*Brazilian Journal of Probability and Statistics*
2013, Vol. 27, No. 4, 401–415
DOI: 10.1214/11-BJPS173


# Errors-In-Variables regression and the problem of moments

**Ali Al-Sharadqah, Nikolai Chernov and Qizhuo Huang**
*University of Alabama at Birmingham*

**Abstract.** In regression problems where covariates are subject to errors (albeit small) it often happens that maximum likelihood estimators (MLE) of relevant parameters have infinite moments. We study here circular and elliptic regression, that is, the problem of fitting circles and ellipses to observed points whose both coordinates are measured with errors. We prove that several popular circle fits due to Pratt, Taubin, and others return estimates of the center and radius that have infinite moments. We also argue that estimators of the ellipse parameters (center and semiaxes) should have infinite moments, too.

## 1 Introduction

In classical regression, where independent variables are error-free, maximum likelihood estimators (MLE) are consistent and have asymptotically minimal variance. In the linear model $y = \alpha + \beta x$, the MLE $\hat{\alpha}$ and $\hat{\beta}$ are unbiased and have minimal variance for every sample size.

But in modern applications (such as pattern recognition and computer vision) both variables $x$ and $y$ are subject to errors, which brings us to the Errors-In-Variables (EIV) regression analysis. More precisely, we study the problem of fitting a curve $P(x, y; \Theta) = 0$, where $\Theta$ represents a vector of unknown parameters, to observed points $(x_1, y_1), \dots, (x_n, y_n)$ that are a random perturbation of some true points $(\tilde{x}_i, \tilde{y}_i)$:

$$x_i = \tilde{x}_i + \delta_i, \qquad y_i = \tilde{y}_i + \varepsilon_i, \qquad i = 1, \dots, n. \tag{1.1}$$

Here $(\delta_i, \varepsilon_i)$ are independent random vectors, usually having normal distribution with zero mean. The true points $(\tilde{x}_i, \tilde{y}_i)$, $1 \le i \le n$, are supposed to lie on the (unknown) true curve, that is, $P(\tilde{x}_i, \tilde{y}_i; \tilde{\Theta}) = 0$ for all $i = 1, \dots, n$; here $\tilde{\Theta}$ denotes the true (but unknown) value of $\Theta$. The true points are either fixed (then their positions are treated as additional parameters) or selected randomly from a certain probability distribution on the true curve. In the former case one gets a *functional model* and in the latter case a *structural model*.

If $\varepsilon_i$ and $\delta_i$ are i.i.d. normal random variables $N(0, \sigma^2)$, then the maximum likelihood estimate (MLE) in the functional model is obtained by minimizing the squares of the geometric (i.e., orthogonal) distances from the observed points to

*Key words and phrases.* Errors-In-Variables regression, moments, fitting circles, fitting ellipses.
Received April 2011; accepted September 2011.

the fitting curve; see Chan (1976) and Chernov (2010). This procedure is called *geometric fit* or *orthogonal distance regression* (ODR).

The minimization of geometric distances (i.e., ODR) has been used since the 1870s [see Adcock (1877)], and lately has become standard in nearly all applications. It is commonly regarded as the best (most accurate and reliable) fitting method.[1] The fact that it produces the MLE gives yet another justification of its high quality. However, in statistical terms, the corresponding parameter estimators often have bizarre features—inconsistency and infinite moments.

**Linear regression**

In the linear model $y = \alpha + \beta x$, minimization of geometric distances gives

$$(\hat{\alpha}_0, \hat{\beta}_0) = \operatorname{argmin} \frac{1}{1+\beta^2} \sum (y_i - \alpha - \beta x_i)^2. \tag{1.2}$$

These estimators are consistent but have infinite absolute moments:

$$\mathbb{E}(|\hat{\alpha}_0|) = \mathbb{E}(|\hat{\beta}_0|) = \infty. \tag{1.3}$$

Anderson (1976) proved this fact assuming that $\delta_i$ and $\varepsilon_i$ are i.i.d. normal random variables, and Chernov (2011) showed that it holds whenever each $(\delta_i, \varepsilon_i)$ just has a continuous distribution with a strictly positive density [but the vectors $(\delta_i, \varepsilon_i)$ must be independent].

On the other hand, one can use classical estimates minimizing vertical distances

$$(\hat{\alpha}_1, \hat{\beta}_1) = \operatorname{argmin} \sum (y_i - \alpha - \beta x_i)^2. \tag{1.4}$$

They happen to have finite moments; see Anderson (1976). But they are inconsistent and are known to be heavily biased toward smaller values of $\beta$; see Anderson (1976) and Chernov (2010). In other words, paradoxically, the better estimators (1.2) have infinite absolute moments (thus, theoretically, their mean squared errors are infinite and their biases cannot even be defined), while the less accurate estimators (1.4) have finite moments (so their biases and mean squared errors are at least finite).

A similar phenomenon was recently observed in a slightly different linear regression model, where adjusted least squares estimators (designed to reduce the bias of the classical least squares) also were found to have infinite moments; see Cheng and Kukush (2006).

[1] In particular, it has been prescribed by a recently ratified standard for testing the data processing software for coordinate metrology; see Ahn (2004).

### Circular regression

If one estimates the center $(a, b)$ and the radius $R$ of a circle by minimizing geometric distances to the observed points, then

$$(\hat{a}_0, \hat{b}_0, \hat{R}_0) = \operatorname{argmin} \sum_{i=1}^{n} \left[\sqrt{(x_i - a)^2 + (y_i - b)^2} - R\right]^2. \tag{1.5}$$

This circular fit has been used since the 1950s [see a survey Chernov (2010)], but the resulting estimators are inconsistent. Chernov (2011) also showed that they have infinite absolute moments:

$$\mathbb{E}(|\hat{a}_0|) = \mathbb{E}(|\hat{b}_0|) = \mathbb{E}(|\hat{R}_0|) = \infty. \tag{1.6}$$

This is true for both functional and structural models provided each vector $(\delta_i, \varepsilon_i)$ has a continuous distribution with a strictly positive density.

The nonlinear minimization problem (1.5) has no closed form solution. To simplify estimation, one can minimize

$$(\hat{a}_1, \hat{b}_1, \hat{R}_1) = \operatorname{argmin} \sum_{i=1}^{n} [(x_i - a)^2 + (y_i - b)^2 - R^2]^2. \tag{1.7}$$

This method is known as the Kåsa method; see Kåsa (1976). It has been widely used since the 1970s. It allows a simple noniterative solution: introducing a new parameter $c = a^2 + b^2 - R^2$ turns the right-hand side of (1.7) into a quadratic polynomial in $a, b, c$, hence finding its minimum is an elementary task. Once we compute $a, b, c$ we can recover $R$ by $R = \sqrt{a^2 + b^2 - c}$.

Kåsa estimates are easy to compute but they are inconsistent and heavily biased toward smaller circles. On the other hand, they have finite moments, as it was recently proved by Zelniker and Clarkson (2006). Hence, the same controversial situation takes place—the better fit (1.5) has infinite moments, while a heavily biased fit (1.7) has finite moments.

Here we investigate several 'intermediate' fitting schemes—circle fits by Pratt and Taubin, and the so-called 'Hyperfit.' They have been designed as improvements to the Kåsa fit, but they are still slightly less accurate than the geometric fit; see Al-Sharadqah and Chernov (2009) and Chernov (2010). We show here that all of them have infinite moments.

Our results confirm that the above paradoxical tendency is rather general in the EIV regression analysis—better fits have infinite moments, while some crude biased fits have finite moments. One can say, ironically, that infinite moments are a 'certificate of quality' for parameters estimators in the EIV analysis.

### Elliptic regression

Suppose we estimate the center $(x_c, y_c)$ and the semi-axes $A \geq B > 0$ of an ellipse by minimizing geometric distances to the observed points. We argue that the estimators of these parameters also have infinite absolute moments. More precisely,

$$\mathbb{E}(|\hat{x}_c|) = \mathbb{E}(|\hat{y}_c|) = \mathbb{E}(\hat{A}) = \infty, \tag{1.8}$$

that is, three out of the four ellipse parameter estimators have an infinite first absolute moment. The minor semiaxis $B$ is always smaller than $A$, and we will see that this leads to $\mathbb{E}(\hat{B}) < \infty$. But its second moment is infinite:

$$\mathbb{E}(\hat{B}^2) = \infty. \tag{1.9}$$

We need to give a word of caution though. In elliptic regression a new issue arises: in many instances the best fitting ellipse would not exist. Strictly speaking, if one fits a quadratic curve (a conic section) to observed points, then the best fitting conic may be (i) an ellipse or (ii) a hyperbola or (iii) a parabola or (iv) a straight line or (v) a pair of straight lines. And while lines and parabolas occur with probability zero (thus they can be ignored), hyperbolas occur with a positive probability and have to be reckoned with.

When the best fitting conic is a hyperbola, then the problem of fitting ellipses has no solution. In that case, for any ellipse one can find another ellipse that fits the given points even better (in the sense of a smaller sum of squares of geometric distances); a sequence of such ellipses that approximate the given points with a progressively better accuracy would converge to a parabola. See a detailed account on this issue by Nievergelt (2004).

In a simple experiment, we have generated random sets of $n = 5$ points with a 2D standard normal distribution and checked whether the best fitting conic (which would just interpolate our 5 points) was an ellipse or a hyperbola. We found, quite surprisingly, that ellipses turned up only in 22% of the cases, while hyperbolas—in 78% of the cases. Thus, hyperbolas actually dominate over ellipses, so the nonexistence of the best fitting ellipse is quite a frequent phenomenon.

For this reason we have to restrict our analysis to the data sets where the best fitting ellipse does exist (i.e., where the best fitting conic is an ellipse, rather than anything else). The expectations in (1.8)–(1.9) have to be understood as *conditional expectations* (i.e., the integrals of the corresponding estimates over the collection of data sets for which the best fitting ellipse exists).

## Consistency

In the EIV regression analysis many popular estimators are inconsistent. For instance, all the above estimators, except (1.2), are inconsistent. There exist consistent estimators for circles [Chernov (2010), Section 7.9] and ellipses [Kukush et al. (2004)], but they have infinite moments, too [see Kukush et al. (2004) and our Figure 1 below].

In fact, the very notion of consistency is commonly redefined in image processing applications, because the sample size $n$ is rarely large and cannot be increased by further sampling; see a well written survey by Kanatani (2004) on this issue. Theoretical studies commonly adopt a "small noise" model, where $n$ is fixed but $\sigma \to 0$; see Anderson (1976) and Chernov (2010), Section 2.5. Under these conditions, all the above estimators *are* consistent, that is, they converge to the true parameter values as $\sigma \to 0$.

**Alternative parametrization**

Sometimes using different parameters can prevent moments from being infinite. For example, lines can be defined by $a_1x + a_2y + a_3 = 0$ with the constraint $\sum a_i^2 = 1$. Circles can be described by $b_1(x^2 + y^2) + b_2x + b_3y + b_4 = 0$ with the constraint $\sum b_i^2 = 1$; ellipses by $c_1x^2 + c_2y^2 + c_3xy + c_4x + c_5y + c_6 = 0$ with $\sum c_i^2 = 1$. Then the corresponding parameters will be always restricted to the interval $[-1, 1]$, so their estimators will have finite moments. Other parameterizations for lines and circles that always lead to finite moments are mentioned by Chernov (2011).

But the above parameters may cause other problems by returning a curve of the wrong type. For example, one may get a hyperbola or a parabola instead of an ellipse if the estimates of $c_i$ do not satisfy the elliptic ancillary constraint $4c_1c_2 > c_3^2$. Besides, in many practical applications it is important to estimate natural geometric characteristics of the fitting object, rather than its abstract "algebraic" parameters.

**Alternative statistical analysis**

The lack of moments raises methodological questions: How can one measure the accuracy of an estimator whose mean square error is infinite (and whose bias is undefined)? Is there any precise meaning to the widely accepted notion that the MLE, such as (1.2) and (1.5), are best estimators, despite having infinite moments?

These questions can be answered in the framework of an unconventional analysis where one assumes that $n$ is fixed and $\sigma \to 0$ and uses the Taylor expansion to construct approximate distributions of estimators. Those distributions have finite moments that very accurately characterize the quality of estimators. For linear regression this approach was first used by Anderson (1976) and Anderson and Sawa (1982); they said the resulting approximations were 'virtually exact.' For more general models this type of theoretical analysis was employed by Al-Sharadqah and Chernov (2009); see also its experimental validation by Al-Sharadqah and Chernov (2011).

## 2 Circle fits

To see why the Kåsa method (1.7) is biased toward smaller circles note that if the observed points are close to the circle, then

$$(x_i - a)^2 + (y_i - b)^2 - R^2 = d_i(d_i + 2R) \approx 2Rd_i, \tag{2.1}$$

where $d_i = \sqrt{(x_i - a)^2 + (y_i - b)^2} - R$ denotes the geometric distance. Then the right-hand side of (1.7) can be approximated by

$$\sum_{i=1}^{n} [(x_i - a)^2 + (y_i - b)^2 - R^2]^2 \approx 4R^2 \sum_{i=1}^{n} d_i^2. \tag{2.2}$$

The factor $R^2$ in (2.2) affects the minimum of that function: the Kåsa fit often minimizes $R$ instead of $d_i$'s.

To improve the Kåsa fit, one can minimize the function

$$\sum_{i=1}^{n} R^{-2}[(x_i - a)^2 + (y_i - b)^2 - R^2]^2 \to \min. \tag{2.3}$$

This fit was proposed first by Chernov and Ososkov (1984), and in a more elegant form by Pratt (1987) who reduced it to an eigenvalue problem which can be solved noniteratively with modern software. The Pratt fit is much more accurate than the Kåsa fit.

Alternatively, the approximation (2.1) allows us to replace (2.3) with

$$\frac{\sum_{i=1}^{n}[(x_i - a)^2 + (y_i - b)^2 - R^2]^2}{\sum_{i=1}^{n}[(x_i - a)^2 + (y_i - b)^2]} \to \min. \tag{2.4}$$

This fit was proposed by Taubin (1991) who reduced the minimization of (2.4) to another eigenvalue problem. The Taubin fit happens to be even more accurate than the Pratt fit, though both fall behind the geometric fit (1.5); see a detailed analysis in Al-Sharadqah and Chernov (2009) and Chernov (2010).

Lastly, the so-called Hyperfit was proposed by Al-Sharadqah and Chernov (2009):

$$\frac{\sum_{i=1}^{n}[(x_i - a)^2 + (y_i - b)^2 - R^2]^2}{\sum_{i=1}^{n}[2(x_i - a)^2 + 2(y_i - b)^2 - R^2]} \to \min. \tag{2.5}$$

It also reduces to an eigenvalue problem. It was shown to have a smaller bias than the other fits.

Our main result here is the following:

**Theorem 1.** *If the joint distribution of all the noise components* $\delta_1, \varepsilon_1, \ldots, \delta_n, \varepsilon_n$ *has a continuous strictly positive density, then the estimators* $\hat{a}, \hat{b}, \hat{R}$ *obtained by the Pratt fit* (2.3) *and the Taubin fit* (2.4) *and the Hyperfit* (2.5) *have infinite moments, that is,* $\mathbb{E}(|\hat{a}|) = \mathbb{E}(|\hat{b}|) = \mathbb{E}(\hat{R}) = \infty$.

The conditions of the theorem are met under the standard assumptions that the noise vectors $(\delta_i, \varepsilon_i)$ are independent and each has a 2D normal distribution. However, our theorem also holds for dependent vectors and arbitrary distributions with a continuous strictly positive density.

Before we prove this theorem, we illustrate it with a numerical experiment. We positioned $n = 10$ 'true' points equally spaced on a semicircle $\{x^2 + y^2 = 1, y \geq 0\}$. We generated $K$ random samples by perturbing the true points with Gaussian noise $\delta_i, \varepsilon_i \sim N(0, \sigma^2)$ at level $\sigma = 0.4$. For each sample we estimated the circle parameters by all the above algebraic fits and by the consistent circle fit, which is an adaptation of the ellipse consistent fit by Kukush et al. (2004) to circles; see

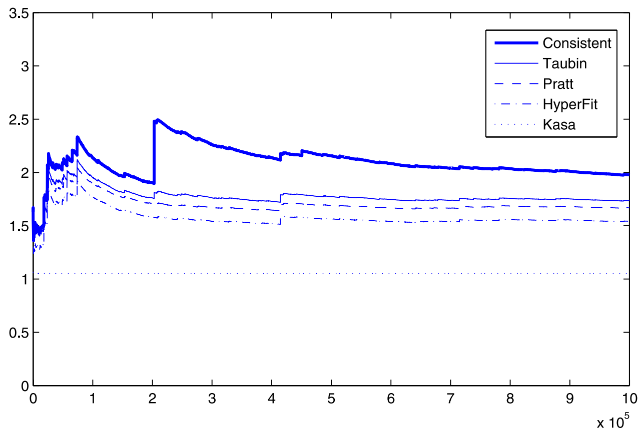


**Figure 1** *Average radius estimate $\hat{R}$ versus the number of random samples.*

Chernov (2010), Section 7.9. Then all the $K$ radius estimates were averaged, for each fit separately.

Figure 1 shows the average estimate $\hat{R}$, as a function of $K$, when $K$ grows from 1 to $10^6$ (for every fit separately). The Kåsa fit corresponds to a flat line near $R \approx 1.1$, this is the only stable fit in our 'pack.' Every other fit yields a wildly oscillating curve that experiences frequent jumps. Such an erratic behavior is characteristic for random variables with an infinite first moment (similar plots can be constructed, for example, for a sample mean from a Cauchy random variable).

## 3 Proof of Theorem 1

### General strategy

An estimator $\hat{\theta}$ of a parameter $\theta$ is a random variable. Its first absolute moment

$$\mathbb{E}(|\hat{\theta}|) = \int_0^\infty \text{Prob}(|\hat{\theta}| > t)\, dt$$

is infinite if the distribution has a power-law tail $\text{Prob}(|\hat{\theta}| > t) \sim t^{-\alpha}$ (as $t \to \infty$) with $\alpha \leq 1$. The reciprocal $\zeta = 1/\hat{\theta}$ then satisfies $\text{Prob}(|\zeta| < t) \sim t^{\alpha}$ as $t \to 0$ with $\alpha \leq 1$. Thus, it is enough to check that $\zeta$ has a positive density which does not vanish at 0.

We employ the following strategy. Suppose we can position data points $(x_1, y_1), \ldots, (x_n, y_n)$ so that the corresponding parameter estimator $\hat{\theta}$ will be infinite, hence its reciprocal $\zeta$, $=1/\hat{\theta}$ vanishes. Next, we note that the estimator $\hat{\theta}$, and hence its reciprocal $\zeta$, are continuous functions of the coordinates $x_1, y_1, \ldots, x_n, y_n$. We will examine how $\zeta$ changes if only one of those coordinate, say, $x_1$, varies.

**Lemma 2.** *Suppose that the derivative $\partial\zeta/\partial x_1$ is bounded, that is, $|\partial\zeta/\partial x_1| \le D$ for some $D > 0$. Then the conditional expectation of $|\hat{\theta}|$, given that the coordinates $y_1, x_2, y_2, \dots, x_n, y_n$ (i.e., all but $x_1$) are fixed, is infinite, that is, $\mathbb{E}(|\hat{\theta}||y_1, x_2, y_2, \dots, x_n, y_n) = \infty$.*

**Proof.** Since the original joint distribution of all the coordinates $x_1, y_1, \dots, x_n, y_n$ has a strictly positive density, the conditional distribution of $x_1$ (given all the other coordinates) also has a strictly positive density:

$$P(dx_1|y_1, \dots, x_n, y_n) = \rho(x_1|y_1, \dots, x_n, y_n)\, dx_1.$$

Now since $|\partial\zeta/\partial x_1| \le D$, the conditional density of $\zeta$ is positive, too, as

$$P(d\zeta|y_1, \dots, x_n, y_n) = |\partial x_1/\partial\zeta| \cdot \rho(x_1|y_1, \dots, x_n, y_n)\, d\zeta$$

and $|\partial x_1/\partial\zeta| \ge 1/D > 0$. Hence, as we have seen above, the conditional expectation of $|\hat{\theta}|$ is infinite. Lemma 2 is proved. □

Next suppose the derivative $\partial\zeta/\partial x_1$ remains bounded when every previously fixed coordinate $y_1$, $x_2, y_2, \dots, x_n, y_n$ is slightly perturbed, that is, the boundedness holds for all $x_i \in I_i$, $2 \le i \le n$, and $y_j \in J_j$, $1 \le j \le n$, for some small intervals $I_i$ and $J_j$. The lengths of $I_i$'s and $J_j$'s do not matter, because the original joint density of $x_1, y_1, \dots, x_n, y_n$ is strictly positive. Hence, we obtain the desired result: $\mathbb{E}(|\hat{\theta}|) = \infty$.

### Circular regression

We fix a small rectangle $\mathbf{B}_1 = [-h, h] \times [1-h^2, 1+h^2]$ centered on the point (0, 1) and two very small squares $\mathbf{B}_2 = [-h^2, h^2] \times [-h^2, h^2]$ and $\mathbf{B}_3 = [-h^2, h^2] \times [-1-h^2, -1+h^2]$ centered on the points (0, 0) and (0, −1), respectively. Here $h$ is a very small number, such as $h = (10n)^{-9}$. Note that $h^2 \ll h$.

Next we choose $(x_1, y_1) \in \mathbf{B}_1$, $(x_2, y_2) \in \mathbf{B}_2$, and all the other points $(x_i, y_i) \in \mathbf{B}_3$ for $i = 3, \dots, n$. Note that $x_1$ is allowed to vary from $-h$ to $h$, while all the other coordinates are confined to much smaller intervals of length $2h^2 \ll 2h$. We will regard $x = x_1$ as the only variable, while all the other coordinates are fixed, and we examine how the fitting circle changes as $x = x_1$ varies from $-h$ to $h$. Let $\hat{a}(x)$ denote the $x$ coordinate of the center of the fitting circle, as a function of $x$, and $\zeta(x) = 1/\hat{a}(x)$. If the circle degenerates to a line, we set $\hat{a} = \infty$ and $\zeta = 0$.

Since $h$ is very small, all our data points lie near the vertical line (the $y$ axis), but the first point $(x, y_1)$ can slightly deviate from it left or right; see Figure 2. It is clear that if $x = h$, then $\hat{a}(x)$ is positive, hence, $\zeta(h) > 0$. Similarly, if $x = -h$, then $\hat{a}(x)$ is negative, that is, $\zeta(-h) < 0$. As $x = x_1$ changes from $-h$ to $h$, the function $\zeta(x)$ changes from negative values to positive values (and remains small). All we need is to show that $\zeta(x)$ has a bounded derivative:

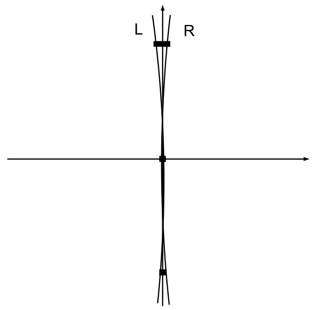


**Figure 2** *A rectangle and two squares around the* $y$ *axis. Two fitting circles: one has center to the left* ($\hat{a} < 0$) *and the other to the right* ($\hat{a} > 0$).

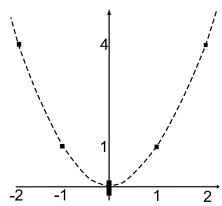


**Figure 3** *A rectangle and four squares near a parabola.*

**Lemma 3.** *For any fixed values* $y_1, x_2, y_2, \ldots, x_n, y_n$ *as above, the function* $\zeta(x_1)$ *has a bounded derivative:* $|\zeta'(x_1)| \le D$ *for all* $x_1 \in [-h, h]$ *and some constant* $D > 0$.

This lemma is proved by direct calculations that will be provided in the next section. It implies that $\mathbb{E}(|\hat{a}|) = \infty$. It is also clear that $\mathbb{E}(\hat{R}) = \infty$. Rotating our construction, say, by $\pi/2$, we obtain $\mathbb{E}(|\hat{b}|) = \infty$ as well. Theorem 1 is proved.

## Elliptic regression

Natural parameters of an ellipse are the coordinates of its center $(x_c, y_c)$ and its semiaxes $A \ge B$. All these parameters grow to infinity as an ellipse degenerates to a parabola. So we choose a parabola, say, $y = x^2$, and five points on it: $(\pm 1, 1)$, $(\pm 2, 4)$ and $(0, 0)$. Then we fix a small rectangle $\mathbf{B}_1$ of size $(2h^2) \times (2h)$ centered on $(0, 0)$ and four very small squares $\mathbf{B}_k$ $(2 \le k \le 5)$ of size $(2h^2) \times (2h^2)$ centered on the other four points, respectively; see Figure 3. Here $h$ is a very small number, such as $h = (10n)^{-9}$.

Next we choose $(x_1, y_1) \in \mathbf{B}_1$, and all the other points $(x_i, y_i)$, $2 \le i \le n$, are placed in the squares $\mathbf{B}_k$, $2 \le k \le 5$, so that each square has at least one data point. Note that $y_1$ is allowed to vary from $-h$ to $h$, while all the other coordinates are confined to much smaller intervals of length $2h^2 \ll 2h$. We will regard $y_1$ as the only variable while all the other coordinates are fixed and examine how the fitting ellipse changes as $y_1$ varies from $-h$ to $h$.

It is not hard to check by elementary geometry that for $y_1 = h$ the best fitting conic is an ellipse, but for $y_1 = -h$ the best fitting conic is a hyperbola. As $y_1$ changes from $h$ to $-h$, the best fitting ellipse grows, degenerates to a parabola (for some $y_1 = y_1^*$) and then becomes a hyperbola. So we only consider the interval $y_1^* < y_1 < h$. At the moment $y_1 = y_1^*$ the ellipse explodes and its major semiaxis $\hat{A}$ becomes infinite. We denote $\zeta = 1/\hat{A}$. We will argue that the following fact holds (but we cannot prove it fully yet, so we call it a conjecture):

**Conjecture 4.** *For any fixed values* $x_1, x_2, y_2, \ldots, x_n, y_n$ *as above, the function* $\zeta(y_1)$ *has bounded derivatives, that is,* $|\zeta'(y_1)| \le D$ *for all* $y_1 \in [-h, h]$ *and some constant* $D > 0$.

Assuming that this conjecture is true, it readily implies that $\mathbb{E}(\hat{A}) = \infty$. It is also clear that $\mathbb{E}(|\hat{y}_c|) = \infty$. Rotating our construction, say, by $\pi/2$, we obtain $\mathbb{E}(|\hat{x}_c|) = \infty$ as well, hence we get (1.8). Moreover, we will argue that as $\hat{A}$ grows to infinity the minor semiaxis also grows but at a slower rate, as $\hat{B} \sim \hat{A}^{1/2}$. This will imply (1.9).

## 4 Proof of Lemma 3

Lemma 3 will be proved here for each circle fit separately.

### Pratt fit

Let $(\rho, \theta)$ be the polar coordinates of the center $(a, b)$. The objective function (2.3) takes all of its small values on circles and lines that pass in the $h$-vicinity of the three basic points: $(0, 0)$, $(0, -1)$ and $(0, 1)$. These circles and lines have parameters restricted to the region where $\rho > 1/(100h)$ and $|\sin\theta| < 100h$. Since $\rho$ attains large values, it is more convenient to deal with its reciprocal $\delta = 1/\rho$. Accordingly, our analysis will be restricted to the region

$$\Omega = \{|\delta| \le 100h \text{ and } |\theta| \le 100h\}. \tag{4.1}$$

We denote $w_i = (x_i - a)^2 + (y_i - b)^2$ and use standard 'sample mean' notation $\bar{w} = \frac{1}{n}\sum w_i$, etc. The conditional minimum of the objective function (2.3), when $a$ and $b$ are kept fixed, is attained at $\hat{R}^4 = \overline{ww} = \frac{1}{n}\sum w_i^2$. Thus, one can eliminate $R$ and rewrite the objective function as

$$\mathcal{F}(a, b) = \frac{\sum_{i=1}^n (w_i - \hat{R}^2)^2}{4n\hat{R}^2} = \frac{\hat{R}^2 - \bar{w}}{2}. \tag{4.2}$$

For convenience we introduce the notation

$$u_i = x_i \cos\theta + y_i \sin\theta \quad \text{and} \quad v_i = -x_i \sin\theta + y_i \cos\theta.$$

Now $\mathcal{F}$ can be expressed in polar coordinates as

$$\mathcal{F}(\rho,\theta) = \hat{R}^2 - (\bar{z} - 2\bar{u}\rho + \rho^2)/2$$
$$= \frac{1}{2}\Big[\sqrt{\overline{zz} - 4\overline{uz}\rho + 2(2\overline{uu} + \bar{z})\rho^2 - 4\bar{u}\rho^3 + \rho^4} \tag{4.3}$$
$$- (\bar{z} - 2\bar{u}\rho + \rho^2)\Big],$$

where $z_i = x_i^2 + y_i^2 = u_i^2 + v_i^2$. To shorten our formulas, it will be convenient to use the following notation:

$$S_{zz} = \sum_{i=1}^{n}(z_i - \bar{z})^2, \qquad S_{uz} = \sum_{i=1}^{n}(u_i - \bar{u})(z_i - \bar{z}), \qquad S_{uu} = \sum_{i=1}^{v}(u_i - \bar{u})^2.$$

We also denote $E = \bar{z}\delta^2 - 2\bar{u}\delta + 1$ and

$$F = \sqrt{(\bar{z}\delta^2 - 2\bar{u}\delta + 1)^2 + S_{zz}\delta^4 - 4S_{uz}\delta^3 + 4S_{uu}\delta^2}.$$

With this notation, it can be shown that

$$\mathcal{F}(\delta,\theta) = (2\delta^2)^{-1}(F - E)$$
$$= \frac{1}{2}(E + F)^{-1}(S_{zz}\delta^2 - 4S_{uz}\delta + 4S_{uu}).$$

Notice that $4 > E + F > \frac{1}{2}$. Also, $S_{zz} = 2 + \chi$, where $\chi$ is a small quantity ($\chi \to 0$ as $h \to 0$), while $S_{uu}$ and $S_{uz}$ are also small quantities. Hence, $\mathcal{F}(\delta,\theta)$ is bounded and continuous in $\Omega$. The second derivative with respect to $x = x_1$ is bounded. Indeed, both $E$ and $F$ are bounded by 2, so the second derivative with respect to $x$ is a fraction function whose numerator is a polynomial of $x$ and $F$ while the denominator has the form $F^k(E + F)^m$ for some integers $k \geq 1$ and $m \geq 1$. Therefore,

$$\frac{\partial \mathcal{F}(\delta,\theta)}{\partial x} \leq M_1, \qquad \frac{\partial^2 \mathcal{F}(\delta,\theta)}{\partial x^2} \leq M_2 \tag{4.4}$$

for some positive constants $M_1$ and $M_2$.

Similarly, the first and the second partial derivatives of $\mathcal{F}(\delta,\theta)$ with respect to $\theta, \delta$ are continuous and bounded by some positive constant $M$ (details are omitted, but they are similar to above justification). Moreover, direct differentiation implies that $\nabla^2\mathcal{F}(\delta,\theta)$ is positive definite because

$$\nabla^2\mathcal{F}(\delta,\theta) = \begin{bmatrix} 1 - \frac{2}{n} + \chi_1 & \chi_2 \\ \chi_2 & 4 + \chi_3 \end{bmatrix},$$

where $\chi_i$'s denote various small quantities (in the sense that $\chi_i \to 0$ as $h \to 0$). Thus, $\mathcal{F}$ is a convex function that has exactly one minimum in $\Omega$ and no other critical points.

Finally, we will show that $|\zeta'| \le 4M$. Indeed, let $(\hat{\delta}, \hat{\theta})$ denote the above unique minimum. Differentiating equations

$$\mathcal{F}_\delta(\hat{\delta}, \hat{\theta}) = 0 \quad \text{and} \quad \mathcal{F}_\theta(\hat{\delta}, \hat{\theta}) = 0$$

(here indices denote partial derivatives) with respect to $x$ gives

$$\mathcal{F}_{\delta\delta}(\hat{\delta}, \hat{\theta})\hat{\delta}' + \mathcal{F}_{\delta\theta}(\hat{\delta}, \hat{\theta})\hat{\theta}' + \mathcal{F}_{\delta x}(\hat{\delta}, \hat{\theta}) = 0,$$
$$\mathcal{F}_{\theta\delta}(\hat{\delta}, \hat{\theta})\hat{\delta}' + \mathcal{F}_{\theta\theta}(\hat{\delta}, \hat{\theta})\hat{\theta}' + \mathcal{F}_{\theta x}(\hat{\delta}, \hat{\theta}) = 0,$$

where $\hat{\delta}'$ and $\hat{\theta}'$ denote the derivatives with respect to $x$. Since all our partial derivatives are uniformly bounded by $M$, we have that $|\hat{\delta}'| \le 2M$ and $|\hat{\theta}'| \le 2M$. Lastly, recall that $\zeta = 1/\hat{a} = \hat{\delta}/\cos\hat{\theta}$, hence,

$$|\zeta'| = \left| \frac{\hat{\theta}' \sin\hat{\theta}}{\cos^2\hat{\theta}}\hat{\delta} + \frac{\hat{\delta}'}{\cos\hat{\theta}} \right| \le 4M$$

(recall that $\hat{\theta} \approx 0$). This completes the proof of Lemma 3.

**Taubin fit**

The above argument, with certain modifications, applies to the Taubin fit. First, eliminating $R$ from its objective function (2.4) gives

$$\begin{aligned} \mathcal{F}(a, b) &= \frac{\sum_{i=1}^n (w_i - \bar{w})^2}{4n\bar{w}} \\ &= \frac{\sum_{i=1}^n [(z_i - \bar{z}) - 2(u_i - \bar{u})\rho]^2}{4n(\bar{z} - 2\bar{u}\rho + \rho^2)}. \end{aligned}$$

Changing variables gives

$$\begin{aligned} \mathcal{F}(\delta, \theta) &= \frac{\sum_{i=1}^n [(z_i - \bar{z})\delta - 2(u_i - \bar{u})]^2}{4n(\bar{z}\delta^2 - 2\bar{u}\delta + 1)} \\ &= \frac{S_{zz}\delta^2 - 4S_{uz}\delta + 4S_{uu}}{4n(\bar{z}\delta^2 - 2\bar{u}\delta + 1)}. \end{aligned}$$

It is clear that the denominator is bounded from below by 3 and from above by $4n$, hence, $\mathcal{F}(\delta, \theta) \le 1$ and is continuous in $\Omega$. In the same way the second partial derivatives $\nabla_{xx}\mathcal{F}$, $\nabla_{x\theta}\mathcal{F}$ and $\nabla_{x\delta}\mathcal{F}$ are continuous and bounded as well. By direct differentiation we find that

$$\nabla^2\mathcal{F}(\delta, \theta) = \begin{bmatrix} 1 - \dfrac{2}{n} + \chi_1 & \chi_2 \\ \chi_2 & 4 + \chi_3 \end{bmatrix},$$

where $\chi_i$'s are again various small quantities. Thus, $\mathcal{F}$ is a convex function that has exactly one minimum in $\Omega$ and no other critical points. The rest of the proof goes unchanged.

**Hyper fit**

The objective function is

$$\mathcal{F}(a, b, R) = \frac{\sum_{i=1}^{n}(w_i - R^2)^2}{4n(2\bar{w} - R^2)} = \frac{\overline{ww} - 2\bar{w}R^2 + R^4}{8\bar{w} - 4R^2}. \tag{4.5}$$

Eliminating $R$ gives the same function (4.2) that we obtained for the Pratt fit earlier. Thus, these two fits return the same center $(a, b)$, so its first moment is infinite, too.

## 5 A case supporting Conjecture 4

We cannot give a full proof of Conjecture 4 yet, but one particular case is tractable and we present it here.

Suppose we choose exactly one point in each square $\mathbf{B}_k$, $2 \le k \le 5$. Then the total number of data points is $n = 5$, and there exists a unique conic interpolating them, which obviously provides the best fit. In that case the type and parameters of that conic can be explicitly determined. For simplicity, let the four points be chosen at the centers of the squares $\mathbf{B}_k$, $2 \le k \le 5$, that is, they are $(\pm 1, 1)$ and $(\pm 2, 4)$. Recall that the first point $(x_1, y_1) \in \mathbf{B}_1$ has a variable $y$-coordinate $|y_1| < h$ and an $x$-coordinate $|x_1| < h^2$. Then for some $(x_1, y_1)$ the interpolating conic will be a hyperbola or a parabola (we will ignore such values of $x_1$ and $y_1$) and for others—an ellipse. We will see that the interpolating conic is an ellipse if and only if $y_1 > x_1^2$.

Since our conic interpolates the four points chosen at the centers of the squares $\mathbf{B}_k$, $2 \le k \le 5$, it is easy to see that its equation is

$$x^2 + cy^2 - (1 + 5c)y + 4c = 0, \tag{5.1}$$

where $c$ is a variable parameter. Since the curve (5.1) also must pass through the point $(x_1, y_1)$, we get

$$c = \frac{y_1 - x_1^2}{4 - 5y_1 + y_1^2}. \tag{5.2}$$

Clearly, $c$ is small (of order $h$). Equation (5.1) can be rewritten as

$$x^2 + c\left(y - \frac{1 + 5c}{2c}\right)^2 = \frac{1 + 10c + 9c^2}{4c}. \tag{5.3}$$

It is now easy to see that the conic (5.1) is an ellipse if and only if $c > 0$ (i.e., $y_1 > x_1^2$), and that this ellipse has semiaxes

$$\hat{A} = \frac{\sqrt{1 + 10c + 9c^2}}{2c}, \qquad \hat{B} = \frac{\sqrt{1 + 10c + 9c^2}}{2\sqrt{c}}. \tag{5.4}$$

Therefore, $\zeta = 1/\hat{A} = 2c/\sqrt{1+10c+9c^2}$, hence, $\partial\zeta/\partial c \approx 2$. It follows from (5.2) that $\partial c/\partial y_1 \approx 1/4$, hence, $\partial\zeta/\partial y_1 \approx 1/2$, which proves Conjecture 4. Also note that (5.4) implies $\hat{B} \sim \hat{A}^{1/2}$, which implies (1.9).

If the points $(x_i, y_i)$, $2 \le i \le 5$, are not exactly as chosen above but lie in very small squares $\mathbf{B}_i$ centered on the above points, the calculations are the same, except the values of all our numerical constants will be slightly different, but the final conclusion $\partial\zeta/\partial y_1 \approx 1/2$ will remain valid.

## Acknowledgments

N. C. is partially supported by NSF Grant DMS-09-69187. The authors are grateful to the anonymous referees for many helpful suggestions.

## References

Adcock, R. (1877). Note on the method of least squares. *Analyst London* **4**, 183–184.

Ahn, S. J. (2004). *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space. LNCS* **3151**. Berlin: Springer.

Al-Sharadqah, A. and Chernov, N. (2009). Error analysis for circle fitting algorithms. *Electronic Journal of Statistics* **3**, 886–911. MR2540845

Al-Sharadqah, A. and Chernov, N. (2011). Statistical analysis of curve fitting methods in Errors-In-Variables models. *Theor. Imovir. ta Matem. Statyst.* **84**, 4–17. MR2857412

Anderson, T. W. (1976). Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. *Journal of the Royal Statistical Society, Ser. B* **38**, 1–36. MR0411025

Anderson, T. W. and Sawa, T. (1982). Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient. *Journal of the Royal Statistical Society, Ser. B* **44**, 52–62. MR0655374

Chan, N. N. (1965). On circular functional relationships. *Journal of the Royal Statistical Society, Ser. B* **27**, 45–56. MR0189163

Cheng, C. L. and Kukush, A. (2006). Non-existence of the moments of the adjusted least squares estimator in multivariate Errors-In-Variables model. *Metrika* **64**, 41–46. MR2242556

Chernov, N. I. and Ososkov, G. A. (1984). Effective algorithms for circle fitting. *Computer Physics Communications* **33**, 329–333.

Chernov N. (2010). *Circular and Linear Regression: Fitting Circles and Lines by Least Squares*. Boca Raton: Chapman & Hall. MR2723019

Chernov, N. (2011). Fitting circles to scattered data: Parameter estimates have no moments. *Metrika* **73**, 373–384. MR2785031

Kanatani, K. (2004). For geometric inference from images, what kind of statistical model is necessary? *Systems and Computers in Japan* **35**, 1–9.

Kåsa, I. (1976). A curve fitting procedure and its error analysis. *IEEE Trans. Inst. Meas.* **25**, 8–14.

Kukush, A., Markovsky, I. and Van Huffel, S. (2004). Consistent estimation in an implicit quadratic measurement error model. *Computational Statistics Data Analysis* **47**, 123–147. MR2087933

Nievergelt, Y. (2004). Fitting conics of specific types to data. *Linear Algebra and Its Applications* **378**, 1–30. MR2031782

Pratt, V. (1987). Direct least-squares fitting of algebraic surfaces. *Computer Graphics* **21**, 145–152. MR0987653

Taubin, G. (1991). Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations, with applications to edge and range image segmentation. *IEEE Trans. Pattern Analysis Machine Intelligence* **13**, 1115–1138.

Zelniker, E. and Clarkson, V. (2006). A statistical analysis of the Delogne–Kåsa method for fitting circles. *Digital Signal Processing* **16**, 498–522.

Department of Mathematics
University of Alabama at Birmingham
Birmingham, Alabama 35294
USA
E-mail: alsha1aa@uab.edu
chernov@uab.edu
qizhuo@uab.edu
URL: http://www.math.uab.edu/~chernov/cl