**Change-Points in High-Dimensional Settings**

Claudia Kirch
(joint work with John A D Aston )

While there is considerable work on change-point analysis in univariate time series, more and more data being collected comes from high dimensional multivariate settings, where the number of components is of the same order or even larger than the number of time points. An appropriate asymptotic framework to investigate statistical procedures for such data assumes that the number of components increases to infinity with the number of time points. In this setup we would like to investigate the properties of univariate tests after the data has been projected onto a vector $\boldsymbol{p}_d$. To this end, we consider the following model:

$$X_{i,t} = \mu_i + \delta_{i,T}\, 1_{\{t>\lfloor \vartheta T\rfloor\}} + e_{i,t}, \quad 1 \le i \le d = d_T, 1 \le t \le T,$$

where (for simplicity) $\{(e_{1,t},\dots,e_{d,T})^T, t = 1,\dots,T\}$ is i.i.d. and $0 < \vartheta < 1$ is the rescaled change-point. We call the vector $\boldsymbol{\Delta}_d = (\delta_{1,T},\dots,\delta_{d,T})^T$ the change and test

$$H_0 : \boldsymbol{\Delta}_d = 0, \qquad H_1 : \boldsymbol{\Delta}_d \neq 0.$$

In this setting, it is apparent that the change $\boldsymbol{\Delta}_d$ is always a one-dimensional object no matter the number of components $d$. This observation suggests that knowledge about where the change-point is located in addition to the underlying covariance structure can significantly increase the signal-to-noise ratio. In applications, certain changes are either expected or of particular interest e.g. an economist looking at the performance of several companies expecting changes caused by a recession will have a good idea which companies will profit or lose. This knowledge can then be used to increase the power in directions close to the search direction $\boldsymbol{p}_d$ while decreasing it for changes that are close to orthogonal to it.

In order to understand this informal statement better and to compare the power behavior of different statistics, we consider contiguous changes, where $\|\boldsymbol{\Delta}_d\| \to 0$ but with such a rate that the power of the corresponding test is strictly between the size and one. We can then compare these **contiguous rates** to understand the power of the test. Concerning a fixed projection $\boldsymbol{p}_d$ it turns out that the contiguous rate is given by

$$T\,\|\Sigma^{-1/2}\boldsymbol{\Delta}_d\|^2 \cos^2(\alpha_{\Sigma^{-1/2}\boldsymbol{\Delta}_d,\Sigma^{1/2}\boldsymbol{p}_d}),$$

where $\Sigma$ is the covariance of the vector $(e_{1,t},\dots,e_{d,T})^T$ and $\alpha_{\boldsymbol{a},\boldsymbol{b}}$ is the smallest angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. From this it is obvious that the **oracle** projection $\boldsymbol{o} = \Sigma^{-1}\boldsymbol{\Delta}_d$ maximizes the contiguous rate. This can be compared to a random projection on the unit sphere after standardizing the data, which is equivalent to projecting with the vector $\boldsymbol{r}_{d,\Sigma} = \Sigma^{-1/2}\boldsymbol{r}_d$, where $\boldsymbol{r}_d$ is a random projection on the unit sphere. Furthermore, we can compare the procedure with a generalization of multivariate change-point procedures for independent components in the above asymptotic framework proposed by Horváth and Hušková [2].

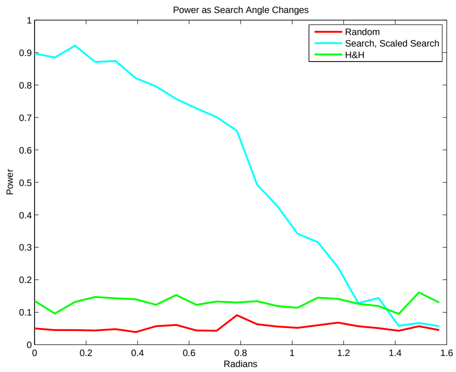



FIGURE 1. Empirical size-corrected power for increasing angles, $\Sigma = \mathrm{Id}$

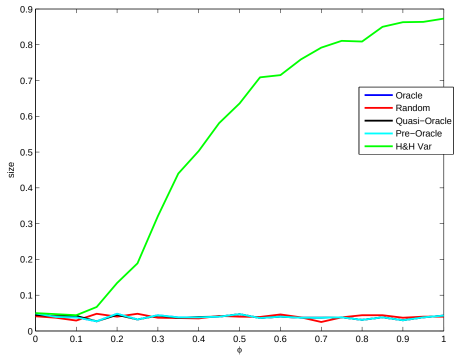



FIGURE 2. Empirical size for increasing contamination by a common factor

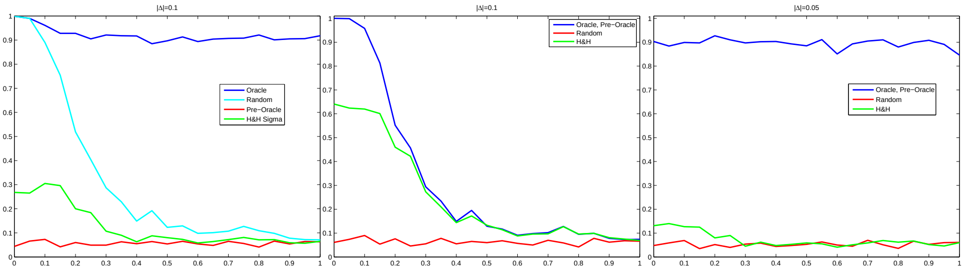
The following table compares the contiguous rates in all three cases:

| | Contiguous Rate |
|---|---|
| Oracle projections | $T\|\Sigma^{-1/2}\boldsymbol{\Delta}_d\|^2$ |
| HH statistic ($\Sigma = \mathrm{Id}$) | $T\|\Sigma^{-1/2}\boldsymbol{\Delta}_d\|^2/\sqrt{d}$ |
| Scaled random projection | $T\|\Sigma^{-1/2}\boldsymbol{\Delta}_d\|^2/d$ (stochastic order) |

It becomes apparent that we lose an order $\sqrt{d}$ between the oracle and the HH statistic as well as another order $\sqrt{d}$ between the HH statistic and the scaled random projection. Figure 1 confirms these theoretical findings and gives an impression on how wide the angle between $\Sigma^{-1/2}\boldsymbol{\Delta}_d$ and $\Sigma^{1/2}\boldsymbol{p}_d$ can be before the HH procedure is better than the projection. Please note, however, that the space covering these angles increases for increasing dimensions.

Usually, in applications $\Sigma$ is not known and needs to be estimated, which is rather problematic particularly in high-dimensional settings without additional parametric or sparsity assumptions. For change-point tests the inverse is needed which results in additional numerical problems for large $d$. Consequently, it is of importance to check the robustness of the procedures with respect to not knowing $\Sigma$.

To this end, we first consider the size of the different procedures. For the projection procedures and a large class of dependency across components only the variance of the projected sequence is needed, which is not difficult to estimate. The HH procedure on the other hand strongly depends on the independence between components or after some possible extensions on the knowledge of $\Sigma^{-1}$. Consequently, it suffers sincere size problems if $\Sigma$ is misspecified. In order to show




$\mathbf{\Delta}_d \sim (1,\ldots,1,0,\ldots,0)^T$, ${}_p\mathbf{o} \sim {}_q\mathbf{o}$ | $\mathbf{\Delta}_d \sim (1,\ldots,1,1,\ldots,1)^T$, $\boldsymbol{o} \sim {}_p\mathbf{o} \sim {}_q\mathbf{o}$ | $\mathbf{\Delta}_d \sim (1,\ldots,1,-1,\ldots,-1)^T$, ${}_p\mathbf{o} \sim {}_q\mathbf{o}$

FIGURE 3. Empirical size-corrected power for increasing contamination by a common factor, $s_j = 1$

this effect we consider the situation where $e_{i,t} = s_i\eta_{i,t} + \Phi\xi_i$, where $\eta_{i,t}$ are independent and standardized and $\xi_i$ is a common standardized disturbance factor across all channels (independent of $\eta$). Figure 2 clearly shows that the projection is much more robust with respect to size.

Considering contiguous rates again we can also investigate the robustness in terms of the power of the different procedures. To this end, we consider the **pre-oracle** ${}_p\mathbf{o} = \mathbf{\Delta}_d$ as well as the **quasi-oracle** ${}_q\mathbf{o} = (\delta_1/\operatorname{var}(e_{1,1}),\ldots,\delta_d/\operatorname{var}(e_{d,1}))^T$. If the Variances are all of the same order, i.e. $0 < c \leq \operatorname{var}(e_{i,1}) \leq C < \infty$, then in the uncorrelated case quasi- and pre-oracle are of the same order, in the general case both of them are always at least as good as the unscaled random projection $\boldsymbol{r}_d$ but can be better, while the HH procedure is always of the same order as the random projection. This fact is confirmed by the simulations in Figure 3.

In summary, projections can greatly increase the power of corresponding change-point tests in high-dimensional settings particularly if the covariance structure is accessible and some information about the location of the change of interest is known. Additionally, such projections are much more robust with respect to both size and power than competing fully multivariate procedures if the covariance structure is misspecified.

## References

[1] J.A.D. Aston, C. Kirch, *Change-poins in high-dimensional settings*, in preparation, 2013.

[2] L. Horváth, M. Hušková, *Change-point detection in panel data*, J. Time Series Anal. **33** (2012), 631-648.