Appendix 3. Bootstrap Confidence Intervals

The iterative two-step least-squares algorithm that we propose allows us to obtain estimates of the values of TFA and CS, as defined in our model. How good are these estimates? What is their precision? To answer these questions, we used a bootstrap procedure. Our choice was dictated by two observations: (*i*) we do not want to make any specific distributional assumptions on the errors in our model; and (*ii*) given the iterative nature of the algorithm, we do not have a close form expression that links the estimated values to the observations. The bootstrap is a very general statistical procedure that allows to learn about sampling variation using the one set of observation at hand (1). By creating a pool of bootstrap data sets, obtained by resampling with replacement from the actual data set, end evaluating the variability of our estimates across these bootstrap datasets, we can learn about the precision of our estimate. Although this conveys the general idea of how to gather information on sample variability by pulling ourselves up on the bootstraps of our current sample, there are a variety of implementations of the bootstrap. Because of the constraints that our model has on both the *P* and *A* parameters, we implemented what is known as parametric bootstrap: we used our data set to gather estimates $\overline{P}$ and $\overline{A}$ for *P* and *A*, respectively. This automatically leads to estimates of the errors in the gene expression values:

$$\overline{Err} = E - \overline{A}\overline{P}$$

We have then created a pool of bootstrap data sets E* by holding our estimates $\overline{P}$ and $\overline{A}$ as true and resampling from the estimated errors. So that

$$Err_i^* = Resample_i(\overline{Err})$$

and

$$E_i^* = \overline{A}\overline{P} + Err_i^*$$

Each of these bootstrap data sets E_i^* is input in the two-step iterative procedure and leads to estimated values for $\overline{A_i^*}$ and $\overline{P_i^*}$. We carried out 200 such bootstrap iterations. Using the quantiles of $\overline{P^*}$ and $\overline{A^*}$ in the bootstrap samples we were able to obtain 95% confidence intervals for all the parameter values.

Reference:

1. Efron, B. & Tibshirani, R. (1993) *An Introduction to the Bootstrap* (Chapman & Hall, London).