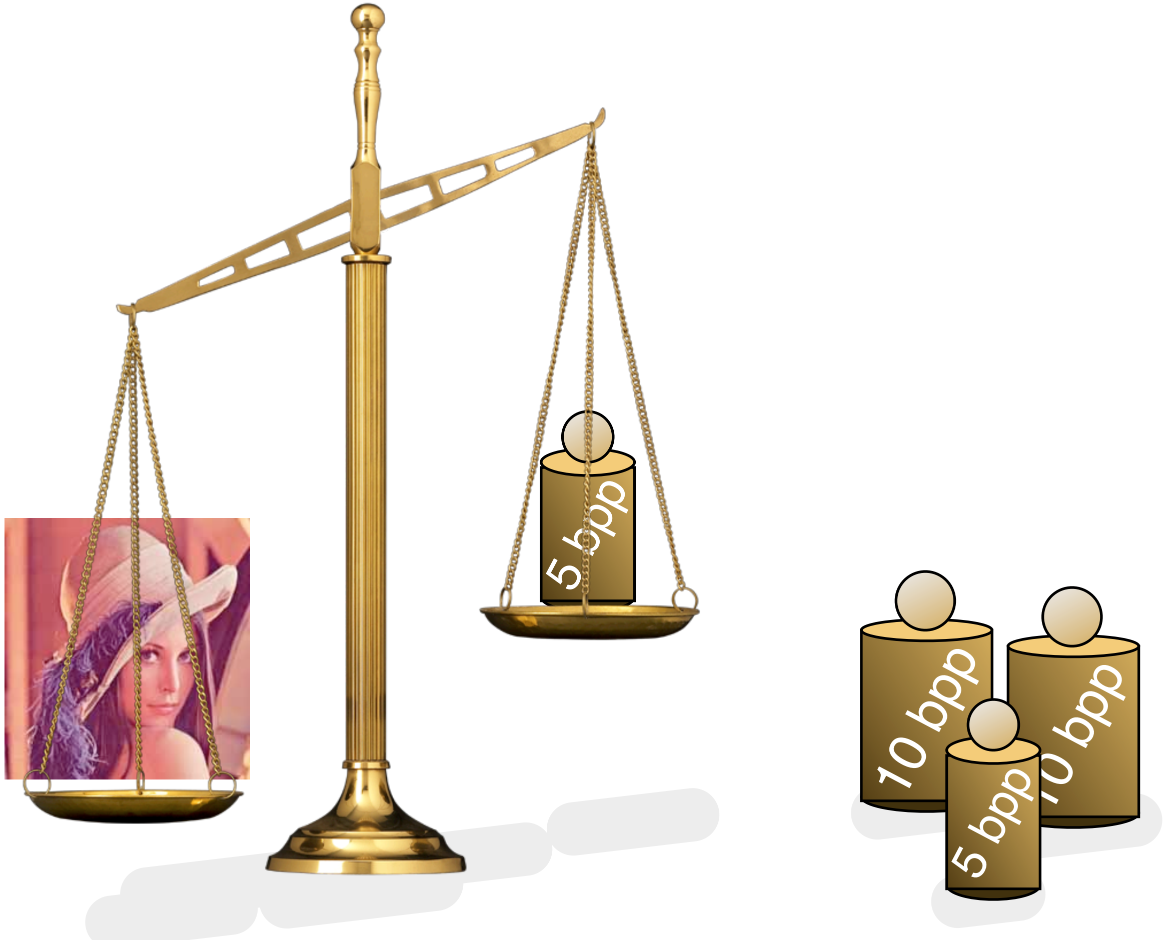
A Two-Factor Error Model for Quantitative Steganalysis

Security, Steganography and Watermarking of Multimedia Contents
SPIE 2006 · San José · California · USA

Rainer Böhme
rainer.boehme@inf.tu-dresden.de
Technische Universität Dresden, Institute for System Architecture
01062 Dresden, Germany

Andrew D. Ker
adk@comlab.ox.ac.uk
Oxford University Computing Laboratory
Parks Road, Oxford OX1 3QD, United Kingdom

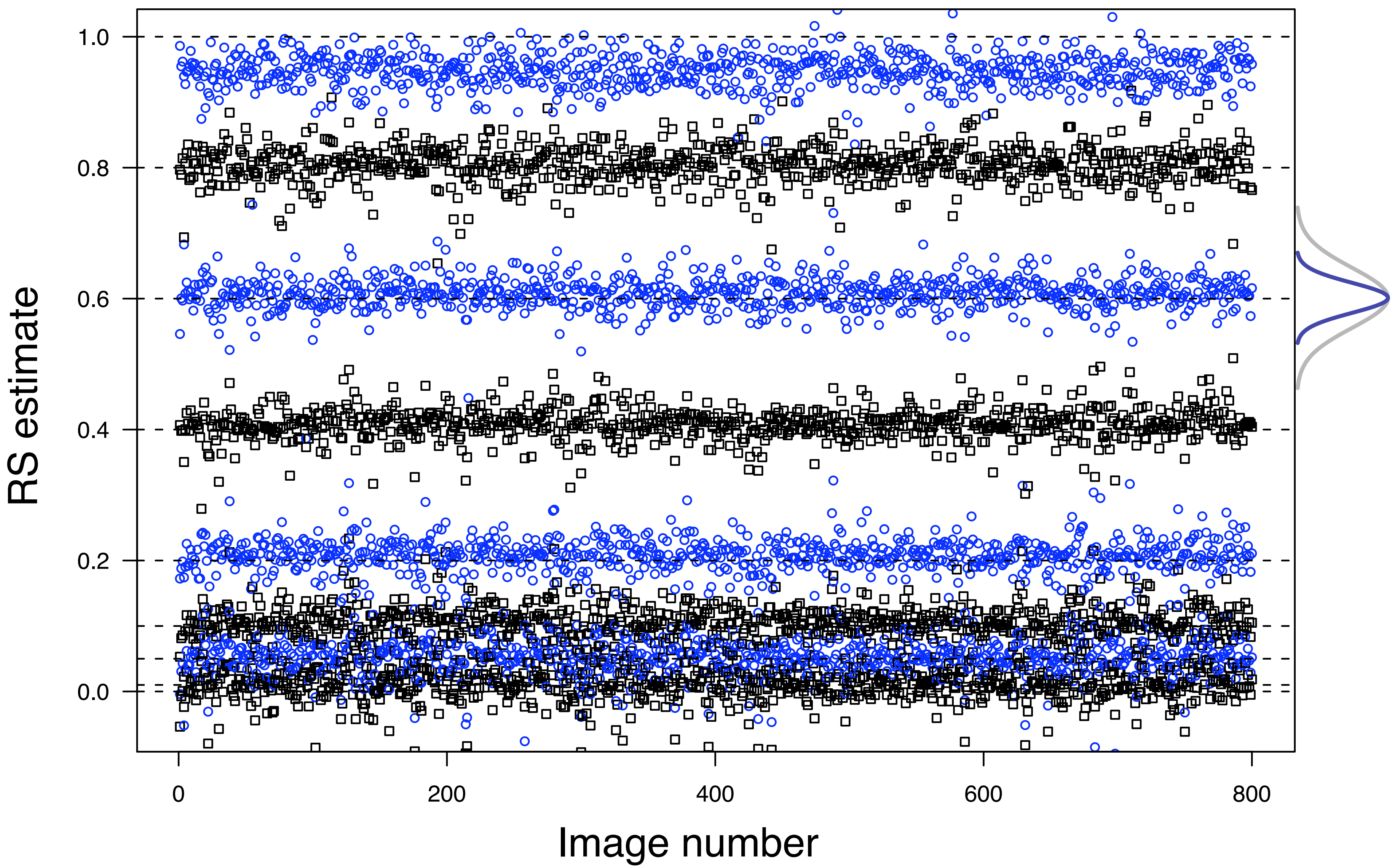
The first authors has received a “Windows on Science” travel stipend from the European Office of Aerospace Research and Development (EOARD). The second author is a Royal Society Research Fellow.



A number of different estimators has been proposed for LSB embedding.

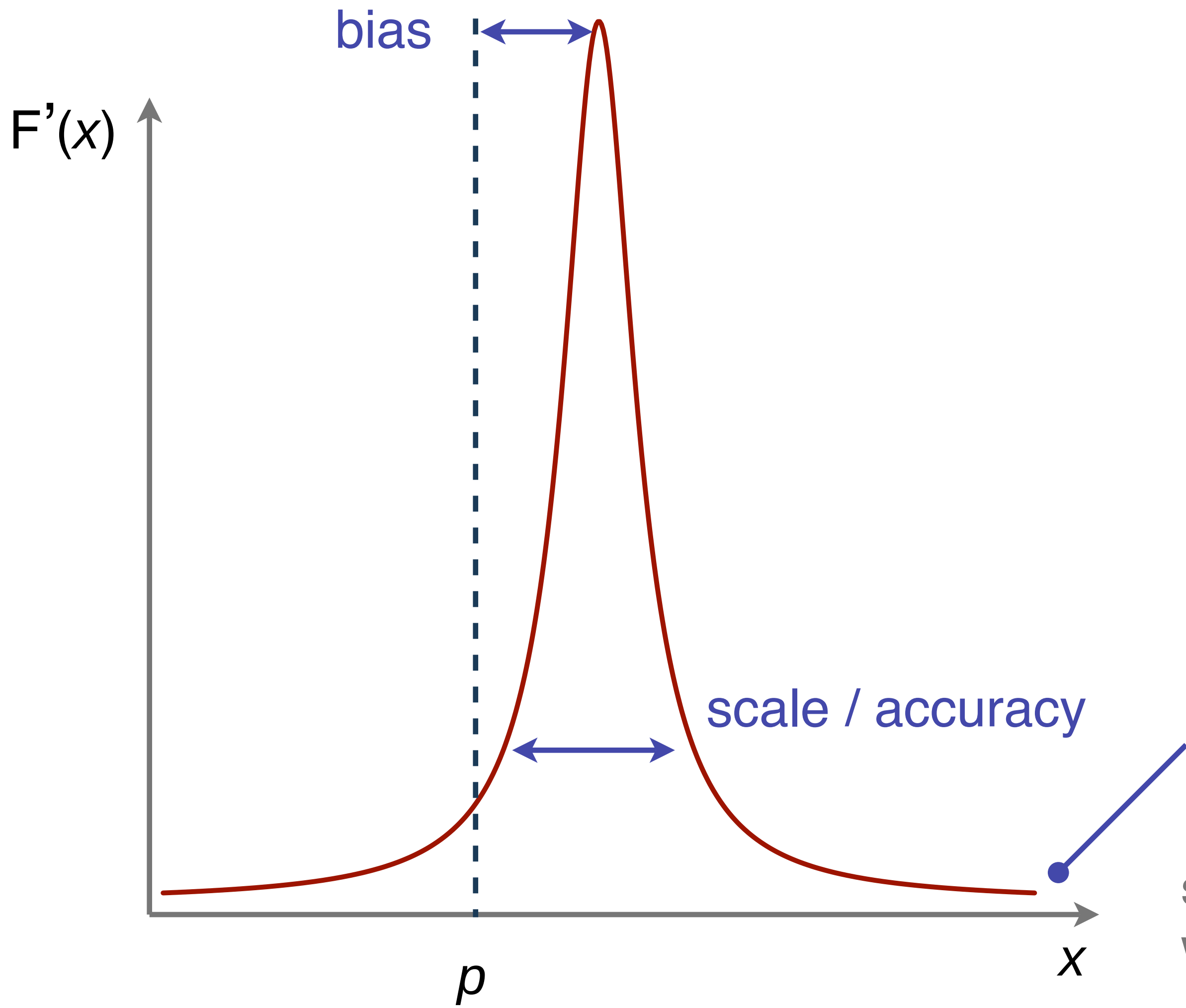
Ker, 2004

Typical Results from Secret Message Length Estimation



Results from 7200 attacks on 800 never-compressed grayscale images

Error Distribution of Estimates



$$\hat{p} = p + e$$

Distribution function

$F(x) = P(e < x) = ???$

$1 - F(x) \sim x^{-k}$

heavy tails

simulation results from images
with randomly chosen message

Error distribution has previously been modelled as Cauchy distribution.

Boehme, 2005

A Two-Factor Error Model

$$\hat{p}_{i,j} = p_{i,j} + e$$

$$\hat{p}_{i,j} - p_{i,j} = X(\mu_i, \sigma_i) + Z(\theta)$$

Within-image error due to random correlation with the message

Between-image error due to characteristics of the image

$$e \sim \mathrm{D}(X) \circ \mathrm{D}(Z)$$

Symbols

- i .. cover image index
- j .. message index
- $\hat{p}$ .. estimation result
- p .. actual embedding rate

Random variables

- X .. within-image error
- Z .. between-image error
- $\mathrm{D}(\cdot)$ .. distribution function operator
- $\circ$ .. convolution operator

New Research Questions

- ? Shape of D(X) and D(Z)
- ? Relative magnitude
- ? Re-examine influencing factors for error components
- ? Similarities and differences between different quantitative steganalysis methods

! *We use a large-scale experiment to explore the relationship for LSB detectors/estimators empirically.*

Experimental Setup

RS	WS	SPA	SPA/LSM	Triples
Fridrich, Goljan & Du, 2001	Fridrich & Goljan 2004	Dumitrescu, Wu & Wang 2002	Lu, Luo, Tang & Shen, 2004	Ker, 2005

800 .. never-compressed images 640 x 458

200 .. secret messages per image

5 .. detectors

Experimental Setup

	RS emb. ratio 0 - 1 bpp	WS emb. ratio 0 - 1 bpp	SPA emb. ratio 0 - 1 bpp	SPA/LSM emb. ratio 0 - 1 bpp	Triples emb. ratio 0 - 1 bpp
grayscale					
red channel					

800 .. never-compressed images 640 x 458

200 .. secret messages per image

5 .. detectors

8 .. embedding ratios (+carriers)

2 .. colour channels: grayscale and red

		RS emb. ratio 0 - 1 bpp	WS emb. ratio 0 - 1 bpp	SPA emb. ratio 0 - 1 bpp	SPA/LSM emb. ratio 0 - 1 bpp	Triples emb. ratio 0 - 1 bpp
grayscale	100%					
	75%					
	50%					
	25%					
red channel	100%					
	75%					
	50%					
	25%					

800 .. never-compressed images 640 x 458

200 .. secret messages per image

5 .. detectors

8 .. embedding ratios (+carriers)

2 .. colour channels: grayscale and red

4 .. image sizes

Experimental Setup

			RS	WS	SPA	SPA/LSM	Triples
			emb. ratio 0 - 1 bpp	emb. ratio 0 - 1 bpp	emb. ratio 0 - 1 bpp	emb. ratio 0 - 1 bpp	emb. ratio 0 - 1 bpp
grayscale	100%						
	75%	scale					
		crop					
	50%	scale					
		crop					
	25%	scale					
		crop					
red channel	100%						
	75%	scale					
		crop					
	50%	scale					
		crop					
	25%	scale					
		crop					

800 .. never-compressed images 640 x 458

200 .. secret messages per image

5 .. detectors

8 .. embedding ratios (+carriers)

2 .. colour channels: grayscale and red

4 .. image sizes

2 .. downsizing methods (scale and crop)

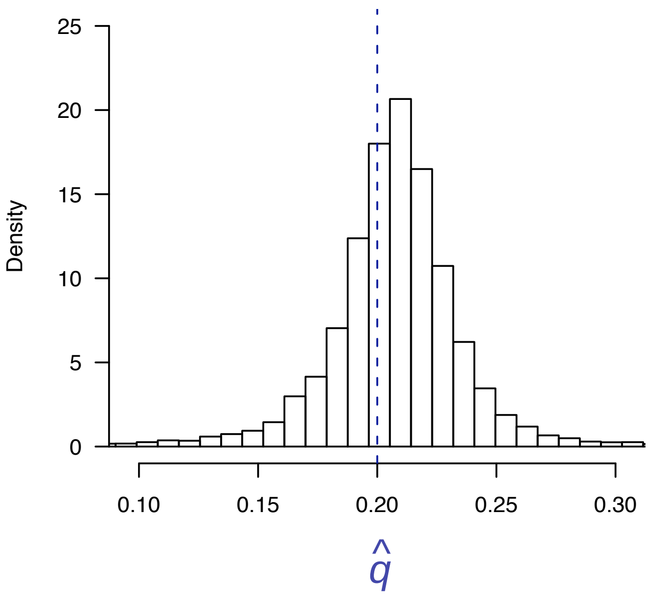
Experimental Setup

2 .. pre-compression methods (raw and JPEG)

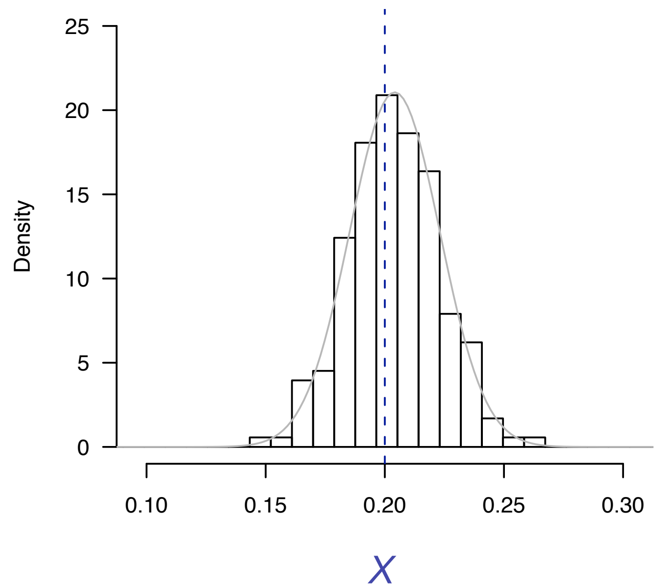
- *800 x 200 attacks per "cell" totalling up to about 200 M attacks*

Error Components for RS Analysis

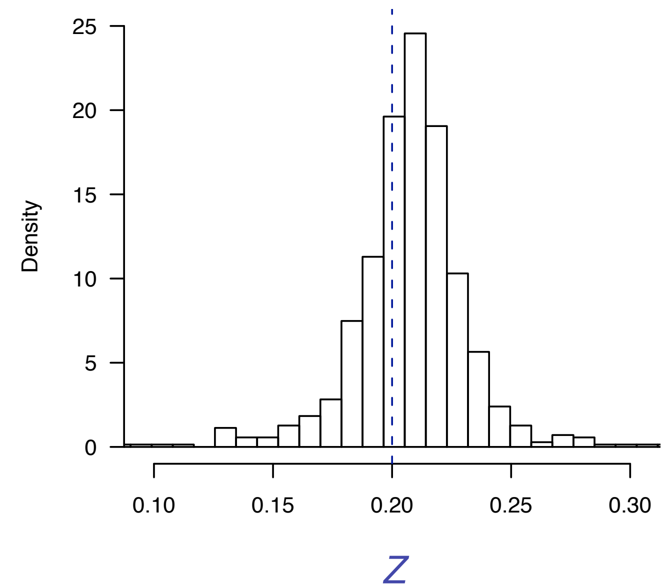
Compound estimation error



Within-image error



Between-image error



Data from 800 never-compressed grayscale images with embedding ratio $p = 0.2$

Within-image error

Empirical evidence for Normality from a series of Shapiro-Wilk tests (see paper).

Between-image error

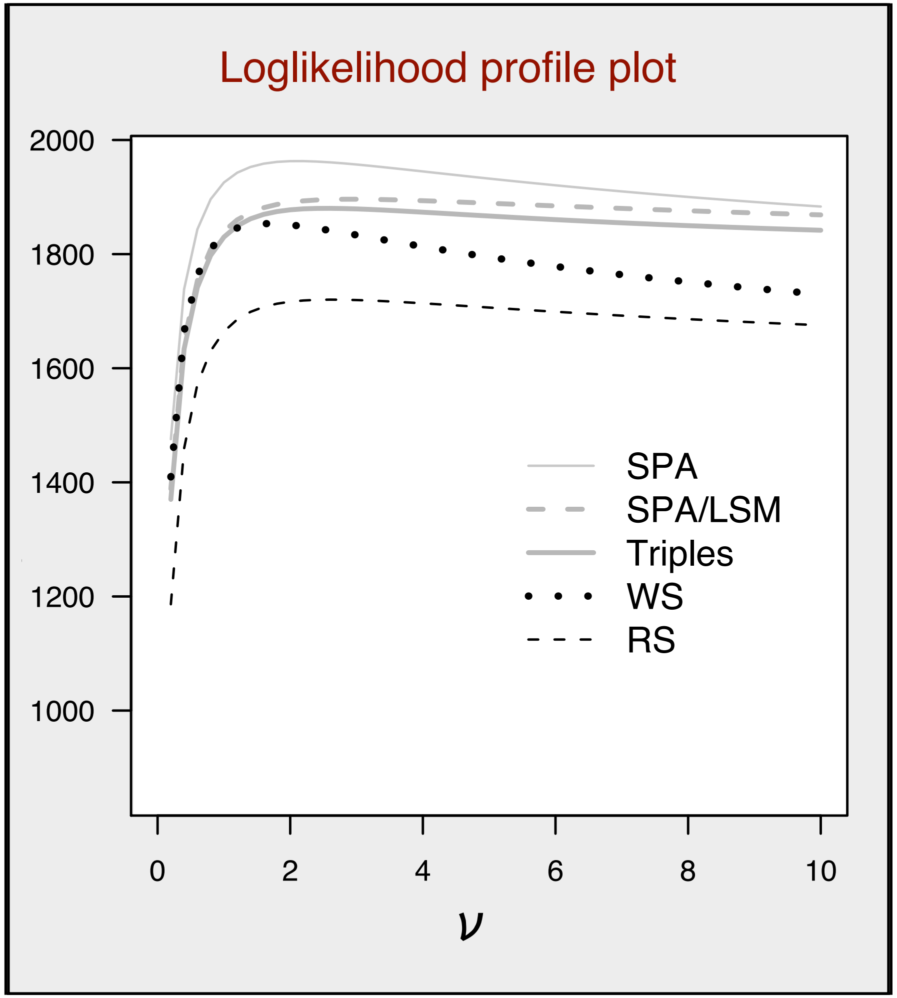
Good fit for heteroscedastic Student *t* distribution:

$$t_\nu(x, \lambda) = \frac{1}{\sqrt{\pi}} \frac{\Gamma((1+\nu)/2)}{\Gamma(\nu/2)} \frac{\lambda^\nu}{(\lambda^2 + x^2)^{(1+\nu)/2}}$$

heavy tails with tail index ν

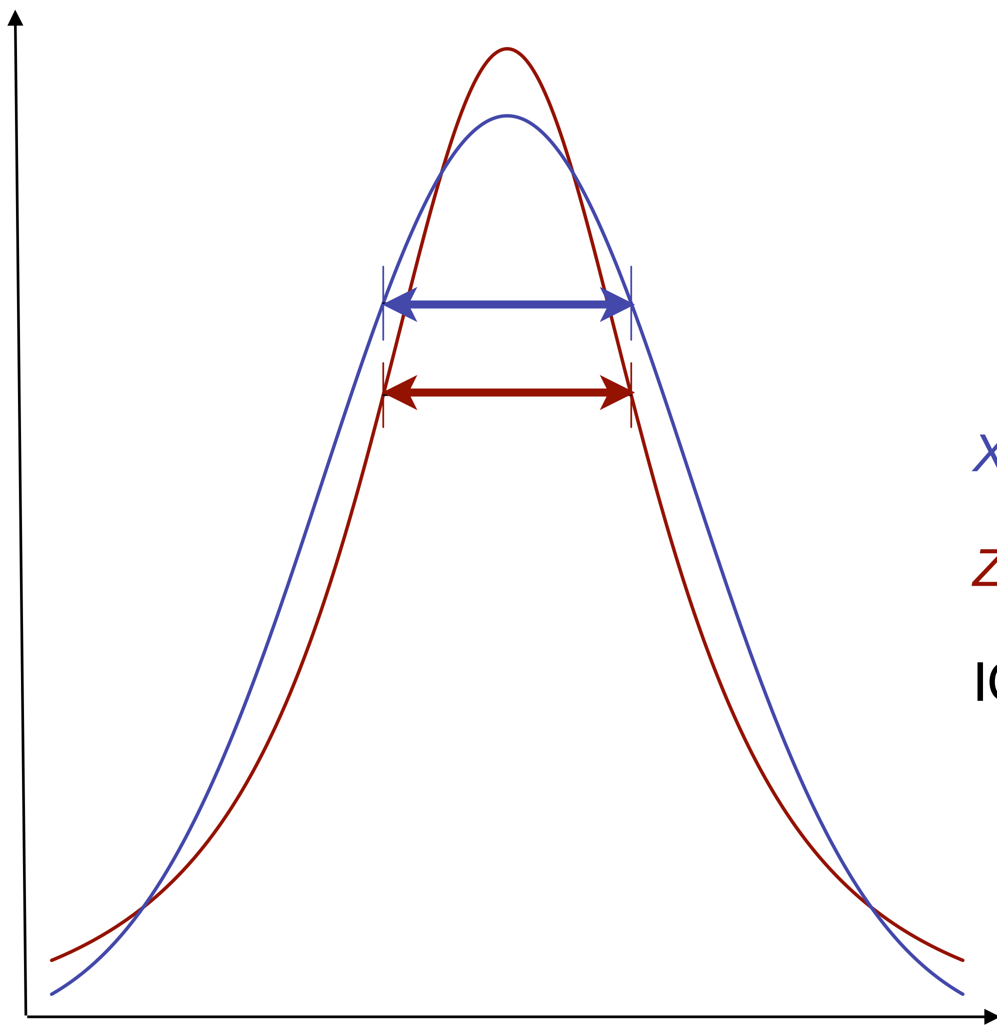
λ .. scale parameter

ν .. degrees of freedom parameter



Robust Comparison of Distribution Spread

Inter-quartile ranges (IQR)



X .. $\sigma = 1.2105$

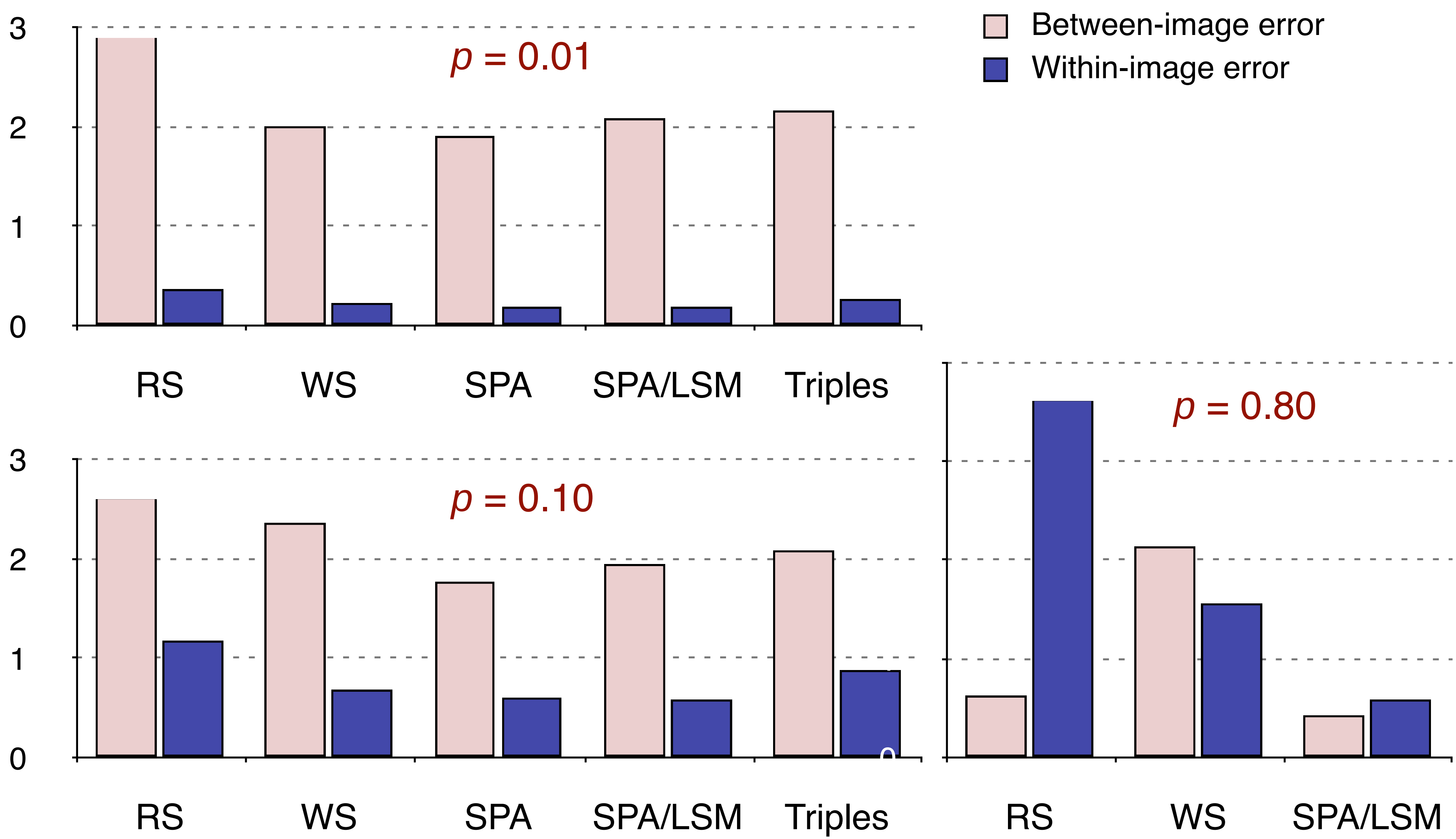
Z .. $\lambda = 1$, $\nu = 2$

$\mathrm{IQR} = Q_{75} - Q_{25} = 1.633$

Is it inappropriate to compare short- and heavy-tailed distributions with moment statistics.

Comparison of Error Magnitudes

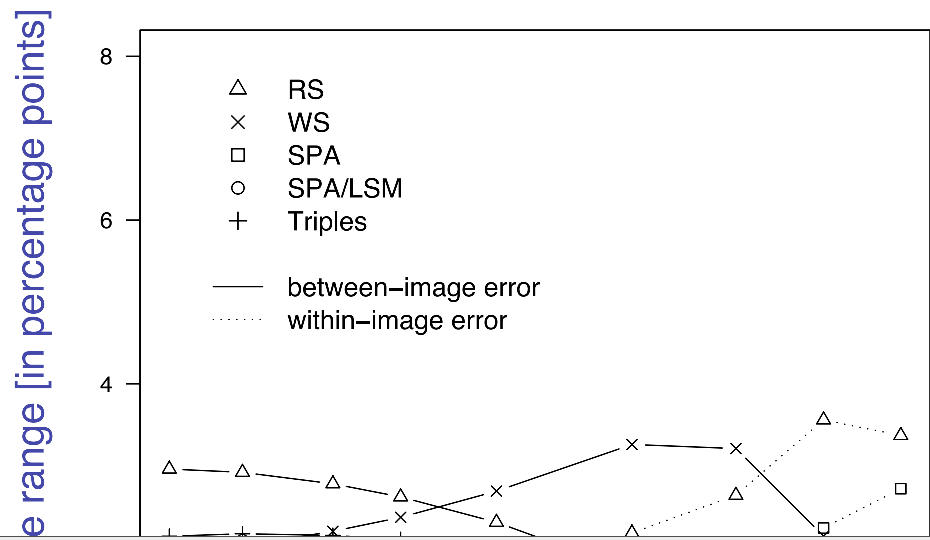
Inter-quartile ranges $Q_{75} - Q_{25}$



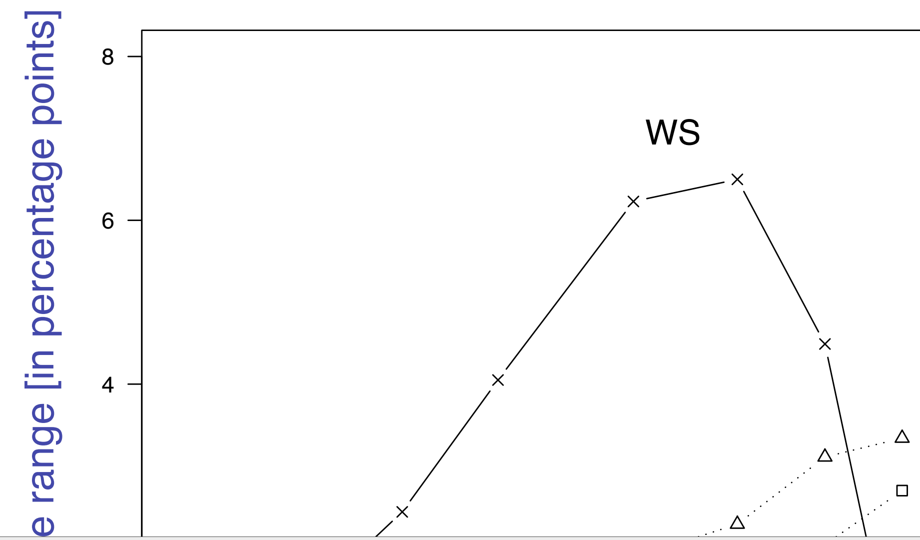
Data from 2 M attacks on 800 never-compressed grayscale images.

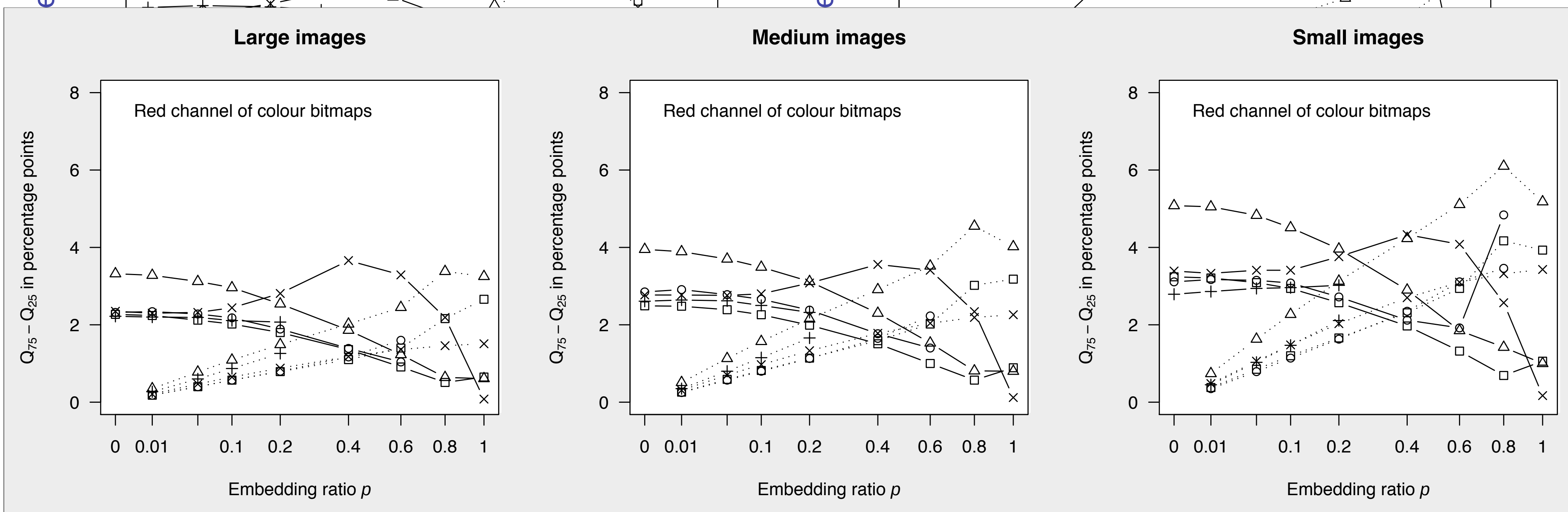
Comparison of Error Magnitudes (cont’d)

Never-compressed grayscale images



JPEG compressed grayscale images





Default model

Estimate detector specific constant and scale bias for between-image errors.

$$\bar{p}_i = a_0 + a_1 \cdot p + e \qquad \text{with} \qquad \bar{p}_i = \frac{1}{200} \sum_j \hat{p}_{i,j}$$

Absolute bias model

Assume that an image-specific bias can be approximated from the detector result if nothing is embedded.

$$\bar{p}_i = a_1 \cdot p + a_2 \cdot \hat{p}_i^{(0)} + e$$

Relative bias model

Test hypothesis that $(\hat{p}_i - p) = \hat{p}_i^{(0)} + p \cdot (1 - \hat{p}_i^{(0)})$

$$\bar{p}_i = \underset{\downarrow\, 1}{a_1} \cdot p + \underset{\downarrow\, 1}{a_2} \cdot (\hat{p}_i^{(0)} - p \cdot \hat{p}_i^{(0)}) + e$$

Assumed residual distribution: $e \sim t_2(0,\lambda)$

Fitted Coefficients for Between-Image Bias

Dependent variable $\bar{p}_i$ with predictors p and $\hat{p}_i^{(0)}$:

		Detector				
Model	Parameter	RS	WS	SPA	SPA/LSM	Triples
Default	a_0	0.01	0.01	0.00	0.00	0.00
	a_1	0.98	1.00	0.99	0.99	0.99
	Residual IQR	2.98	2.30	1.99	1.65	2.03
Absolute bias	a_1	1.01	1.01	1.00	1.00	0.99
	a_2	0.79	0.83	0.82	0.89	0.95
	Residual IQR	1.00	1.65	0.48	0.25	0.14
Relative bias	a_1	1.01	1.01	1.00	1.00	0.99
	a_2	0.99	0.99	1.00	1.00	1.00
	Residual IQR	0.49	1.55	0.06	0.06	0.08
		84%	33%	97%	96%	96%

All coefficients significant on the [illegible]el.
Data from 27 k attacks on never-compressed grayscale images.

Data fitted with heteroscedastic *t* regression methods. (see Taylor & Verbyla, 2005)

Correlation of Image-Specific Bias

Correlation coefficients of $\hat{p}_i^{(0)}$ between detectors over images.

	Detector				
	RS	**WS**	**SPA**	**SPA/LSM**	**Triples**
RS	1.00				
WS	0.64	1.00			
SPA	0.76	0.86	1.00		
SPA/LSM	0.60	0.86	0.89	1.00	
Triples	0.45	0.66	0.64	0.70	1.00

Data from 4000 attacks on never-compressed grayscale images.

Correlation coefficients for JPEG images are somewhat lower.

All correlation matrices have only one eigenvalue larger than 1.

Correlation coefficients estimated with the *t*-copula method.

(see Dematra & McNeil, 2005)

Magnitude of within-image error

Ordinary least squares regression to data with $p > 0 = const$

$$\log \hat{\sigma}_i = b_0 + b_1 \cdot x + e \qquad e \sim N(0, \sigma_e)$$

Magnitude of between-image error

Maximum likelihood fit for heteroscedastic t regression:

$$\hat{p}_i^{(0)} = a_0 + \varepsilon \qquad \varepsilon \sim t_2(0, \lambda)$$

$$\log \lambda^2 = b_0 + b_1 \cdot x + e \qquad e \sim t_2(0, \cdot)$$

Image-specific bias

Maximum likelihood fit for heteroscedastic t regression:

$$\hat{p}_i^{(0)} = a_0 + a_1 \cdot x + \varepsilon \qquad \varepsilon \sim t_2(0, \lambda)$$

$$\log \lambda^2 = b_0 + e \qquad e \sim t_2(0, \cdot)$$

Result Summary: Factors Influencing Detection Accuracy

Predictor	Within-image error: dispersion	Between-image error: dispersion	Between-image error: bias
Embedding ratio	increases for all detectors	reduces for all detectors but WS	reduces for all detectors
Local variance	increases for all detectors R^2: 12 (Triples) – 39% (RS)	increases for all detectors R^2: 1 (WS) – 3% (RS)	no direct effects
Saturation			
.. % at hist. ends	reduces for all detectors R^2: 4 (WS) – 19% (Triples)	increases all but Triples R^2: 1 (RS) – 2% (SPA)	under-estimation for all detectors R^2: 1 (Triples) – 6% (RS)
.. % at hist. mode	reduces for all detectors R^2: 15 (WS) - 36% (Triples)	increases for all detectors R^2: 3 (RS) – 8% (WS)	under-estimation for all detectors R^2: 4 (Triples) – 7% (RS)

Results for local variance and saturation are estimated on data with $p = 0.05$ bpp.

Concluding Remarks

> Rationale and evidence for (at least) two error components in quantitative LSB steganalysis

> Separation of components allows for more prudent analysis of the sources for estimation errors

> Don’t ignore within-image errors, and
don’t benchmark stego-estimators with moment statistics.

! *Thanks for your attention!*

Q&A

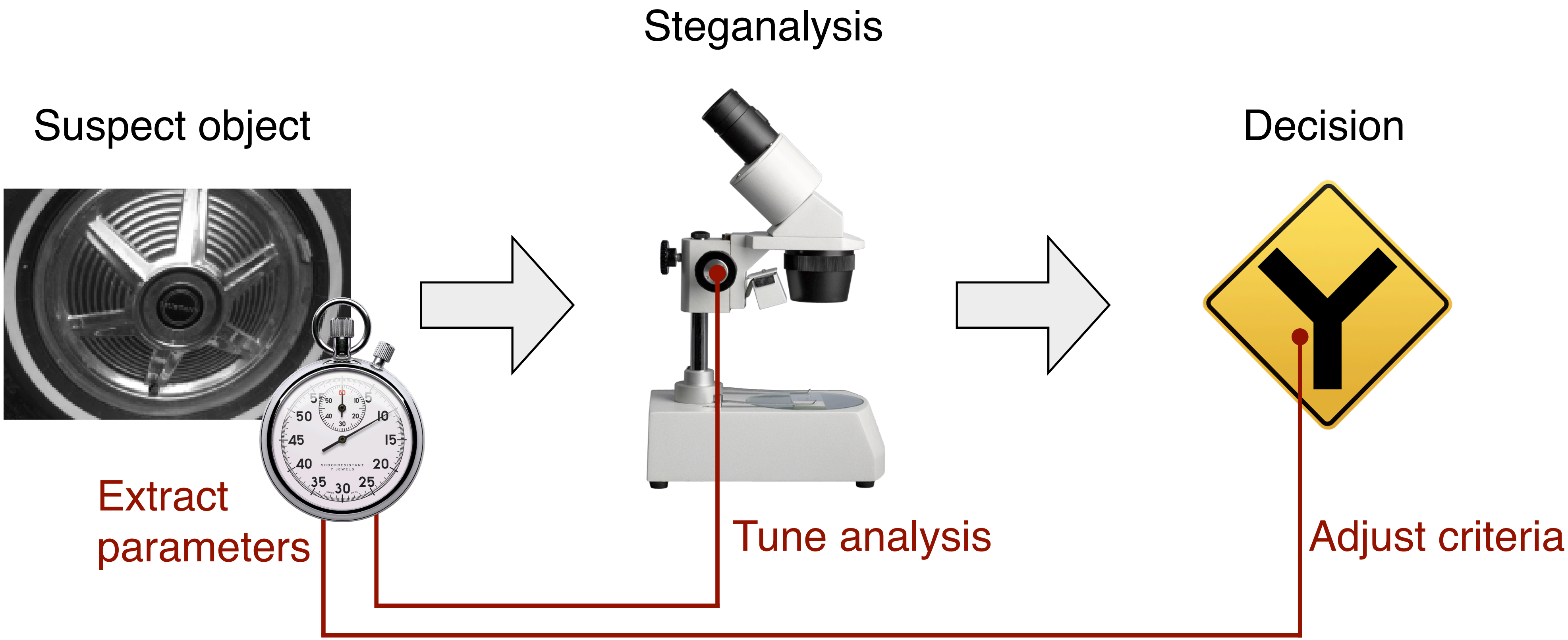
TECHNISCHE UNIVERSITÄT DRESDEN

Rainer Böhme
rainer.boehme@inf.tu-dresden.de
Technische Universität Dresden, Institute for System Architecture
01062 Dresden, Germany

Andrew D. Ker
adk@comlab.ox.ac.uk
Oxford University Computing Laboratory
Parks Road, Oxford OX1 3QD, United Kingdom

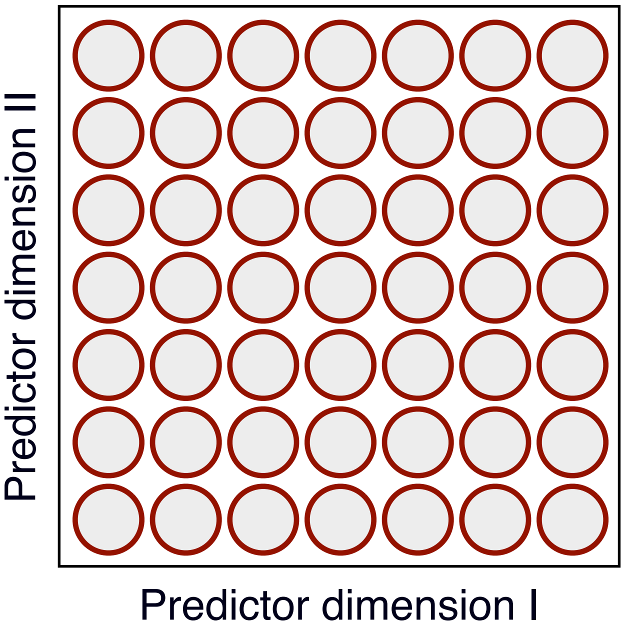
Towards Tailored Steganalysis

A TWO-FACTOR ERROR MODEL FOR QUANTITATIVE STEGANALYSIS



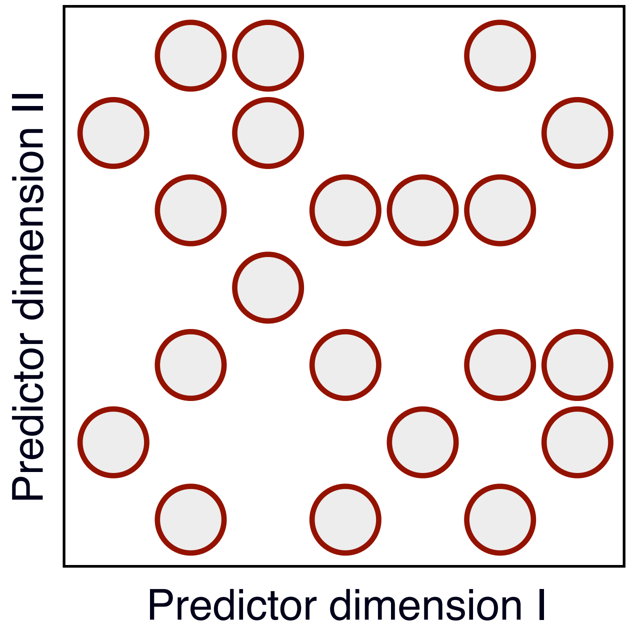
Appendix: Notes on Test Data Structure

Exhaustive tests



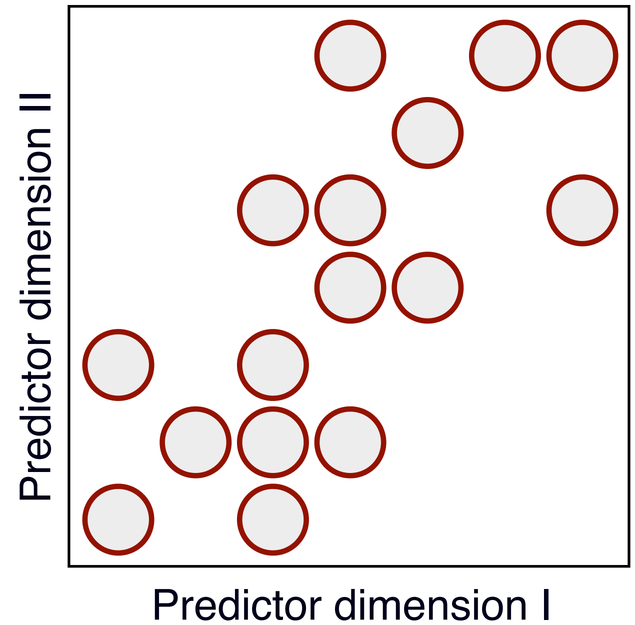
No model required since all breakdowns can be tabulated

Typical case



Full range of all predictor dimensions covered

Confounded predictors



Identification of individual influence from dependent predictors is error-prone

Structure of the Talk

1 **Quantitative Steganalysis**
RS Analysis · WS Analysis · Error Distribution

2 **Methodology**
Linear and Nonlinear Regression · Example Models

3 **Influence of Image Properties**
Image Size · Macro Characteristics

4 **Concluding Remarks**
Outlook · Limitations · Summary

1

Quantitative Steganalysis

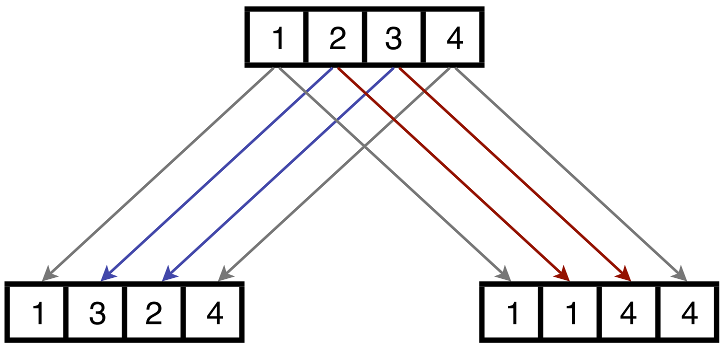
RS Analysis · WS Analysis · Error Distribution

Define two flipping functions

F_{+1}: (0↔1, 2↔3, …)

F_{-1}: (1↔2, 3↔4, …)

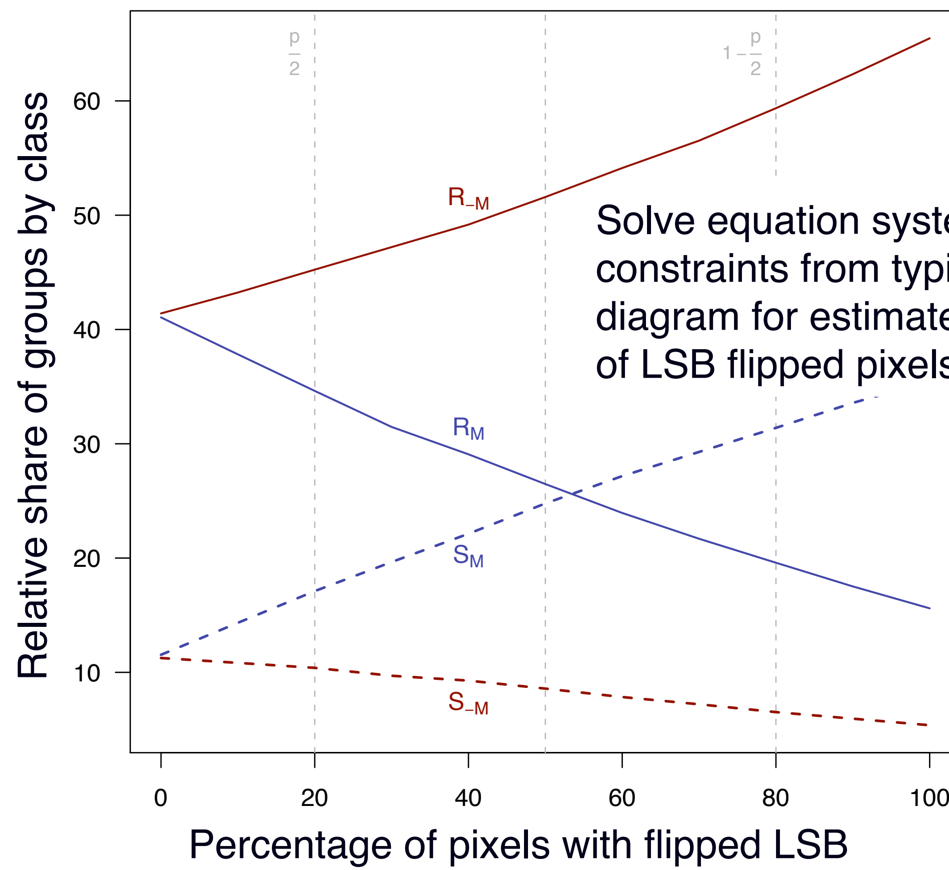
Cut image in groups of pixels, compute dual statistics



Compute sum of absolute differences

f(F(x))   f(x)   f(F(x))

Empirical RS diagram:
relation between embedding ratio and share of groups



Solve equation system with constraints from typical RS diagram for estimated number of LSB flipped pixels

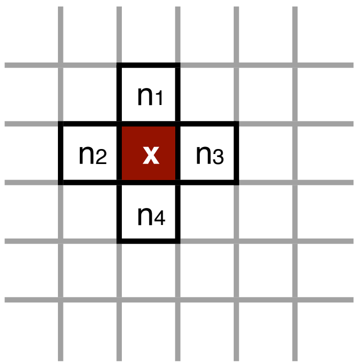
Classify groups per inequation

⋮ > ⋮ < ⋮ ⇒ Count as ***regular group*** R_M and R_{-M} , respectively.

⋮ < ⋮ > ⋮ ⇒ Count as ***singular group*** S_M and S_{-M} , respectively.

Ref.: Fridrich, Goljan & Du 2001

Recall: Weighted Stego Image Analysis (WS)



Estimate cover image

as arithmetic average $\mu(x)$ of the four closest neighbors of pixel x in the stego image.

Weight influence of pixels

since the accuracy of cover image estimation varies with local variance $\mathbf{v}(x)$.

Infer secret message length

from the difference of the observed stego image and an the estimated cover image

$$\hat{q} = -\frac{2}{\Sigma\, \mathbf{v}(x)} \sum_{x \in X} \frac{(F(x) - x) \cdot (x - \mu(x))}{1 + \mathbf{v}(x)^{\alpha}}$$

q : estimated embedding ratio

α : a constant used to control the influence of weighting

F : LSB flipping function