1 Computational Time

Here we assess the gains we achieve by implementing the interweaving srategy. Figure 1 shows the average amount of time in minutes it takes each of the base and interweaving samplers to obtain 1000 effectively independent draws from the marginal posterior distributions of both of V and W, i.e. given the ESS for either V or W and the time in minutes the sampler ran:

$$\text{Time per 1000 effective draws} = \frac{\text{Time}}{\text{ESS}} \times 1000.$$

For $T = 10$ the state sampler is clearly the fastest across all regions of the parameter space. The state-dist interweaving sampler is comparable when the signal to noise ratio is above one while the state-error interweaving sampler is comparable to the state sampler when the signal to noise ratio is below one for both V and W. When the signal to noise ratio is very close to one both the dist-error and triple interweaving samplers are similar to the state sampler.

When $T = 100$ the dist-error and triple interweaving samplers are the clear winners. The state sampler begins to slow down for W when the signal to noise ratio is above one and for V when the signal to noise ratio is below one. The major difference for the interweaving samplers in time per effective draw from $T = 10$ to $T = 100$ comes from how often the sampling algorithm can use adaptive rejection sampling to draw $W|V, \gamma_{0:T}, y_{1:T}$ and $V|W, \psi_{0:T}, y_{1:T}$. The target density for W is more likely to be log concave when the signal to noise ratio is low, but as T increases this constraint becomes more lenient. Similarly the target density for V is more likely to be log concave when the signal to noise ratio is high, but as T decreases this constraint becomes more lenient. The upshot is that for T small, at least one of the draws of $W|V, \gamma_{0:T}, y_{1:T}$ and $V|W, \psi_{0:T}, y_{1:T}$ cannot use adaptive rejection sampling unless the signal to noise ratio is very close to one, while when T is large adaptive rejection sampling can be almost always be used for both draws — the only exceptions being when the signal to noise ratio is either very large or very small.

Adaptive rejection sampling is relatively fast while the t approximation used to rejection sample from the target densities is often very inefficient. The result is the difference between the interweaving samplers for $T = 10$ and $T = 100$. When $T = 10$, adaptive rejection sampling fails for at least on of the two target densities unless the signal to noise ratio is very close to one, while for $T = 100$ the signal to noise ratio has to be very extreme before adaptive rejection sampling fails. Thus we see the dist-error and triple interweaving algorithms becoming significantly faster across most of the parameter space.

When $T = 1000$ the advantage is less clear. Specifically for low signal to noise ratios, the the dist-error and triple interweaving algorithms begin having trouble with efficiently drawing W. The state-error interweaving algorithm is similar in this case, except it appears to be a little faster when the signal to noise ratio is fairly close to one or above one while it is a little slower when the ratio is significantly below one. When the ratio is significantly below one, the state sampler actually appears to be the most efficient. The issue here is that all of the algorithms have mixing problems for either V or W for almost all signal to noise ratios when T is large, but the state sampler is a bit quicker to compute. The interweaving algorithms are superior whenever they mix better than the state sampler.

It is also worth asking whether interweaving buys us anything over an equivalent alternating or random kernel algorithm. Figure 2 attempts to answer this question again with timer per effective draw. The three algorithms are very similar with three exceptions. First the random kernel algorithm appears to be a little bit slower for some values of T and of the signal to noise ratio. Second, for $T = 100$ the GIS version of the algorithm appears to be at least as fast and sometimes faster than the corresponding alternating algorithm. Finally, this relationship reverses for $T = 1000$ — the alternating version of the algorithm appears to be at least as fast and sometimes faster as the corresponding GIS algorithm.

A quick comparison in Figure 3 shows that the triple alternating sampler may be the most generally fast sampler when T is large. For some values of the signal to noise ratio one algorithm or another might be faster than the triple alternating sampler for either V or W, but for most of the parameter space the triple
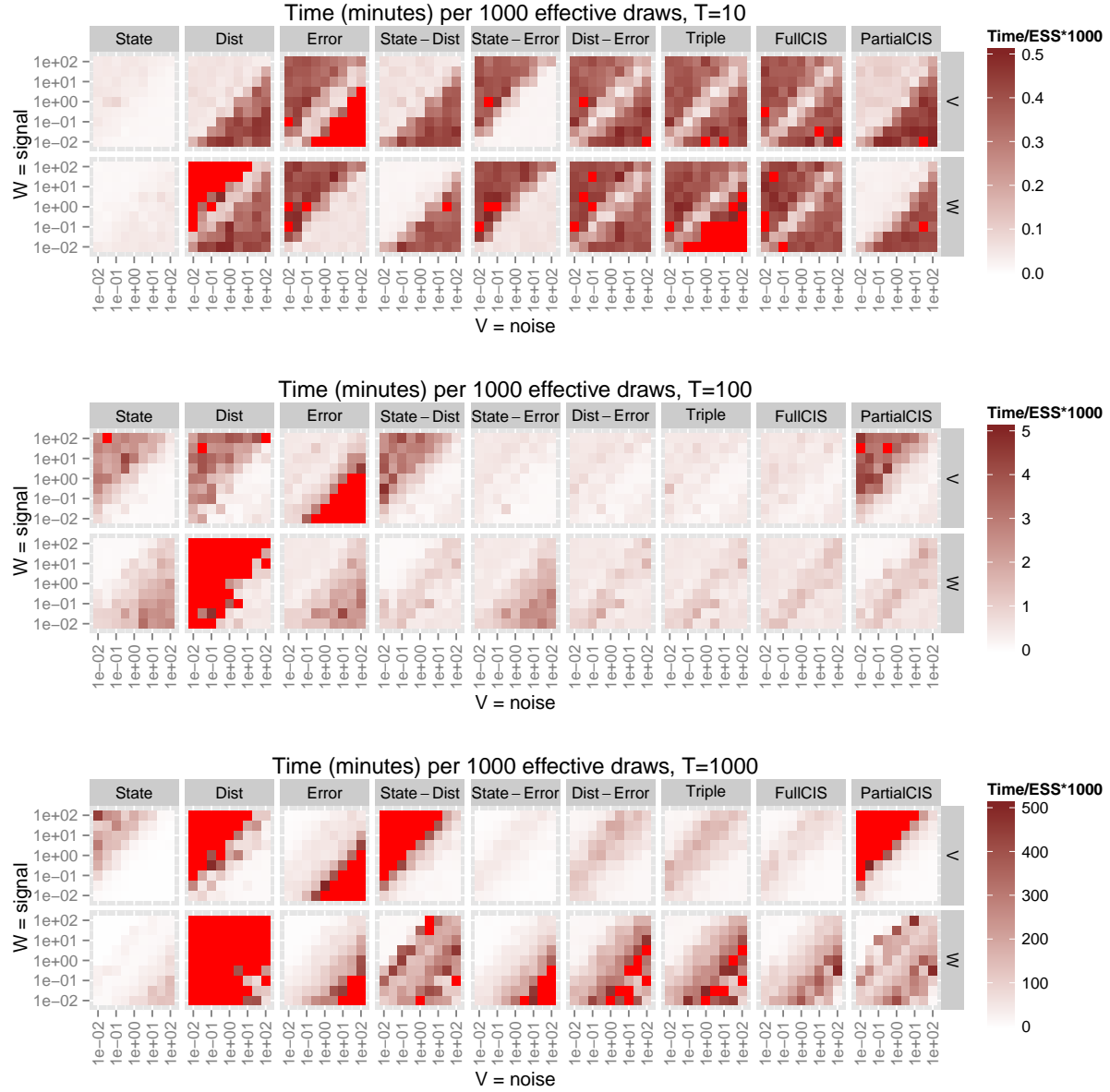
Figure 1: Time in minutes per 1000 effective draws in the posterior sampler for V and W, for $T = 10$, $T = 100$, and $T = 1000$, in the state, scaled disturbance and scaled error samplers and for all four GIS samplers based on any two or three of these. Horizontal and vertical axes indicate the true values of V and W respectively for the simulated data. The signal-to-noise ratio is constant moving up any diagonal. In the upper left the signal is high, in the lower right the noise is high. For plotting purposes, times larger than the top of the scale are displayed in bright red.

alternating sampler is pretty close to the others. The one exception seems to be for W when the signal to noise ratio is low where the state sampler is significantly faster, but any other sampler is significantly slower than both the state and triple alternating samplers. The state sampler is comparable, except its significantly slower than several other options for V when the signal to noise ratio is larger. The state-error alternating sampler is similar to the triple alternating sampler except it is faster when both samplers are relatively fast
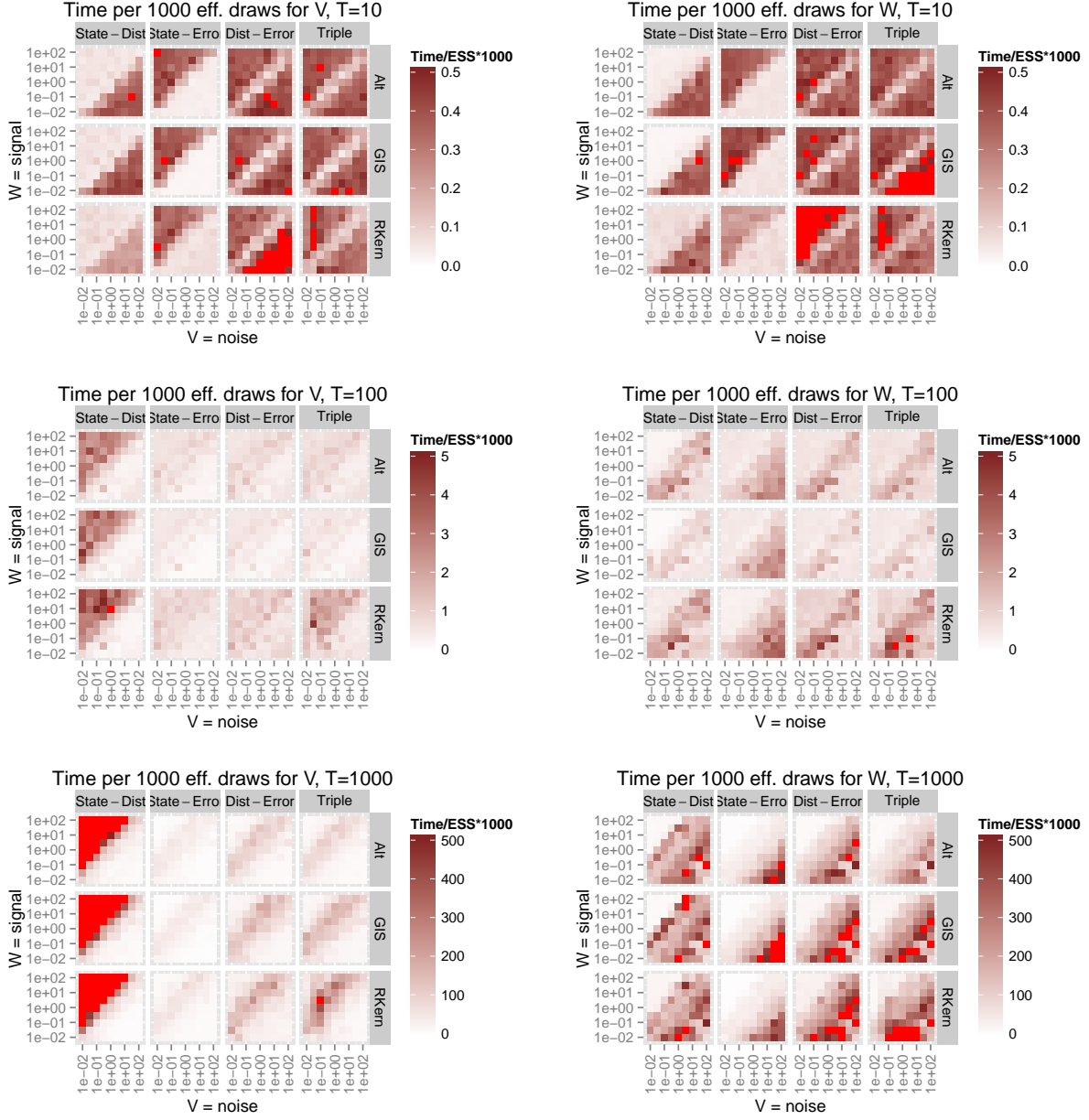
Figure 2: Time in minutes per 1000 effective draws in the posterior sampler for V and W, for $T = 10$, $T = 100$, and $T = 1000$, in the alternating, GIS, and random kernel samplers. Horizontal and vertical axes indicate the true values of V and W respectively for the simulated data. The signal-to-noise ratio is constant moving up any diagonal. In the upper left the signal is high, in the lower right the noise is high. For plotting purposes, times larger than the top of the scale are displayed in bright red.

and slower when both samplers are relatively slow. The upshot is that if the signal to noise ratio is probably not too much smaller than one and T is large, the state-error alternating sampler is the fastest for both V and W while when the ratio is probably not much larger than one the standard state sampler is the fastest for V and W. If the signal to noise ratio could be anything, the triple alternating sampler is the most likely

to be fairly fast, but a better strategy would be to run only run that sampler until it can obtain a rough value of the signal to noise ratio which can then be used to intilligently choose between the state-error alternating sampler and the state sampler.
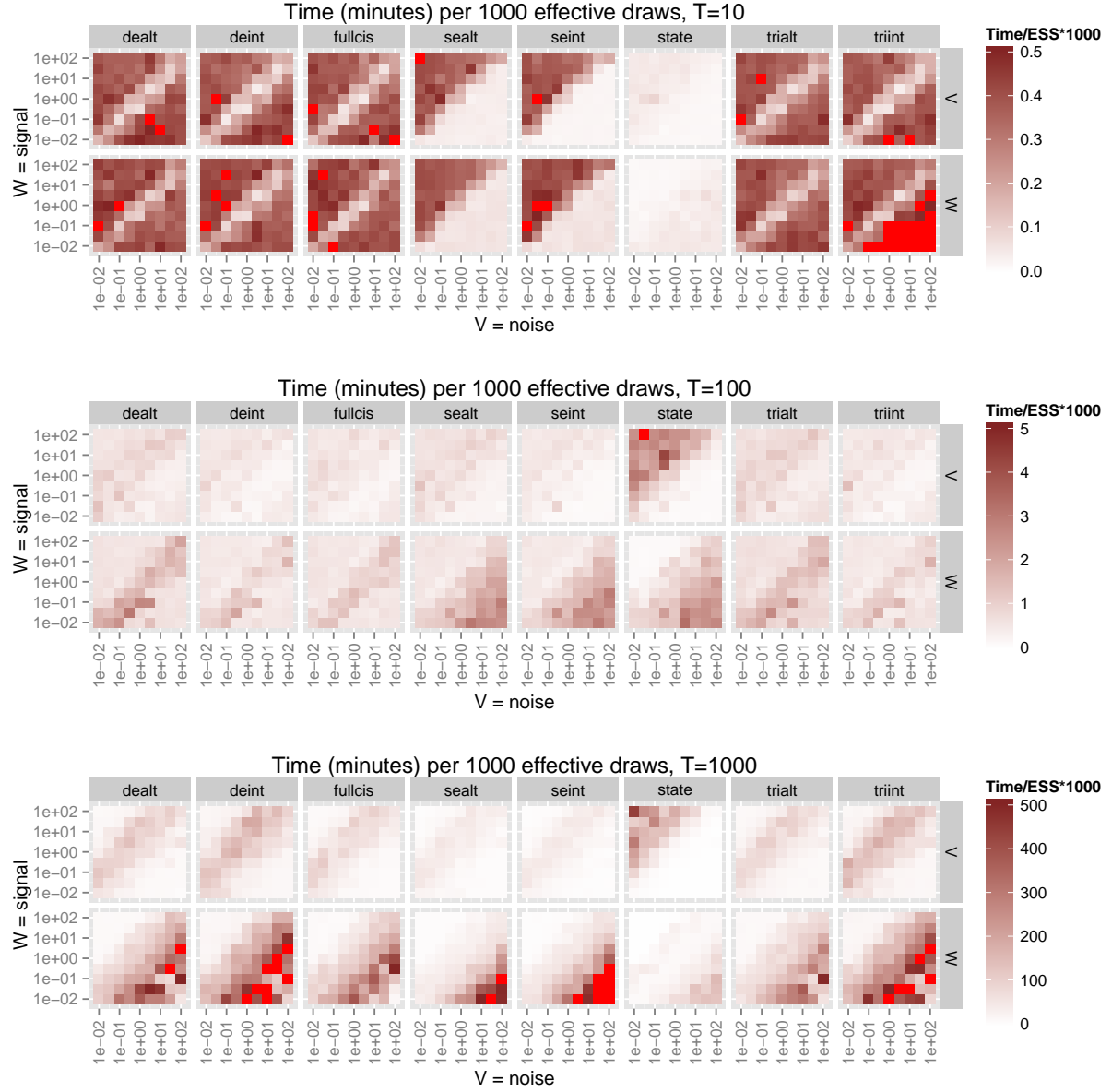

Figure 3: Time in minutes per 1000 effective draws in the posterior sampler for V and W, for $T = 10$, $T = 100$, and $T = 1000$, in the state, scaled disturbance and scaled error samplers and for all four GIS samplers based on any two or three of these. Horizontal and vertical axes indicate the true values of V and W respectively for the simulated data. The signal-to-noise ratio is constant moving up any diagonal. In the upper left the signal is high, in the lower right the noise is high. For plotting purposes, times larger than the top of the scale are displayed in bright red.
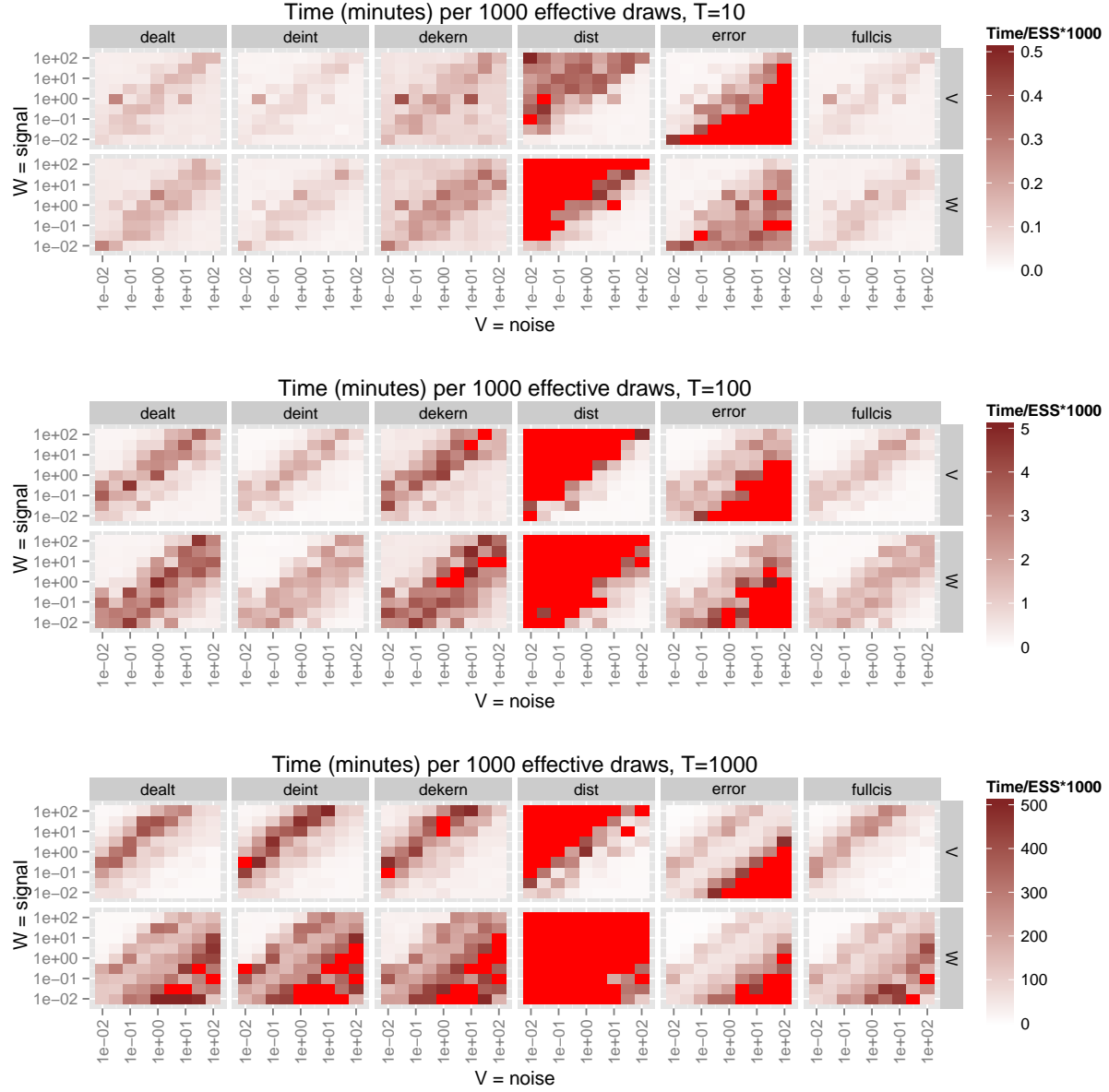
Figure 4: Time in minutes per 1000 effective draws in the posterior sampler for V and W, for $T = 10$, $T = 100$, and $T = 1000$, in the state, scaled disturbance and scaled error samplers and for all four GIS samplers based on any two or three of these. Horizontal and vertical axes indicate the true values of V and W respectively for the simulated data. The signal-to-noise ratio is constant moving up any diagonal. In the upper left the signal is high, in the lower right the noise is high. For plotting purposes, times larger than the top of the scale are displayed in bright red.