KS, of Cramer and von Mises (which is equivalent to the energy test of Székely and Rizzo (2004) on ranks), referred to as CVM, and of Anderson and Darling, referred to as AD.

We examined the distributions depicted in Figure 2. The three scenarios in the third row were examined in Jiang et al. (2014). The remaining scenarios were chosen to have different numbers of intersections in the densities, ranging from 2 to 18, in order to examine the effect of partition size m on power when the optimal partition size increases, as well as verify that the regularized statistic has good power. The scenarios also differ by the range of support of where the differences in the distributions lie (specifically, in the first and third scenario in the first row the difference between the distributions is very local), since this makes the comparison between the two aggregation methods interesting. We considered symmetric as well as asymmetric distributions. Gaussian shift and scale setups were considered in Appendix G. Such setups are less interesting in the context of this work, because if the two distributions differ only in shift or scale then specialized tests such as Wilcoxon rank-sum for shift will be preferable, but it is important to know that the suggested tests do not break down in this case. We used 20000 simulated data sets, in each of the configurations of Figure 2.

Table 1 and Figure 3 show the power for the setups in Figure 2. These results show that if the number of intersections of the two densities is at least four, tests statistics with $m \geq 4$ have an advantage. Since the classical competitors, KS, CVM and AD, are based on $m = 2$, they perform far worse in these setups. Moreover, although HHG and DS have better power than the classical tests, HHG is essentially an $m \leq 3$ test, and DS penalizes large ms severely, therefore their power is still too low when fine partitioning is advantageous. The minimum p-value statistic, which does not require to preset m, is remarkably efficient: in Figure 2 we see that in all settings considered,