

How much information is in any five-year testing period?



Matt Gerstenberger & Jeremy Zechar



Motivation

- **RELM & CSEP prospective tests have been targeted at 5-year testing length for reasons of practicality**
- **Many of the tests are statistically powerful and can distinguish between two models in 5-years**

- However –

- **Taking the non-stationarity of seismicity into account: Are the results meaningful?**
- **Is 5-years representative of anything?**

Statistical Power (based on Zechar et al (2010))

The probability of correctly rejecting a false null, i.e., the ability of a test to distinguish two forecasts

- Power is relevant to a single pair of compared models**
- Simulate catalogs based each; test with these cats**
- How many simulated earthquakes necessary to distinguish between models?**
- Use S-Test and M-Test**
- $M > 4.95$**

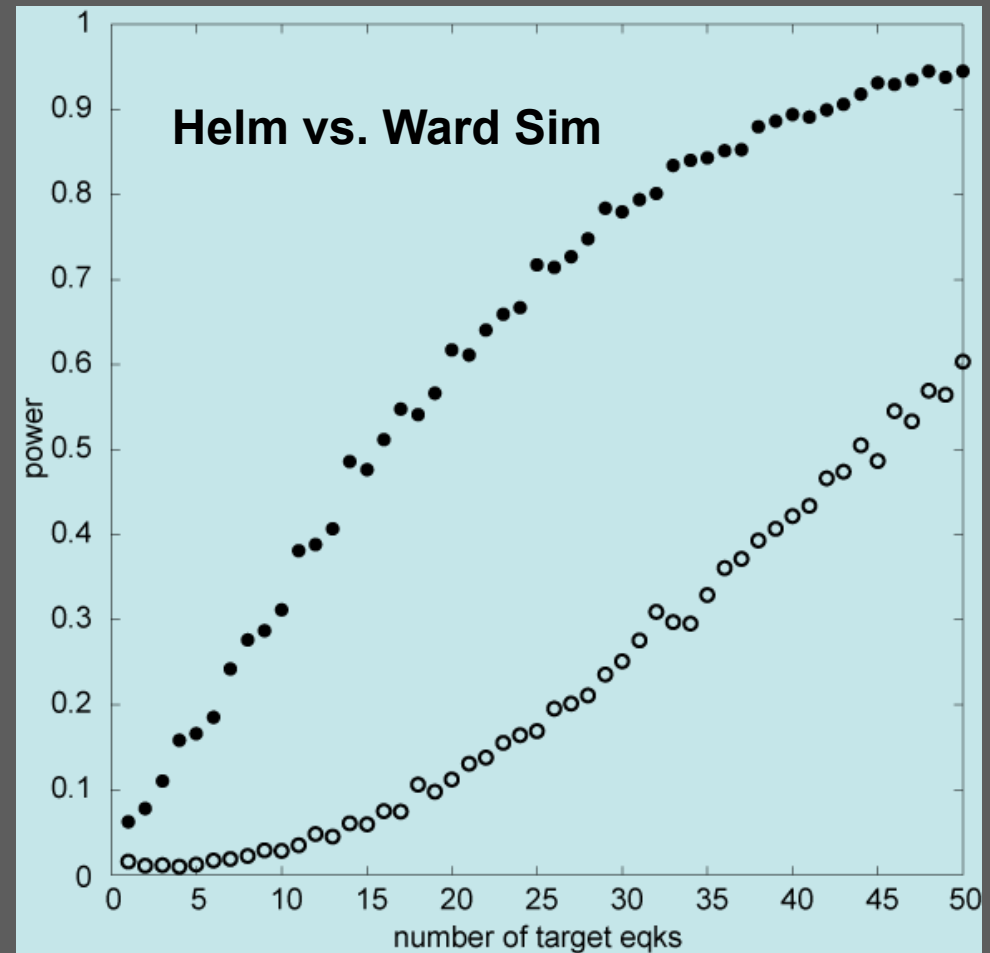
Statistical Power (based on Zechar et al (2010))

- **S-Test**

- NSHM96 -> Uni: 1 eqs
- NSHM96 -> Helm: 5 eqs
- Helm -> NSHM96: 13 eqs

- **M-Test**

- NSHM96 -> EmpiricalANSS: 56 eqs
- EmpiricalANSS -> NSHM96: 14 eqs



The Approach

- Compare the result of testing models using a 1932-2005 (74 year) catalog to smaller catalogs
- Compared to simulated catalogs using two types of bootstrapping
 - Standard bootstrapping with replacement
 - Block bootstrapping

Blocked bootstrap catalogs

- **Starting with any random event in the 1932-2005 catalog, select the next consecutive:**
 - 90 days
 - 1 year
 - 5 years
 - 10 years
 - 20 years
 - 50 years
- **To create one catalog**
- **Repeat 10,000 times**
- **Catalogs not independent**

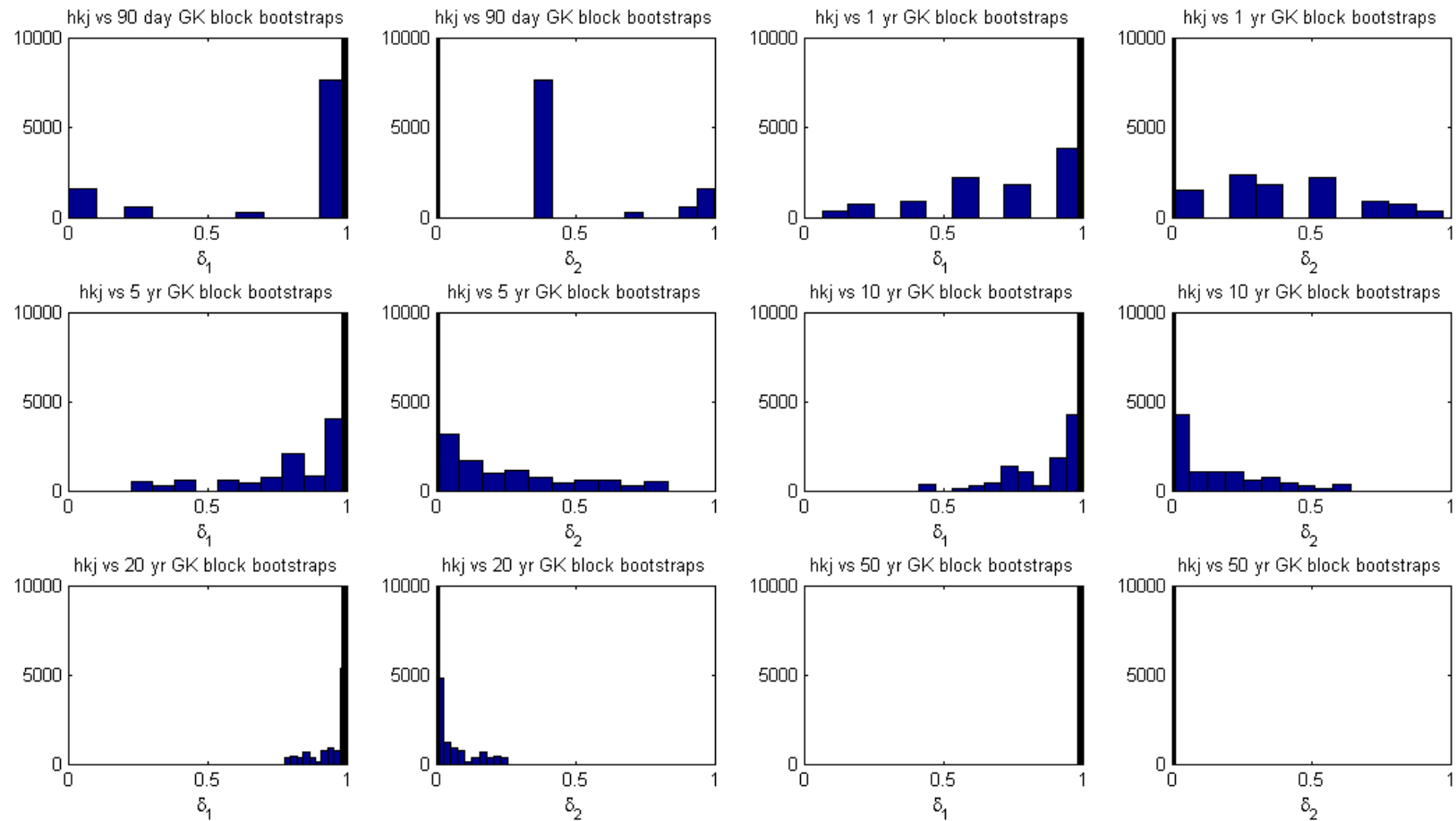
Bootstrap with replacement catalogs

- **Construct a 5yr catalog using randomly selected periods of length:**
 - 1 day
 - 10 day
 - 30 day
 - 90 day
 - 1 year
 - 2 years
- **Any period can be selected more than once**
- **Different length will represent different clustering scales (variability w/in 5yrs)**

The Models

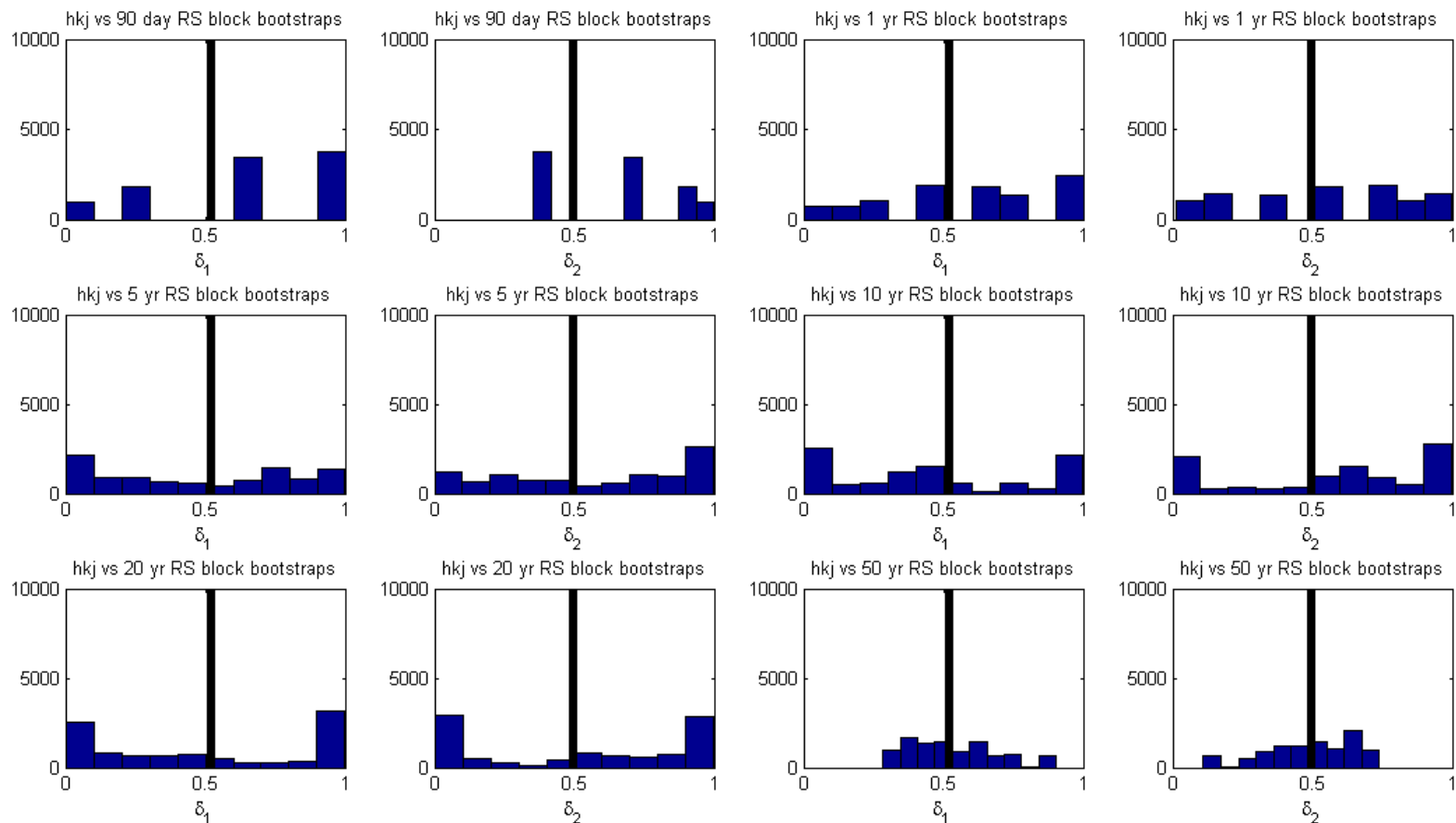
- **For now we have only looked at**
 - Helmstetter-Kagan-Jackson (HKJ)
 - Helmstetter-Kagan-Jackson+Aftershocks (HKJA)
- **We have declustered using**
 - Gardner-Knopoff (GK)
 - Reasenberg (RS) – HKJ uses RS

N-test, block bootstrap, HKJ with GK declus



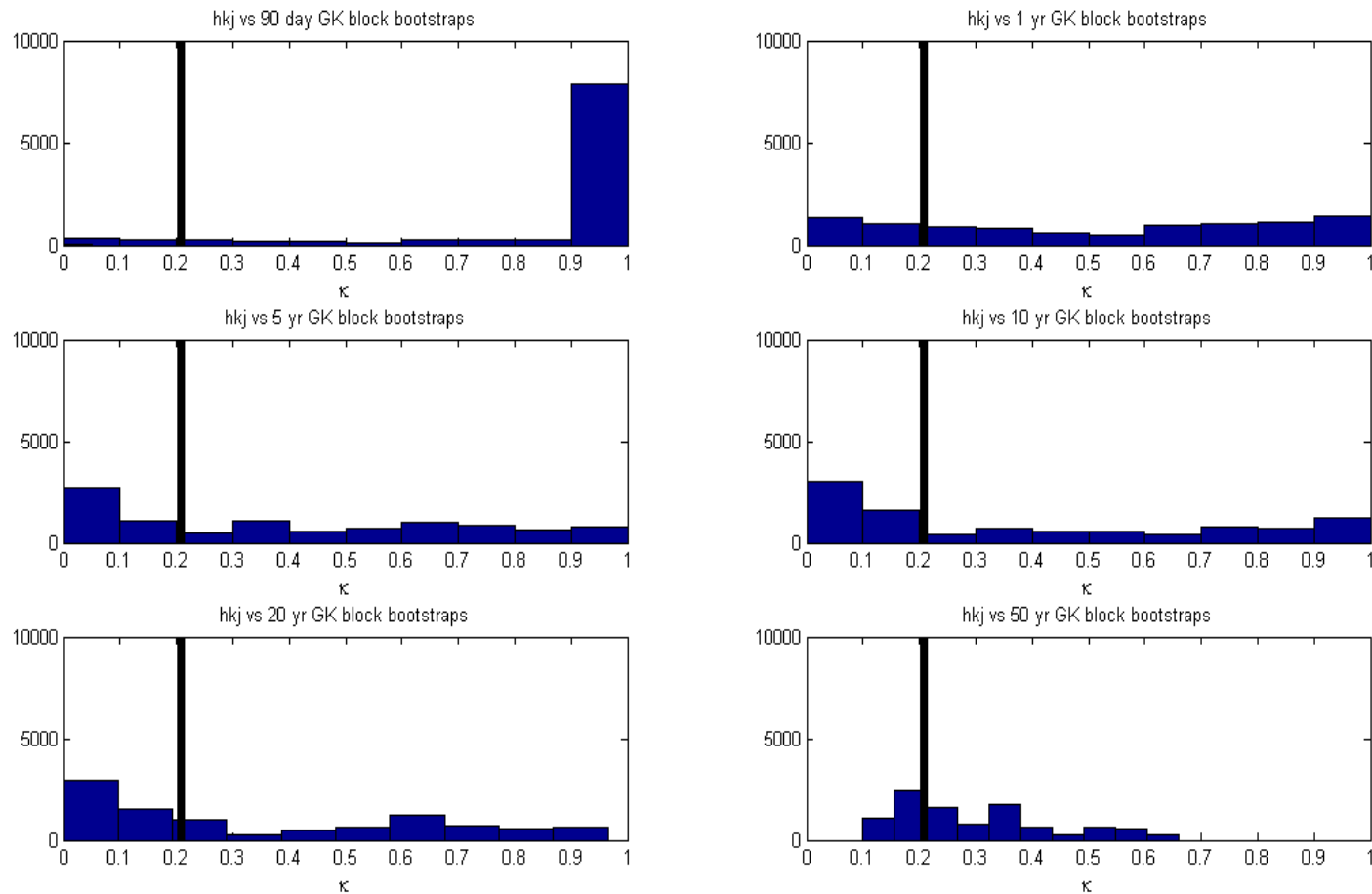
Low d_1 = under-prediction; low d_2 = over-prediction
Black line is result from 1932-2005

N-test, block bootstrap, HKJ with RS declus



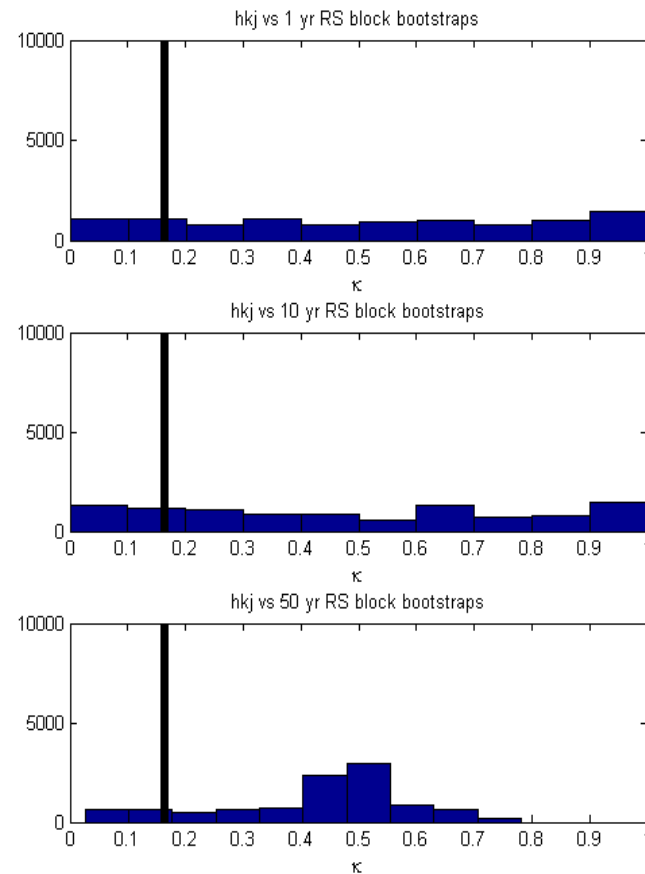
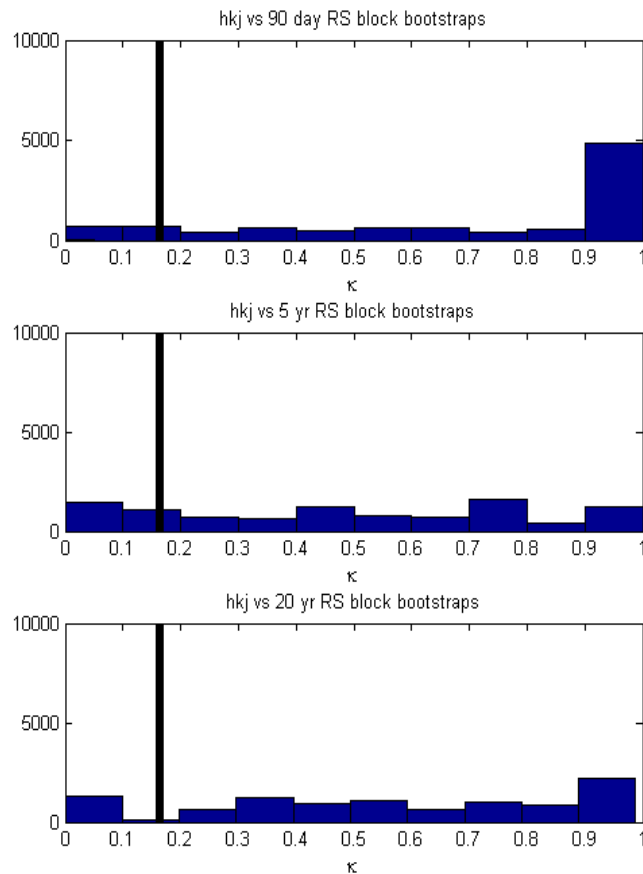
Low d_1 = under-prediction; low d_2 = over-prediction
Black line is result from 1932-2005

M-test, block bootstrap, HKJ with GK declus



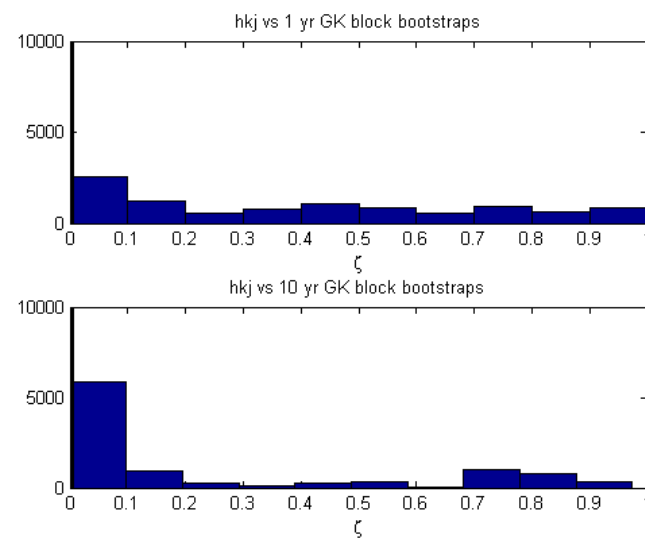
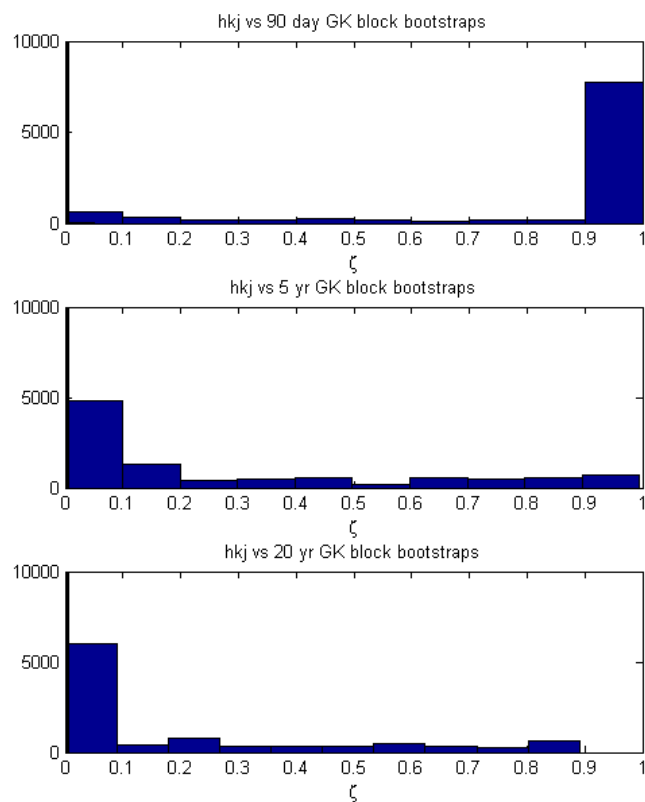
Black line is result from 1932-2005

M-test, block bootstrap, HKJ with RS declus



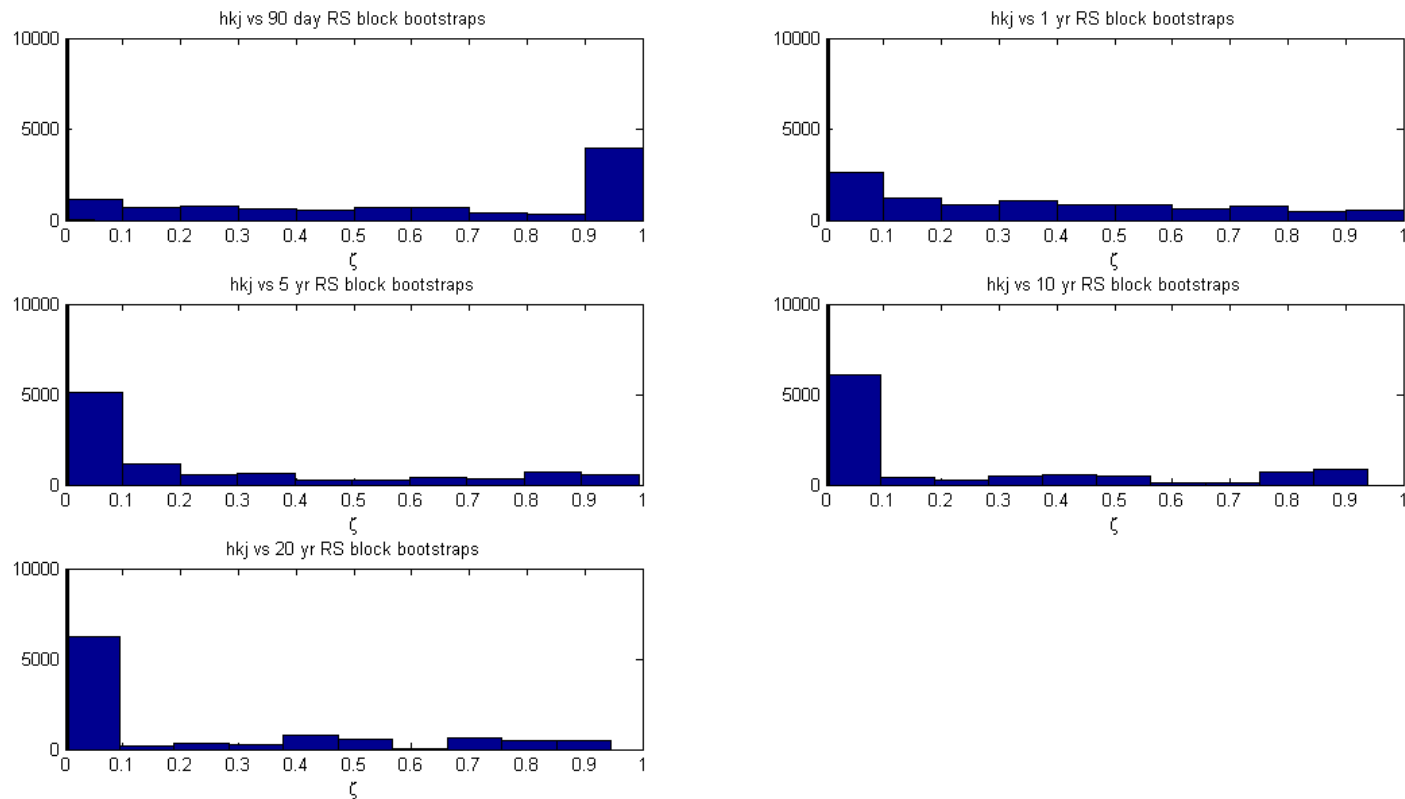
Black line is result from 1932-2005

S-test, block bootstrap, HKJ with GK declus



Black line is result from 1932-2005

S-test, block bootstrap, HKJ with RS declus



Black line is result from 1932-2005

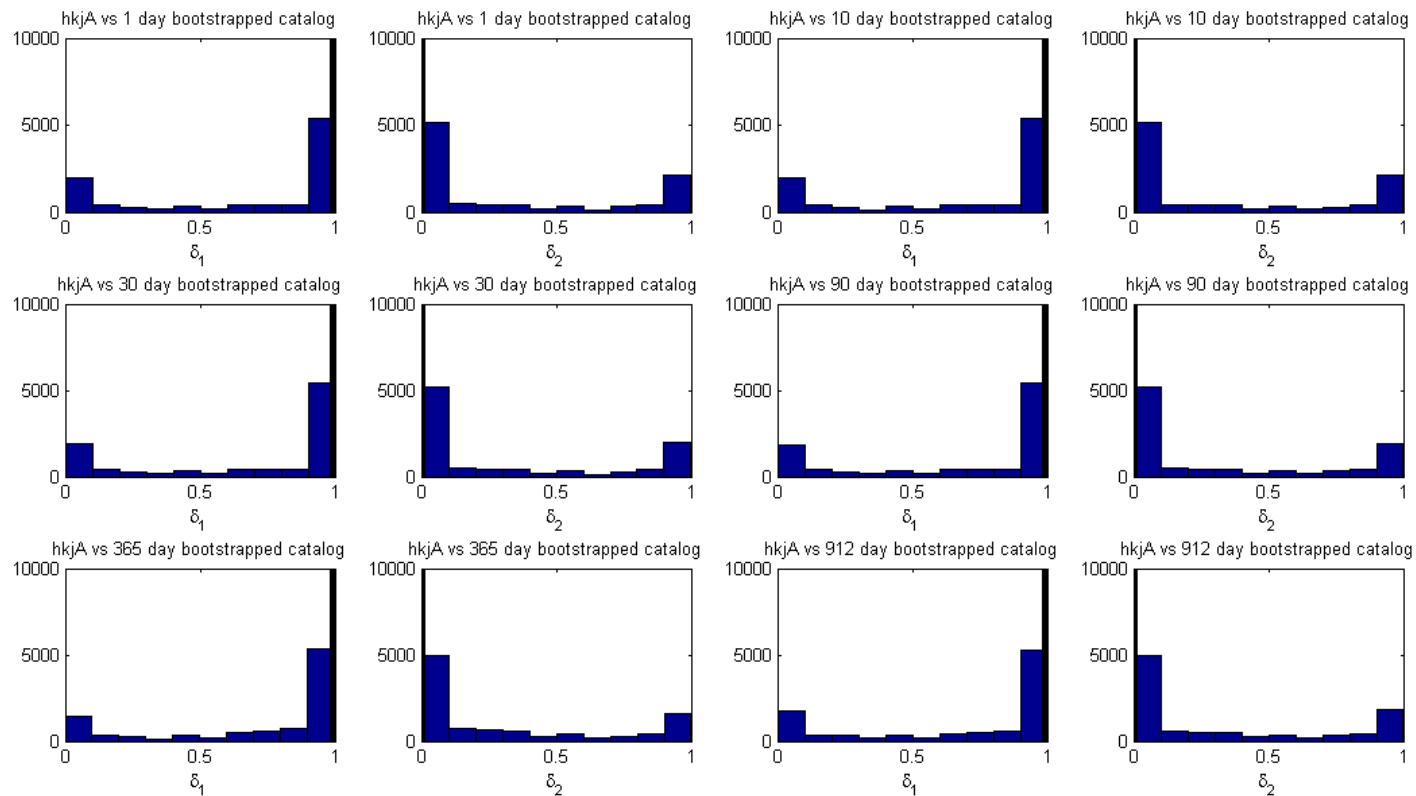
Summary

- **N-Test: 20 & 50 year tests, model is rejected. However, most likely result in any 5-yr period is non-rejection**
- **S-Test, M-Test: systematic increase in rejection rate as block size increase. Greater than 5% level**
- **HKJ fails long-term S-Test. Rejections grow with increasing time. Evidence of “slightly wrong” information?**

What next?

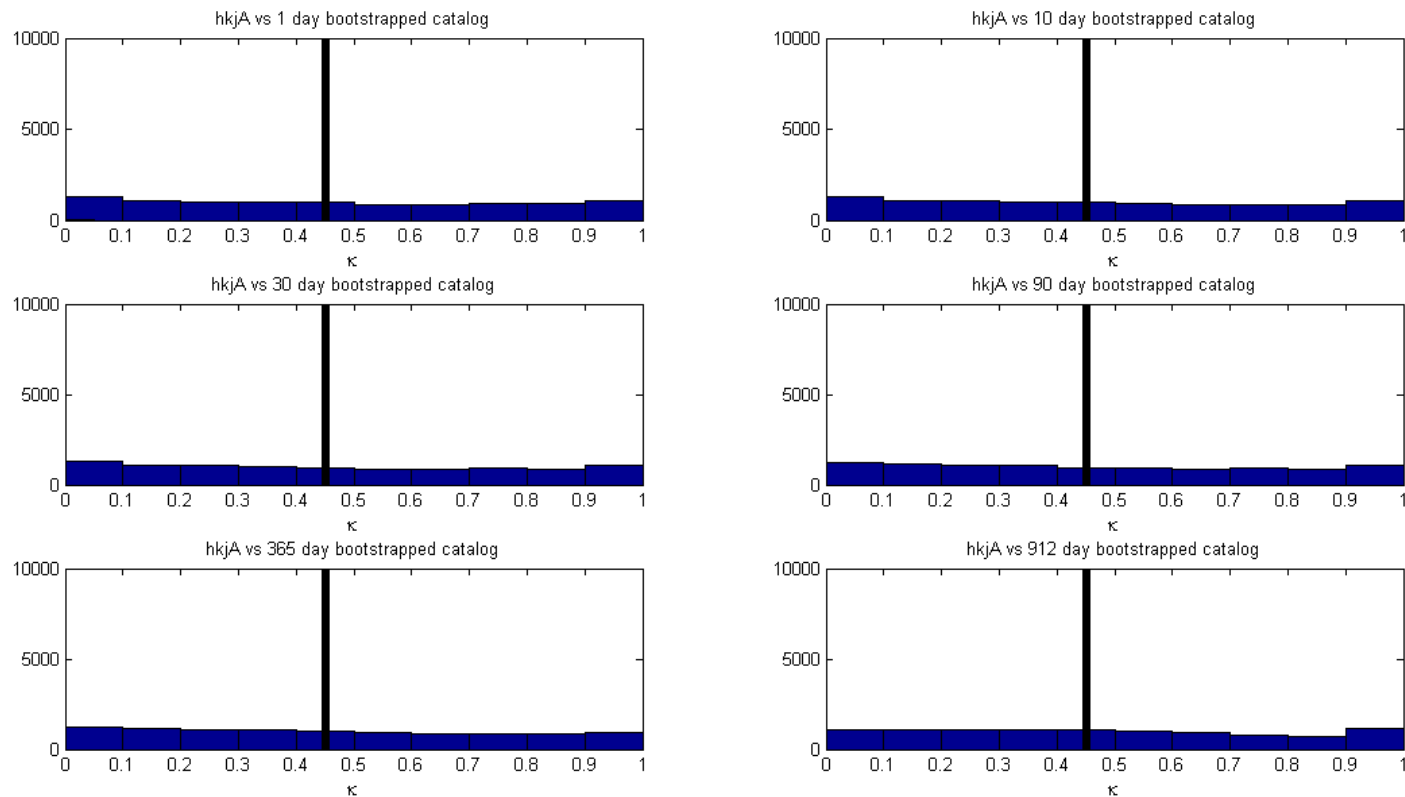
- **Testing of more models!**

N-test, bootstrap, HKJ+A No Declus



Low d1 = under-prediction; low d2 = over-prediction
Black line is result from 1932-2005

M-test, bootstrap, HKJ+A No Declus



Black line is result from 1932-2005