Econometric Reviews, 26(5):503–528, 2007

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474930701512105

Taylor & Francis
Taylor & Francis Group

ASSESSING THE PRECISION OF TURNING POINT ESTIMATES IN POLYNOMIAL REGRESSION FUNCTIONS

Florenz Plassmann □ *Department of Economics, State University of New York at Binghamton, Binghamton, New York, USA*

Neha Khanna □ *Department of Economics and Environmental Studies Program, State University of New York at Binghamton, Binghamton, New York, USA*

□ *Researchers often report point estimates of turning point(s) obtained in polynomial regression models but rarely assess the precision of these estimates. We discuss three methods to assess the precision of such turning point estimates. The first is the delta method that leads to a normal approximation of the distribution of the turning point estimator. The second method uses the exact distribution of the turning point estimator of quadratic regression functions. The third method relies on Markov chain Monte Carlo methods to provide a finite sample approximation of the exact distribution of the turning point estimator. We argue that the delta method may lead to misleading inference and that the other two methods are more reliable. We compare the three methods using two data sets from the environmental Kuznets curve literature, where the presence and location of a turning point in the income-pollution relationship is the focus of much empirical work.*

Keywords Asymmetric confidence interval; Environmental Kuznets curve hypothesis; MCMC; Quantiles.

JEL Classification C16; C51; C52; Q50.

1. INTRODUCTION

Economists often use polynomial regression functions to assess the empirical evidence when economic theory predicts a nonmonotonic relationship between two variables. For example, in testing the human capital model of wage profiles it is common to estimate log wage

Received March 25, 2006; Accepted April 26, 2006
Address correspondence to Florenz Plassmann, Department of Economics, State University of New York at Binghamton, Binghamton, New York 13902-6000, USA; E-mail: fplass@binghamton.edu

equations that include age and its square as right-hand side variables. Likewise, a standard hypothesis regarding voter participation is that households at the very low and very high end of the income scale have lower participation rates in the electoral process than middle-income households, and regression models of voter participation include a quadratic income term on the right-hand side. Another example, the environmental Kuznets curve (EKC) hypothesis, posits an inverted U-shaped relationship between income and pollution, and the standard practice is to regress some indicator of environmental degradation, such as emissions or deforestation, on a quadratic or cubic function of income.

In each case, the regression model yields an estimated turning point. It is necessary to evaluate the precision of the estimated turning point to determine whether it is indeed representative of the data rather than being an artifact of the polynomial functional form. However, many authors do not assess the precision of their turning point estimate (see, for example, Abbot and Beach, 1993; Kimhi and Rapaport, 2004; Miles, 1997; Neumark and Taubman, 1995; Tokle and Huffman, 1991, from the labor economics and agricultural economics literature; see Greene and Nikolaev, 1999, from the public economics literature, and Barrett and Graddy, 2000; Ferreira, 2004; Torras and Boyce, 1998; Stern and Common, 2001, from the environmental economics literature). Others linearize the turning point estimator and approximate its distribution using the normal distribution (the delta method)—see, for example, Cole et al. (1997), Grossman and Krueger (1995), Harbaugh et al. (2002), and Millimet et al. (2003). List and Gallet (1999) mention 95 percent confidence intervals and Co et al. (2004) report standard errors for their turning point estimates, but neither paper includes any supporting details.

Turning point estimators for polynomial regression equations are nonlinear functions of highly correlated random variables, and their distributions are generally nonsymmetric. We argue that the delta method that is commonly used to approximate these distributions with the symmetric normal distribution is likely to lead to misleading inference. We compare the delta method to two alternative methods: the first alternative is based on the assumption that the errors are (asymptotically) normally distributed and uses the exact distribution of the turning point estimator of quadratic regression functions. The other relies on Markov chain Monte Carlo (MCMC) techniques to provide an estimate of the exact finite sample distribution of the turning point estimator. We discuss and evaluate the three methods in the next section and then compare them using two data sets from the EKC literature, where the presence and location of the turning point in the income-pollution relationship remains the focus of empirical work. We argue that, for these data, the two alternative methods lead to more reliable inference than the delta method.

2. THREE METHODS TO ASSESS THE PRECISION OF TURNING POINT ESTIMATES IN POLYNOMIAL REGRESSION FUNCTIONS

2.1. Quadratic Models

Quadratic regression functions are a simple and widely used approach to estimate nonmonotonic relationships.[1] Assume that the relationship between two variables, y and x, is estimated with the quadratic equation $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, \ldots, n,$ where ε_i is a nonsystematic disturbance. The extremum, or turning point, τ, of the equation is $\tau = \beta_1/(-2\beta_2)$. A commonly used estimator of τ is $\hat{\tau} = \hat{\beta}_1/(-2\hat{\beta}_2)$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimators of β_1 and β_2, respectively.[2] To distinguish estimators and estimates, we denote the point estimates of β_1, β_2, and τ by $\hat{\beta}_1^0$, $\hat{\beta}_2^0$, and $\hat{\tau}^0$, respectively.

Method 1. Approximating the distribution of $\hat{\tau}$ using a Taylor expansion (delta method).

The generally used estimates of the moments of $\hat{\tau}$ are based on a first-order Taylor series expansion of $\hat{\tau}$ around β_1 and β_2 (see, for example, Greene, 2002, pp. 913–914, and also Papke and Wooldridge, 2005). This expansion yields the approximation of $\hat{\tau}$ as

$$\begin{aligned}
\hat{\tau}^T &= -\frac{\beta_1}{2\beta_2} - \frac{1}{2\beta_2}\left(\hat{\beta}_1 - \beta_1\right) + \frac{\beta_1}{2(\beta_2)^2}\left(\hat{\beta}_2 - \beta_2\right) \\
&= -\frac{\beta_1}{2\beta_2} - \frac{1}{2\beta_2}\hat{\beta}_1 + \frac{\beta_1}{2(\beta_2)^2}\hat{\beta}_2 \\
&= a + b\hat{\beta}_1 + 2ab\hat{\beta}_2 \\
&\approx \hat{\tau}, \qquad (1)
\end{aligned}$$

where $a = -\beta_1/(2\beta_2)$ and $b = -1/(2\beta_2)$. One can therefore approximate the mean and variance of $\hat{\tau}$ with $E(\hat{\tau}^T) = a + bE(\hat{\beta}_1) + 2abE(\hat{\beta}_2)$ and $Var(\hat{\tau}^T) = b^2 Var(\hat{\beta}_1) + (2ab)^2 Var(\hat{\beta}_2) + 4ab^2 Cov(\hat{\beta}_1 \hat{\beta}_2)$, using the sample

[1]An alternative is to estimate different slope coefficients for the two parts of the relationship (either as a spline function or with two separate regressions) and to test whether the slope coefficients are different from each other. This approach restricts the regression equation by less than a single equation polynomial regression, but unlike the polynomial regression, it requires separate estimation of the position of the turning point.

[2]Alternatively, one can reparameterize the equation to obtain $y_i = a(x_i - b)^2 + c + \varepsilon_i$, where $a = \beta_2$, $b = -\beta_1/(2\beta_2)$, and $c = \beta_0 - \beta_1^2/(4\beta_2)$. Now the turning point is one of the parameters of the model (b) that can be estimated directly, but the new equation is nonlinear in the parameters. Millimet et al. (2003) show that this reparameterization can greatly improve the accuracy of the estimates when the regressors can be chosen (optimal design).

estimates $\hat{\beta}_1^0$ and $\hat{\beta}_2^0$ to determine the values of the unknown parameters a and b.

The distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ determine the distribution of $\hat{\tau}^T$. For example, if the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are bivariate normally distributed, then $\hat{\tau}^T$ is normally distributed as well and the approximated 95 percent confidence interval around the turning point estimate is symmetric.[3]

Method 2. Estimating the exact distribution of $\hat{\tau}$ if $\hat{\beta}_1$ and $\hat{\beta}_2$ are bivariate normally distributed.

If $\hat{\beta}_1$ and $\hat{\beta}_2$ are bivariate normally distributed as $N[(\mu_1, \mu_2)^T, ((\sigma_1^2, \sigma_{12}^2)^T(\sigma_{12}^2, \sigma_2^2)^T)]$, then the joint distribution of $\hat{\beta}_1$ and $-2\hat{\beta}_2$ is $N[(\mu_1, -2\mu_2)^T, ((\sigma_1^2, -2\sigma_{12}^2)^T(-2\sigma_{12}^2, 4\sigma_2^2)^T)]$. The cumulative distribution function (cdf) of $\hat{\tau} = \hat{\beta}_1/(-2\hat{\beta}_2)$, $F(t)$, is given by (see Fieller, 1932)[4]

$$F(t) = G\left(\frac{\mu_1 + 2\mu_2 t}{2\sigma_1\sigma_2 d(t)}, \frac{\mu_2}{\sigma_2}; \frac{2\sigma_2 t - \rho\sigma_1}{2\sigma_1\sigma_2 d(t)}\right) + G\left(\frac{-2\mu_2 t - \mu_1}{2\sigma_1\sigma_2 d(t)}, -\frac{\mu_2}{\sigma_2}; \frac{2\sigma_2 t - \rho\sigma_1}{2\sigma_1\sigma_2 d(t)}\right) \tag{2}$$

with

$$d(t) = \left(\frac{t^2}{\sigma_1^2} - \frac{\rho t}{\sigma_1\sigma_2} + \frac{1}{4\sigma_2^2}\right),$$

$$G(h, k; \gamma) = \frac{1}{2\pi\sqrt{(1-\gamma^2)}} \int_h^\infty \int_k^\infty \exp\left(-\frac{x^2 - 2\gamma xy + y^2}{2(1-\gamma^2)}\right) dx\, dy,$$

and $\rho = -\sigma_{12}^2/(\sigma_1\sigma_2)$. In general, $F(t)$ is not symmetric. Although the bivariate normal integral in $G(h, k, \gamma)$ has no closed form solution, many computer packages make its evaluation straightforward (for example, with the function *binormal* in STATA9) and we use the exact values throughout the paper. If such programs are unavailable, then Equation (2) can be approximated by

$$H(t) = \Phi\left\{\frac{-2\mu_2 t - \mu_1}{2\sigma_1\sigma_2 d(t)}\right\} \pm \Phi\left\{\frac{\mu_2}{\sigma_2}\right\}, \tag{3}$$

[3]This result as well as the results in the following sections hold asymptotically if $\hat{\beta}_1$ and $\hat{\beta}_2$ are asymptotically normally distributed.

[4]Our Equation (2) differs slightly from Fieller's equations 15–18 because we consider the distribution of the ratio of $\hat{\beta}_1$ and $-2\hat{\beta}_2$ while Fieller considers the distribution of the ratio of $\hat{\beta}_1$ and $\hat{\beta}_2$.

where Φ is the cdf of the univariate standard normal distribution (see Hinkley, 1969) and the second term on the right-hand side has the same sign as the expression $\mu_1/\sigma_1^2 + \rho\mu_2/(2\sigma_2^2)$ (see Kotz et al., 2000, p. 328).

Given estimates of the moments of $\hat{\beta}_1$ and $\hat{\beta}_2$, it is straightforward to use either Equations (2) or (3) to estimate quantiles of $\hat{\tau}$ (for example, the quantiles corresponding to the 2.5th and 97.5th percentiles that define the standard 95 percent confidence interval). Because the distribution of $\hat{\tau}$ is generally not symmetric, it is inappropriate to approximate the 95 percent confidence interval by adding and subtracting 1.96 times the estimated standard error.

The density function of $\hat{\tau}$ can be written as the product of a Cauchy density and a term that contains the normal density and the normal integral. Fieller (1932, pp. 432 and 435) and Marsaglia (1965, p. 196) show that the moments of $\hat{\tau}$ do not exist, which has the interesting implication that $\hat{\tau}^0 = \hat{\beta}_1^0/(-2\hat{\beta}_2^0)$ is not an accurate estimate of the turning point. If $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased estimators of β_1 and β_2, then the median (the 50th percentile) of the distribution of $\hat{\tau}$ is a better estimate of $\beta_1/(-2\beta_2)$.

When is it appropriate to use the normal approximation through the delta method (Method 1) to assess the precision of the turning point estimate in quadratic regression models? Shanmugalingam (1982) provides numerical evidence that the distribution of $\hat{\tau}$ is symmetric in t only if $\rho = CV_2/CV_1$, where $CV_1 = \sigma_1/\mu_1$ and $CV_2 = -\sigma_2/\mu_2$ are the coefficients of variation of $\hat{\beta}_1$ and $-2\hat{\beta}_2$. Fieller (1932) and Hinkley (1969) show that $\hat{\tau}$ is approximately normally distributed if CV_2 is small. Figure 1 compares the probability density function of $\hat{\tau}$, $f(t)$, with the normal density function of $\hat{\tau}^T$ that we obtain with the delta method (Method 1) for the assumed values $\mu_1 = 9.9$, $\mu_2 = -20$, $\sigma_1 = 5$, $\sigma_2 = 10$, and $\sigma_{12}^2 = -49.5$.[5] These values imply $CV_1 = 0.5051$, $CV_2 = 0.5$, and $\rho = CV_2/CV_1 = 0.99$, so that $f(t)$ is symmetric around the turning point at $x = 0.2475$. The high value of ρ is typical for the correlation between the coefficients of polynomial regression functions. The figure shows that the exact distribution has a slimmer peak and wider tails than the normal approximation through the delta method, so that the approximation underestimates the true range of $\hat{\tau}$. This should not be surprising, because the exact distribution is related to the leptokurtic Cauchy distribution.

Figure 2 shows the two density functions for $\mu_1 = 15$, with all other parameter values being the same as before (thick lines, left-hand scale). These values yield a turning point at $x = 0.3750$ and $\rho = 0.99 \neq CV_2/CV_1 = 1.5$, with $CV_1 = 0.33$ and $CV_2 = 0.5$. Although the change is fairly minor, the exact density function $f(t)$ is not symmetric any more (it even has

[5]The density function is derived in Fieller (1932, p. 432) and (Hinkley, 1969, p. 636). For convenience, we report it in Appendix 1.

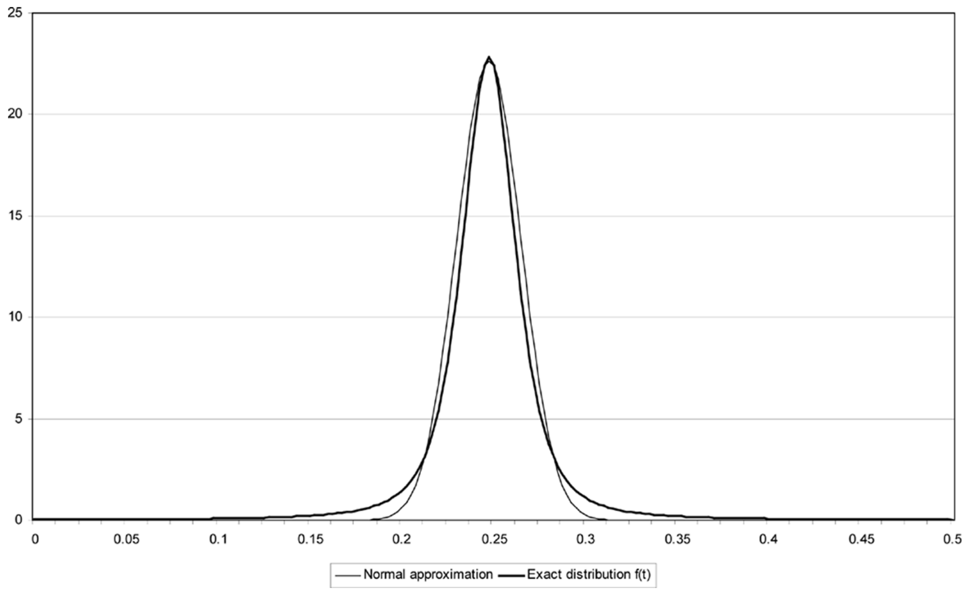

FIGURE 1 Comparing the normal approximation of $f(t)$ with the exact distribution if $\rho = CV_2/CV_1 = 0.99$ and $CV_2 = 0.5$.

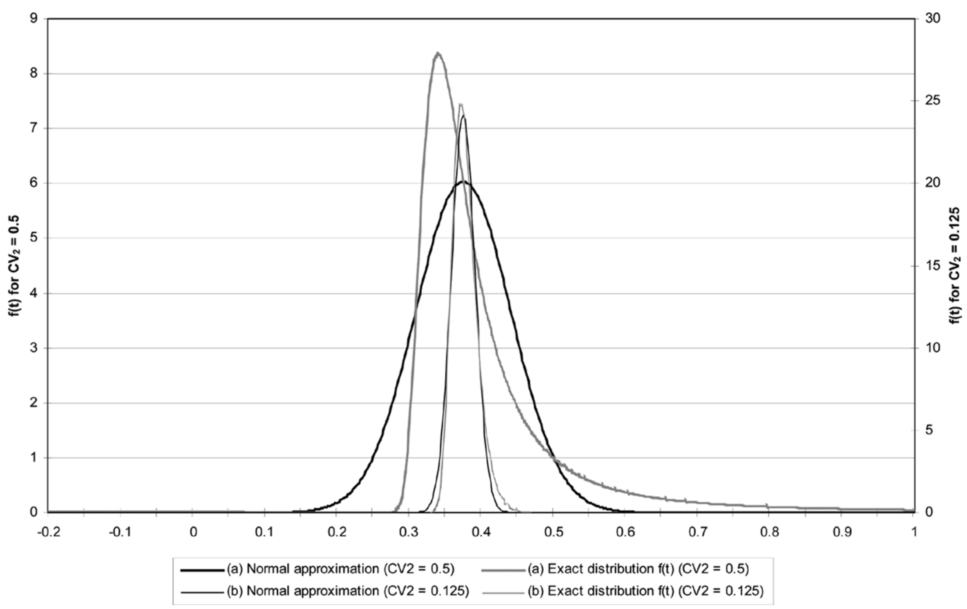

FIGURE 2 Comparing the normal approximation of $f(t)$ with the exact distribution if $\rho = 0.99$, $CV_2/CV_1 = 1.5$, and (a) $CV_2 = 0.5$ (thick lines, left-hand scale), (b) $CV_2 = 0.125$ (thin lines, right-hand scale).

a second mode at −0.0955), and the normal approximation is clearly ill suited to describe the properties of $\hat{\tau}$.[6]

Figure 2 also shows the two density functions for a smaller value of CV_2, where $\mu_1 = 15$, $\mu_2 = -20$, $\sigma_1 = 1.25$, $\sigma_2 = 2.5$, and $\sigma_{12}^2 = -3.09375$, which yields $\rho = 0.99 \neq CV_2/CV_1 = 1.5$ with $CV_1 = 0.0833$ and $CV_2 = 0.125$ (thin lines, right-hand scale). Although the correlation coefficient ρ and the ratio of the coefficients of variation remain unchanged compared to the case of $CV_2 = 0.5$, the lower value of CV_2 causes the distribution of $\hat{\tau}$ to approach the normal distribution. If CV_2 decreases while CV_1 remains constant (instead of CV_2/CV_1 remaining constant), then $f(t)$ approaches the normal density function even faster.[7] Whether it is appropriate to approximate the true distribution with the delta method depends on the values of all parameters of the joint distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$, and there are no general threshold values of CV_2 and CV_2/CV_1 below which the delta method is sufficiently precise. The adequacy of the delta method must therefore be assessed on a case-by-case basis. Because it is straightforward to evaluate the exact distribution of $\hat{\tau}$, we conclude that there is little reason to use the delta method to assess the precision of turning point estimate in quadratic regression models.

Method 3. Finite sample estimate of the exact distribution of $\hat{\tau}$ based on MCMC output.

MCMC methods are iterative techniques that use Markov chains to perform Monte Carlo integrations of integrals of interest. In a Bayesian context, they are frequently used to obtain estimates of the posterior distributions of unknown model parameters. MCMC methods generate large numbers of samples from posterior distributions (for example, the posterior distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$) that can be used to draw inferences about the properties of these distributions. The samples can also be used to estimate posterior distributions of functions of correlated model parameters, for example, the posterior distribution of $\hat{\tau} = \hat{\beta}_1/(-2\hat{\beta}_2)$, which provides information about the quantiles of $\hat{\tau}$.[8]

The setup of MCMC methods requires assumptions about the distribution of the data and the prior distributions of the parameters. If the prior distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ are normal, then MCMC provides

[6]The second mode is barely visible in Figure 2 because $f(-0.0955) = 0.01996$.

[7]See Marsaglia (1965), Shanmugalingam (1982), and Wolgrom (2001) for additional figures that illustrate how the distribution of the ratio of two correlated normally distributed random variables may deviate from normality and symmetry.

[8]The literature on MCMC methods has grown tremendously since the early 1990s and we refer the reader to this literature for methodological details (see, for example, Gilks et al., 1996). MCMC has also become part of standard econometrics textbooks; see, for example, Greene (2002, pp. 444–447).

a numerical finite sample approximation of the exact distribution of $\hat{\tau}$ (see Method 2) and there is no intrinsic reason to prefer MCMC over Method 2. If the distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ are not normal, then Method 2 is not applicable while application of MCMC is often still straightforward. In addition, MCMC works well for the assessment of turning points of higher order polynomial functions.

2.2. Assessing Turning Points of Higher Order Polynomial Regression Functions

Unlike Method 2, Methods 1 and 3 can be used to assess the precision of turning points of higher order polynomial equations. If, for example, the relationship between y and x is estimated with the cubic equation $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$, then the standard estimators of the two turning points are

$$\hat{\tau}^c_{1,2} = \frac{-\hat{\beta}_2 \pm \sqrt{\hat{\beta}_2^2 - 3\hat{\beta}_1\hat{\beta}_3}}{3\hat{\beta}_3}, \tag{4}$$

where $\hat{\beta}_3$ is the estimator of β_3. Method 2 is not applicable if the numerator is not normally distributed, which is not the case if $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are trivariate normal.[9] It is straightforward, however, to determine the first order Taylor series approximations of $\hat{\tau}^c_1$ and $\hat{\tau}^c_2$ around the parameters β_1, β_2, and β_3 (see Appendix 2), and to assess the precision of the estimate with the delta method. It is equally straightforward to use MCMC output to obtain numerical approximations of the exact distributions of $\hat{\tau}^c_1$ and $\hat{\tau}^c_2$. The higher the order of the polynomial equation, the more complex is the algebraic Taylor series approximation of the turning point. Application of MCMC, on the other hand, remains simple.

3. COMPARING THE THREE METHODS WITH DATA FROM THE EKC LITERATURE

We compare the three methods by assessing the precision of turning point estimates obtained under the EKC hypothesis. Under the EKC hypothesis, pollution increases with income as long as income is relatively low and decreases once income has crossed a certain threshold. Most

[9]Hamblen (1956) derives the distribution of the two real roots of $\hat{\beta}_1 + 2\hat{\beta}_2 x_i + x_i^2 = 0$ (that is, $\hat{\beta}_3$ is a constant equal to 1/3) for bivariate normal $\hat{\beta}_1$ and $\hat{\beta}_2$. This distribution is not normal because of the need to account for the probability that the roots are complex. There does not seem to be a simple expression of the distributions of the real roots of $\hat{\beta}_1 + 2\hat{\beta}_2 x_i + 3\hat{\beta}_3 x_i^2 = 0$ if $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are trivariate normal, although Vom Scheidt and Bharucha-Reid (1983) derive a normal approximation (see also Bharucha-Reid and Sambandham, 1986).

studies have tested the EKC hypothesis with multicountry panel data sets, assuming that there is a single global relationship between income and pollution (for example, Grossman and Krueger, 1995; Harbaugh et al., 2002). Because there is evidence that this relationship might differ over time and across countries, some recent studies have analyzed a microlevel version of the EKC hypothesis, using panel and cross-sectional data of a single country (for example, Kahn, 1998; Khanna, 2002; List and Gallet, 1999; Millimet et al., 2003; Plassmann and Khanna, 2006). We apply our three methods to turning point estimates obtained from multicountry as well as single-country data.

Our first data set is the GEMS (Global Environment Monitoring System) panel data set used by Harbaugh et al. (2002) that contains information on ambient sulfur dioxide (SO_2) concentrations in 102 cities in 45 countries between 1971 and 1992. This is an updated version of the data set used by Grossman and Krueger (1995) and the cornerstone of the EKC literature. Our second data set is the one used by Plassmann and Khanna (2006) that contains information about the number of days during which the 1990 ambient coarse particulate matter (PM_{10}) concentrations exceeded their National Ambient Air Quality Standard (NAAQS) at 704 locations in the United States. We refer the reader to the original articles for summary statistics of both data sets.

We estimate quadratic and cubic relationships between pollution and income under the specifications reported in Harbaugh et al. (2002) and Plassmann and Khanna (2006), using ordinary least squares (for data set 1) and an MCMC method, the Gibbs sampler, for both data sets. For the Gibbs sampler analyses, we assume that the distribution of the SO_2 concentrations (data set 1) is normal and that the distribution of the numbers of days during which the concentrations of PM_{10} exceed the NAAQS (data set 2) is Poisson; we further assume that the prior distribution of the coefficients is multivariate-normal in both analyses.[10] We summarize the setup of our Gibbs sampler analyses in Appendix 3. We obtained the Gibbs sampler estimates from 10,000 runs after a burn-in of 10,000 runs for data set 1, and from 9,000 runs after a burn-in of 1,000 runs for data set 2.

It is important to emphasize that any inference of the precision of a turning point estimate requires an adequately specified statistical

[10]The concentrations of PM_{10} remained at or below the NAAQS at 632 locations and exceeded the NAAQS for a maximum of 8 days at only one location, so a count analysis is appropriate. Combining the Poisson regression model with normally distributed coefficients yields the Poisson-lognormal model. The closed form solution of the Poisson-lognormal distribution is unknown, which makes maximum likelihood analysis of this model cumbersome, but application of MCMC is straightforward. We describe the model selection process and report the results of alternative analysis of these data with normal, Poisson, and negative binomial models in the supplement to Plassmann and Khanna (2006).

model.[11] If the true relationship is of either lower or higher order than the one estimated, then any inference based on the misspecified model is likely to be invalid. For example, consider a linear regression model that is erroneously estimated with a quadratic term. The turning point $\tau = \beta_1/(-2\beta_2)$ is not defined because the true value of β_2 is zero. Although it is of course possible to calculate a value $\hat{\tau}^0 = \hat{\beta}_1^0/(-2\hat{\beta}_2^0)$ as long as $\hat{\beta}_2^0 \neq 0$, this value clearly does not represent an estimate of the (nonexisting) turning point, and inference based on any of the three methods discussed above is invalid. Similarly, if the true relationship is quadratic but a cubic regression equation is estimated, then the turning points in Equation (4) do not exist because the true value of β_3 is zero, and neither the delta method nor MCMC lead to valid inference.

Similar problems can arise if the true relationship is of higher order than the one estimated. For example, sometimes a straight line approximates a cubic relationship better than a quadratic function does. Estimating a cubic relationship with a quadratic regression function will then either lead to an estimate of $\hat{\beta}_2$ that is (close to) zero or to a nonzero estimate with a large standard error. In both cases, one would conclude that there is no turning point although the true relationship is nonlinear and may contain up to two turning points within the range of the data.

There are various tools to guide the task of choosing the appropriate polynomial. A simple t-test can help to choose between linear and quadratic specifications, while an F-test (or a χ^2-test for nonnormal disturbances and large samples) helps to distinguish between higher order polynomials. Auxiliary regressions of residuals against a function with higher order polynomial term(s) may indicate remaining systematic patterns. Comparison of the graphs implied by the coefficient estimates of different polynomial functions can provide insights into the degree of additional curvature gained at the expense of lost degrees of freedom. As it is always the case in analyses of observational data, there is no universal method to determine the appropriate functional relationship, and a combination of these methods is most likely to be successful.

3.1. Data Set 1: Harbaugh et al. (2002)

Harbaugh et al. (2002) analyze the EKC hypothesis using income and lagged income as covariates (in addition to several nonincome variables and fixed effects for 267 measurement sites), and they determine the

[11]We thank an anonymous referee for drawing our attention to this issue and to the examples.

TABLE 1 Comparing least squares and Gibbs sampler estimates for data set 1

	Quadratic specification		Cubic specification	
	Least squares (1)	Gibbs sampler (2)	Least squares (3)	Gibbs sampler (4)
		Coefficient estimates		
GDP	−0.1165	−0.1196	−0.3025	−0.3070
	(0.0518)	(0.0519)	(0.1286)	(0.1297)
GDP	0.0054	0.0055	0.0278	0.0280
	(0.0021)	(0.0021)	(0.0136)	(0.0136)
GDP			−0.0007	−0.0007
			(0.0004)	(0.0004)
(Lagged GDP)	0.0091	0.0092	0.3987	0.3868
	(0.0524)	(0.0525)	(0.1270)	(0.1269)
(Lagged GDP)	−0.0018	−0.0019	−0.0470	−0.0462
	(0.0024)	(0.0024)	(0.0144)	(0.0144)
(Lagged GDP)			0.0015	0.0015
			(0.0005)	(0.0005)
		Aggregate coefficients		
(= GDP + Lagged GDP)	−0.1074	−0.1103	0.0962	0.0799
	(0.0420)	(0.0410)	(0.0987)	(0.1007)
(= GDP^2 + (Lagged GDP))	0.0036	0.0036	−0.0192	−0.0182
	(0.0017)	(0.0016)	(0.0087)	(0.0087)
(= GDP^3 + (Lagged GDP))			0.0008	0.0008
			(0.0002)	(0.0003)
		Turning points (at per capita GDP in $1,000)		
Turning point 1	14.9654	15.3435	12.7507	12.9001
Turning point 2			3.1177	2.6516
		R and test statistics		
Adjusted R	0.2149		0.2197	
F-test statistic and	4.88		6.31	
p value	0.0077		0.0019	

Estimates in parentheses are the standard errors of the coefficients. Estimates of the nonincome covariates are not shown.

The coefficient estimates in column 3 correspond to the estimates in Table 4, column 7 in Harbaugh et al. (2002).

The *F*-test statistic for the quadratic model is for the null hypothesis that the coefficients of GDP and (lagged GDP) are zero, and the F-test statistic for the cubic model is for the null hypothesis that the coefficients of GDP and (lagged GDP) are zero.

coefficients that enter the calculation of the turning point as sums of the income and lagged income coefficients. Columns 1 and 2 of Table 1 show the least squares and Gibbs sampler estimates of the four income coefficients of the quadratic specification, and columns 3 and 4 show the estimates for the six income coefficients of the cubic specification (the estimates in column 3 are identical to the estimates reported in Harbaugh et al., 2002, Table 4, column 7: fixed effects model with the

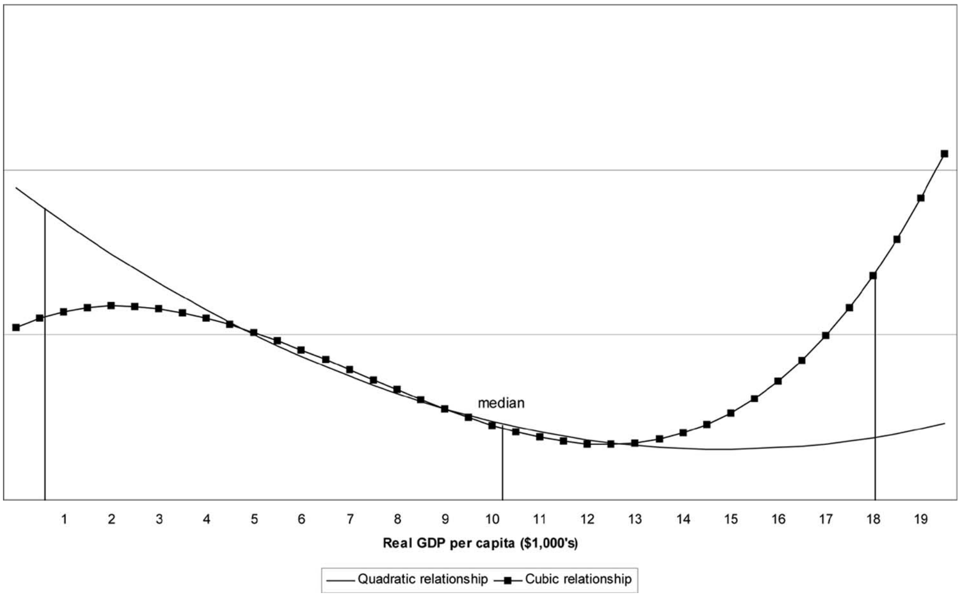

FIGURE 3 Income-pollution relationships implied by the quadratic and cubic Gibbs sampler analyses of data set 1.

left-hand side pollution variable measured in logs).[12] Although the least squares and Gibbs sampler coefficient estimates are very similar, the small differences lead to notable differences in the turning point estimates that we calculated from the aggregate coefficients. The aggregation of coefficients therefore introduces additional uncertainty into the estimate of the turning point. Nevertheless, we use the aggregate coefficients in the following analyses to make our results comparable to those in Harbaugh et al. (2002).

The last row in Table 1 reports the F-test statistics and the corresponding *p*-values for the hypotheses that the coefficients of the two additional terms (GDP^2 and $(\text{lagged GDP})^2$ for the quadratic specification and GDP^3 and $(\text{lagged GDP})^3$ for the cubic specification) are both zero. Both tests indicate that the relationship between income and pollution is nonlinear, while the second test suggests that the cubic specification is more appropriate. Figure 3 graphs the relationships implied by the quadratic and cubic Gibbs sampler analyses; the three vertical lines

[12]Harbaugh et al. (2002, p. 546) state "there is no *a priori* reason to prefer any one of the specifications in [their] Table 4 to the others." We use column 7 because (1) it yields the best fit among the seven specifications they report, given the number of covariates included, and (2) the sampling distributions of two of its three turning point estimators are highly skewed and illustrate our point very well.

indicate the range and the median of the GDP data.[13] Both functional forms track the negative relationship to the left and at the median of the data but differ notably in their description of the relationship at the upper and lower end of the data. We also estimated a quartic model (not shown) whose graph is fairly similar to that of the cubic specification and whose F-test statistic of 4.63 indicates that GDP^4 and $(\text{lagged GDP})^4$ are barely significant at the 10 percent level. Overall we conclude that the cubic specification fits the data best, but that the quadratic specification has sufficient curvature (the estimates of β_2 are statistically significant at the 10% level) to make an examination of its turning point worthwhile.

Table 2 shows the OLS and Gibbs sampler estimates of the quadratic and cubic specifications.[14] The small differences in the least squares and Gibbs sampler point estimates of the coefficients and the elements of the covariance matrix lead to noticeable differences in the estimates of the density functions of $\hat{\tau}$. We note five interesting results.

First, the point estimates of the turning points are very similar across the three methods. For Method 1, we determined the point estimates for the quadratic and cubic specifications as $\hat{\tau}^0 = \hat{\beta}_1^0/(-2\hat{\beta}_2^0)$ and from Equation (4), respectively. For Methods 2 and 3, we estimated the location of the turning point as the median of the estimated density functions (recall that the moments of $\hat{\tau}$ do not exist if $\hat{\beta}_1$ and $\hat{\beta}_2$ are normally distributed).

Second, Method 1 yields slightly different estimates of the empirical standard errors of the turning points for the OLS and Gibbs sampler estimates (especially for the cubic specification), which result from the differences in the estimates of the covariance matrices. Method 3 yields a very large empirical standard error for the turning point of the quadratic specification (column 2) because the Gibbs sampler sampled several values of $\hat{\beta}_2$ close to zero. If $\hat{\beta}_2$ is normally distributed, then its support includes 0, and values close to zero are not impossible. The fact that the Gibbs sampler output from 10,000 runs contained 526 values of $\hat{\beta}_2$ between -0.001 and 0.001 suggests that small values of the turning point are not as uncommon as the Method 1 standard errors imply (note that least squares and the Gibbs sampler yield virtually identical estimates of the standard

[13]Because we are interested in a comparison of shapes, we shifted the cubic function vertically so that the two functions have identical means over the per capita GDP range $[0.77, 18.10]$ and ignored the presence of nonincome covariates and regional fixed effects in the equations. We therefore do not show units on the vertical axis. An alternative is to evaluate the equations at the medians and means of the nonincome covariates. Because this practice ignores the impact of the fixed effects, it overstates the differences between the curves. We think that, in the presence of such fixed effects, our method is more appropriate to assess the differences between functional forms.

[14]We report the Gibbs sampler estimates for the quadratic specification to illustrate that the Methods 2 and 3 estimates of the quantiles of $\hat{\tau}$ (column 2) are very similar.

TABLE 2 Coefficient and turning point estimates of data set 1

	Quadratic specification					Cubic specification		
	Least squares estimates (1)		Gibbs sampler estimates (2)			Least squares estimates (3)	Gibbs sampler estimates (4)	
GDP + lagged GDP ($\hat{\beta}_1^0$)	−0.1074		−0.1103			0.0962	0.0799	
	(0.0420)		(0.0410)			(0.0987)	(0.1007)	
GDP^2 + (lagged GDP)^2 ($\hat{\beta}_2^0$)	0.0036		0.0036			−0.0192	−0.0182	
	(0.0017)		(0.0016)			(0.0087)	(0.0087)	
GDP^3 + (lagged GDP)^3 ($\hat{\beta}_3^0$)						0.0008	0.0008	
						(0.0002)	(0.0003)	
Cov ($\hat{\beta}_1, \hat{\beta}_2$)	−0.00006		−0.00006			−8.1*E*-04	−8.2E-04	
Cov ($\hat{\beta}_1, \hat{\beta}_3$)						2.2*E*-04	2.2E-05	
Cov ($\hat{\beta}_2, \hat{\beta}_3$)						−2.1*E*-06	−2.2E-06	
Adjusted R^2	0.2149					0.2197		
Turning point 1 (trough)	*Method* 1	*Method* 2	*Method* 1	*Method* 2	*Method* 3	*Method* 1	*Method* 1	*Method* 3
Turning point estimate	14.9654	14.8436	15.3435	15.2325	15.2467	12.7507	12.9001	12.9162
Finite sample standard error	(3.4703)		(3.4123)		(128.32)	(1.1377)	(0.9233)	(0.8302)
Quantiles(%)								
2.5	8.1637	7.6681	8.6553	8.6900	8.6748	10.5208	11.0903	10.7918
5.0	9.2568	9.6782	9.7301	10.4125	10.0326	10.8791	11.3812	11.2292
10.0	10.5182	11.0093	10.9705	11.5755	11.4349	11.2927	11.7168	11.6461
50.0	14.9654	14.8436	15.3435	15.2325	15.2467	12.7507	12.9001	12.9162
90.0	19.4125	22.1921	19.7164	22.5625	22.9297	14.2086	14.0833	14.1241
95.0	20.6740	27.4556	20.9568	27.8074	28.0242	14.6222	14.4189	14.4995
97.5	21.7671	36.2345	22.0317	36.5134	36.1533	14.9805	14.7098	14.9087

Turning point 2 (peak)			
Turning point estimate	3.1177	2.6516	2.794
Finite sample standard error	(2.4001)	(2.6409)	(8.3248)
Quantiles(%)			
2.5	−1.5866	−2.5246	−10.0022
5.0	−0.8305	−1.6927	−5.8312
10.0	0.0419	−0.7327	−2.6095
50.0	3.1177	2.6516	2.7937
90.0	6.1934	6.0360	5.2623
95.0	7.0658	6.9960	5.8992
97.5	7.8219	7.8279	6.4402

Estimates in parentheses are the standard errors of the coefficients. Coefficient estimates of nonincome variables are not shown. The turning point estimates of Methods 2 and 3 are the medians of the density functions. Note that we measure income in 1,000$. For both specifications, we obtained the Gibbs sampler estimates from 10,000 runs after a burn-in of 10,000 runs. We determined the quantiles for Method 3 directly from the Gibbs sampler output, not from the smoothed histograms shown in Figures 4 and 5.

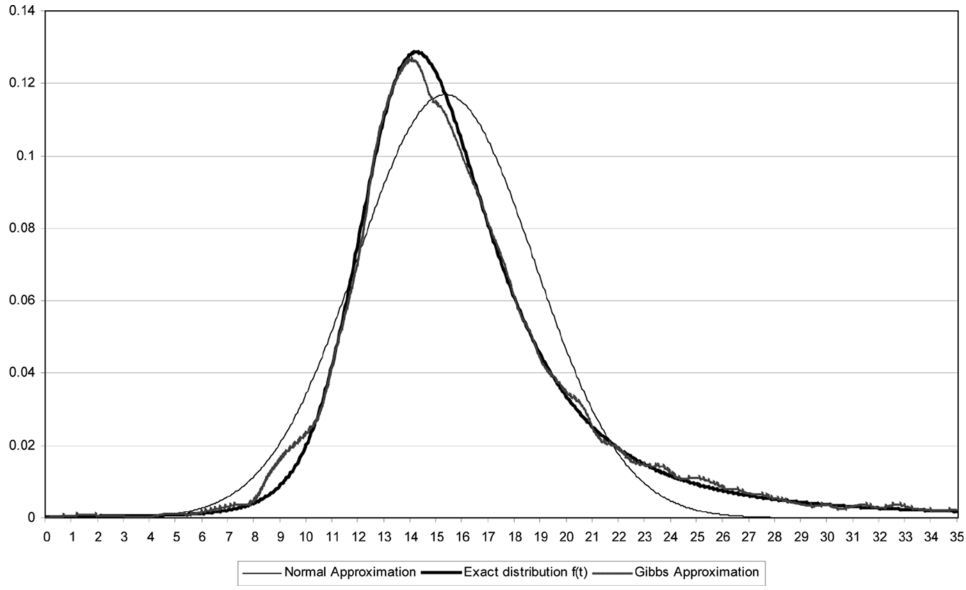

FIGURE 4 Density functions of the turning point estimators for the quadratic model of data set 1.

error of $\hat{\beta}_2$). We take this large standard error of the turning point of the quadratic specification as a further indication that the quadratic model is misspecified and that the cubic specification is more appropriate for these data.

Third, for the quadratic specification, the quantiles estimated with Method 1 are very different from those estimated with Methods 2 and 3.[15] This is not surprising because the ratio CV_2/CV_1 in column 1 is $-0.4665/-0.3911 = 1.1928$ ($-0.4560/-0.3717 = 1.2266$ in column 2), while the correlation coefficient ρ in column 1 is 0.8682 (0.8750 in column 2). Figure 4 shows the graphs of the three density functions in column 2; it indicates that the true distribution of the turning point is not symmetric and that Method 1 greatly underestimates the right tail of the distribution.

Fourth, the two sets of quantiles in column 2 that we estimated with Methods 2 and 3 are fairly similar.[16] They vary from 8.67 to 8.69 at the

[15]The quantiles estimated with Method 2 differ slightly across columns 1 and 2 because we aggregated the coefficients of GDP and its lags, which amplifies the small differences in the estimates of the covariance matrix of $\hat{\beta}_1$ and $\hat{\beta}_2$.

[16]We determined the quantiles for Method 3 directly from the Gibbs sampler output. These quantiles would have been even more similar to those obtained from Method 2 if we had determined them from the smoothed histograms that we show in Figures 4 to 5. (We smoothed the empirical density function that we obtained from the Gibbs sampler output using a normal kernel and chose the window width (0.2) so that the Gibbs sampler density in these figures had the most overlap with the exact density.) In Figure 2 we report the quantiles of the nonsmoothed distribution to avoid any bias caused by the arbitrary choice of window width.

2.5th percentile and from 36.15 to 36.51 at the 97.5th percentile. The largest income value in data set 1 is 18.095, which is less than half of the upper range of both 95 percent confidence intervals. (We follow Harbaugh et al., 2002 and measure income in 1,000$.) The quadratic specification therefore provides little evidence that the relationship between the ambient concentrations of SO_2 and GDP per capita is nonmonotonic. Rather, the estimates suggest that the concentrations of SO_2 decrease monotonically at a decreasing rate as income increases.

Fifth, for the cubic specification, Methods 1 and 3 imply very similar distributions for the first and very different distributions for the second turning point estimator. Figure 5 shows the graphs of the four density functions in column 4. A comparison of these density functions suggests that Method 1 is appropriate for turning point 1, the trough (although the Gibbs sampler indicates that the normal distribution somewhat underestimates the tails), and that it greatly underestimates the left tail of the distribution of turning point 2, the peak. It is therefore not surprising that the two methods yield very similar estimates of the empirical standard error of the first turning point, but that Method 1 greatly underestimates the empirical standard error of the second turning point.

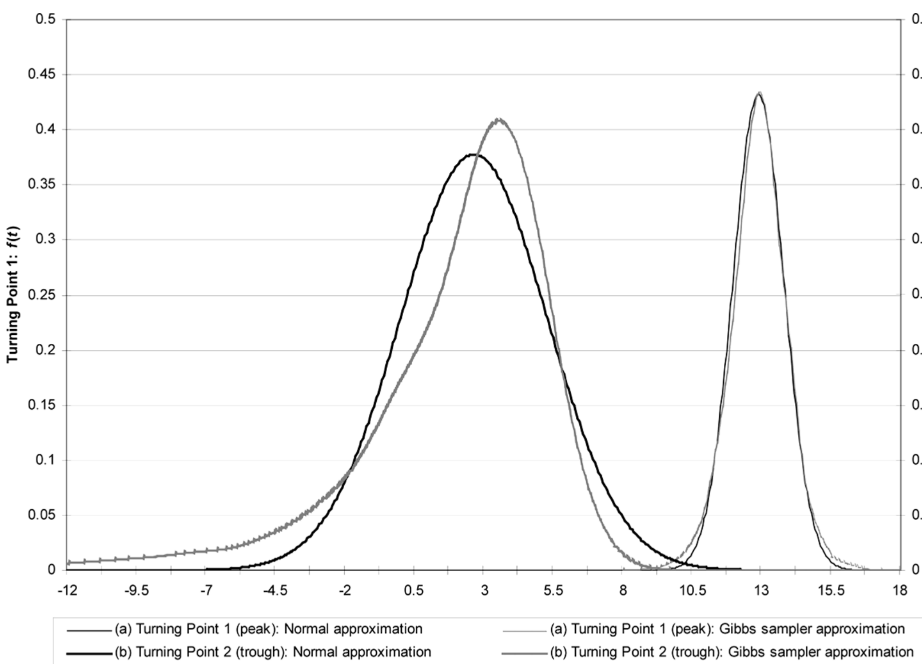

FIGURE 5 Density functions of the turning point estimators of the cubic model of data set 1. (a) First turning point (thin lines, left-hand scale), and (b) second turning point (thick lines, right-hand scale).

In light of the results from the quadratic model, we take the small confidence interval around the trough as a sign that the relationship between concentrations of SO_2 and GDP is negative until per capita GDP reaches about \$13,000 and that the concentrations of SO_2 remain fairly constant once per capita GDP exceeds this level. To determine whether the ultimate increase in pollution is more than an artifact of the cubic specification, it would be necessary to examine additional functional forms.

Harbaugh et al. (2002) use Method 1 to derive standard errors for their turning point estimates but do not assess the precision of their estimates. Instead, they analyze several different model specifications to determine whether their turning point estimates are robust. Our analysis of the precision of the turning point estimates implies that the 95% confidence intervals span almost the entire sample income range, especially in the case of the quadratic specification. Like Harbaugh et al. (2002), we conclude that the quadratic and cubic models provide little support for an inverted-U shaped relationship between SO_2 concentrations and national income.

3.2. Data Set 2: Plassmann and Khanna (2006)

Table 3 contains the results of quadratic and cubic Poisson-lognormal analyses of the number of days during which the concentrations of PM_{10} exceeded their NAAQS. Recall that we undertake the Gibbs sampler analysis under the assumption that the coefficients are multivariate normally distributed so that Method 2 is applicable to our analysis.

The estimate of β_2 in the quadratic specification (column 1) is significantly different from zero at the 99% level, which suggests that the relationship is nonlinear. The value of the χ^2-test statistic for $\beta_3 = 0$ in the cubic specification (column 2) is 4.38, which exceeds the critical value at 95% but not at 99% significance. Figure 6 compares the graphs of the two relationships and indicates that the cubic specification has greater curvature over the data range. We conclude that there is some evidence that the true relationship may be cubic, but that the quadratic specification is a sufficiently close approximation to permit a meaningful analysis of its turning point.

Our main conclusions regarding the three methods are identical to those in the previous section, and we summarize them briefly. First, the quantiles obtained by Methods 2 and 3 under the quadratic specification are very similar. Second, Methods 2 and 3 suggest that the distributions of all turning points are skewed to the right.[17] Third, Method 1 overestimates both tails of the distributions of the two peaks and greatly underestimates

[17]For the quadratic specification, the ratio of the coefficients of variation is 0.9012 and the correlation coefficient is 0.9989.

TABLE 3 Coefficient and turning point estimates of data set 2

		Quadratic specification (1)			Cubic specification (2)
ln(income) ($\hat{\beta}_1^0$)		10.5741			−5.6151
		(4.0140)			(10.3541)
(ln(income))^2 ($\hat{\beta}_2^0$)		−0.5986			1.2649
		(0.2048)			(1.3651)
(ln(income))^3 ($\hat{\beta}_3^0$)					−0.0730
					(0.0600)
Cov ($\hat{\beta}_1, \hat{\beta}_2$)		−0.8211			−13.3188
Cov ($\hat{\beta}_1, \hat{\beta}_3$)					0.5318
Cov ($\hat{\beta}_2, \hat{\beta}_3$)					−0.0802
Pseudo R^2		0.2185			0.2238
Turning point 1 (peak)	*Method* 1	*Method* 2	*Method* 3	*Method* 1	*Method* 3
Turning point estimate	8.8320	8.8336	8.8316	8.5621	8.6935
Finite sample standard error	(0.3642)		(2.6583)	(0.9548)	(9.9295)
Quantiles(%)					
2.5	8.1202	6.8950	7.0291	6.6907	4.0089
5.0	8.2346	7.5671	7.6801	6.9915	5.7066
10.0	8.3667	8.0475	8.0722	7.3382	6.9856
50.0	8.8320	8.8336	8.8316	8.5621	8.6935
90.0	9.2981	9.1733	9.1656	9.7856	9.2346
95.0	9.4300	9.2411	9.2367	10.1323	9.3305
97.5	9.5446	9.2947	9.3018	10.4335	9.3933
Turning point 2 (trough)					
Turning point estimate				2.9960	3.2061
Finite sample standard error				(3.6598)	(172.5259)
Quantiles (%)					
2.5				−4.1735	−34.7389
5.0				−3.0217	−16.9505
10.0				−1.6929	−5.9332
50.0				2.9960	3.2074
90.0				7.6849	13.5923
95.0				9.0137	23.1000
97.5				10.1655	42.0712

Estimates in parentheses are the standard errors of the coefficients. Coefficient estimates of nonincome variables are not shown. The turning point estimates of Methods 2 and 3 are the medians of the density functions. Note that we measure income in logs of 1,000$. For both specifications, we obtained the Gibbs sampler estimates from 9,000 runs after a burn-in of 1,000 runs. We determined the quantiles for Method 3 directly from the Gibbs sampler output, not from the smoothed histograms shown in Figures 7 and 8.

the variance of the distribution of the trough of the cubic model. We show the distributions of the two peaks implied by the three methods in Figures 7 and 8.

Fourth, the peaks of both specifications are very close to the lower end of the data range. The further away a turning point is from the bulk of the data, the more likely it is an artifact of the functional relationship. By overestimating the upper tail of the distributions, Method 1 suggests

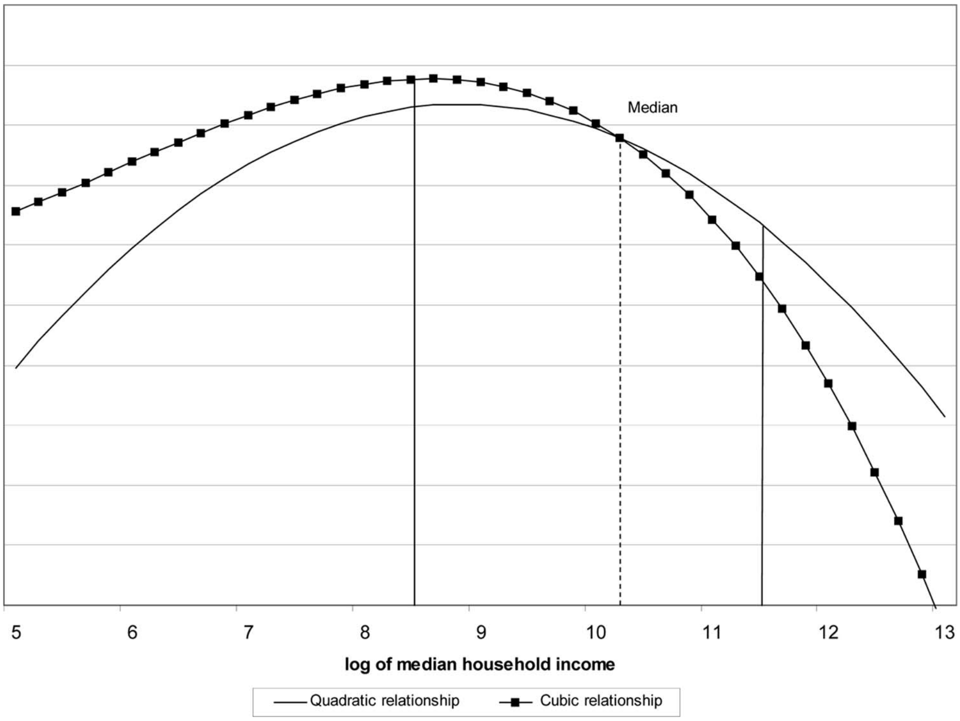

FIGURE 6 Income-pollution relationships implied by the quadratic and cubic analyses of data set 2.

that the turning point may nevertheless be close to the median of the income data (10.25), and Method 1's 95% confidence interval for the cubic specification even includes this median. On the basis of Method 1, there is insufficient evidence *against* a turning point, and it is possible to argue that the true relationship could be nonmonotonic. However, the upper limits of the 95% confidence intervals of Methods 2 and 3 for both specfications are between 9.29 and 9.39, while only 6% of the observations have a log income below 9.4. Because over 90% of the data are to the right of the turning point and are used to estimate the curvature of the decreasing part of the relationship (the right leg of the EKC), it is fairly unlikely that the true relationship is nonmonotonic over the range of our data. We conclude that the relationship most likely decreases monotonically at an increasing rate.

4. CONCLUSION

Whether the widely used normal approximation (delta method) is appropriate for assessing the precision of turning point estimates depends on the characteristics of the sampling distribution of the turning point

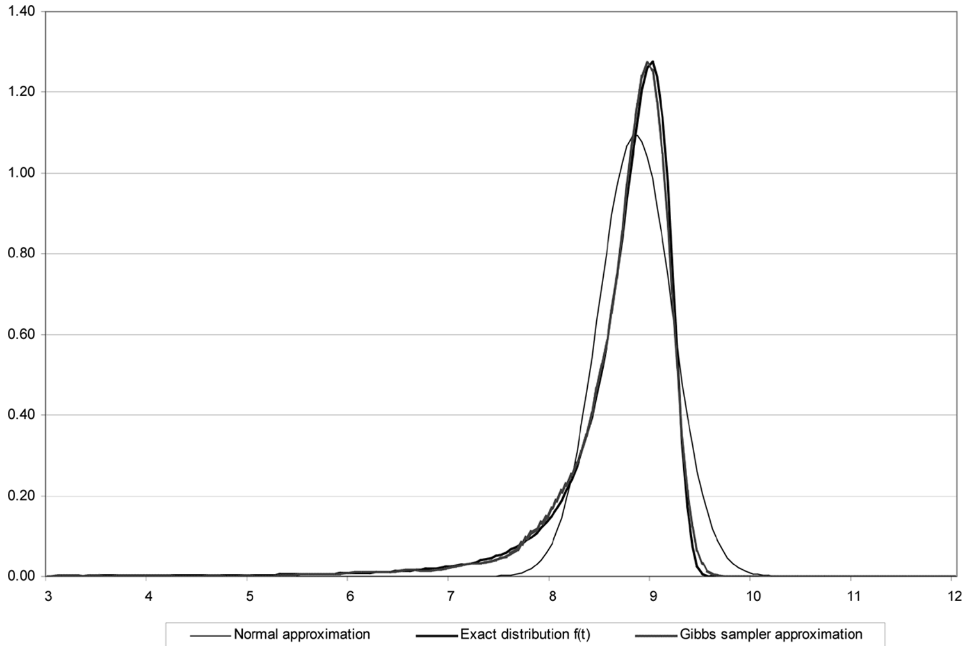

FIGURE 7 Density functions of the turning point estimators of the quadratic model of data set 2.

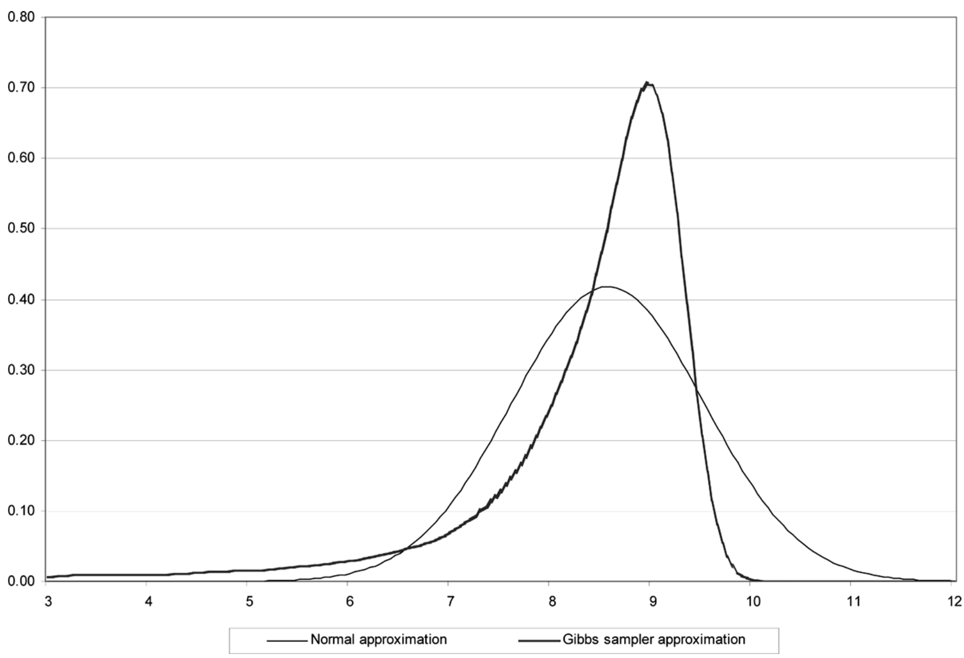

FIGURE 8 Density functions of the turning point estimator of the peak in the cubic model of data set 2.

estimator. Five of the six turning point estimators that we examine have asymmetric sampling distributions, and the normal approximation leads to misleading confidence intervals in these cases. The confidence intervals for the turning points in the Harbaugh et al. data set are even wider than those obtained with the normal approximation and do not support the EKC hypothesis for SO_2 concentrations and GDP per capita. For the Plassmann and Khanna data set, the delta method implies a symmetric 95% confidence interval for the peak that includes the median of the data under the cubic specification, while the 95% confidence intervals of the asymmetric distributions include less than 10% of the data. The delta method therefore leads to misleading inference.

Comparing the coefficients of variation and the correlation coefficient of the two components of the turning point estimator in quadratic regression functions provides some information about the symmetry of the sampling distribution and whether the delta method might be appropriate. However, modern computer packages make it straightforward to determine the exact sampling distribution of the quadratic turning point estimator, and the Gibbs sampler makes it equally straightforward to determine a finite sample approximation of the sampling distribution of the turning point estimators of higher order polynomial regression functions. Given that the delta method may lead to misleading results, we conclude that it is prudent to use either the exact sampling distribution or the Gibbs sampler approximation for statistical inference of the turning points of polynomial regression functions.

APPENDIX 1. THE DENSITY FUNCTION OF $\hat{\tau}$

The probability density function (pdf) of $\hat{\tau}$ is given by Fieller (1932, p. 432) and Hinkley (1969, p. 636); we have adjusted the equation to show the pdf of the ratio of $\hat{\beta}_1$ and $-2\hat{\beta}_2$ rather than the pdf of the ratio of $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$f(t) = \frac{b(t)d(t)}{\sqrt{2\pi}\sigma_1\sigma_2 a^3(t)}\left[\Phi\left(\frac{b(t)}{\sqrt{1-\rho^2}a(t)}\right) - \Phi\left(-\frac{b(t)}{\sqrt{1-\rho^2}a(t)}\right)\right] + \frac{\sqrt{1-\rho^2}}{\pi\sigma_1\sigma_2 a^2(t)}\exp\left(-\frac{c}{2(1-\rho^2)}\right), \tag{A.1}$$

where

$$a(t) = \sqrt{\frac{t^2}{\sigma_1^2} - \frac{\rho t}{\sigma_1\sigma_2} + \frac{1}{4\sigma_2^2}}$$

$$b(t) = \frac{\mu_1 t}{\sigma_1^2} - \frac{\rho(\mu_1 - 2\mu_2 t)}{2\sigma_1\sigma_2} - \frac{\mu_2}{2\sigma_2^2}$$

$$c = \frac{\mu_1^2}{\sigma_1^2} + \frac{\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_2^2}{\sigma_2^2}$$

$$d(t) = \exp\left(\frac{b^2(t) - ca^2(t)}{2(1-\rho^2)a^2(t)}\right).$$

APPENDIX 2. FIRST-ORDER TAYLOR SERIES APPROXIMATION OF EQUATION (4)

$$\tau_{1,2}^{c,T} = e + b^*\hat{\beta}_1 + c^*\hat{\beta}_2 + d^*\hat{\beta}_3, \tag{A.2}$$

where

$$a = \frac{-\beta_2 \pm \left((\beta_2)^2 - 3^*\beta_1\beta_3\right)^{1/2}}{3\beta_3}$$

$$b = \frac{\pm\left((\beta_2)^2 - 3^*\beta_1\beta_3\right)^{-1/2}}{2}$$

$$c = \frac{\pm\left((\beta_2)^2 - 3^*\beta_1\beta_3\right)^{-1/2}\hat{\beta}_2}{3\beta_3} - \frac{1}{3\beta_3}$$

$$d = \frac{\beta_2 \pm \left((\beta_2)^2 - 3^*\beta_1\beta_3\right)^{1/2}}{3(\beta_3)^2} \pm \frac{\left((\beta_2)^2 - 3^*\beta_1\beta_3\right)^{-1/2}\beta_1}{2\beta_3}$$

$$e = a - b\beta_1 - c\beta_2 - d\beta_3.$$

If $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are unbiased estimators of β_1, β_2, and β_3, and trivariate normally distributed, then $\tau_1^{c,T}$ and $\tau_2^{c,T}$ are normally distributed with means $E[\tau_{1,2}^{c,T}] = e + bE[\hat{\beta}_1] + cE[\hat{\beta}_2] + dE[\hat{\beta}_3]$ and variances

$$Var(\tau_{1,2}^{c,T}) = b^2 Var(\hat{\beta}_1) + c^2 Var(\hat{\beta}_2) + d^2 Var(\hat{\beta}_3) + 2bcCov(\hat{\beta}_1\hat{\beta}_2) + 2bdCov(\hat{\beta}_1\hat{\beta}_3) + 2cdCov(\hat{\beta}_2\hat{\beta}_3).$$

APPENDIX 3. SETUP OF THE GIBBS SAMPLER ANALYSES

Implementation of the Gibbs sampler requires knowledge of the full-conditional distributions of all parameters of interest. Such full-conditional distributions are derived from the joint distribution of the data and the model parameters. In both models, one needs to determine the distributions of the parameters that describe the impact of the two (three) income covariates, β_1 and β_2 (and β_3 in the cubic model) and the other covariates. The data set of Harbaugh et al. contains four other covariates, and Plassmann and Khanna's data set contains nine other covariates.

1. Setup of the Gibbs Sampler for the Quadratic Normal Model

Denote the log of the ambient concentrations of particle matter by the vector $Y = (Y_1, \ldots, Y_n)$, where n is the number of observations, and the covariates by the $A \times n$ matrix $X = (X_1, \ldots, X_n)$, where each X_i is an $A \times 1$ vector. We assume that Y follows an n-variate normal distribution with density function $f_Y(y; \mu, \Sigma)$, where $\mu = (\mu_1, \ldots, \mu_n)$, is the mean vector with $\mu_i = X_i\beta$ and Σ is the covariance matrix. To simplify the example, we assume that the Y_i have identical variances σ^2 and are independent of each other, so that Σ is a diagonal matrix with σ^2 on the main diagonal. We follow the standard practice and assume that σ^{-2} follows a gamma distribution with density function $f_\sigma(\sigma^{-2}; a, b)$. With respect to the specification of the priors of the parameters of interest, we assume that β follows a 12-variate normal distribution with density function $f_\beta(\beta; c, \Omega)$ with mean vector c and covariance matrix Ω. These assumptions yield the posterior density function

$$f(Y, \beta, \sigma \mid X) = \prod_{i=1}^{n} f_{Y_i}(y_i \mid \mu_i, \sigma^2) \cdot f_\sigma(\sigma^{-2} \mid a, b) \cdot f_\beta(\beta \mid c, \Omega). \tag{A1}$$

Because we assume conditionally conjugate priors, the full-conditional distribution of β is n-variate normal with $\beta \mid \cdot \sim N[V^{-1}(\Omega^{-1}c + \sigma^{-2}X_iY_i), V^{-1})]$, where $V^{-1} = \sigma^{-2}X_i^TX_i + \Omega^{-1}$, and the full-conditional distribution of σ^{-2} is gamma with $\sigma^{-2}|\cdot \sim G\left[\frac{1}{2}(n+a), \frac{1}{2}\left(\sum_{i=1}^{n}(Y_i - X_i\beta)^T (Y_i - X_i\beta) + ab\right)\right]$. To close the model, we set $a = 0.1$, $b = 0.1$, $c = (0, \ldots, 0)$, $\Omega = diag(100)$, and $\sigma^2 = 100$. We used the program *WinBugs* (v1.3) (available at http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml) to generate the samples and for convergence diagnostics.

2. Setup of the Gibbs Sampler for the Poisson-Lognormal Model

Denote the number of days on which the ambient concentrations of particle matter exceeded the NAAQS by the vector $Z = (Z_1, \ldots, Z_n)$, where n is the number of observations, and the covariates by the $A \times n$ matrix $X = (X_1, \ldots, X_n)$, where each X_i is an $A \times 1$ vector. We assume that each Z_i follows a Poisson distribution with density function $f_Z(z_i; \mu_i)$, where $\mu_i = \exp(X_i\beta)$. We make the same assumptions about β as in the normal model. These assumptions yield the posterior density function

$$f(Z, \beta \mid X) = \prod_{i=1}^{n} f_{z_i}(z_i \mid \mu_i) \cdot f_\beta(\beta \mid c, \Omega). \tag{A2}$$

The density function of the full-conditional distribution of β is proportional to

$$g(\beta \mid Z, X) = \prod_{i=1}^{n} P(\exp(X_i\beta)) \cdot \phi_{11}(c, \Omega), \tag{A3}$$

where P is the density function of the univariate Poisson distribution and ϕ_{11} is the density function of the A-variate normal distribution. It is not possible to relate g to a known distribution for which standard sampling algorithms are available, and we used the Metropolis–Hastings method described in Chib et al. (1998) to sample β. We used the program *Bayesian Output Analysis* (available at http://www.public-health.uiowa.edu/boa) for convergence diagnostics.

ACKNOWLEDGMENTS

We thank Bill Harbaugh for generously sharing the data used in Harbaugh et al. (2002), and we thank David Maradan and two anonymous referees for helpful comments. All errors are ours.

REFERENCES

Abbot, M. G., Beach, C. M. (1993). Immigrant earnings differentials and birth-year effects for men in Canada: post-War-1972. *Canadian Journal of Economics* 26(3):505–524.

Barrett, S., Graddy, K. (2000). Freedom, growth, and the environment. *Environment and Development Economics* 5(4):433–456.

Bharucha-Reid, A. T., Sambandham, M. (1986). *Random Polynomials*. Orlando, Florida: Academic Press.

Chib, S., Greenberg, E., Winkelmann, R. (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics* 86:33–54.

Co, C. Y., List, J. A., Qui, L. D. (2004). Intellectual property rights, environmental regulations, and foreign direct investment. *Land Economics* 80(2):153–173.

Cole, M. A., Rayner, A. J., Bates, J. M. (1997). The environmental Kuznets curve: an empirical analysis. *Environment and Development Economics* 2:401–416.

Ferreira, S. (2004). Deforestation, property rights, and international trade. *Land Economics* 80(2):174–193, May.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika* 24(3–4):428–440.

Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

Greene, W. H. (2002). *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Greene, K., Nikolaev, O. (1999). Voter participation and the redistributive state. *Public Choice* 98(1–2):213–226.

Grossman, G. M., Krueger, A. H. (1995). Economic growth and the environment. *Quarterly Journal of Economics* 110(2):353–377.

Hamblen, J. W. (1956). Distributions of roots of quadratic equations With random coefficients. *Annals of Mathematical Statistics* 27(4):1136–1143.

Harbaugh, W. T., Levinson, A., Wilson, D. M. (2002). Reexamining the empirical evidence for an environmental Kuznets curve. *Review of Economics and Statistics* 84(3):541–551.

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika* 56(3):635–639.

Kahn, M. E. (1998). A household level environmental Kuznets curve. *Economics Letters* 59(3):269–273.

Khanna, N. (2002). The income elasticity of non-point source air pollutants: revisiting the environmental Kuznets curve. *Economics Letters* 77(3):387–392.

Kimhi, A., Rapaport, E. (2004). Time allocation between farm and off-farm activities in Israeli farm households. *American Journal of Agricultural Economics* 86(3):716–721.

Kotz, S., Balakrishnan, N., Johnson, S. (2000). *Continuous Multivariate Distributions*. Vol. 1. New York: John Wiley.

List, J. A., Gallet, C. (1999). The environmental Kuznets curve: does one size fit all? *Ecological Economics* 31:409–423.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association* 60(309):193–204.

Melas, V. B., Pepelyshev, A., Cheng, R. C. H. (2003). Designs for estimating an extremal point of quadratic regression models in a hyperball. *Metrika* 58(2):193–208.

Miles, D. (1997). A household level study of the determinants of incomes and consumption. *Economic Journal* 107(4):1–25.

Millimet, D. L., List, J. A., Stengos, T. (2003). The environmental Kuznets curve: real progress or misspecified models? *Review of Economics and Statistics* 85(4):1038–1047.

Neumark, D., Taubman, P. (1995). Why do wage profiles slope upwards? Tests of the general human capital model. *Journal of Labor Economics* 13(4):736–761.

Papke, L. E., Wooldridge, J. M. (2005). A computational trick for delta method standard errors. *Economics Letters* 86:413–417.

Plassmann, F., Khanna, N. (2006). Household income and pollution: implications for the debate about the environmental Kuznets curve hypothesis. *Journal of Environment and Development* 15(1):1–19.

Shanmugalingam, S. (1982). On the analysis of the ratio of two correlated normal variables. *Statistician* 31(3):251–258.

Stern, D., Common, M. (2001). Is there an environmental Kuznets curve for sulfur? *Journal of Environmental Economics and Management* 41(3):162–178.

Tokle, J. G., Huffman, W. E. (1991). Local economic conditions and wage labor decisions of farm and rural nonfarm couples. *American Journal of Agricultural Economics* 73(3):652–670.

Torras, M., Boyce, J. K. (1998). Income, inequality, and pollution: a reassessment of the environmental Kuznets curve. *Ecological Economics* 25(2):147–160.

Vom Scheidt, J., Bharucha-Reid, A. T. (1983). On the distribution of the roots of random algebraic polynomials. *Wissenschaftliche Beiträge Ingenieurhochschule Zwickau*. Sonderheft, pp. 44–82.

Wolgrom, G. (2001). More results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 69(5):1381–1389.