A Framework for Exploring Systematic Error in a Quantitatively Complete Measurement Model

Rick Danielson[a], Will Perrie[a], and Joël Chassé[b], Fisheries and Oceans Canada,
[a]Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada
[b]Gulf Fisheries Centre, Moncton, New Brunswick, Canada

2. From Error to Signal

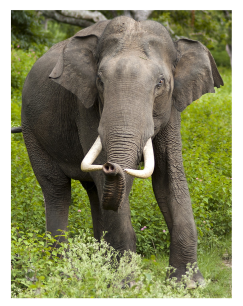
To build a model that is complete and nonlinear from the outset, we begin by equating *C*, the calibrated measurements, and *t*, the simplest signal model (ignoring noise for the moment). It seems reasonable to characterize geophysicists as attentive to unresolved scales in *C* when developing their numerical models. Similarly, statisticians are attentive to limitations of the linear model *t*. Thus, when we equate measures and model, a geophysicist might subtract representation error from *C* and a statistician might add equation error to *t*:

$$\begin{aligned} \text{Calibrated} \quad C - \epsilon_C^{REP} &= t + \epsilon_C^{EQ} \\ \text{Uncalibrated} \quad U - \epsilon_U^{REP} &= \alpha_U + \beta_U t + \epsilon_U^{EQ} \end{aligned}$$

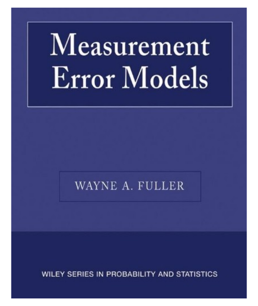
Equation error accommodates nonlinear signal in the relationship between *C* and *U* (via *t*).

Representation error accommodates scales in *C* or *U*, but not *both*.

Measure - representation error   Model + model error



Ceci n'est pas un éléphant
(en.wikipedia.org/wiki/Asian_elephant)
All models are wrong



Although we write these errors separately, we also need to ask to what extent geophysicists and statisticians are concerned with the same error.

In the literature, it is rare to find a separate equation error or representation error term included in the uncalibrated measurement equation. An accommodation of representation error in both *C* and *U* is sometimes seen as a debate. Perhaps we are just "ripping off the bandaid", in that if we want to accommodate representation error in *C* and *U*, then we might need to accommodate equation error in *C* and *U* as well.

Including a separate equation error in *U* can be justified by its linear relationship to *t* (note that *C* and *U* are equally justified, even though their numerical values differ). Moreover, there seems to be no other way for this model to incorporate the impact of hidden or confounding variables. The only measurements are *C* and *U*, and these may be impacted differently. (There is a third justification that is more practical, as it relates to the model solution.)

```
n 17. aug. 2017 10:49, GRAHAM DUNN wrote:
Hi Rick

I had wondered whether your correlated errors were equation errors
rather than measurement errors (or a combination of both) but I
don't think it has any practical implications (just a different
interpretation). I can't see any reason why equation errors should
not be correlated.

With best wishes
Graham


Graham Dunn
Professor of Biomedical Statistics
Centre for Biostatistics
University of Manchester
```

Equation error can be **correlated** and **combined** with other errors

Nonlinearity in all measures

Is that still called **linear regression?**

In any case, the interpretation of error in terms of signal is challenging, but in 2017, Graham Dunn provided a five-line comment. This took some time to absorb, but basically, he said that equation error could be mixed in with other errors and equation error could be correlated.

If we carry this advice forward, and place measurements on the LHS and model terms on the RHS, then we have linear association, correlated and uncorrelated equation error, and representation error, which is uncorrelated by definition. All are signal components insofar as they are needed to describe *C* and *U* even without measurement error. Without loss of generality, measurement error in *C* and *U* can be divided among the signal terms on the RHS, or written as its own term (as we wish):

$$\begin{aligned} C &= t + \epsilon_C^{EQ} + \epsilon_C^{REP} \\ U &= \alpha_U + \beta_U t + \epsilon_U^{EQ} + \epsilon_U^{REP} \end{aligned}$$

Measures = linear association
+ correlated equation error
+ uncorrelated equation error
and representation error } Signal
+ measurement error (in all terms) Noise

Since we are interested in terms that can be evaluated numerically, we follow Prof. Dunn's advice and let correlated equation error stand alone and combine uncorrelated equation error and representation error into a separate term:

$$\begin{aligned} C &= t + \epsilon + \epsilon_C \\ U &= \alpha_U + \beta_U t + \epsilon + \epsilon_U \end{aligned}$$

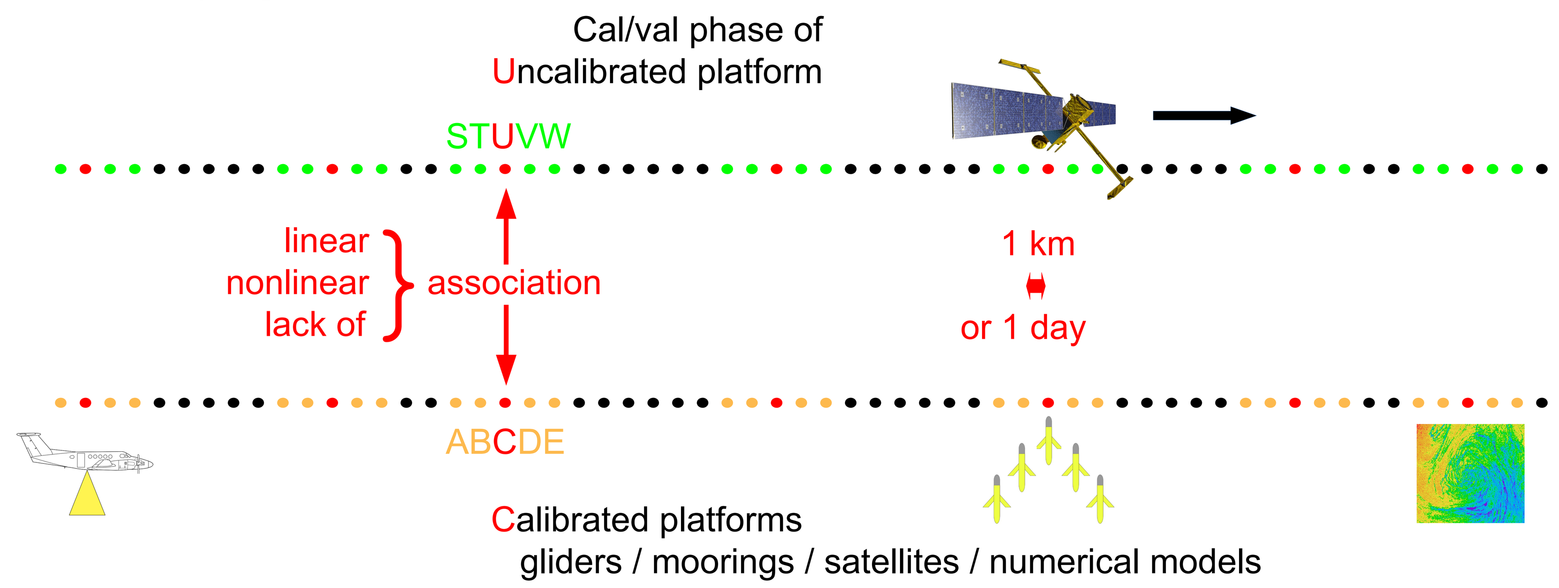
Measures = linear association
+ nonlinear association
+ lack of association

The core of our wavelike measurement model is here. This form makes explicit a "systematic error" term ε that we would call error cross-correlation, but whose *genuine* interpretation is nonlinear association. Similarly, our "random error" terms have the *genuine* interpretation of a lack of association. Although signal defines our interpretation of terms, again, we can conceive of spurious components in each term (https://arxiv.org/abs/2110.08969 is a more formal derivation).

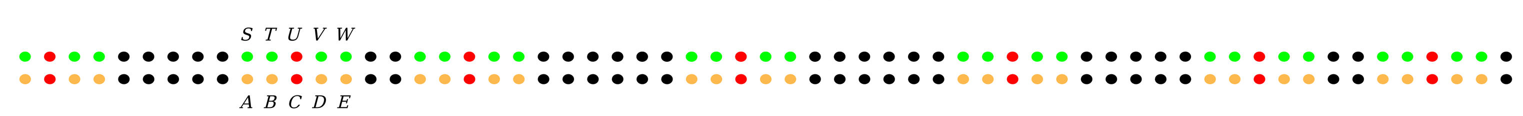
Can systematic and random error, as defined in VIM (2008), also be interpreted in a way that follows Pearson's (1902) emphasis of "genuine error"? Respectively, we submit that "nonlinear association" and "lack of association" are their genuine interpretations. First, we develop these out of a purely statistical notion of a "measurement model", or regression model. Then, we examine a playful controlled experiment in the hydrological context.

1. Sampling Signal as a Wave

Our wavelike measurement model employs a collocated sampling strategy that is denoted by the coloured dots (moving left to right). Of interest is the linear, nonlinear, and lack of association between a less familiar "uncalibrated" platform or instrument, and one that is more familiar, like an in situ platform. It is notable that wavelike sampling is possible for large datasets that are gridded in space or time; not every scientific discipline has the modelling freedom that comes with such extensive sampling.



We employ a trichotomy (e.g., truth + systematic + random error) to describe individual model terms. Model heritage can be traced to an introduction by Pearson (1902) of three observers as a means to identify the errors of each. The geophysical focus on Pearson's approach re-emerged as triple collocation in 1988, and has helped motivate our wavelike model. However, instead of three independent observers (or three sets of red dots), our method-of-moments solutions are identifiable using "predictive samples" for just two datasets. Whereas a collocation of three or more datasets is rare, our wavelike model still requires successive sampling of two datasets:



$$\begin{aligned}
S &= \alpha_S + \beta_S\, t + \lambda_S\left(\lambda_T\left(\epsilon + \epsilon_U\right) + \epsilon_T\right) + \epsilon_S \\
T &= \alpha_T + \beta_T\, t + \lambda_T\left(\epsilon + \epsilon_U\right) + \epsilon_T \\
U &= \alpha_U + \beta_U\, t + \epsilon + \epsilon_U \\
V &= \alpha_V + \beta_V\, t + \lambda_V\left(\epsilon + \epsilon_U\right) + \epsilon_V \\
W &= \alpha_W + \beta_W\, t + \lambda_W\left(\lambda_V\left(\epsilon + \epsilon_U\right) + \epsilon_V\right) + \epsilon_W \\
A &= \alpha_A + \beta_A\, t + \lambda_A\left(\lambda_B\left(\epsilon + \epsilon_C\right) + \epsilon_B\right) + \epsilon_A \\
B &= \alpha_B + \beta_B\, t + \lambda_B\left(\epsilon + \epsilon_C\right) + \epsilon_B \\
C &= t + \epsilon + \epsilon_C \\
D &= \alpha_D + \beta_D\, t + \lambda_D\left(\epsilon + \epsilon_C\right) + \epsilon_D \\
E &= \alpha_E + \beta_E\, t + \lambda_E\left(\lambda_D\left(\epsilon + \epsilon_C\right) + \epsilon_D\right) + \epsilon_E
\end{aligned}$$

S,T,V,W are autocorrelated *instruments* (proxies) of *U*

A,B,D,E are autocorrelated *instruments* (proxies) of *C*

1 km or 1 day

or "predictive samples" ("forecasts" or "revcasts")

Here, α, β, and *t* capture linear association between samples, with symmetric first-order autoregressive errors for each dataset separately. Errors are also correlated in ε, which is a term not always included in canonical models (Fuller 2006). Geophysical theory (equations of motion) allows for a nonlinear relationship between any two samples (like *B-C* or *C-U*). The key question is how a wavelike model accommodates nonlinearity in what looks to be a set of linear relationships?

4. Recommendations and References

(1) Progress or lack of progress (challenges) so far
Mahalanobis (1947) seems to address the metrological practice of *expanded uncertainty*, but is unclear about its motivation. Meteorologists might pick up where Mahalanobis (1947) left off [less with reference to Type-A/B cyclogenesis (Bosart 1994) and more with reference to what Stigler (2018) refers to "high serial correlations" in atmospheric soundings]. Pearson (1902) and Mahalanobis (1947) seem to be advocating for genuine interpretations, but equation error (Fuller 2006) only develops later.

(2) Indication of what is needed to improve the information content in environmental measurements or models
Metrologists might distinguish quantitative models to address a) "what is the process?" and b) "how do we measure the process?" Even for the same process, these are not necessarily the same model. We are familiar with a hierarchy of process models (Held 2015), but perhaps not with a hierarchy of measurement (or regression) models. One tenet of statistical inquiry is a model with two terms: observation = truth + error (Salsburg 2017), but metrological convention admits a further division into systematic and random error. We submit that the corresponding nonlinear model with three terms (observation = linear + nonlinear + unassociated) also has quantitative solutions.

(3) Proposals of new or existing tools not currently in use, for improving information content
Shameless self-promotion: "numerical consensus" methods to solve this nonlinear model are at https://github.com/JuliaAtmosOceanHydro/MeasurementModelDemos

(4) Description of potential means to achieve that improvement
One may appeal to colleagues who wish to learn about nonlinear processes that, at the moment, they might be training a neural network to emulate. (Outside our fields, some colleagues also deal with nonlinear processes that are downright spooky.)

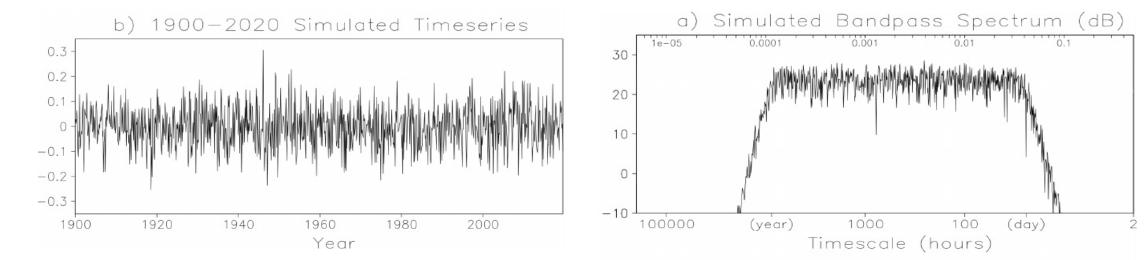
(5) Opportunities for collaboration between communities to achieve these improvements
There seems to be an opportunity to bring statisticians and data scientists closer together (Donoho 2017).

References are provided at
https://docs.google.com/document/d/1ukxYVb0Ec8DERQIzqh99DiByYRM97QEPES5sniROTiM

3. A Control Experiment

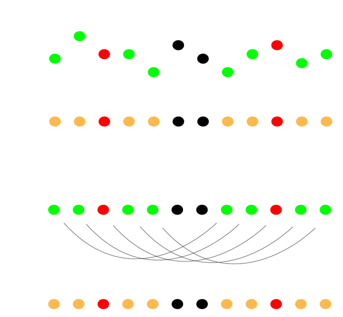
Returning to Prof. Dunn's question of practical implications, there is no doubt about a different interpretation: measurements, variance, and covariance all have a nonlinear component, and thus, Pearson correlation also has a nonlinear component. Such interpretations can be quantified. By random samples of uniform and Gaussian distributions, we can simulate anomalous river height or river slope at 10000 well separated reaches that have been observed hourly since the turn of the year 1900. The anomaly is such that there is only variability between a day and a year:

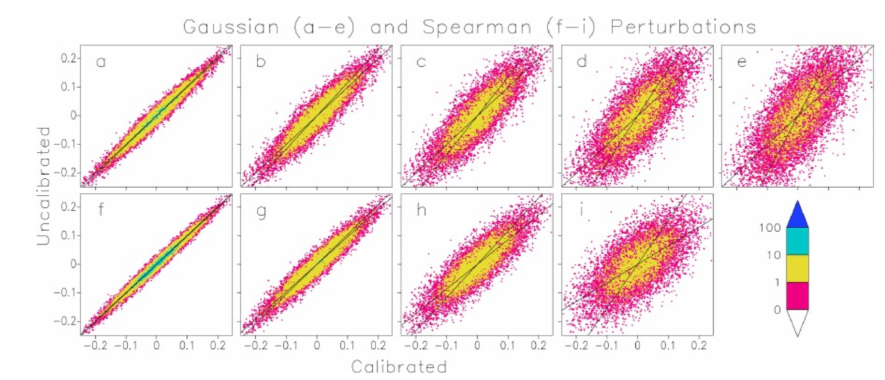
- For each river, signal is slow and wavelike (daily-to-yearly)
- 10000 well-separated rivers sampled (*STUVW* and *ABCDE*) at 1-hr intervals
- Perfect calibration : $\beta_U = 1$ (*STUVW* = *ABCDE*)



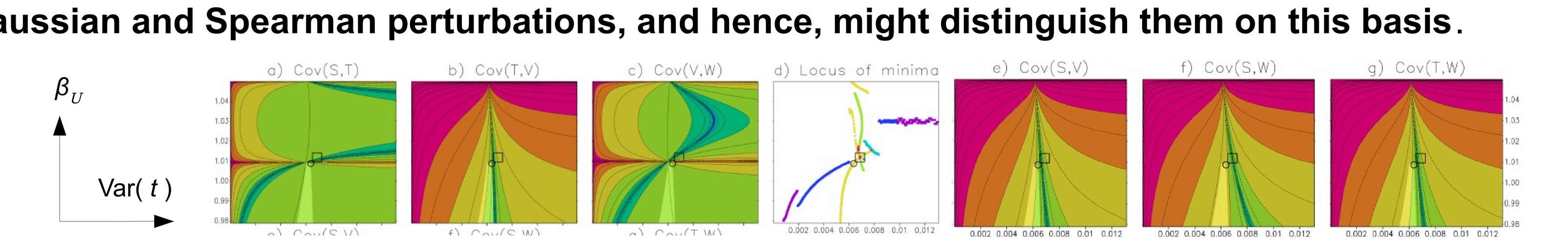
For each river reach, we generate about a million samples of a uniform random distribution, bandpass filter to get a timeseries in the range -0.2 to 0.2, and from the middle of each timeseries (like the one above), we take 5 consecutive samples to define the slowly varying signal in both *C* and *U*, which are perfectly calibrated. This experiment takes up the challenge of trying to simulate a confounding impact in the presence of measurement error, so two types of perturbations are employed, and both are designed to be resolved by a lack of association between *C* and *U*:

- Gaussian perturbations (only of *U*) are fast and turbulent
- Spearman perturbations (*C* vs *U*) are slow and wavelike (look Gaussian)
- Can we distinguish Gaussian (noise) and Spearman (signal) perturbations?





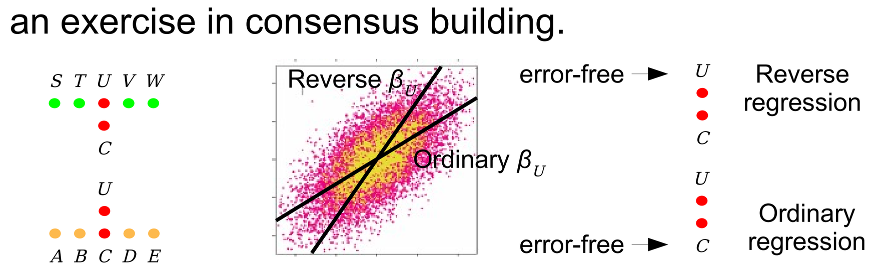
Gaussian perturbations are added to *U* in varying degrees. Spearman perturbations are a random reordering of signal in *U* that might appear to treat two river reaches as exchangeable, but are meant to simulate a confounding impact without imposing a systematic relationship between *C* and *U*. Although Gaussian and Spearman perturbations might appear to be similar in the upper and lower panels, by design, Gaussian and Spearman perturbations impose no turbulent variations across *STUVW*. Spearman perturbations can be said to perturb *U* alone, but because they are a perturbation of relative order, it follows that if we consider *C* to be unperturbed and equal to the baseline signal, and *U* alone to be perturbed, then **it is only predictive samples involving *STUVW*, but not *ABCDE*, that access both Gaussian and Spearman perturbations, and hence, might distinguish them on this basis**.



Model solutions by the method of moments require only six sample sets, *ABCDEU* or *CSTUVW*. Each set yields 6 variance equations, 15 covariance equations, and 17 unknowns. Analytic solutions exist for 15 of these unknowns, but the remaining two unknowns are the variance of *t* and β_U, the two key parameters. We minimize the distance of a solution to these two parameters using minima or pathways in each of six covariance equations that involve the variables other than *U* and *C*. Together with the constraint that variance is positive, as expected, we refer to this equitable accommodation of weak or imperfect constraints as an exercise in consensus building.

Wavelike Solutions

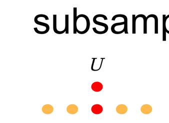
- Noisy subsamples
- Consensus solutions
- Better bounds on β_U

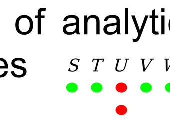


Particle-like Solutions

- No subsamples
- Analytic solutions
- Worse bounds on β_U

Our two wavelike solutions happen to be directly comparable to ordinary and reverse regression solutions that are analytic. So, in the case of our wavelike solutions, *STVW* and *ABDE*, they provide subsamples around *U* and *C* that allow us to estimate Gaussian error, and in the case of analytic solutions by reverse and ordinary regression, we would assume that those samples don't exist and that there is no error in the *U* or *C* variable bracketed by them.

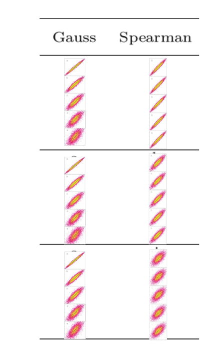




σ_t^2	β_U (O/WAV/R)	Var(C)	LA	NA	UR	Var(U)	LA	NA	UR
6.46	0.95/0.96/1.08	6.57	98.25	0.60	1.15	6.75	87.88	0.58	11.54
6.47	0.95/0.96/1.21	6.57	98.50	0.31	1.20	7.52	79.25	0.27	20.48
6.44	0.95/0.96/1.41	6.57	97.95	0.68	1.37	8.79	67.63	0.51	31.86
6.45	0.95/0.96/1.70	6.57	98.14	0.41	1.45	10.57	56.46	0.25	43.29
6.43	0.95/0.96/2.06	6.57	97.80	0.60	1.60	12.85	46.39	0.31	53.31
6.46	0.85/0.86/1.21	6.57	98.36	0.35	1.28	6.75	70.75	0.34	28.90
6.47	0.85/0.86/1.35	6.57	98.50	0.20	1.30	7.52	63.75	0.18	36.07
6.46	0.85/0.86/1.57	6.57	98.35	0.27	1.38	8.79	54.56	0.20	45.24
6.43	0.85/0.86/1.89	6.57	97.87	0.57	1.56	10.57	45.30	0.36	54.34
6.42	0.85/0.86/2.30	6.57	97.62	0.65	1.73	12.85	37.29	0.33	62.38
6.35	0.63/0.64/1.62	6.57	96.68	1.15	2.17	6.75	39.06	1.12	59.83
6.37	0.63/0.64/1.80	6.57	96.89	0.95	2.16	7.52	35.23	0.84	63.93
6.38	0.63/0.65/2.11	6.57	97.04	0.83	2.13	8.79	30.21	0.62	69.17
6.40	0.63/0.65/2.54	6.57	97.34	0.57	2.09	10.57	25.26	0.36	74.39
6.43	0.63/0.65/3.08	6.57	97.80	0.22	1.97	12.85	20.90	0.11	78.98

Our lower bound on β_U = 1 (WAV) is **no better than ordinary regression** (O)

No distinction : mixed Spearman–Gaussian proxy



σ_t^2	β_U (O/WAV/R)	Var(C)	LA	NA	UR	Var(U)	LA	NA	UR
5.79	0.95/1.04/1.08	6.57	88.13	3.28	8.59	6.75	92.47	3.20	4.33
5.30	0.95/1.04/1.21	6.57	80.71	11.22	8.07	7.52	75.68	9.81	14.51
4.88	0.95/1.03/1.41	6.57	74.26	18.41	7.32	8.79	58.78	13.76	27.46
4.42	0.95/1.02/1.70	6.57	67.23	26.30	6.47	10.57	43.44	16.34	40.22
3.94	0.95/1.01/2.06	6.57	60.00	34.18	5.82	12.85	31.36	17.44	51.20
4.68	0.85/1.16/1.21	6.57	71.24	2.57	26.20	6.75	92.67	2.50	4.84
4.30	0.85/1.13/1.35	6.57	65.45	10.69	23.86	7.52	73.75	9.34	16.91
3.88	0.85/1.11/1.57	6.57	59.09	19.23	21.68	8.79	55.00	14.38	30.62
3.55	0.85/1.13/1.89	6.57	54.06	23.80	22.14	10.57	43.15	14.79	42.06
3.05	0.85/1.12/2.30	6.57	46.39	33.32	20.29	12.85	29.56	17.02	53.42
2.65	0.63/1.54/1.62	6.57	40.28	1.63	58.08	6.75	92.34	1.59	6.07
2.41	0.63/1.39/1.80	6.57	36.65	12.03	51.32	7.52	63.62	10.51	25.87
2.27	0.63/1.45/2.11	6.57	34.47	13.13	52.40	8.79	55.33	9.81	34.86
2.03	0.63/1.48/2.54	6.57	30.94	17.78	51.28	10.57	42.18	11.04	46.78
1.57	0.63/1.51/3.08	6.57	23.95	27.50	48.55	12.85	27.62	14.05	58.33

Our upper bound on β_U = 1 (WAV) is **better than reverse regression** (R)

Spearman proxy

Gaussian proxy

Linear (LA), nonlinear (NA), and lack of association (UR) results are expressed as normalized *U* and *C* variance budgets, averaged over 100 simulations. By design, *ABDE* are poorly employed, because they don't sample any Gaussian perturbations, which makes this model solution perfectly consistent with ordinary regression, which assumes *C* is error-free. By contrast, *STVW* are well employed in sampling Gaussian perturbations. This model solution functions an upper bound on β_U, but provides a better bound than reverse regression. **The lack of association terms in *C* and *U* allow us to distinguish Spearman and Gaussian perturbations**, and we suspect that a model with all three association categories is needed to do so.