

# Supplementary materials

## Contents

<b>1</b>	<b>Notations and Definitions</b>	<b>1</b>
<b>2</b>	<b>Behavior of raw wpd</b>	<b>1</b>
2.1	Null design . . . . .	2
2.2	Alternate designs . . . . .	3
<b>3</b>	<b>Normalisation</b>	<b>7</b>
3.1	Methodology . . . . .	8
3.2	Null design . . . . .	8
3.3	Alternate designs . . . . .	8
3.4	Sample size . . . . .	9
3.5	Number of permutations . . . . .	9

## 1 Notations and Definitions

Consider two cyclic granularities  $A$  and  $B$ , such that  $A = \{a_j : j = 1, 2, \dots, J\}$  and  $B = \{b_k : k = 1, 2, \dots, K\}$  with  $A$  placed across x-axis and  $B$  across facets. Let the pairwise distances between pairs  $(a_j b_k, a_{j'} b_{k'})$  be denoted as  $d_{(jk, j'k')} = JSD(a_j b_k, a_{j'} b_{k'})$ . Pairwise distances could be within-facets or between-facets. Figure 1 illustrates how the within-facet or between-facet distances are defined. Pairwise distances are within-facets ( $d_w$ ) when  $b_k = b_{k'}$ , that is, between pairs of the form  $(a_j b_k, a_{j'} b_k)$  as shown in panel (3) of Figure 1. If categories are ordered (like all temporal cyclic granularities), then only distances between pairs where  $a_{j'} = (a_{j+1})$  are considered (panel (4)). Pairwise distances are between-facets ( $d_b$ ) when they are considered between pairs of the form  $(a_j b_k, a_j b_{k'})$ .

## 2 Behavior of raw wpd

Most of the behavior of the measure wpd was studied via simulation. The simulations explore how wpd performs under various designs and parameters and its limitations. To study the behavior of wpd, simulations were carried out for four different designs and the following factors that could potentially have an impact on the values of wpd:

- $nx$  (number of levels of x-axis)
- $n_{facet}$  (number of levels of facets)
- $\lambda$  (tuning parameter)
- $\omega$  (increment in each panel design)

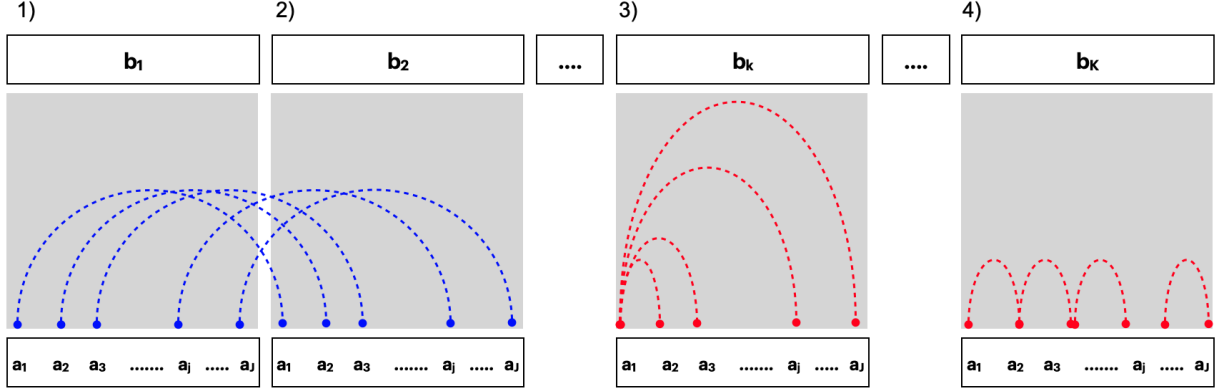


Figure 1: Within and between-facet distances shown for two cyclic granularities A and B, where A is mapped to x-axis and B is mapped to facets. The dotted lines represent the distances between different categories. Panel 1) and 2) show the between-facet distances. Panel 3) and 4) are used to illustrate within-facet distances when categories are un-ordered or ordered respectively. When categories are ordered, distances should only be considered for consecutive x-axis categories. Between-facet distances are distances between different facet levels for the same x-axis category, for example, distances between  $(a_1, b_1)$  and  $(a_1, b_2)$  or  $(a_1, b_1)$  and  $(a_1, b_3)$ .

- *dist* (normal/non-normal distributions with different location and scale)
- *n* (sample size for each combination of categories)
- *nsim* (number of simulations)
- *nperm* (number of permutations of data)
- *designs*
  - $D_{null}$  (No difference in distribution)
  - $D_{var_f}$  (Difference in distribution only across facets)
  - $D_{var_x}$  (Difference in distribution only across x-axis)
  - $D_{var_{all}}$  (Difference in distribution in both facets and x-axis)

## 2.1 Null design

This section explores the behavior of wpd in designs where there is no difference in distribution between x and facet categories. We have considered different initial distributions to study the impact of initial distribution under the null setup. Since the measure wpd is essentially set up to detect differences in distributions irrespective of underlying distribution, it would be ideal if it has minimal dependency on the type, location and scale of the initial distribution. To that end, some pre-processing of the data is preferred to bring it to the same scale, location and type. The Normal Score Transform or NQT has been applied in various fields of geo-statistics in order to make most asymmetrical distributed real world measured variables more treatable and normal-like.

Following the work of Krzysztofowicz (1997) the empirical NQT involves the following steps: 1. Sort the sample of measured variable  $X$  from the smallest to the largest observation  $x_{(1)}, \dots, x_{(i)}, \dots, x_{(n)}$ . 2. Estimate the cumulative probabilities  $p_{(1)}, \dots, p_{(i)}, \dots, p_{(n)}$ . using a plotting position like  $i/(n+1)$  such that  $p_{(i)} = P(X \leq x_{(i)})$ . 3. Transforming each observation  $x_{(i)}$  of  $X$  into observation  $y(i) = Q^{-1}(p(i))$  of the standard normal variate  $Y$ , with  $Q$  denoting the standard normal distribution and  $Q^{-1}$  its inverse, applying discrete mapping.

Further, we want to study the distribution of wpd for different  $nx$  and  $nfacet$ .

### 2.1.1 Simulation setup

Using two types of distributions, viz. normal and gamma (non-normal), we generated observations for each combination of  $nx$  and  $nfacet$  from the following sets:  $nx = \{2, 3, 5, 7, 14, 20, 31, 50\}$  and  $nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$  to cover a wide range of levels from very low to moderately high. Each combination is being referred to as a *panel*. That is, data is being generated for each of the panels  $\{nx = 2, nfacet = 2\}, \{nx = 2, nfacet = 3\}, \{nx = 2, nfacet = 5\}, \dots, \{nx = 50, nfacet = 31\}, \{nx = 50, nfacet = 50\}$ . For each of the 64 panels,  $ntimes = 500$  observations are drawn for each combination of the categories. That is, if we consider the panel  $\{nx = 2, nfacet = 2\}$ , 500 observations are generated for each of the combination of categories from the panel, namely,  $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$ . The values of  $\lambda$  is set to 0.67, since we want to up-weight the within-facet distances and that of  $\omega$  is set to 0, since there is no significant differences between distributions in the null case. Observations were generated for each type of distribution changing the shape and scale to study the effect of shape, scale and type of distribution on wpd. The set of distributions considered for this purpose is  $N(0, 1), N(5, 1), N(0, 5), \Gamma(0.5, 1), \Gamma(2, 1)$ . Each of the scenario is run  $nsim = 200$  times to see the distribution of wpd values for each scenario.

Figure 2 and 3 show distribution of wpd for different initial normal and gamma distribution to study the effect of changing location and scale on the distribution of wpd. It seems like the distribution changes across different facet and x levels but look the same for each panel, which implies wpd value is unaffected by the change in location and scale of the the normal distribution. Figure 4 shows how mean and sd of the distribution of wpd changes with the increasing x and facet levels. It seems like both mean and standard deviation are affected more by change in the x-axis levels than the facet levels. Figure 5 gives another way to look at the effect of changing facet and x levels on the distribution of wpd for an initial distribution  $N(0, 5)$ . Clearly, the location of the distribution shifts to the right for increasing x levels and scale is different for low and high values of facet levels with a longer left tail for lower facet levels. We have considered alternate distribution, location and scale for this simulation setting, but do not present the results as they are behaviorally similar.

## 2.2 Alternate designs

This section explores the behavior of wpd in designs where there is infact difference in distribution between facet categories ( $D_{var_f}$ ) or across x-categories ( $D_{var_x}$ ) or both ( $D_{var_{all}}$ ). Since it is established in the last section that after preprocessing the data through NQT, the initial distribution does not have a role to play in the values of wpd, we proceed with a  $N(0,1)$  distribution only. Supposably, the tuning parameter ( $\lambda$ ) and increment parameter ( $\omega$ ) should impact the values of wpd.

### 2.2.1 Simulation setup

Using  $\omega = \{1, 2, \dots, 10\}$  and  $\lambda = seq(from = 0.1, to = 0.9, by = 0.05)$ , observations are drawn from a  $N(0,1)$  distribution for each combination of  $nx$  and  $nfacet$  from the following sets:  $nx = nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$ .  $ntimes = 500$  is assumed for this setup as well. Furthermore, to generate different distributions across different combination of facet and x levels, the following method is deployed - suppose the distribution of the combination of first levels of x and facet category is  $N(\mu, \sigma)$  and  $\mu_{jk}$  denotes the mean of the combination ( $a_j b_k$ ), then  $\mu_{j.} = \mu + j\omega$  (for design  $D_{var_x}$ ) and  $\mu_{.k} = \mu + k\omega$  (for design  $D_{var_f}$ ). Table 1 shows an example of how initial distributions are assigned in a panel with  $nfacet = 3$  and  $nx = 2$  for different designs using  $\omega = 1$ .

We now discuss the effect of different tuning parameters  $\lambda$  with the help of two alternate designs ( $D_{var_x}$  and  $D_{var_f}$ ) and  $\omega$ . ?? displays different values of  $\lambda$  for a relatively small and higher  $\omega$  for  $nx = nfacet = \{2, 3, 5, 7, 9\}$ . The  $\lambda$  for which the two designs intersect will then be chosen as the optimal  $\lambda$  that could then be weighed appropriately to up-weight the within-facet distances and down-weight the between-facet distances. The design  $D_{var_x}$  increases linearly with increasing  $\lambda$ , whereas  $D_{var_f}$  decreases linearly with increasing  $\lambda$ . This is expected as wpd has a linear relationship with  $\lambda$  by construction. The two designs

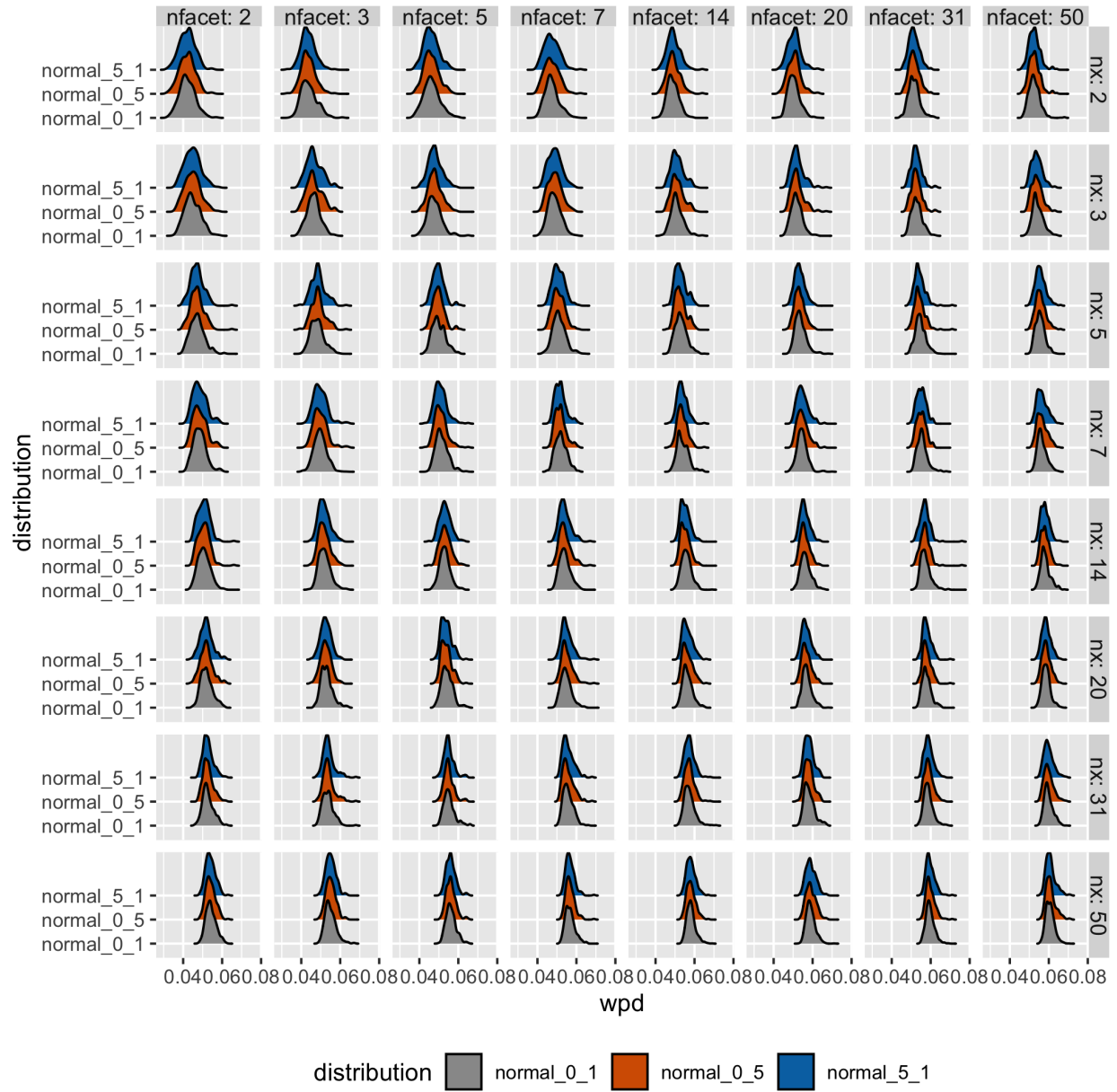


Figure 2: Ridge plots of raw wpd is shown for  $N(0,1)$ ,  $N(5,1)$  and  $N(0,5)$  distribution. The densities change across different facet and x levels but look same for each panel, which implies wpd value is unaffected by the change in mean and standard deviation of the normal distribution

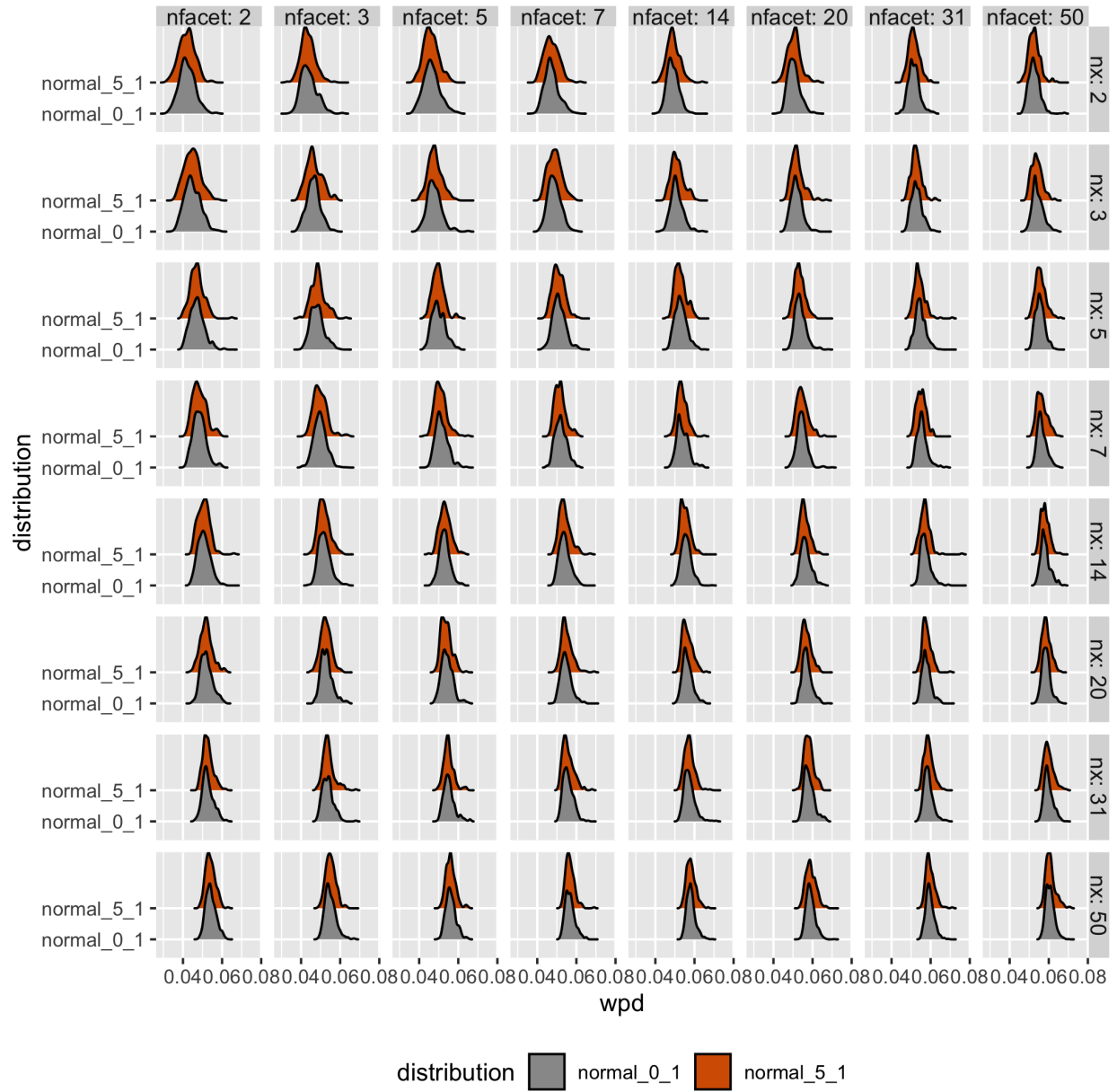
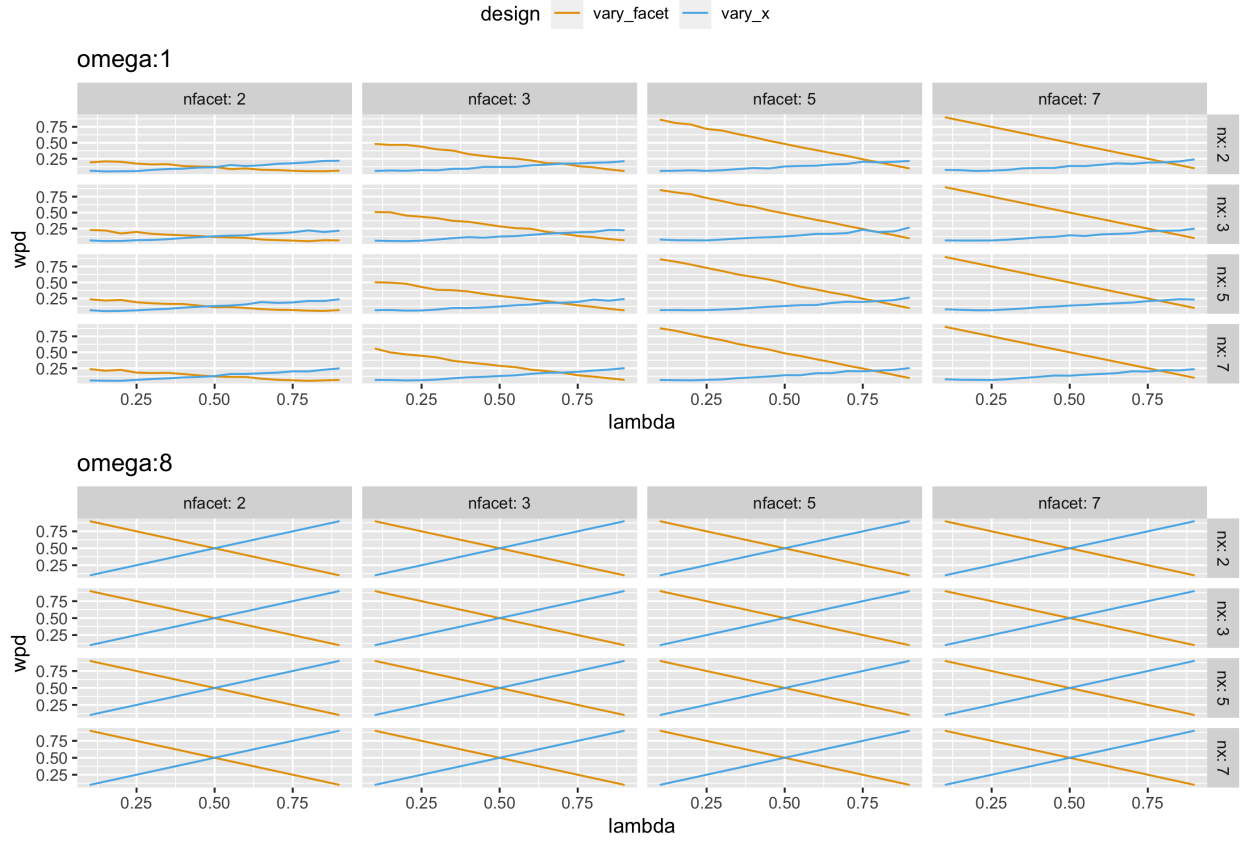


Figure 3: Ridge plots of raw wpd is shown for  $\text{Gamma}(0.5,1)$ ,  $\text{Gamma}(2,1)$  distribution. The densities change across different facet and x levels but look same for the two distributions, which implies wpd value is unaffected by the change in the shape parameter of the gamma distribution

Table 1: Simulation setup for a panel with 3 facet levels and 2 x-axis levels for different designs starting from an initial distribution  $N(0, 1)$  for the combination  $(a_1, b_1)$ .

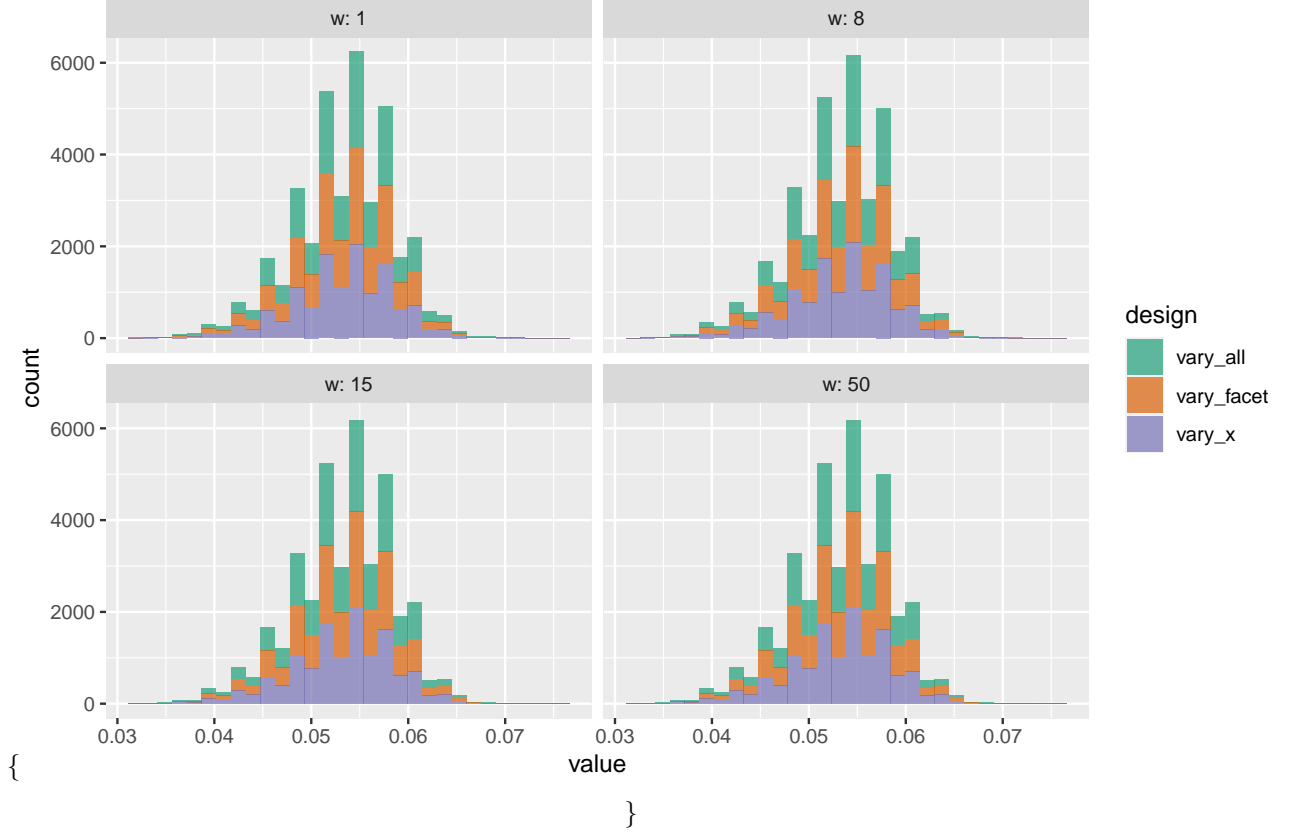
x	facet	$D_{var_f}$	$D_{var_x}$	$D_{var_{all}}$
$a_1$	$b_1$	$N(0, 1)$	$N(0, 1)$	$N(0, 1)$
$a_2$	$b_1$	$N(0, 1)$	$N(1, 1)$	$N(1, 1)$
$a_1$	$b_2$	$N(1, 1)$	$N(0, 1)$	$N(2, 1)$
$a_2$	$b_2$	$N(1, 1)$	$N(1, 1)$	$N(3, 1)$
$a_1$	$b_3$	$N(2, 1)$	$N(0, 1)$	$N(4, 1)$
$a_2$	$b_3$	$N(2, 1)$	$N(1, 1)$	$N(5, 1)$

mostly intersect at  $\lambda = 0.5$  for a higher value of  $\omega$ . Figure 7 shows how the value of  $\lambda$  changes with increasing  $\omega$  for  $\omega = \{1, 2, \dots, 10\}$  and  $\lambda = seq(from = 0.1, to = 0.9, by = 0.05)$ .



We now discuss the values  $w_{pd}$  takes for different designs. Ideally the  $w_{pd}$  should have the highest values for  $D_{var_{all}}$ , followed by  $D_{var_x}$  and  $D_{var_y}$  for  $\lambda \geq 0.5$ . We test this for lower and higher values of  $\omega = c(1, 8, 15, 50)$

\begin{figure}



{ Distribution of wpd under different designs are overlapped to see how they measure relative to each other. We would have expected values of wpd to be such that vary\_facet is less than vary\_x is less than vary\_all but distribution of vary\_x to lie above vary\_facet. } \end{figure}

### 2.2.2 Discussion

The distribution of wpd is different for different levels of facets and x-axis levels. This is because the statistic maximum which is used to define wpd is affected by the number of categories. The measure would have higher values if  $C_i$  or  $C_j$  has higher levels. However, we would ideally want a higher value of the measure only if there is significant difference between distributions across facet or x-axis categories, and not because the number of categories are higher. Therefore, in order to compare wpd across different combinations of facet and x-axis levels, we need to eliminate the impact of different levels of the facets and x-axis first and get a normalized measure. Henceforth we call the measure discussed as  $wpd_{raw}$  and the normalised measure as  $wpd_{norm}$ . The measure  $wpd_{norm}$  could potentially lead to comparison of the measure across different panels and also identifying only the interesting panels from a dataset.

## 3 Normalisation

In an attempt to make the densities ?? same across different levels, we want to start by making their scale and location same. In that regard, we shuffle the data multiple times and compute the sampling distribution of for each combination of facet and x-axis levels.

The null hypothesis is that the two cyclic granularities do not differ on the outcome (i.e., that the outcome is observed independently of treatment assignment). When we permute the outcome values during the test, we therefore see all of the possible alternative treatment assignments we could have had and where the mean-difference in our observed data falls relative to all of the differences we could have seen if the

outcome was independent of treatment assignment. While a permutation test requires that we see all possible permutations of the data (which can become quite large), we can easily conduct “approximate permutation tests” by simply conducting a vary large number of resamples. That process should, in expectation, approximate the permutation distribution.

Finally, we run some simulation experiments to see if normalisation works. If normalisation has worked, it should work for both x level and facet-levels.

Thus these median maximum pairwise distances need to be normalized for different harmonies in a way that eliminates the effect of different levels, consequently enabling comparison across different harmonies. The Fisher–Tippett–Gnedenko theorem in the field of Extreme Value Theory states that the maximum of a sample of iid random variables after proper re-normalization can converge in distribution to only one of Weibull, Gumbel or Freschet distribution, independent of the underlying data or process.

### 3.1 Methodology

The mean and sd for each combination of facet and x levels could be computed by shuffling the data repeatedly and obtaining the distribution of  $MMPD_{raw}$  for different combinations. If  $MMPD_{raw}$  is then adjusted by the location and scale of this distribution, we could assume that the distribution of the resultant  $MMPD_{norm}$  will have same distribution across different combinations of x and facet levels. All data is repeatedly shuffled  $nperm$  times in random manner to obtain the measure  $MMPD_{raw}$ . Then the  $MMPD_{norm}$  is obtained by scaling the observed value by the mean and sd of the distribution of  $MMPD_{raw}$  obtained from the permuted data.

Step 7 in the Algorithm defined in Section ?? need to be revised in order to compute  $MMPD_{norm}$  instead of  $MMPD_{raw}$ . Step 1-6 stays the same. The entire algorithm will have to repeated for all harmony pairs considered in the context.

A simulation study is conducted to see if this methodology is working.

### 3.2 Null design

### 3.3 Alternate designs

The behavior of the measure is monitored and understood through simulation experiments. A single simulation consisting of computational operations on a panel generating MMPD the value of which represents to what extent a pair of cyclic granularities would be interesting when displayed in the design.

In comparing different simulation runs, we distinguish between different distributions and simulation scenarios (also iterations may be).

#### 3.3.1 Simulated designs

D1. Same distribution of the measured variable across all x-axis and facet categories D2. Different distribution across facet categories but same across x-axis categories D3. Different distribution across x-axis categories but same across facet categories D4. Different across both x-axis and facet categories

Each of these case scenarios tried against Normal and a non-normal (Gamma) distribution to check if underlying distribution has a role to play.

#### 3.3.2 Environment

R version 4.0.1 (2020-06-06) is used with platform: x86\_64-apple-darwin17.0 (64-bit) running under: macOS Mojave 10.14.6



### **3.4 Sample size**

How smaller or different sample size play a role in the distribution of the measure.

### **3.5 Number of permutations**

Find the number of permutations which are sufficient for consistent estimation of mean and standard deviation

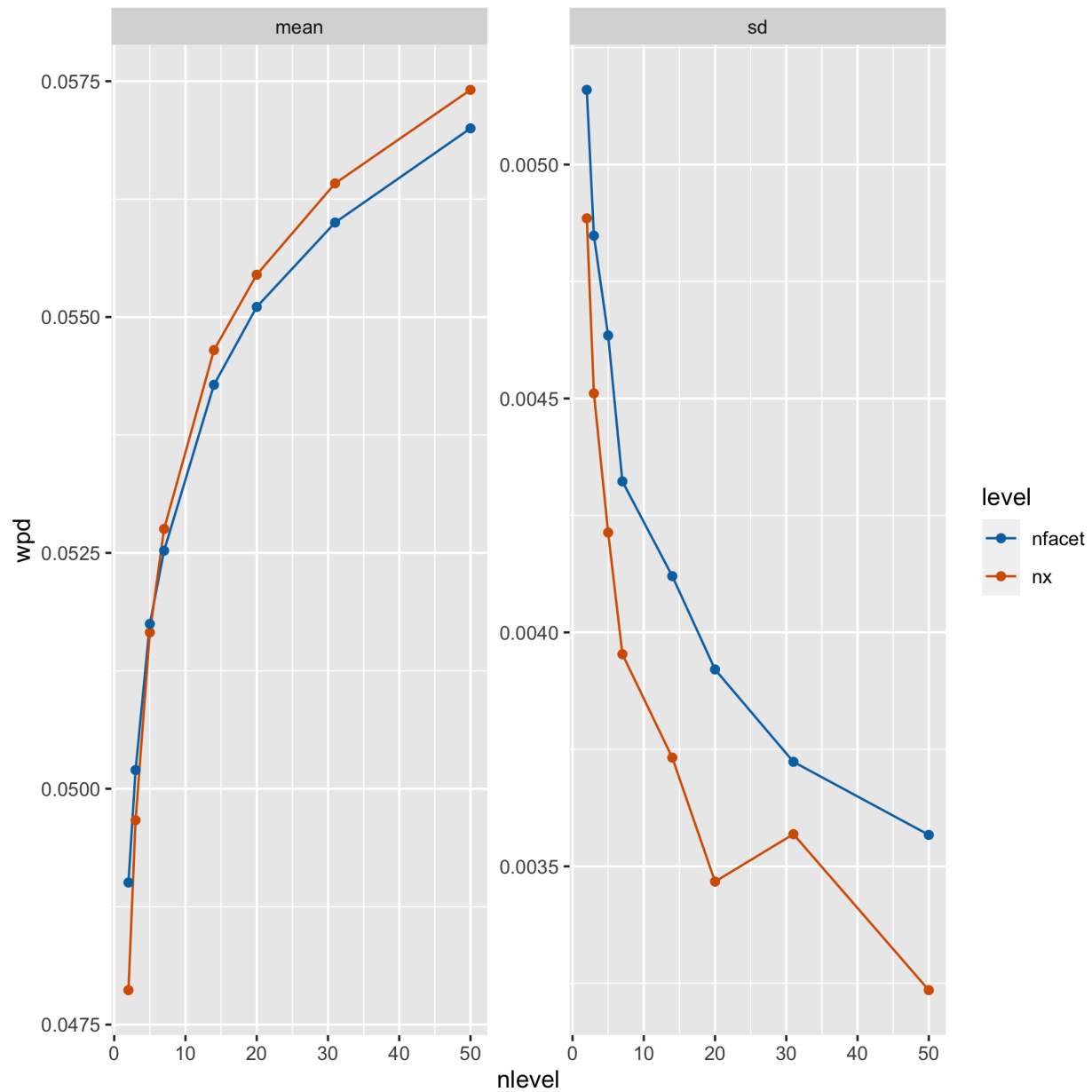


Figure 4: Movement of mean and sd for raw wpd is shown for different number of levels (nlevel) of x-axis and facets through line plots. Mean increases and sd decreases more sharply for increasing x-axis levels as compared to facet levels. It seems like both mean and standard deviation are affected more by change in the x-axis levels than the facet levels.

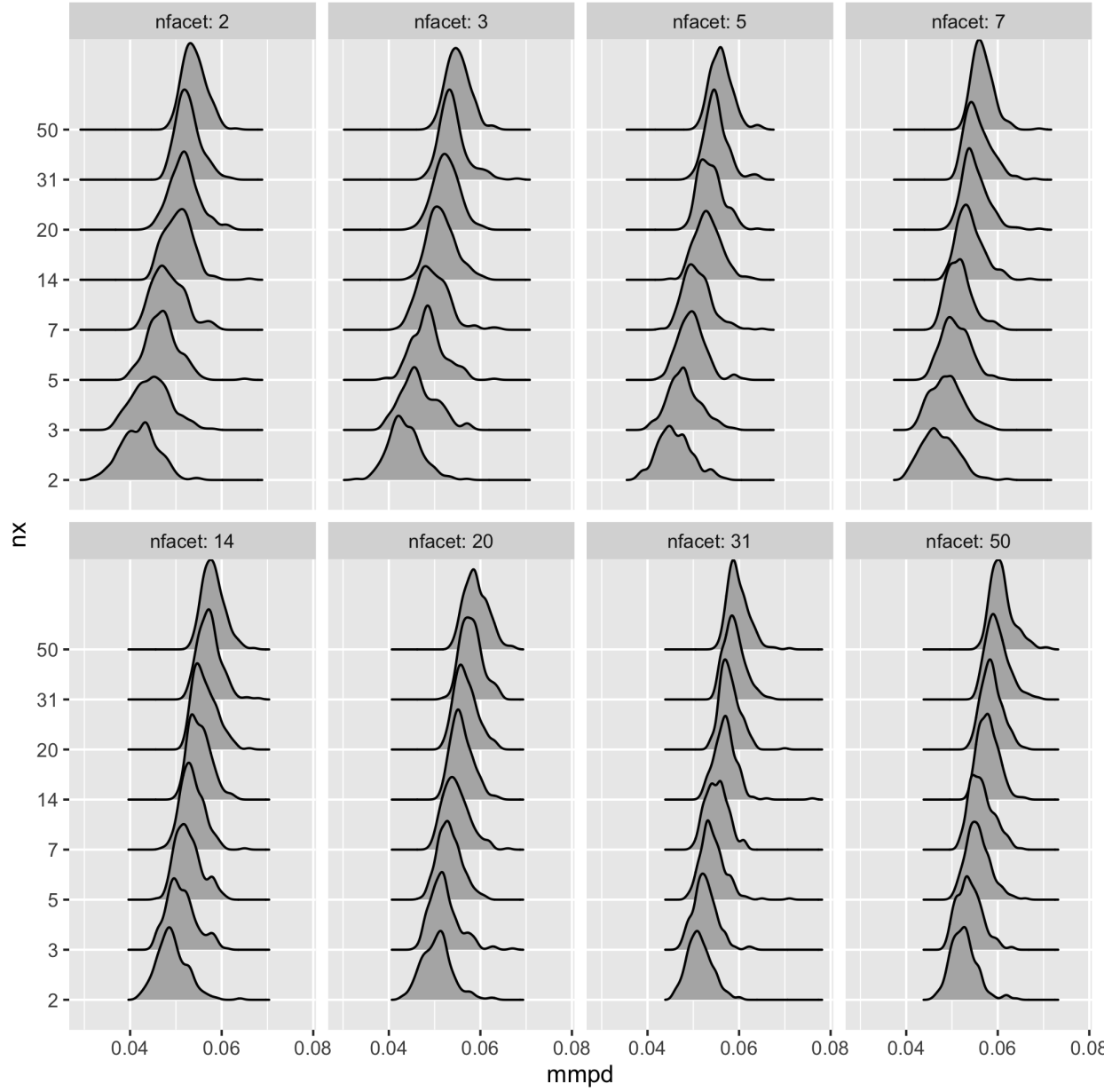


Figure 5: Ridge plots of raw wpd is shown for  $N(0,5)$  distribution. For each panel, it could be seen that the location shifts to the right for increasing  $x$  levels. Across each panel, the scale of the distribution seems to change for low/moderately lower values and higher values of  $n$ facets and left tails are longer for lower facet levels.

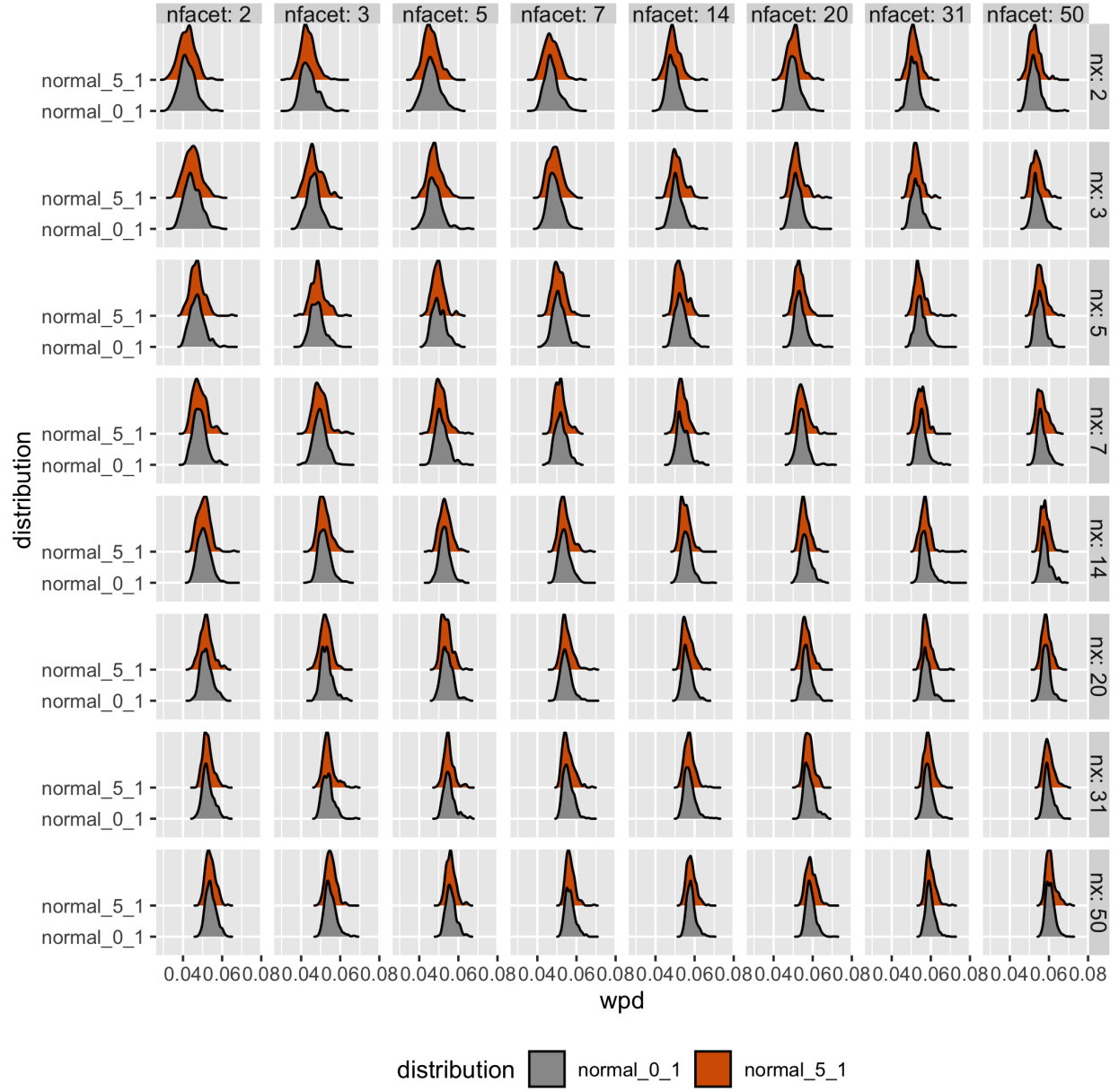


Figure 6: Ridge plots of raw wpd is shown for G(0.5,1) and G(2,1) distribution after quantile transformation looks similar and hence is unaffected by change in location.

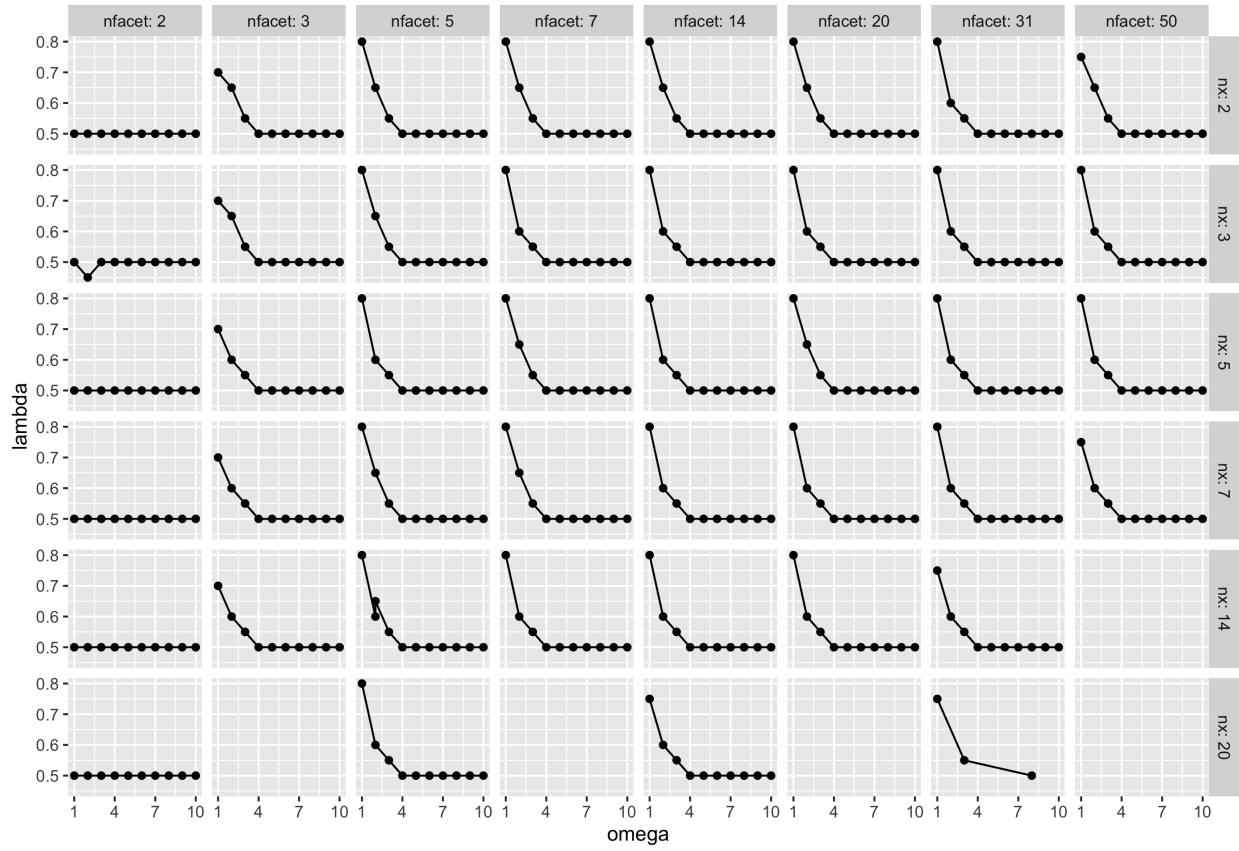


Figure 7: For most panels it is observed that the most common value of the tuning parameter for which the designs interact is 0.5, which implies any value greater than 0.5 could be chosen to up-weight the within-facet distances and down-weight the between-facet distances for most situations.

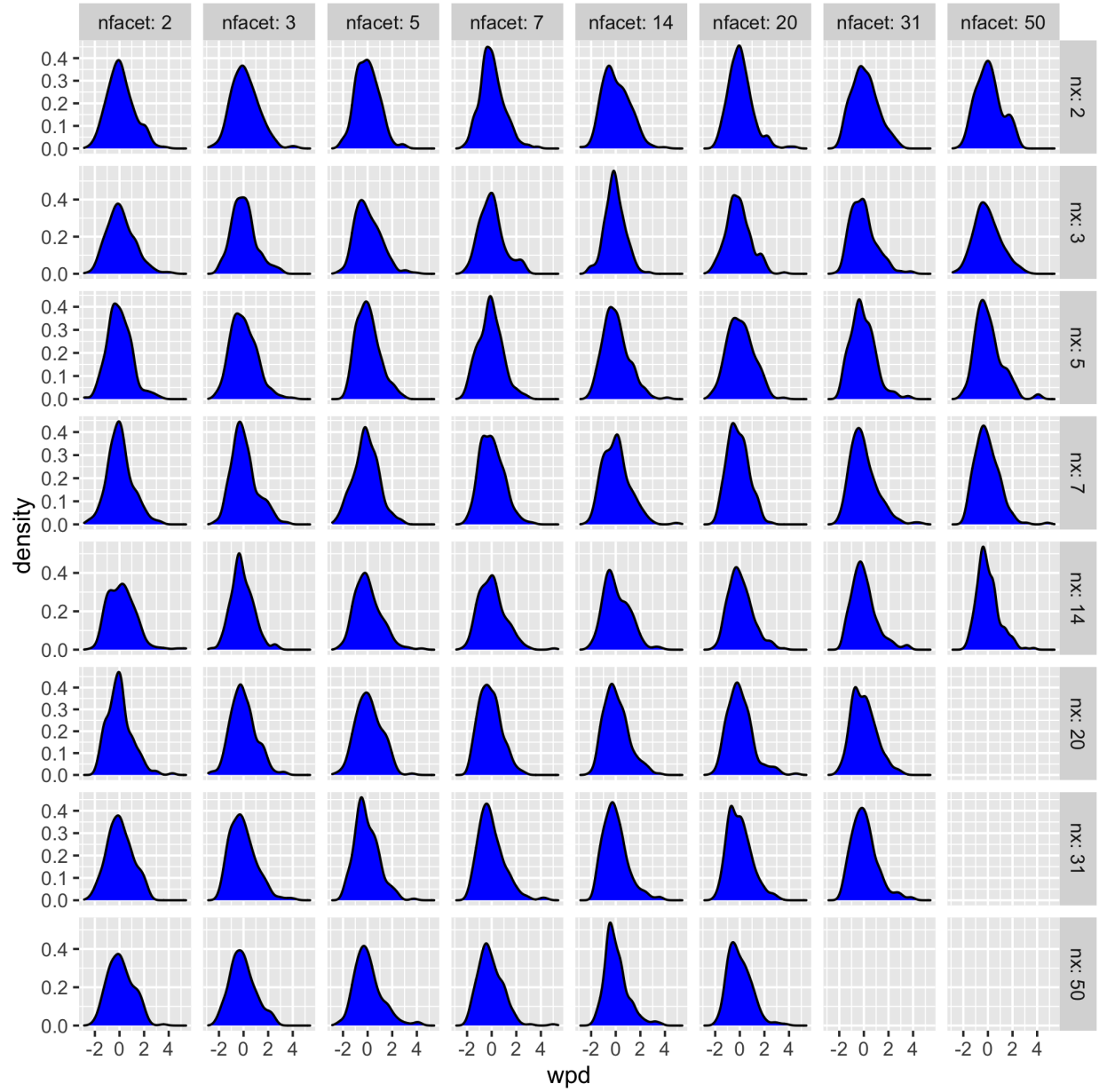


Figure 8: Density plots of normalised wpd is shown for Normal(0,1) distribution. The mean and sd of the distribution looks almost similar for different facet and x-axis categories.