

Supplementary Material

0.1 A detailed exploration of likelihood discrepancies

To explore the affect of varying the fixed off-diagonal rank on the likelihood surface, Figure 1 provides an analog of Figure 3 of the primary manuscript for a variety of ranks at a fixed HODLR level of 4 under otherwise identical simulation conditions. Like in the case of Figure 3, the color scale and contours are the same for each sub-graphic. The interpretation here is also similar: while there is slight variation between the MLEs as the off-diagonal rank is varied, every MLE for a given level is very near the MLE for the other levels, and the discrepancy in the likelihood value for each MLE is modest.

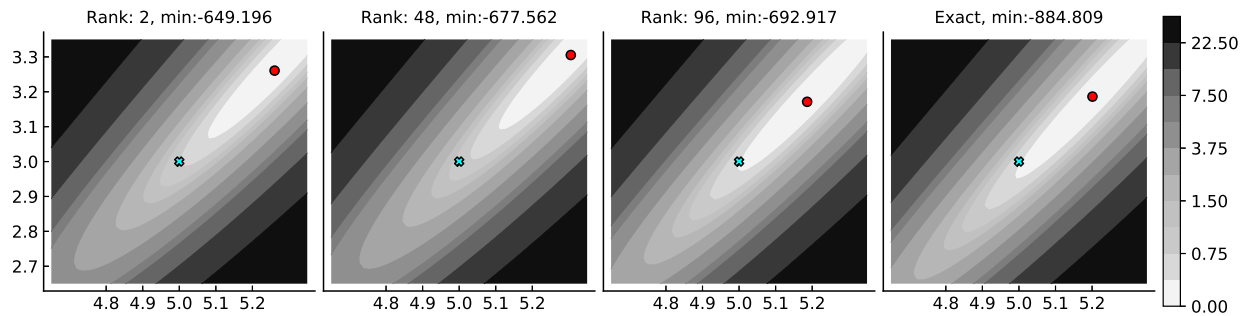


Figure 1: Centered log-likelihood surface for $n = 2^{12}$ data points randomly sampled on the box $[0, 100]^2$ with Matérn covariance with parameters $\theta_0 = 3$, $\theta_1 = 5$, and $\nu = 1$, with ν assumed known. The blue x is the “true” parameters of the simulation, and the red circle is the minimizer. The value at the minimizer is shown with the level in the title.

0.2 Relative Efficiency of Stochastic Optimization

As was discussed in the Numerical Results section, optimization to a relative tolerance of $\varepsilon = 10^{-8}$ using the stochastic gradient and Hessian may require substantially more evaluations of the likelihood and derivatives being used, although we do not see any evidence to suggest that the difference between the two numbers grows with n in any systematic way. Below, Tables 1 and 2 summarize the number of function evaluations required for the optimization of $l(\boldsymbol{\theta})$ and $l_H(\boldsymbol{\theta})$ that was performed to produce Figures 4 and 5.

Table 1: Minimum, median, and maximum number of function calls required for optimization to relative tolerance $\varepsilon = 10^{-8}$ for the Hessian-free application.

n	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}
$l_H(\boldsymbol{\theta})$	21,29,33	18,40,44	15,28,39	12,25,38	18,29,38	3,22,35	2,26,31
$l(\boldsymbol{\theta})$	2,17,22	9,17,21	2,17,24				

Table 2: Minimum, median, and maximum number of function calls required for optimization to relative tolerance $\varepsilon = 10^{-8}$ for the second-order trust region method.

n	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}
$l_H(\boldsymbol{\theta})$	4,8,11	5,8,10	4,8,11	6,7,12	5,7,13	5,9,23	3,9,13
$l(\boldsymbol{\theta})$	4,5,6	5,5,5	4,5,5				

0.3 Sample confidence ellipsoids for simulated data

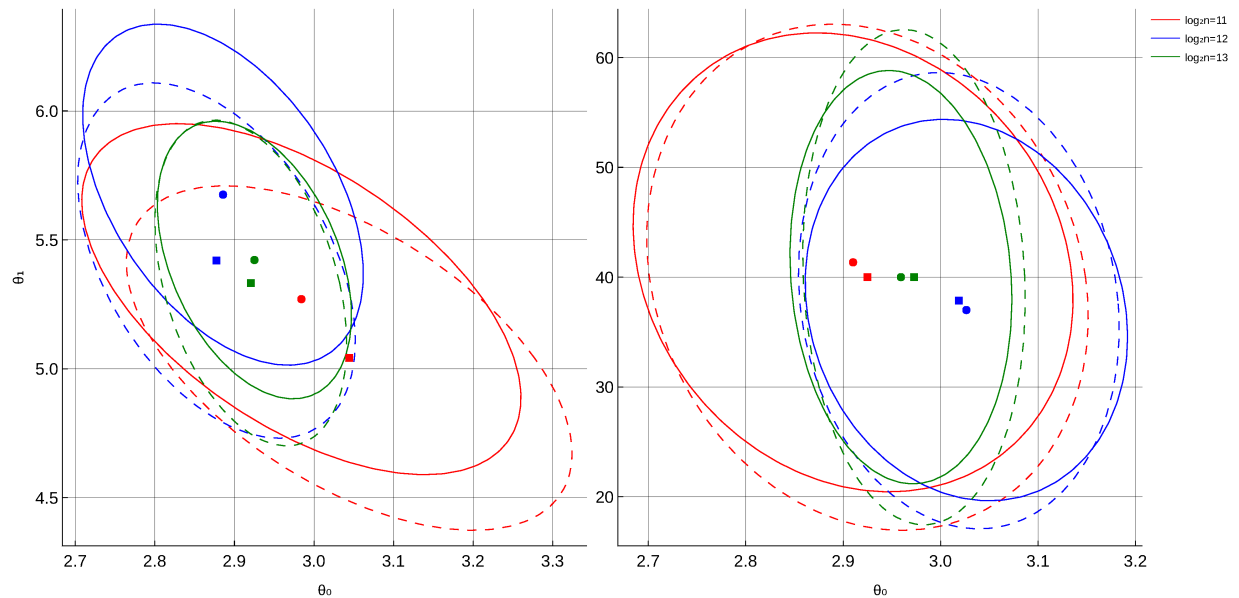


Figure 2: Approximated 95% confidence ellipses based on the exact (dashed) or stochastically approximated expected (solid) Fisher matrices at the exact (square) and approximated (circle) MLEs for the data from the first trial of Figures 4 and 5 respectively.

Figure 2 shows a comparison of the 95% confidence ellipses generated by the exact and approximated information matrices at the exact and approximated minimizers. Generally, the agreement is good, and there is no systematic over- or under-estimation of uncertainty in the approximated model, although there are differences in both the centers of the ellipses and the inferred negative correlation of the two parameters between the exact and approximated regions.