Simulation Details

E. Shannon Neeley

In the second simulation, we varied the parameters of the simulation to see how changing various aspects of the process could affect the normalization procedures. We varied the number of proteins (columns), the number of samples (rows), and the standard deviations (SD) for the λ_j, δ_p, and γ_p. Table 1 shows the MSE for all methods but for the contrasts comparing correlated columns and differential expression for 13 different simulation scenarios.

The simulations show that VS method nearly always outperforms both other methods. VS and ML are more comparable in scenarios 11 and 13, when the variability in γ_p is small or 0. The change in MSE in scenarios 1, 2, and 3 show that γ_p is able to be better estimated when there are more proteins used. HK normalization usually performs worse than ML and VS normalization, especially when there is a lot of variability in the sample loadings as in scenario 7.

VS normalization is slightly worse at estimating protein expression in scenario 6 when the standard deviation of the λ_i term has been reduced to 2, twice the variability of the random noise in the data. This is likely due to a breakdown in the method of parameter estimation as opposed to the failure of the model. One of the assumptions in the estimation is that the λ_j term dominates the c_{jp} term in the VS model so when this assumption is violated the γ_j are not estimated as well. In practice, however, we expect the assumption to hold. The sample loadings are quite variable with respect to the data. The MSE for this scenario is still smaller than the other methods.

If the variability of the λ_j gets too large, as in scenario 8, then none of the methods do a very good job of

Sim	Proteins	Samples	σ_λ	σ_δ	σ_γ	Correlation Min(MSE) = 0.002			Expression Min(MSE) = 2.00		
						HK	ML	VS	HK	ML	VS
1	30	200	4	2	0.10	0.019	0.026	0.002	4.29	3.56	2.48
2	70	200	4	2	0.10	0.016	0.020	0.002	3.65	2.49	1.72
3	10	200	4	2	0.10	0.018	0.043	0.013	4.14	3.34	2.40
4	30	75	4	2	0.10	0.018	0.027	0.006	4.12	3.00	2.61
5	30	800	4	2	0.10	0.018	0.031	0.001	3.98	3.43	2.54
6	30	200	2	2	0.10	0.007	0.005	0.003	3.85	3.41	2.78
7	30	200	10	2	0.10	0.100	0.162	0.002	9.13	5.90	2.25
8	30	200	1000	2	0.10	0.886	1.377	0.140	40361.10	15238.19	321.23
9	30	200	4	1	0.10	0.019	0.027	0.003	4.51	3.75	2.22
10	30	200	4	5	0.10	0.017	0.024	0.005	4.23	3.50	2.47
11	30	200	4	2	0.05	0.008	0.006	0.003	3.18	2.52	2.24
12	30	200	4	2	0.50	0.305	0.380	0.003	66.91	55.89	3.62
13	30	200	4	2	0.00	0.007	0.003	0.003	2.39	1.79	1.92

Table 1: Results from 13 simulations scenarios that compare housekeeping (HK), median loading (ML), and variable slope (VS) normalization with median loading (ML) normalization.

estimation, though VS normalization is much better than the others. VS normalization also performs much better when there is a lot of variability in the γ_p (scenario 12). Scenario 13 shows that if the VS model is wrong such that $\gamma_p \equiv 1$ for all p, the estimation procedure does not grossly compromise the results. In practice we do not think this scenario is realistic.