Small Variance Factors

Here are some of my thoughts on what to do if you get a very small VF from a Least Squares solution. Bruce Harvey, Sep 2011.

Generally a VF that is too big indicates we have an error and it is simply a matter of finding and correcting the error(s). A small VF is another matter. This is a problem that arises with GPS data, and occasionally with other types of data and in research environments – it is a bit of a black art. Questions about GPS variances and VFs are important and have been an issue since the early days of GPS surveying. This is the sort of thing where professional experience and expertise guide your solution. It is not a simple technical question. In that way it is somewhat similar to cadastral boundary determination when there is conflicting evidence and you have to decide what to do.

Background information about the Variance Factor?

First, a quick recall of some of my textbook's coverage of Variance Factors especially when they are less than 1. [Harvey (2009) Practical Least Squares and Statistics for Surveyors].

Here we call the estimated variance factor VF. It is also called the a posteriori variance factor or $\hat{\sigma}_o^2$.

$$VF = \hat{\sigma}_0^2 = \frac{v^T P v}{n-u} \approx \frac{\sum \left(v/s \right)^2}{n-u} = \frac{\left(v_1/s_1 \right)^2 + \left(v_2/s_2 \right)^2 + \left(v_3/s_3 \right)^2 + \ldots + \left(v_n/s_n \right)^2}{n-u}$$

Usually the a priori estimate of the VF, σ_o^2, is set to 1, which assumes the input VCV is reliable and does not need scaling. The statistical test is also called the variance ratio test, chi square test, or global model test. The two-tail test limits on the VF are:

$$\frac{1}{F_{1-\alpha/2, r_2, r_1}} < VF < F_{1-\alpha/2, r_1, r_2}$$

Typical values for acceptable limits of the VF are about: $0.6 < VF < 1.6$, but depend on confidence level chosen and degrees of freedom.

There are several possible reasons for VF to fail the statistical test. Many of these causes will cause larger residuals (corrections to observations) and thus the VF to be too large. Here we are going to look at reasons why a VF might be too small. A small VF means the observations fit better than we expect, the observations are "too good".

The main reason for a VF that is too small is a poor estimate of the **quality** (standard deviations and correlations) of the observations (i.e. the stochastic model). A more realistic input covariance matrix might be required. If the cause of the failure of the variance factor test is purely an error in the input standard deviations then: if the VF is too small then the input standard deviations are too large. But we also need to consider what happens if we ignore correlations between the input observations in the input Q matrix.

The effect of errors in the input statistical information on the estimated parameters (e.g. coordinates) is usually small but the effect on the estimated precision of the coordinates and the VF can be severe.

What effect do unchecked radiations and similar observations have on the VF? In a traverse survey a radiation to a point i.e. distance and direction will have v = 0, if there are no checks on those observations. So $(v/s)^2 = 0$ and the contribution of those two observations to the VF might appear to be zero, thus causing a smaller VF? No. Those two unchecked observations will also have zero redundancy numbers. $VF \approx \Sigma(v/s)^2 /(n-u)$. The two extra n are matched by two extra u, so n-u doesn't change.

One simple case that can cause a VF to be too small might be an error in the units you use in your program. Some researchers writing their own program (or similar) might have coordinates in units of metres and enter their standard deviations in units of mm. If you don't allow for that in the calculations then VF will be too small.

Another case is when we fit curves to data (or similar problems). If we increase the number of parameters in our model, such as using a very high order polynomial, then those parameters may absorb some of the noise (random errors). That will yield a very a small VF and poor redundancy. Adding too many parameters is not likely to happen with surveyors using existing network software solutions, but it can happen in a research environment.

We also need to be careful about how we are pre-processing our data. Is there something that we are doing that is forcing the data to fit together better than it should? For example, when we pre-process GPS observations around a triangle are the 'baselines' vectors forced to a perfect close? If so, then shouldn't we just process independent baselines / vectors? Or are we making some invalid assumptions about our models.

What should you do with the Variance Factor, in general?

Some people use the VF to scale the estimated VCV matrix of the parameters, some don't. If you buy a Least Squares program you should make sure you know whether the program is or is not automatically scaling the estimated standard deviations and error ellipses by √(VF).

Let's consider the various cases:

a) If the input variances and correlations of the observations are known to be correct then test whether VF is close to 1 (assuming a priori VF = 1).

If the test passes don't scale the estimated VCVs. It probably wouldn't make much difference anyway because multiplying by the square root of a number near 1 doesn't alter standard deviations and error ellipses of adjusted parameters very much.

If test fails something is wrong, so investigate.

b) If input variances and correlations of the observations are not well known then there are different opinions saying what you should do.

If the network's redundancy is small it is not clear what you should do. But if there are many observations compared to the number of parameters then either scale VCV of the parameters by the VF or change the input variances of the observations and do another solution. Iterate this until the VF test passes, provided of course that there are no gross or systematic errors in the observations or functional model or calculation errors.

People at one extreme say: always change the input variances of observations until the VF test passes. Whilst people at the other extreme say: if VF test fails something is wrong - investigate. In my opinion your decision depends on how well you know the input VCV of the observations and the functional model, and the size of the redundancy (degrees of freedom).

Small Variance Factors, in GPS?

Here we look at a problem from an experienced surveyor who is adjusting a large GPS net with a few hundred reduced static GPS+Glonass baselines and about 50 survey marks. Only 'independent' baselines were used. He did two adjustments – unconstrained (check of network consistency) and constrained. The constrained adjustment used selected SCIMS MGA coordinates with weighted heights. We usually find that solutions that hold more points fixed or with weighted constraints will have larger residuals and larger VF than solutions with minimum constraints or a freenet datum. That is because a constrained solution forces the data to fit to more points while a minimum constraint solution only requires data to "fit with itself" i.e. loop miscloses. Also, an error in coordinates of constrained points, or an error in scale in the data or the coordinates, will cause larger residuals and larger VF.

We can be unsure about what to use as the a priori input precisions for the data in both unconstrained and constrained adjustments. There appears to be many differing opinions. We should consider the consequences of our choice of options. While the a priori input precisions have only a small effect on the final coordinates, we wish to be realistic in the choice of a priori values because they affect outlier detection and the standard deviations of the adjusted coordinates (and things like error ellipses).

One option is to scale the a priori input precisions by √(VF) to bring the adjustment to VF = 1. In this example that would require horiz = 0.002m + 0.3 ppm and vert = 0.004m + 0.3ppm. Doing this for the unconstrained adjustment gives (optimistic?) values of input standard deviations of a few mm for each line.

Another option is to use the instrument manufacturer's specifications/advertised qualities for input standard deviations for that type of observation and processing, i.e. horiz = 0.003m + 0.5ppm, vert = 0.006m + 0.5ppm. Doing so yields a VF of about 0.45 (for the unconstrained solution). So in this case it appears the data fits (with itself) better than the manufacturer expects.

Another option is to use values recommended by the state or national authorities on these matters (e.g. horiz = 0.005m + 0.7ppm, vert = 0.015m + 2ppm). Doing so gives a VF of around 0.12 (for the unconstrained solution) and 0.23 (for the constrained solution) both of which are very small.

If input standard deviations of observations are chosen that give a very small VF, then that indicates all the v/s are small and that makes it difficult to find outliers. One option for outlier detection in this case is to disregard the usual test flag limits and simply investigate those observations with the largest |v/s|.

We can investigate loop miscloses, for example sum the ΔH around each loop of observed lines, or ΔX etc. Large miscloses will highlight possible gross errors. Also, (if there are many loops) the miscloses can be used to estimate the standard deviations of the observations. See, for example, Harvey (2009) Ch2, Q23. However the method in that example assumes each line has equal weight which means they are all similar length or the ppm component is insignificant. Even if we assume each line has equal weight in a loop, it can still be a useful independent indicator of the standard deviations of the observations. Some loop miscloses were calculated for the above data set and the results were found to be much as expected – if anything a bit better than expected. So no gross errors are likely and the observations really are precise and the input standard deviations are mostly consistent with the loop miscloses.

Correlations between the observations are often ignored. If the correlations are reliably known they should be used. If we scale the standard deviations then we should probably also scale the covariances so that the correlations don't change. If software allows for input of covariances or correlations between observations, then it is a useful exercise to compare two solutions. One solution using the covariances, and the other solution ignoring them (i.e. set to 0). I suspect there will be little change to the adjusted coordinates but the VF and Qx might change by a small amount. Adjustments were carried out on the above test data with the correlations disabled and there was very little change to the adjusted cords: 0.002m in the most far flung mark of the unconstrained adjustment. Mostly changes less than 0.001m. The VF went from 0.3 to 0.28.

Final thoughts:

If you have considered all the above points and investigated them, and you are using reasonable input standard deviations based on experience and you have carefully investigated the data preprocessing, loop miscloses and consistency with known coordinates of some marks… and your VF is still significantly <1, then I suggest you "live with it". Accept the low VF with what you consider to be reasonable input standard deviations and move on, your coordinates will be fine and your error ellipses etc will be a little conservative.

Further reading: Harvey (2009) for least squares matters, and Chris Rizos' GPS material at http://www.gmat.unsw.edu.au/snap/gps/gps_survey/chap9/chap9.htm and … /chap5/chap5.htm