Appendix B: A Test of the Accuracy of Volume Estimation

The ability to analyze the volumes of regions is an attractive feature of PSP. Volume estimation, however, is nontrivial. To achieve a high degree of accuracy, a sophisticated numerical--either sampling- or nonsampling-based--method would have to be developed, which would increase computational time substantially. In this initial version of PSP, we decided to postpone such work and opted for a method that was reasonably accurate yet highly efficient. A region's volume is estimated using an ellipsoid defined by the eigen decomposition of the estimated covariance matrix of the region's sample. This ellipsoid approximation should be biased toward overestimation because an ellipsoid is an "evenly packed" object in every possible direction. Because of this, the amount of bias should be related to the degree of concavity, or "bowedness" of the region. Because no method exists for adjusting the bias toward the volume of an arbitrary multidimensional object with the same variance-covariance measures, it was necessary to measure the severity of the bias in order to determine how it might limit analyses using the measure.

We therefore designed a very stringent simulation test to assess the accuracy of volume estimation. A region in a multidimensional space was randomly generated in the following way. On an evenly spaced grid in this space, a cell was selected as a starting point of the region. The region was then expanded (i.e., grew) into one of its orthogonally neighboring cells in a random direction, to form a two-celled region. The region was expanded again into a neighboring cell, adjacent to any part of the region, not just the new cell. This process was repeated until this randomly-shaped region, *R*, filled a preset number or cells. Its volume, V_R, was estimated by PSP using the equation:

$$V_R = V_d (d+2)^{d/2} \left| \mathbf{S} \right|^{1/2}$$

where V_d is the volume of a d-dimensional sphere of radius 1, and $\mathbf{S}$ is the covariance matrix estimated from the MCMC output.

Accuracy was measured by comparing the estimated volumes of two of these arbitrarily-shaped regions. The ratio of these two volumes was used as the measure of accuracy, for which the known, true ratio (1) served as the reference point. It is important to understand that it is not possible to compare the estimated volume of a region with its "true" volume. The reason is that the scale of the parameter space is arbitrary, and so the value of a region's volume is not meaningful by itself. The value is meaningful only in comparison with another region's volume, and this is the test that we performed.

The simulation was carried out using a two-factorial design that yielded a variety of situations in which PSP's volume estimation might be used in real applications. The number of dimensions of the space was the first variable. There were three levels (4, 7, 10), which were chosen to cover a range of models in use in the discipline. The second factor was the number of cells used to form each of the two randomly-shaped regions (3, 6, 9). All three levels were completely crossed (e.g., 3&3, 3&6, 3,&9, etc.), yielding a total of six conditions, to increase the diversity of possible shapes being compared, which was meant to further increase the realism and difficulty of the test. Note that because of the algorithm's scale-adapting ability, as discussed in Appendix A, comparison of regions differing in their number of cells (e.g., 3 vs 9) is not an obstacle. The pair of regions being compared on each trial was simply scaled to have the same volume so that the true ratio should be one in all six comparisons.

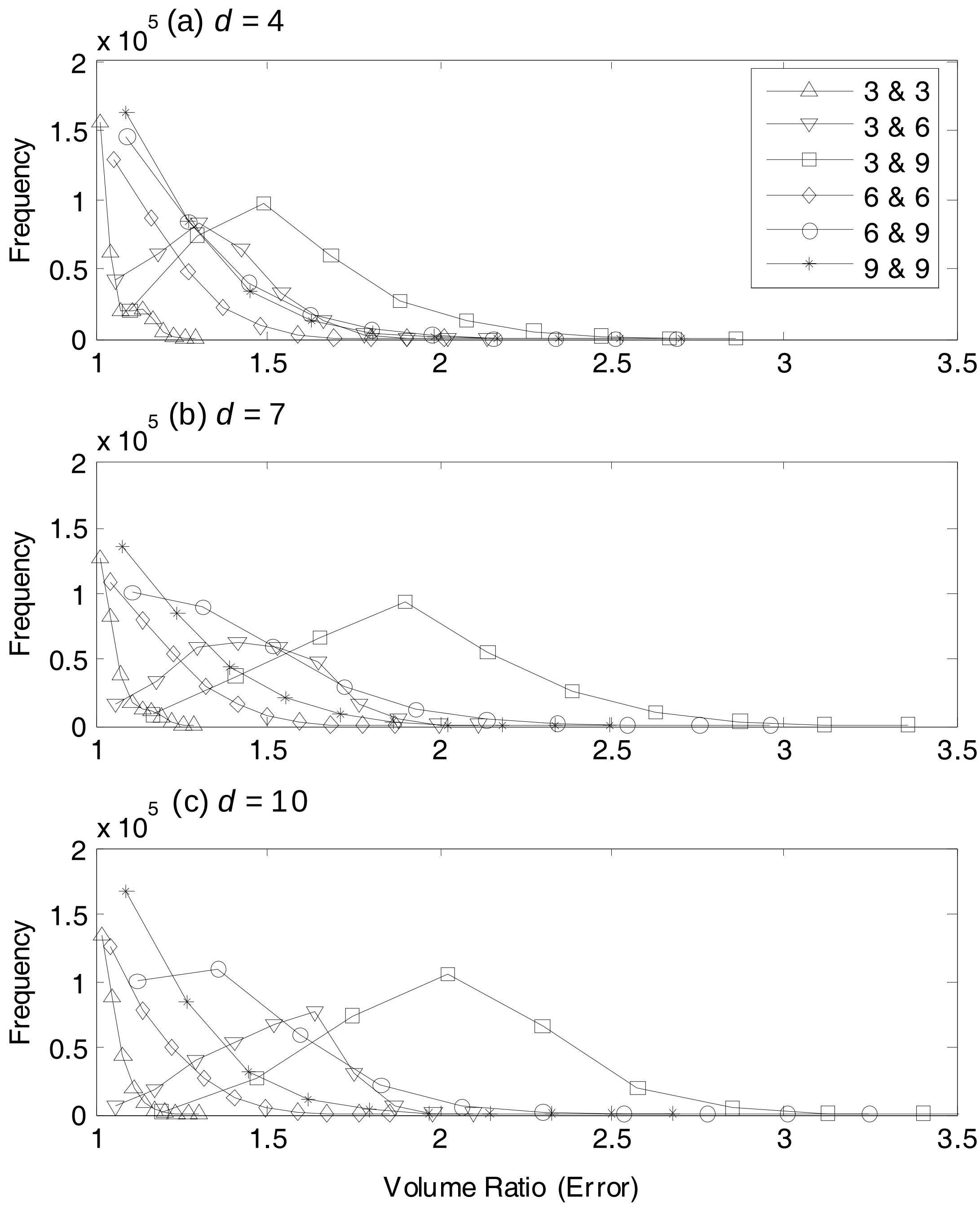
Because the region-generation method described above was highly unconstrained (requiring only continuity between cells), 6-cell and 9-cell regions had the potential to have extremely bizarre shapes (e.g., multi-pronged objects in various directions and dimensions). As a result, conditions 3&6 and 3&9 were likely to be the most challenging pairings. Volume estimation was pushed to the limits in these cases.

In each of the pairings across dimensionality, the region-generation and volume-estimation/comparison process was repeated 300,000 times. Regions were sampled 60,000 on each trial, which is equivalent to the five multiple runs used in the Merge and TRACE analyses. The results are shown in Figure 1A. Each panel shows the distribution of volume ratio estimates for a given dimensionality, d, across the different cell-number pairs. The ratio of the larger estimate to the smaller was used, so values greater than one indicate a bias.

Across most pairing trials, volume estimation was good, with the peaks of the distributions very near 1, and the tails falling off sharply so that few values exceeded 1.5. This was especially true for the same-cell-number pairings (3&3, 6&6, 9&9), whose distributions changed little across dimensionality. Mean ratios for these three conditions, collapsed over dimensionality, were 1.05, 1.16, and 1.21, respectively.

For the pairings in which cell number differed, bias was greater, with the distributions shifted to the right, most notably in the 3&9 condition. In addition, for these three cases, bias increased with dimensionality, as can be seen by comparing the distributions across graphs. Both of these increases in bias are probably due to the greater degree of concavity possible in the 6-celled and 9-celled regions relative to the 3-celled regions.

The results of this test indicate that although ellipsoid estimation can be biased, it is remarkably accurate much of the time, and fairly robust given the wide variety of region shapes



possible. Biases are frequently negligible. When they are more substantial, we have a good idea as to their cause, so improvements in volume estimation are currently being targeted to these situations. Even without making such improvements, however, the current volume estimation process can be trusted. For the models used in the present study, volume differences between regions were *exponentially* large. To view them all on the same graph, we had to plot them on a log scale (e.g., Figure 8). Thus, it goes without saying that the volume estimation method is more than sufficiently accurate to use rank-ordered estimates in analyzing the volumes of parameter regions no matter what their shape. In many instances, comparison on an interval scale would be appropriate.