Supplementary Material for

A Cross-validated Ensemble Approach to Robust Hypothesis Testing of Continuous Nonlinear Interactions:

Application to Nutrition-Environment Studies

Contents

A Derivation for Ensemble Kernel Matrix

Given the ensemble hat matrix $\widehat{\mathbf{A}}$ in Section 4, we consider how to identify the ensemble kernel matrix $\widehat{\mathbf{K}}$ by solving:

$$\widehat{\mathbf{K}}(\widehat{\mathbf{K}}+\lambda_{\mathbf{K}}\mathbf{I})^{-1}=\widehat{\mathbf{A}}.$$

Specifically, if denote $(\mathbf{U}_A, \mathbf{U}_K)$ and $(\{\delta_{A,k}\}_{k=1}^n, \{\delta_{K,k}\}_{k=1}^n)$ the eigenvector and eigenvalues of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{K}}$, respectively, then the above system reduces to:

$$\mathbf{U}_A diag\Big(\delta_{A,k}\Big)\mathbf{U}_A^T=\mathbf{U}_K diag\Big(\frac{\delta_{K,k}}{\delta_{K,k}+\lambda_{\mathbf{K}}}\Big)\mathbf{U}_K^T$$

and adopts closed form solution $\mathbf{U}_K=\mathbf{U}_A$ and $\delta_{K,k}=\lambda_{\mathbf{K}}\frac{\delta_{A,k}}{1-\delta_{A,k}}$. Therefore the ensemble kernel matrix $\widehat{\mathbf{K}}$ is estimated as:

$$\widehat{\mathbf{K}}=\lambda_{\mathbf{K}}*\mathbf{U}_A diag\Big(\frac{\delta_{A,k}}{1-\delta_{A,k}}\Big)\mathbf{U}_A^T.$$

Choice of ensemble tunning parameter $\lambda_{\mathbf{K}}$

Notice that we have left the "ensemble tunning parameter" $\lambda_{\mathbf{K}}$ unspecified. In practice, $\lambda_{\mathbf{K}}$ serves only as a constant scaling factor for the kernel matrix $\mathbf{K}$, whose exact value does not impact either the prediction or the p-value calculation, since both procedures are scale invariant with respect to the kernel matrix. Therefore it can be set to a value of our choice. One common choice for $\lambda_{\mathbf{K}}$ is to set $\lambda_{\mathbf{K}}=min\Big(1,(\sum_{k=1}^n\frac{\delta_{A,k}}{1-\delta_{A,k}})^{-1}\Big)$ such that $tr(\widehat{\mathbf{K}})\leq 1$, this is because the Rademacher complexity of the overall ensemble can be upper-bounded as a function of $tr(\widehat{\mathbf{K}})$ [6]. Another interesting choice is $\lambda_{\mathbf{K}}=O\Big(min(\{\hat{\lambda}_d\}_{d=1}^D)\Big)$. Intuitively, this means the tunning parameter for ensemble kernel matrix $\widehat{\mathbf{K}}$ should grow in the same rate as the tunning parameter for the best-peforming base kernel, as the ensemble kernel matrix is expected to perform as well or better than the best-performing base kernel. Further, such choice provides guarantee on the generalization performance of the ensemble by bounding the ensemble kernel matrix's decay rate of the tail sum of the eigenvalues [7].

B Discussion on theoretical aspects of CVEK

B.1 Oracle Selection

We remind readers that the CVEK's ensemble strategy (Step 2) is similar to that of Jackknife Model Averaging (JMA) [5] in the case of leave-one-out cross validation and the Super Learner [10] in the case of K-fold cross validation. Its oracle property in model selection has been established both asymptotically and in finite-sample [9]. That is, on average, given the same set of estimated kernel predictors $\{\hat{h}_d\}_{d=1}^{D}$, the behavior of the ensemble made by the cross-validated selector converges in $O(\frac{1}{n})$ rate to the "oracle ensemble" made by an oracle that has access to infinite amount of validation data. Consequently, under the null hypothesis, given a set of base kernels $\{\hat{h}_d\}_{d=1}^{D}$ with diverse mathematical properties, the oracle selection property of CVEK guarantees the selection of a model ensemble that best describes the data, thereby resulting in correct Type I error by mitigating model misspecification under the null.

B.2 Generalization

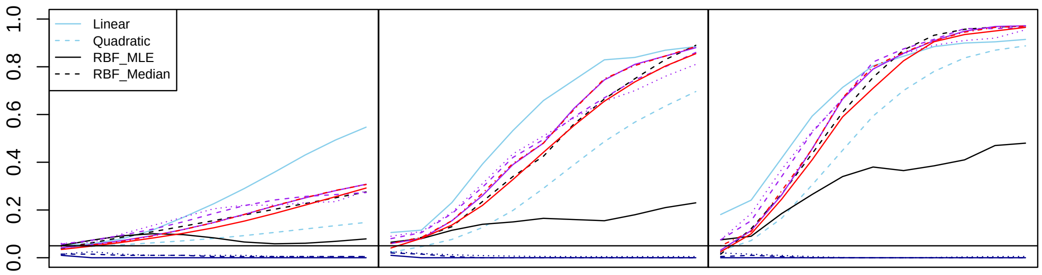
In limited samples, the oracle property does not ensure a powerful test, and the estimator's *generalization* property, i.e. how fast the estimator's behavior approaches its asymptotic counterpart, must also be taken into consideration to guarantee good power under the alternative. This is because even under a oracle selector, an ensemble of flexible estimators in finite samples will still overfit the interaction effect under the alternative, leading to a test with low power. To explore this issue, we notice that CVEK's ensemble form corresponds to that of the *Ensemble of Kernel Predictors* (EKP) [4], for which, under arbitrary choice of base kernels, the generalization error of the ensemble estimator converges at a rate of at least $O(\frac{1}{\sqrt{n}})$. However, for most epidemiological studies with small to moderate samples, a nonparametric rate of $O(\frac{1}{\sqrt{n}})$ may be too slow to merit powerful inference. It is therefore of great practical interest to understand if the ensemble's generalization performance can be improved beyond the rate of $O(\frac{1}{\sqrt{n}})$ with a careful selection of base kernels. To this end, we draw upon a classical result from the RKHS literature that the generalization error of a kernel-based estimator is bounded above by its RKHS's local Rademacher complexity [2], a measure of the richness of the class of candidate functions that is characterized by the rate of eigenvalue decay of its kernel function [7]. Consequently, since the learned ensemble estimator $\widehat{\mathbf{h}} = \sum_{d=1}^{D} \widehat{u}_d \mathbf{h}_d$ lies in convex combination of the RKHSs generated by the base kernels, the generalization property of the CVEK estimator is explicitly characterized by the generalization property of the "selected" base estimators (i.e. the $\hat{\mathbf{h}}_d$ assigned non-zero weight by the ensemble) in terms of their respective rates of eigenvalue decay. Moreover, it can be shown that for special classes of kernels, the generalization error rate can indeed be improved upon: finite-rank kernels (e.g. linear and polynomial kernels) are able to achieve $O(\frac{1}{n})$, while kernel families with exponential rate eigenvalue decay (e.g. Gaussian RBF kernel) can achieve a rate of $O(\frac{log(n)}{n})$ [3].

B.3 Choice for Base Kernels

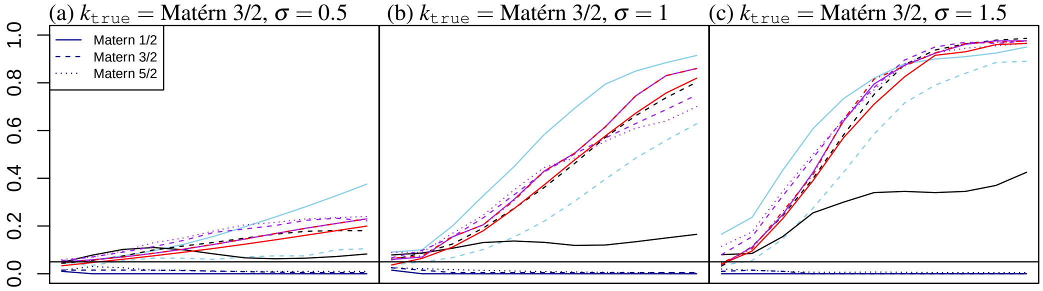
In combination of the CVEK's oracle selection property, the above result suggests that, if there exists sets of parametric kernels in the library, CVEK behaves as a parametric model by achieving a rate of $O(\frac{1}{n})$ if the data-generation function is indeed parametric. For more complex data-generation mechanism, CVEK is able to achieve a rate of $O(\frac{log(n)}{n})$ by including sets of Gaussian RBF kernels in the library. Consequently, we recommend practitioners to construct the kernel library with a mix of parametric kernels (linear, polynomial) and smooth kernels with exponential rate in eigendecay (e.g. a collection of Gaussian RBF kernel with different fixed spatial smoothness parameters). We notice that despite the extreme smoothness of these base kernels, the resulting ensemble is in fact very flexible and hence does not risk underfitting the data generation function. This is because by Bochner's theorem, fitting the model with an ensemble of Gaussian RBF kernels is equivalent to approximating the spectral density of the function with convex mixtures of Gaussian densities, which is shown to be capable of approximating arbitrary continuous density with compact support in $\mathbb{R}^d$ [1]. More generally, practitioners should be careful about using an ensemble containing only flexible kernels, since in order to represent a large function space, these kernel functions usually have heavy-tailed spectral densities and therefore slow eigenvalue decay [7]. Finally, we notice that there also exists interesting kernel families that lie outside the usual scope of theoretical analysis, but tend to do well in practice. One such example is the Neural Network kernel [8]. We investigate the performance of these kernels and compare to the performance of the Gaussian RBF ensemble in Section 4.

C Additional Simulation Results

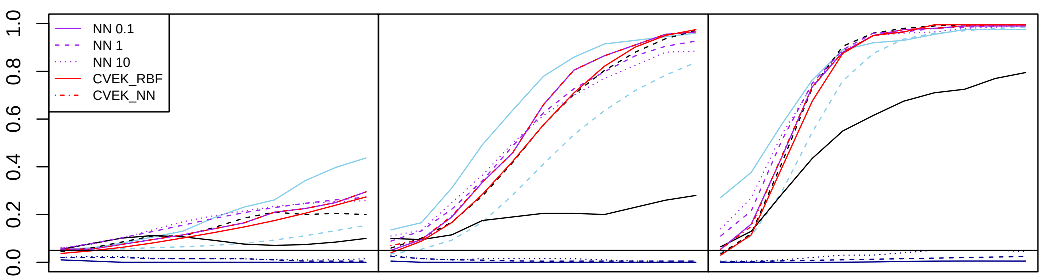
C.1 Additional Figures



(a) $k_{\mathtt{true}}$ = Matérn 3/2, $\sigma = 0.5$ (b) $k_{\mathtt{true}}$ = Matérn 3/2, $\sigma = 1$ (c) $k_{\mathtt{true}}$ = Matérn 3/2, $\sigma = 1.5$

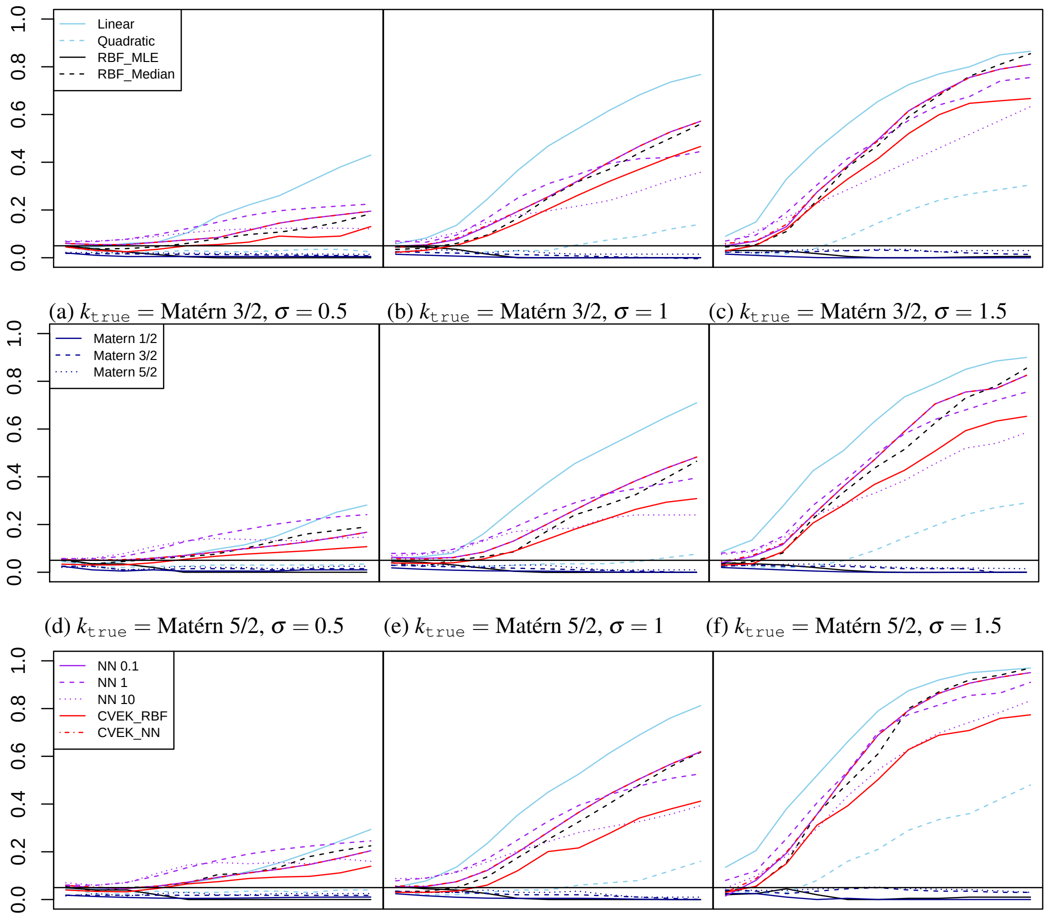


(d) $k_{\mathtt{true}}$ = Matérn 5/2, $\sigma = 0.5$ (e) $k_{\mathtt{true}}$ = Matérn 5/2, $\sigma = 1$ (f) $k_{\mathtt{true}}$ = Matérn 5/2, $\sigma = 1.5$



(g) $k_{\mathtt{true}}$ = Gaussian RBF, $\sigma = 0.5$ (h) $k_{\mathtt{true}}$ = Gaussian RBF, $\sigma = 1$ (i) $k_{\mathtt{true}}$ = Gaussian RBF, $\sigma = 1.5$

Figure C.1: Estimated $\widehat{P}(p < 0.05)$ (y-axis) as a function of Interaction Strength $\delta \in [0, 1]$ (x-axis) for $n = 200$ and $p_1 = p_2 = 6$.
Sky Blue: Linear (Solid) and Quadratic (Dashed) Kernels, **Black**: RBF-Median (Solid) and RBF-MLE (Dashed), **Dark Blue**: Matérn Kernels with $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, **Purple**: Neural Network Kernels with $\sigma = 0.1, 1, 10$, **Red**: CVEK based on RBF (Solid) and Neural Networks (Dashed). Horizontal line marks the test's significance level (0.05). When $\delta = 0$, $\widehat{P}$ should be below this line.

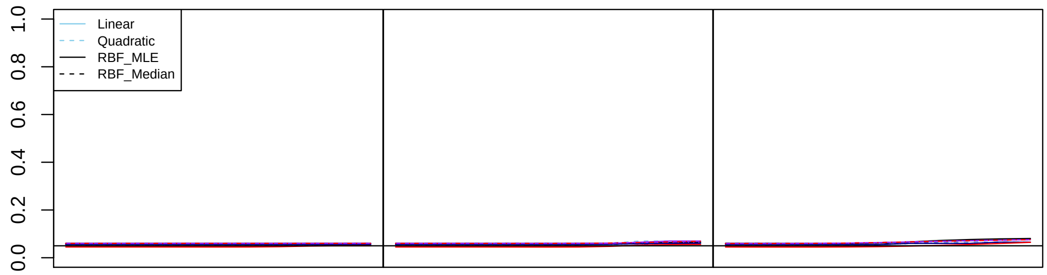


(a) $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 0.5$ (b) $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 1$ (c) $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 1.5$

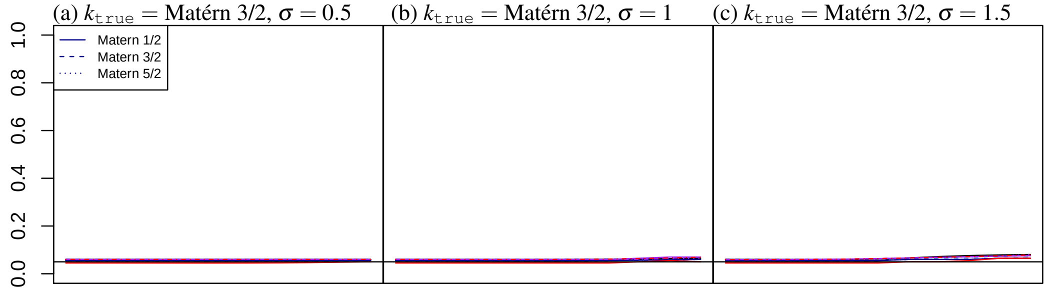
(d) $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 0.5$ (e) $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 1$ (f) $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 1.5$

(g) $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 0.5$ (h) $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 1$ (i) $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 1.5$

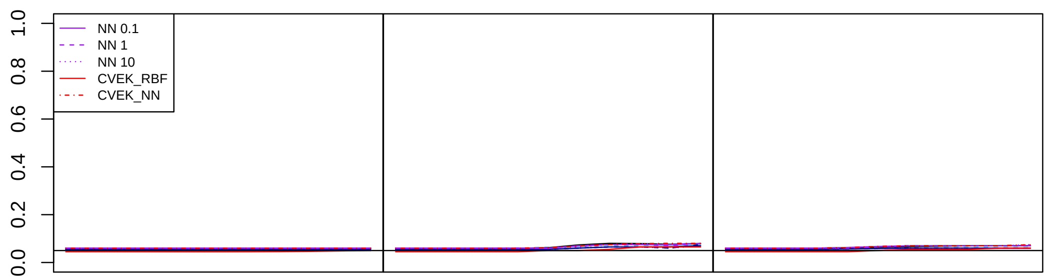
Figure C.2: Estimated $\widehat{P}(p < 0.05)$ (y-axis) as a function of Interaction Strength $\delta \in [0, 1]$ (x-axis) for $n = 200$ and $p_1 = p_2 = 10$.
Sky Blue: Linear (Solid) and Quadratic (Dashed) Kernels, **Black**: RBF-Median (Solid) and RBF-MLE (Dashed), **Dark Blue**: Matérn Kernels with $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, **Purple**: Neural Network Kernels with $\sigma = 0.1, 1, 10$, **Red**: CVEK based on RBF (Solid) and Neural Networks (Dashed).
Horizontal line marks the test's significance level (0.05). When $\delta = 0$, $\widehat{P}$ should be below this line.



(a) $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 0.5$ (b) $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 1$ (c) $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 1.5$

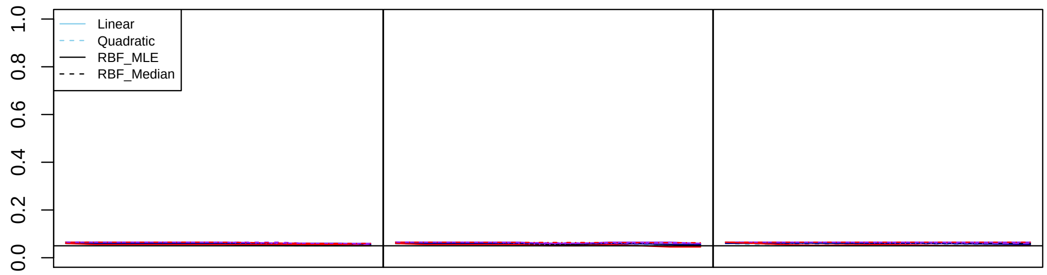


(d) $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 0.5$ (e) $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 1$ (f) $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 1.5$

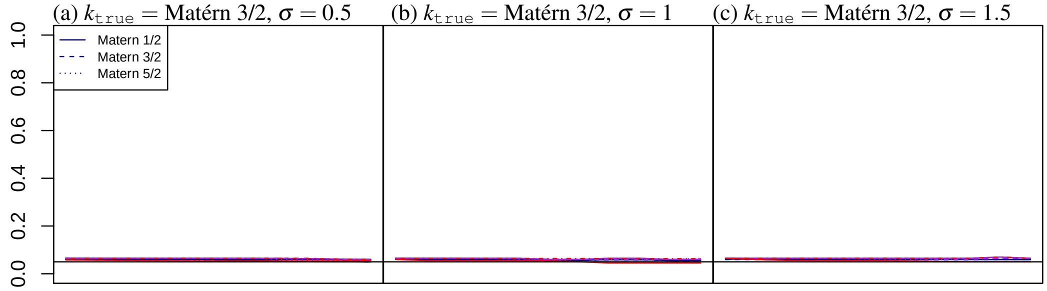


(g) $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 0.5$ (h) $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 1$ (i) $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 1.5$

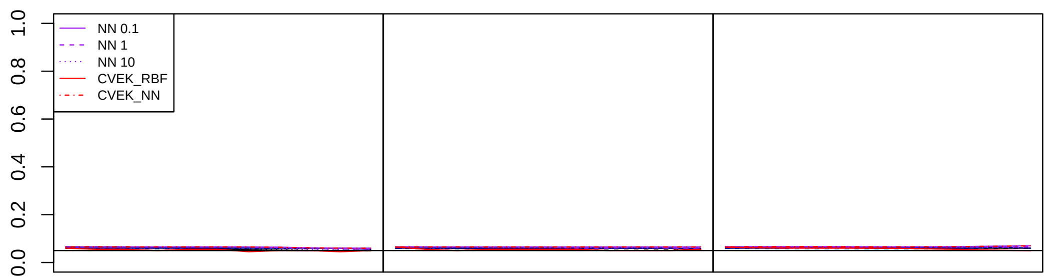
Figure C.3: Estimated $\widehat{P}(p < 0.05)$ (y-axis) as a function of Interaction Strength $\delta \in [0,1]$ (x-axis). For $n = 200$, $p_1 = p_2 = 3$ and residual distribution $t_{df=5}$.
Sky Blue: Linear (Solid) and Quadratic (Dashed) Kernels, **Black**: RBF-Median (Solid) and RBF-MLE (Dashed), **Dark Blue**: Matérn Kernels with $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, **Purple**: Neural Network Kernels with $\sigma = 0.1, 1, 10$, **Red**: CVEK based on RBF (Solid) and Neural Networks (Dashed).
Horizontal line marks the test's significance level (0.05). When $\delta = 0$, $\widehat{P}$ should be below this line.



(a) $k_{\mathtt{true}}$ = Matérn 3/2, $\sigma = 0.5$ (b) $k_{\mathtt{true}}$ = Matérn 3/2, $\sigma = 1$ (c) $k_{\mathtt{true}}$ = Matérn 3/2, $\sigma = 1.5$



(d) $k_{\mathtt{true}}$ = Matérn 5/2, $\sigma = 0.5$ (e) $k_{\mathtt{true}}$ = Matérn 5/2, $\sigma = 1$ (f) $k_{\mathtt{true}}$ = Matérn 5/2, $\sigma = 1.5$



(g) $k_{\mathtt{true}}$ = Gaussian RBF, $\sigma = 0.5$ (h) $k_{\mathtt{true}}$ = Gaussian RBF, $\sigma = 1$ (i) $k_{\mathtt{true}}$ = Gaussian RBF, $\sigma = 1.5$

Figure C.4: Estimated $\widehat{P}(p < 0.05)$ (y-axis) as a function of Interaction Strength $\delta \in [0, 1]$ (x-axis). For $n = 200$, $p_1 = p_2 = 3$ and residual distribution $t_{df=10}$.
Sky Blue: Linear (Solid) and Quadratic (Dashed) Kernels, **Black**: RBF-Median (Solid) and RBF-MLE (Dashed), **Dark Blue**: Matérn Kernels with $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, **Purple**: Neural Network Kernels with $\sigma = 0.1, 1, 10$, **Red**: CVEK based on RBF (Solid) and Neural Networks (Dashed).
Horizontal line marks the test's significance level (0.05). When $\delta = 0$, $\widehat{P}$ should be below this line.

C.2 Additional Observations

In this section we document the value of estimated $\widehat{P}(p < 0.05)$ from the simulation presented in Section 6 (Simulation Experiment) of the paper. Recall that the simulation data is generated from below mechanism:

$$y_i = h_1(\mathbf{x}_{i,1}) + h_2(\mathbf{x}_{i,2}) + \delta * h_{12}(\mathbf{x}_{i,1}, \mathbf{x}_{i,2}) + \varepsilon_i$$

where h_i's are functions with unit norm sampled from the reproducing kernel Hilbert spaces (RKHSs) generated by $k_{\texttt{true}}$, and the data is then fitted using Gaussian process with $k_{\texttt{model}}$.

Each table documents the $\widehat{P}(p < 0.05)$ resulted from fixing $k_{\texttt{true}}$ to a Matérn kernel with specific value of smoothness parameter ν and complexity parameter σ, and then varying the strength of the interaction $\delta \in [0, 1]$ and the model kernel $k_{\texttt{model}}$.

Our general observations are:

1. The value of test power increases as the value $k_{\texttt{true}}$'s complexity parameter σ becomes larger. This is possibly caused by the fact that the interaction becomes easier to detect as the pure interaction function $h_{12} \in \mathscr{H}_{12}$ becomes more complex as in it varies more quickly.

2. Given the data-generation mechanism:

   (a) Polynomial kernels (Linear and Quadratic kernels) exhibits underfit, and result in inflated Type I error but also low power.

   (b) Lower-order Matérn kernels (Matern 1/2 and 3/2) tend to exhibits overfit for smoother $k_{\texttt{true}}$'s, and result in deflated Type I error and diminished low power. This conclusion cautions us against the approach of extending model complexity by naively relaxing model's smoothness (i.e. differentiability) constraint.

   (c) Gaussian RBF Kernels in general can perform well, but only if the hyperparameter is chosen carefully. Specifically, selecting the hyperparameter σ by maximizing model likelihood does not perform well in small sample. On the other hand, the naive approach of selecting σ by setting σ to population median performs surprisingly well.

   (d) Neural Network kernels also work well in general. Their performance is also impacted by the hyperparameters. However not as sensitive as Gaussian RBF.

C.3 Detailed Numeric Results for Figure 1

Table C.1: $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 0.5$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.100	0.135	0.170	0.300	0.420	0.515	0.600	0.695	0.740
Quadratic	0.045	0.055	0.125	0.190	0.285	0.395	0.515	0.695	0.800
RBF_MLE	0.055	0.060	0.115	0.155	0.215	0.315	0.390	0.505	0.585
RBF_Median	0.040	0.050	0.115	0.205	0.280	0.420	0.580	0.665	0.790
Matern 1/2	0.045	0.030	0.005	0.005	0.010	0.015	0.015	0.010	0.015
Matern 3/2	0.060	0.065	0.015	0.010	0.005	0.010	0.010	0.020	0.020
Matern 5/2	0.060	0.075	0.035	0.010	0.010	0.010	0.015	0.025	0.025
NN 0.1	0.040	0.055	0.135	0.205	0.310	0.445	0.590	0.740	0.840
NN 1	0.055	0.080	0.150	0.220	0.320	0.475	0.550	0.675	0.775
NN 10	0.060	0.095	0.165	0.225	0.360	0.475	0.565	0.680	0.760
CVEK_RBF	0.035	0.040	0.110	0.165	0.280	0.380	0.535	0.700	0.800
CVEK_NN	0.040	0.055	0.140	0.205	0.315	0.445	0.585	0.725	0.825

Table C.2: $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 1$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.200	0.370	0.620	0.735	0.815	0.870	0.875	0.900	0.905
Quadratic	0.060	0.250	0.595	0.805	0.880	0.930	0.940	0.960	0.965
RBF_MLE	0.050	0.215	0.555	0.780	0.875	0.925	0.945	0.965	0.980
RBF_Median	0.025	0.235	0.645	0.865	0.930	0.960	0.975	0.980	0.980
Matern 1/2	0.015	0.025	0.090	0.205	0.235	0.300	0.325	0.370	0.385
Matern 3/2	0.030	0.080	0.145	0.205	0.260	0.300	0.360	0.390	0.410
Matern 5/2	0.040	0.095	0.180	0.230	0.270	0.310	0.365	0.400	0.415
NN 0.1	0.050	0.250	0.650	0.850	0.930	0.960	0.970	0.980	0.980
NN 1	0.055	0.290	0.675	0.860	0.930	0.965	0.980	0.980	0.985
NN 10	0.075	0.305	0.685	0.850	0.935	0.955	0.980	0.985	0.985
CVEK_RBF	0.030	0.205	0.635	0.850	0.920	0.965	0.975	0.975	0.980
CVEK_NN	0.045	0.240	0.660	0.855	0.935	0.960	0.975	0.980	0.980

Table C.3: $k_{\texttt{true}}$ = Matérn 3/2, $\sigma = 1.5$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.235	0.370	0.655	0.775	0.835	0.870	0.880	0.900	0.910
Quadratic	0.080	0.350	0.720	0.860	0.920	0.930	0.955	0.960	0.970
RBF_MLE	0.050	0.420	0.835	0.945	0.970	0.985	0.985	0.985	0.990
RBF_Median	0.030	0.405	0.880	0.960	0.975	0.975	0.980	0.985	0.985
Matern 1/2	0.000	0.040	0.170	0.325	0.430	0.495	0.560	0.620	0.675
Matern 3/2	0.015	0.045	0.225	0.360	0.475	0.530	0.600	0.700	0.750
Matern 5/2	0.020	0.075	0.260	0.430	0.515	0.575	0.630	0.735	0.780
NN 0.1	0.040	0.460	0.870	0.955	0.975	0.975	0.975	0.980	0.980
NN 1	0.040	0.465	0.860	0.960	0.970	0.980	0.980	0.980	0.980
NN 10	0.060	0.505	0.860	0.940	0.975	0.980	0.980	0.980	0.985
CVEK_RBF	0.020	0.420	0.880	0.960	0.975	0.975	0.980	0.980	0.985
CVEK_NN	0.050	0.485	0.870	0.965	0.975	0.980	0.980	0.980	0.980

Table C.4: $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 0.5$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.095	0.100	0.145	0.235	0.330	0.430	0.550	0.700	0.740
Quadratic	0.045	0.060	0.085	0.150	0.215	0.295	0.415	0.585	0.710
RBF_MLE	0.060	0.060	0.100	0.145	0.180	0.225	0.285	0.375	0.440
RBF_Median	0.040	0.050	0.095	0.145	0.245	0.310	0.410	0.575	0.675
Matern 1/2	0.045	0.035	0.005	0.005	0.005	0.010	0.010	0.010	0.015
Matern 3/2	0.065	0.055	0.015	0.005	0.000	0.005	0.010	0.010	0.020
Matern 5/2	0.065	0.065	0.040	0.005	0.005	0.005	0.010	0.020	0.020
NN 0.1	0.040	0.050	0.110	0.190	0.245	0.360	0.460	0.640	0.775
NN 1	0.045	0.055	0.120	0.175	0.255	0.320	0.440	0.590	0.680
NN 10	0.045	0.065	0.135	0.200	0.265	0.370	0.465	0.595	0.680
CVEK_RBF	0.035	0.035	0.080	0.140	0.195	0.315	0.420	0.625	0.745
CVEK_NN	0.040	0.040	0.110	0.185	0.245	0.355	0.450	0.635	0.770

Table C.5: $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 1$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.220	0.410	0.660	0.785	0.845	0.890	0.910	0.935	0.950
Quadratic	0.080	0.265	0.645	0.820	0.895	0.930	0.955	0.975	0.975
RBF_MLE	0.060	0.235	0.545	0.770	0.865	0.920	0.945	0.975	0.980
RBF_Median	0.035	0.250	0.685	0.875	0.925	0.965	0.975	0.980	0.980
Matern 1/2	0.020	0.030	0.110	0.205	0.245	0.295	0.335	0.365	0.395
Matern 3/2	0.040	0.105	0.155	0.220	0.255	0.325	0.370	0.400	0.420
Matern 5/2	0.045	0.120	0.205	0.230	0.285	0.345	0.370	0.410	0.425
NN 0.1	0.055	0.280	0.680	0.855	0.935	0.955	0.970	0.985	0.985
NN 1	0.060	0.310	0.685	0.850	0.925	0.960	0.980	0.985	0.985
NN 10	0.070	0.320	0.695	0.865	0.925	0.955	0.980	0.985	0.985
CVEK_RBF	0.035	0.230	0.660	0.845	0.925	0.955	0.975	0.980	0.980
CVEK_NN	0.050	0.295	0.675	0.855	0.930	0.960	0.970	0.985	0.985

Table C.6: $k_{\texttt{true}}$ = Matérn 5/2, $\sigma = 1.5$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.265	0.460	0.750	0.855	0.890	0.915	0.925	0.940	0.945
Quadratic	0.085	0.440	0.840	0.915	0.945	0.965	0.970	0.975	0.985
RBF_MLE	0.055	0.505	0.870	0.955	0.975	0.985	0.990	0.990	0.990
RBF_Median	0.020	0.600	0.920	0.965	0.975	0.980	0.985	0.990	0.995
Matern 1/2	0.000	0.060	0.230	0.430	0.550	0.640	0.665	0.760	0.795
Matern 3/2	0.015	0.100	0.285	0.465	0.570	0.660	0.705	0.790	0.825
Matern 5/2	0.020	0.130	0.315	0.495	0.610	0.680	0.720	0.815	0.840
NN 0.1	0.045	0.575	0.905	0.970	0.975	0.980	0.980	0.980	0.990
NN 1	0.055	0.590	0.905	0.970	0.980	0.980	0.980	0.990	0.990
NN 10	0.085	0.580	0.900	0.965	0.980	0.980	0.980	0.990	0.990
CVEK_RBF	0.020	0.590	0.920	0.970	0.975	0.980	0.985	0.995	0.995
CVEK_NN	0.055	0.580	0.900	0.970	0.980	0.980	0.980	0.990	0.995

Table C.7: $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 0.5$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.095	0.150	0.225	0.325	0.480	0.605	0.705	0.815	0.875
Quadratic	0.045	0.055	0.145	0.215	0.365	0.525	0.635	0.770	0.850
RBF_MLE	0.055	0.080	0.130	0.200	0.265	0.350	0.440	0.560	0.635
RBF_Median	0.055	0.055	0.125	0.235	0.350	0.500	0.595	0.735	0.840
Matern 1/2	0.040	0.050	0.005	0.005	0.015	0.015	0.020	0.020	0.020
Matern 3/2	0.060	0.075	0.030	0.015	0.015	0.020	0.030	0.030	0.030
Matern 5/2	0.065	0.080	0.035	0.015	0.015	0.025	0.030	0.030	0.035
NN 0.1	0.040	0.065	0.150	0.260	0.420	0.560	0.650	0.830	0.890
NN 1	0.055	0.085	0.165	0.255	0.380	0.535	0.600	0.720	0.840
NN 10	0.060	0.095	0.165	0.260	0.410	0.550	0.630	0.720	0.800
CVEK_RBF	0.035	0.035	0.135	0.225	0.365	0.525	0.640	0.790	0.865
CVEK_NN	0.040	0.060	0.150	0.260	0.420	0.555	0.645	0.825	0.890

Table C.8: $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 1$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.300	0.630	0.885	0.945	0.960	0.965	0.985	0.985	0.990
Quadratic	0.085	0.615	0.945	0.970	0.975	0.990	0.995	1.000	0.995
RBF_MLE	0.060	0.595	0.930	0.980	0.985	0.990	0.990	0.990	0.990
RBF_Median	0.025	0.730	0.970	0.980	0.985	0.995	0.995	0.995	1.000
Matern 1/2	0.000	0.095	0.365	0.585	0.675	0.755	0.820	0.865	0.880
Matern 3/2	0.030	0.155	0.410	0.620	0.710	0.775	0.840	0.880	0.890
Matern 5/2	0.035	0.205	0.440	0.645	0.735	0.775	0.840	0.880	0.905
NN 0.1	0.045	0.730	0.965	0.980	0.985	0.995	0.995	0.995	0.995
NN 1	0.100	0.710	0.960	0.980	0.980	0.990	0.995	0.995	0.995
NN 10	0.095	0.700	0.955	0.980	0.980	0.990	0.990	0.990	0.990
CVEK_RBF	0.030	0.725	0.960	0.980	0.985	0.990	0.990	0.995	0.995
CVEK_NN	0.050	0.715	0.970	0.980	0.985	0.990	0.995	0.995	0.995

Table C.9: $k_{\texttt{true}}$ = Gaussian RBF, $\sigma = 1.5$

$k_{\texttt{model}}/\delta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Linear	0.225	0.350	0.555	0.730	0.815	0.855	0.900	0.905	0.915
Quadratic	0.095	0.350	0.695	0.845	0.905	0.920	0.945	0.960	0.960
RBF_MLE	0.045	0.510	0.880	0.945	0.970	0.980	0.985	0.985	0.990
RBF_Median	0.035	0.600	0.905	0.975	0.980	0.980	0.985	0.985	0.985
Matern 1/2	0.000	0.025	0.155	0.295	0.410	0.525	0.600	0.695	0.750
Matern 3/2	0.005	0.025	0.165	0.355	0.490	0.590	0.680	0.800	0.830
Matern 5/2	0.005	0.025	0.205	0.385	0.515	0.625	0.730	0.820	0.875
NN 0.1	0.045	0.595	0.910	0.965	0.980	0.980	0.980	0.980	0.980
NN 1	0.060	0.580	0.885	0.955	0.970	0.970	0.980	0.980	0.980
NN 10	0.120	0.520	0.875	0.940	0.970	0.975	0.975	0.975	0.975
CVEK_RBF	0.025	0.585	0.915	0.970	0.980	0.980	0.985	0.985	0.985
CVEK_NN	0.055	0.590	0.905	0.970	0.980	0.980	0.980	0.985	0.985

References

[1] Athanassia Bacharoglou. Approximation of probability distributions by convex mixtures of Gaussian measures. *Proceedings of The American Mathematical Society - PROC AMER MATH SOC*, 138:2619–2619, July 2010.

[2] Peter L. Bartlett, Olivier Bousquet, and Shahar Mendelson. Local Rademacher complexities. *The Annals of Statistics*, 33(4):1497–1537, August 2005.

[3] Corinna Cortes, Marius Kloft, and Mehryar Mohri. Learning Kernels Using Local Rademacher Complexity. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2760–2768. Curran Associates, Inc., 2013.

[4] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Ensembles of Kernel Predictors. In *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*, pages 145–152, Corvallis, Oregon, 2011. AUAI Press.

[5] Bruce E. Hansen and Jeffrey S. Racine. Jackknife model averaging. *Journal of Econometrics*, 167(1):38–46, March 2012.

[6] Gert R. G. Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I. Jordan. Learning the Kernel Matrix with Semidefinite Programming. *Journal of Machine Learning Research*, 5(Jan):27–72, 2004.

[7] Shahar Mendelson. On the Performance of Kernel Classes. *Journal of Machine Learning Research*, 4(Oct):759–771, 2003.

[8] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine*

Learning. University Press Group Limited, January 2006. Google-Books-ID: vWtwQgAA-CAAJ.

[9] Mark van der Laan and Sandrine Dudoit. Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples. *U.C. Berkeley Division of Biostatistics Working Paper Series*, November 2003.

[10] Mark J. van der Laan, Eric C. Polley, and Alan E. Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6:Article25, 2007.