeake Bay Foundation. *The Chesapeake Bay Program: Municipal Facilities in the Bay Watershed using Nutrient Reduction Technology (NRT)*. (2003b). Website: <http://www.chesapeakebay.net/status/> Accessed: 4.2.04.
- "Civil Monetary Penalty Inflation Adjustment Rule." *Federal Register*. (February 13 2004). Website: <http://www.epa.gov/earth1r6/6en/w/penalties-mar-15-2004.pdf> Accessed 8.20.04.
- Clean Water Network. "Clean Water Act Enforcement is Weak and Sporadic." Website: <http://www.cwn.org/docs/facts/enforce.htm> Accessed 12.11.01.
- Connecticut Department of Environmental Protection. "DEP Continues Watershed Approach to Removing Nitrogen from the Long Island Sound." (2002). Website: <http://dep.state.ct.us/whatshap/press/2002/mf0214.htm> Accessed 8.20.04.

- Connecticut Department of Environmental Protection. "Connecticut's Nitrogen Control Program." (2001). Website: <http://www.dep.state.ct.us/wtr/lis/nitrogengp/ngpfs.pdf> Accessed. 6.15.04.
- Daigger, G. T. and J. A. Buttz. *Upgrading Wastewater Treatment Plants*. Lancaster, PA: Technomic Publishing Company, Inc. 2<sup>nd</sup> edition. 1998.
- Dasgupta, S., H. Hettige, and D. Wheeler. "What Improves Environmental Compliance? Evidence from Mexican Industry." *Journal of Environmental Economics and Management*. 39 (2000): 39-66.
- Davis, M. L. and D. A. Cornwell. *Introduction to Environmental Engineering*. New York: McGraw Hill, Inc. 2<sup>nd</sup> edition. 1991.
- Deegan, Glenn E. "Judicial Enforcement of State and Municipal Compliance with the Clean Water Act: Can the Courts Succeed?" *Boston College Environmental Affairs Law Review*. 19 p. 765. (1992).
- Doonan, J., P. Lanoie, B. Laplante. "Environmental Performance of Canadian Pulp and Paper Plants: Why Some do Well and Others do not?" CIRANO. Scientific series. (2002). Website: <http://www.cirano.qc.ca/pdf/publication/2002s-24.pdf> Accessed 1.30.03
- Driscoll, C.T., D. Whitall, J. Aber, E. Boyer, M. Castro, C. Cronan, C. L. Goodale, P. Groffman, C. Hopkinson, K. Lambert, G. Lawrence, S. Ollinger. 2003. "Nitrogen Pollution in the Northeastern United States: Sources, Effects, and Management Options. *BioScience* 53 No. 4 (2003): 357-374.
- Farber, D. A. "Taking Slippage Seriously: Noncompliance and Creative Compliance in Environmental Law." *Harvard Environmental Law Review*. 23 p. 297 (1999).
- Fahrenthold, D. A. "EPA Cleanup Proposal Would Limit Sewage Plant Pollutants." *Washington Post*. (July 26, 2004). Page B01.
- Foulon, J., P. Lanoie, and B. Laplante. "Incentives for Pollution Control: Regulation or Information?" *Journal of Environmental Economics and Management*. 44 (2002): 169-187.
- Franckx, L. "The Use of Ambient Inspections in Environmental Monitoring and Enforcement When the Inspection Agency Cannot Commit Itself to Announced Inspection Probabilities." *Journal of Environmental Economics and Management*. 43 (2002): 71-92.

- Fuller, D. A. "Compliance, Avoidance, and Evasion: Emissions Control Under Imperfect Enforcement in Steam-Electric Generation." *RAND Journal of Economics*. 18 No. 1 (1987): 124-137.
- Gray, N.F. *Water Technology: An Introduction for Environmental Scientists and Engineers*. London: Arnold Publishers. (1999).
- Greene, W. H. *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall. (1997).
- Hammer, M. J. and M. J. Hammer, Jr. *Water and Wastewater Technology*. 3<sup>rd</sup> edition. New Jersey: Prentice Hall. (1996).
- Harford, J. D. "Measurement Error and State-Dependent Pollution Control Enforcement." *Journal of Environmental Economics and Management*. 21 (1991):67-81.
- Harford, J. D. and W. Harrington. "A Reconsideration of enforcement Leverage when Penalties are Restricted." *Journal of Public Economics*. 45 (1991):391-395.
- Harrington, W. "Enforcement Leverage When Penalties are Restricted." *Journal of Public Economics*. 37 (1988):29-53.
- Hartman, Raymond S., Mainul Huq, David Wheeler. "Why Paper Mills Clean Up: Determinant of Pollution Abatement in Four Asian Countries." *World Bank Working Paper 1710*. (1995).  
<http://www.worldbank.org/html/dec/Publications/Workpaper/WPS11700series/wps1710/wps1710.pdf> Accessed 1.30.03
- Heyes, A. and J. Doucet. "2 Stage Enforcement and Regulatory Polarisation: A Simple Model with Application to the USEPA." Department of Economics Working Paper, Royal Holloway College, University of London. (1999).
- Hipel, K., X. Yin, and D. M. Kilgour. "Can a Costly Reporting System Make Environmental Enforcement More Efficient?" *Stochastic Hydrology and Hydraulics*. 9 (1995):151-170.
- Horowitz, J. K. and S. Bandyopadhyay. "Overcompliance in Point-Source Water Pollution." Website: [http://are.berkeley.edu/~peter/AERE\\_Papers/horowitz.pdf](http://are.berkeley.edu/~peter/AERE_Papers/horowitz.pdf)
- Kagan, R.A., N. Gunningham, and D. Thornton. "Explaining Corporate Environmental Performance: How Does Regulation Matter?" *Law and Society Review*. 37 No.1 (2003): 51-89.
- Laplante, B. and P. Rilstone. "Environmental Inspections and Emissions of the Pulp and Paper Industry in Quebec." *Journal of Environmental Economics and Management*. 31 (1996): 19-36.



- Liu, P.C. “Regulator Inspection and Violator Deterrence Under Clean Water Act Regulation of Pulp and Paper Mill Water Pollution.” PhD Dissertation, Stanford University. (August 1995).
- Liu, D. H. F. and B. G. Liptak. (eds.). *Wastewater Treatment*. New York: Lewis Publishers (CRC Press LLC). (1999).
- “Long Island Sound Study.” USEPA Long Island Sound Office. (2001). Website: <http://www.seagrant.sunysb.edu/Pages/LISHealth01.pdf> Accessed 4.04.04.
- Magat, W. A. and W. K. Viscusi. “Effectiveness of the EPA’s regulatory enforcement: The case of industrial effluent standards.” *Journal of Law and Economics*. 33 (1990): 331-360.
- Malik, A. S. “Avoidance, Screening and Optimum Enforcement.” *RAND Journal of Economics*. 21 No. 3 (Autumn 1990): 341-352.
- Malik, A. S. “Self-reporting and the Design of Policies for Regulating Stochastic Pollution.” *Journal of Environmental Economics and Management*. 24 No.3 (1993): 241-257.
- Manning W.G. and J. Mullahy. “Estimating log models: to transform or not to transform?” *Journal of Health Economics*. 20 (2001):461-494.
- Maryland Department of the Environment(1). “Governor Ehrlich’s Bay Restoration Fund for Enhanced Nutrient Removal.” Website: <http://www.mde.state.md.us/Water/CBWRF/ENR.asp> Accessed. 8.25.04.
- Maryland Department of the Environment(2). “Biological Nutrient Removal Program.” Website: [http://www.mde.state.md.us/Programs/WaterPrograms/WQIP/wqip\\_bnr.asp](http://www.mde.state.md.us/Programs/WaterPrograms/WQIP/wqip_bnr.asp) Accessed 8.25.04.
- McClelland, J. and J. Horowitz. “The Costs of Water Pollution Regulation in the Pulp and Paper Industry.” *Land Economics*. 75 No. 2 (1999):220-32.
- Oljaca, N., A. G. Keeler, and J. Dorfman. “Penalty functions for environmental violations: Evidence from water quality enforcement.” *Journal of Regulatory Economics*. 14 No. 3 (1998): 255-64.
- “Nutrient Pollution.” Chesapeake Bay Program. Website: <http://www.chesapeakebay.net/info/nutr1.cfm> Accessed: 4.2.04
- Pianin, E. “Widespread Water Violations Decried.” *Washington Post*. (August 7, 2002).

- Sarbanes, P. S. "Statement of Senator Paul S. Sarbanes. Hearing before the Senate Committee on Environment and Public Works Subcommittee on Fisheries, Wildlife and Water." (2002). Website: [http://epw.senate.gov/107th/Sarbanes\\_022802.htm](http://epw.senate.gov/107th/Sarbanes_022802.htm) Accessed 8.26.04.
- Spanos, A. *Statistical Foundations of Econometric Modeling*. Cambridge: Cambridge University Press. (1986).
- State NPDES Program Authority. Website: [http://cfpub1.epa.gov/npdes/images/State\\_NPDES\\_Prog\\_Auth.pdf](http://cfpub1.epa.gov/npdes/images/State_NPDES_Prog_Auth.pdf) Accessed 8.20.04.
- Stoddard, A., J. Harcum, J. Simpson, J. R. Pagenkopf, R. K. Bastian. *Municipal Wastewater Treatment: Evaluating Improvements in National Water Quality*. New York: John Wiley & Sons, Inc. (2002).
- United States Environmental Protection Agency (USEPA). "Municipalities and Wastewater Treatment Plants." (2002). Website: <http://cfpub2.epa.gov/npdes/home.cfm> Accessed 6.15.03.
- USEPA. Guidance on the Privatization of Federally Funded Wastewater Treatment Works. Office of Water. EPA-832-B-00-002. (2000). <http://www.epa.gov/OW-OWM.html/pdfs/prigud.pdf> Accessed 3.19.03
- United States Environmental Protection Agency (USEPA). "Region 10's National Pollutant Discharge Elimination System Permit Program." *Office of Inspector General Audit Report: Water*. E1HWF7-10-0012-8100076. (1998).
- United States Environmental Protection Agency (USEPA). "Combined Sewer Overflow O&M Fact Sheet." (1999a). <http://www.epa.gov/OWM/mtb/o&m.pdf> Accessed 2.14.03
- United States Environmental Protection Agency (USEPA). "Toxicity Reduction Evaluation Guidance for Municipal Wastewater Treatment Plants." USEPA Office of Wastewater Management. EPA-833B-99-002. (1999b).
- United States Environmental Protection Agency (USEPA). "NPDES Permit Writers' Manual." USEPA Office of Water. EPA-833-B-96-003. (December 1996). <http://www.epa.gov/npdes/pubs/> Accessed 3.18.03
- United States Environmental Protection Agency (USEPA). "Water Permitting 101." Website: <http://www.epa.gov/npdes/pubs/101pape.htm> Accessed 5.8.02.
- United States General Accounting Office (USGAO). "Environmental Enforcement: EPA Cannot Ensure the Accuracy of Self-Reported Compliance Monitoring Data." *GAO/RCED-93-21*. (1993).

United States General Accounting Office (USGAO). “Many Violations Have Not Received Appropriate Enforcement Attention.” *GAO/RCED-96-23*. (1996).

United States General Accounting Office (USGAO). “More Consistency Needed Among EPA Regions in Approach to Enforcement.” *GAO/RCED-00-108*. (2000).

United States General Accounting Office (USGAO). “Observations on Compliance and Enforcement Activities Under the Clean Water Act.” *GAO/T-RCED-91-80*. (1991).

United States General Accounting Office (USGAO). “Stronger Enforcement Needed to Improve Compliance at Federal Facilities.” *GAO/RCED-89-13*. (1988).

## **List of Experts**

Johnson, Gary. Senior Environmental Engineer, Municipal Facilities, Water Management Bureau, Connecticut Department of Environmental Protection.

Luckman, Steve. Division Chief, Municipal Surface Discharge Permits Division, Wastewater Permits Program, Water Management Administration, Maryland Department of the Environment.

Randall, Clifford W. Director, Occoquan Watershed Monitoring Program and Professor, Department of Civil and Environmental Engineering, Virginia Tech.

Yu, Ta-Shon. Director of Technical Services, Water Quality Infrastructure Program, Water Management Administration, Maryland Department of the Environment.

## Appendix: Residual plots

