Geochronology Discuss.,
https://doi.org/10.5194/gchron-2020-4-RC4, 2020

Interactive comment on "Robust Isochron Calculation" *by* Roger Powell et al.

Anonymous Referee #4

Received and published: 23 April 2020

[11pt]extarticle[]graphicx[]color

framed

alltt [T1]fontenc [latin9]inputenc geometry verbose,tmargin=1in,bmargin=1in,lmargin=1in,rma float amsmath amsthm setspace [unicode=true,pdfusetitle, bookmarks=true,bookmarksnumbered=true,bookmarksopen=false, breaklinks=false,pdfborder=0 0 1,backref=false,colorlinks=false] hyperref

[bottom]footmisc

[style=authoryear,natbib=true,firstinits=true,backend=biber]biblatex

sortnamelast-first

inflFncs.bib

txfonts upgreek

upquote

Review of “Robust Isochron Calculation” by Powell et al. (gchron2020-4)

April 23, 2020

1 Summary

Prior to writing this review I have read the reviews by the three Anonymous Reviewers AN1, AN2, AN3, abd have also looked over a number the paper’s references. The application of robust regression straight line fits with errors in variables to isochron calculation is an interesting one, and it seems potentially important in analyzing such data. It is striking how sketchy the body of this paper is compared with the very well written and more detailed treatment of robust regression applied to isochron calculation by Powell et al. (2002). One concludes that the main contribution of the current paper is the computational algorithm in Appendix A, and that the paper is aimed at a very small audience of specialists who are quite familiar with the paper’s main reference. I could therefore just comment on the computational algorithm. But in case the authors and the journal want to a paper that is accesible by a wider audience, in which case reviewer AN3’s comments are relevant, I offer the comments below with that possibility in mind.

2 Improvements Needed to Serve a Wider Audience

2.1 Empirical Motivation

The use of robust statistical methods to deal with non-normal distributions associated with outliers is important. But the first step in justifying a paper that is applications oriented, as is this one, is to provide empirical evidence demonstrating that the chronic data indeed exhibits such characteristics. This should minimally be done in Section 3 for the data of Figure 6, by making a normal qqplot of the residuals from the linear fit. But, it would be better to do it at the beginning of the paper, in one way or another, for example: (a) display a qqplot of the residuals from the ultimate fit (point to Section 3 with regard to the fit), or (b) using residuals from an exploratory robust fit with a repeated medians estimate, as discussed in Siegel, (1982). "Robust Regression Using Repeated Medians". *Biometrika*, 69, 1, 242-244.

2.2 The Errors in Variables Problem

As was pointed out by AN1, the authors are trying to compute a robust straight line fit taking into account that their problem is an errors-in-variables (EIV) problem. While the authors do not state this clearly, it is immediately obvious from the paper's two York references and McClean reference that the focus is on an EIV problem. The authors need to provide up front the basic equations that describe the EIV problem and state the well-known results that classical least square (LS) estimate is biased (stating the bias formula), unless there is no error in the independent variable, and furthermore this bias does not go away as the sample size goes to infinity. (a "consistent" estimator is one that converges to the true parameter value, in a probabilistic sense as the sample size goes to infinity).

The main place that one sees clearly that the paper is trying solve a classic EIV problem is in equations (A1) - (A4) of the Appendix. But then in order to solve the problem, the values of the variances $\sigma^2_{x_k}$ and $\sigma^2_{y_k}$ and the correlations $\rho_{x_k y_k}$ need to be specified. One can infer from the comments in Section 2.5 and the ellipses in Figure 6, that those values were used for each x_x, y_k pair, but there is no substantive discussion of this. For the sake of reproducible research, the authors should make available the data set and values used for those variances and correlations.

However, based on Figure 6, one might assume that the $\sigma^2_{x_k}$ are relatively small enough for that data to be ignored and just use classical LS (no EIV). In fact the authors should compare the their result with that of classical LS. Of course for other chronic data sets the errors in the independent variable may be much larger?

2.3 The Explanation in Section 2.2

Like reviewer AN2, it took a repeated reading to figure out that the r_k are standardized versions of the regression residuals e_k. The explanation in Section 2.2. is quite confusing in several ways, the first of which is the sentence "The residual for data point, k, used, denoted r_k, is the scaling factor . . . ". It seems clear from reading Appendix A, that what is meant is that the r_k are the scaled residuals given be (A4), where the numerator e_k there is the raw residuals (except the sign should be changed in the numerator), and the denominator is the weight given to the residual to take care of the errors-in-variables aspect. The writers should just state this clearly in the body of the text, by properly introducing the EIV problem and how the denominator scaling is actually computed (one can infer this in Appendix A, but readers will appreciate a clear up-front explanation).

The sentence in the last line of page 3 states: "Although it is not obvious from the form of $\rho(r_k)$, HUBER is equivalent to bringing data points in to $\pm h$ when $\mid r_k \mid > h$.", is indeed not obvious. But the authors should make it obvious by plotting the shape of the $\psi(r_k) = \rho'(r_k)$, and emphasizing that $r_k = e_k/\sigma_{e_k}$, where that behavior will be totally obvious.

2.4 Concerning the mswd Criterion

The mswd is finally defined in equation (2), and I think the paper would be better served to introduce it early on in connection with a very brief but clear review of the nonlinear

LS EIV minimization problem statement. In any event the equation (2) should have a right-hand side where r_k is replaced by e_k/σ_{e_k} to continue to emphasize that the r_k are the standardized residuals. This is quite in keeping with equation (1) of Wendt & Carl (1991).

In equation (3) I think that the authors mean that **nmad(r)** is the standardized median of the **mad(r)**, the latter of which is more clearly written as $mad\{r_k\} = med\{|r_k - med\{r_k\}|\}$, and provide the normalization constant 1.4826. This is a very good thing to do, i.e., use a robust scale of the standardized residuals r_k.

If my assumption above is correct, the authors should reconsider the sentence "Given that s is based on a median, its magnitude depends on that half of the data that have the smallest absolute values of r", maybe modify it, maybe delete it.

2.5 Statistical Efficiency and Choice of Huber Psi Function Parameter

The terms "efficiency" first appears in the second line of Section 2.2 and the term "efficient" appears in the next to last paragraph of Section 2.2, and both appearances are followed by the phrase "(see below)". Based on that, I expected later on to see a definition of efficiency that is commonly used in statistics , and in robust statistics in particular. But the discussion of efficiency that finally occurs in Section 2.4 does not provide such a definition, and is not of little use. The efficiency (EFF) of a robust estimator $\hat{\theta}$ at a data distribution F is defined as that ratio of the variance of a "best" estimator at data distribution F to the variance of the robust estimator at data distribution F

$$\mathsf{EFF}(\hat{\theta}; F) = \mathsf{var}(\hat{\theta}_{best}; F)/\mathsf{var}(\hat{\theta}_{robust}; F)$$

usually expressed as a percent. In the context of the manuscript, the data distribution F is taken to be a normal distribution and the best estimator for a normal distribution is the

LS estimator. It is common practice to choose the tuning constant, i.e., h in the paper so that the estimator has 95% normal distribution efficiency. And there is a trade-off between normal distribution efficiency and robustness toward non-normal distributions that result in fat-tailed distributions and outliers. One can achieve more of the latter robustness by decreasing the normal distribution efficiency, and vice versa. I provide this level of detail because it is important with regard to the comments I make **below**.A an asymptotic version of efficiency of a robust estimator, in the context of the trade-off between normal distribution efficiency and robustness, can be found in Section 3.4 of Maronna et al. (2019).

2.6 Focus on Normal Mixture Distributions

A focus on specific normal mixture distributions is a seemingly natural thing to do. But it is not of much help unless the researcher is pretty certain the the data (the error term in the straight line regression model in the paper) really conforms to such normal mixture distributions, based on a sufficiently large data set size. This could be easily checked by examining normal qqplots of the residuals from the fit. Two component normal mixture models that generate outliers results in a normal qqplot with two approximately linear pieces, and often one finds three linear pieces that require a three component normal mixture models. The authors could easily check this for the residuals in their real data example.

If such analysis does not indicate that a normal mixture distribution is reasonable, then there is no point in focusing on a a normal mixture distribution.Furthermore, the point of robust estimation is that for data whose distribution has a central region that is normal, the variation in the tails and extent of outliers does not have much influence on the variability of the robust estimator.

2.7 Choice of Robust Loss Function and Associated Psi Function

As Powell et al. (2002) clearly indicates, the lead author of the current paper (and possibly co-authors) is quite familiar with some of the basic robustness material in Hampel et al. (2006), including the possibility of using not only the Huber psi function, but also a redescending psi function. The focus on the Huber psi function in the current manuscript is quite reasonable both intuitively and theoretically. With regard to the latter, the Huber psi function has a well-known optimal property of (asymptotically in sample size) minimizing the maximum variance over a mixture model where the central model is normal and contamination component can be any symmetric distribution (which therefore does not induce estimator bias). Furthermore , the optimization problem is convex in this case. The tuning parameter h in the present manuscript is conventionally chosen to achieve a specified normal distribution efficiency, typically 95%.

On the other hand, under the more realistic model that the contamination component is completely unrestricted (and therefore can lead to estimator bias), there is an optimal bias robust regression estimator, whose loss function is bounded (unlike the unbounded Huber loss function) and whose psi function smoothly redescends. This estimator is described in Section 5.8.1 of Maronna et al. (2019). This estimator also has a tuning parameter that is set at a user's desired normal distribution efficiency, again typically at 95%. Although the optimization problem is no longer convex, there exists a very good algorithm for computing the optimal bias robust estimator for a specified normal distribution efficient. This method is available in the R package RobStatTM available at CRAN uses that algorithm.

For the kind of data in Figure 6 of the manuscript, the Huber choice should do well since the independent variable leverage is not very large, but the optimal bias robust regression may do better and would be worth trying. For other kinds of data sets with large leverage outliers or very fat-tailed residuals, the optimal bias robust estimator may

provide a better solution. However, adapting this estimator to the EIV problem solved by the algorithm in Appendix A would be challenging.

Minor Comment

It seems quite inappropriate to use HUBER to refer to the M-estimator (Maximum-likelihood type estimator) using his favorite rho and psi function. It would be better to use M-estimator is the basic concept, and say that the paper focused on the Huber rho and psi. Similarly, use of YORK to refer to the algorithm he published seems awkward, particularly since he was not the only person to introduce that algorithm.

2.8 The Computational Algorithm

It seems that computation algorithm in Appendix A is the main contribution of the paper. While the Huber M-estimator optimization problem is convex and the numerical algorithm is known to converge, this EIV formulation is no longer convex, and though it may work well, the iterative algorithm might converge to a bad local minimum, or possibly fail to converge. It would be good to have a convergence proof, but that may be exceedingly difficult or impossible. In any event a good starting point helps a lot, and for the general multi-factor regression the L1 start is a good one.

However, for the simple straight line regression problem, the Siegel (1982) repeated median estimator, available in the R package `mblm` is likely to be even better as an initial condition for the Appendix A algorithm (the repeated median has low efficiency but has a breakdown point of 1/2, as compared with the L1 breakdown point of zero). One thought about the possible use of the repeated median estimator is the following. Use the repeated median estimator to define the slope needed in (A3) to define the weighted residuals in (A2), and hold it fixed while solving the robust regression problem. In this case you can use any of the robust regression computational methods available in the R package RobStatTM, including robust regression based on the Huber psi or the optimal bias robust psi mentioned int he previous subsection. And one can then do a few outer loop iterations to update the fixed values for (A3), and maybe even one or two iterations would suffice.