UCI Paul Merage School of Business

Leadership for a Digitally Driven World™

MFIN 290: Financial Econometrics

Lecture 3-2

Last Time

Heteroskedasticity

Consequences

Corrections

- GLS/FGLS
  - What it is/How to test if "okay"
- Correcting Standard Errors
  - White Robust
  - Newey West

Testing for Heteroskedasticity

Introduction to Time Series

Introduction to Time Series

- We have motivated time series a bit already with non-sphericality of the disturbances.
- The time series discussion in this course is intended to familiarize you with the concepts, some key results, relevant statistical tests (and how they work), and what NOT to do…
- We'll begin by discussing some common examples of a parameterized variance covariance matrix: $\Omega = \Omega(\theta)$

White Noise Process

- Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be a sequence with the following properties
- $E[\varepsilon_t] = 0$
- $E[\varepsilon_t]^2 = \sigma^2$
- $E[\varepsilon_t \varepsilon_s] = 0 \; \forall \; t \neq s$
- Mean zero, variance sigma squared, all *autocovariances* are zero.
- This is called a "white noise process". If we also assume that
- $\varepsilon_t \sim N(0, \sigma^2)$
- We call it a "Gaussian White Noise" process

Moving Average Process

- Let
- $y_t = \mu + \varepsilon_t + \lambda\varepsilon_{t-1}$
- Where λ, μ are constants and ε_t is a white noise process.

- We call this a "First Order Moving Average Process"

- First order, because there is only one lag / one λ...

MA(1) Moments

- Mean:
- $E[y_t] = E[\mu] + E[\varepsilon_t] + E[\lambda\varepsilon_{t-1}]$
- $E[y_t] = E[\mu] + 0 + \lambda 0 = \mu$
-

MA(1) Moments

- Variance of y_t:
- $E\left[y_t - E[y_t]\right]^2 = E[y_t - \mu]^2$
- $= E[\mu + \varepsilon_t + \lambda\varepsilon_{t-1} - \mu]^2 = E[(\varepsilon_t + \lambda\varepsilon_{t-1})^2]$
- $= E[(\varepsilon_t + \lambda\varepsilon_{t-1})(\varepsilon_t + \lambda\varepsilon_{t-1})] = E[\varepsilon_t^2 + 2\lambda\varepsilon_t\,\varepsilon_{t-1} + \lambda^2\varepsilon_{t-1}^2]$

$$= \sigma^2 + 2\lambda 0 + \lambda^2\sigma^2$$
$$= (1 + \lambda^2)\sigma^2$$
$$\ldots and\ t\ doesn't\ show\ up\ here, so\ this\ is\ for\ all\ t$$

- Note: $Var(y_t) > Var(\varepsilon_t) = \sigma^2$

MA(1) Autocovariance

- Define the first autocovariance as the covariance of y_t with its first lag:

- $E[(y_t - \mu)(y_{t-1} - \mu)]$
- $= E[(\varepsilon_t + \lambda\varepsilon_{t-1})(\varepsilon_{t-1} + \lambda\varepsilon_{t-2})]$
- $= E[\varepsilon_t\varepsilon_{t-1} + \lambda\varepsilon_t\varepsilon_{t-2} + \lambda\varepsilon_{t-1}\varepsilon_{t-1} + \lambda^2\varepsilon_{t-1}\varepsilon_{t-2}]$

MA(1) Autocovariance

- Define the first autocovariance as the covariance of y_t with its first lag:

- $E[(y_t - \mu)(y_{t-1} - \mu)]$
- $= E[(\varepsilon_t + \lambda\varepsilon_{t-1})(\varepsilon_{t-1} + \lambda\varepsilon_{t-2})]$
- $= E[\varepsilon_t\varepsilon_{t-1} + \lambda\varepsilon_t\varepsilon_{t-2} + \lambda\varepsilon_{t-1}\varepsilon_{t-1} + \lambda^2\varepsilon_{t-1}\varepsilon_{t-2}]$
- $= E[\varepsilon_t\varepsilon_{t-1}] + E[\lambda\varepsilon_t\varepsilon_{t-2}] + E[\lambda\varepsilon_{t-1}\varepsilon_{t-1}] + E[\lambda^2\varepsilon_{t-1}\varepsilon_{t-2}]$
- $= 0 + \lambda 0 + \lambda\sigma^2 + \lambda^2 0$
- $= \lambda\sigma^2$

MA(1) Autocovariance

- Higher autocovariances are all zero:

- $E[(y_t - \mu)(y_{t-2} - \mu)] = E[(\varepsilon_t + \lambda\varepsilon_{t-1})(\varepsilon_{t-2} + \lambda\varepsilon_{t-3})]$
- $= E[\varepsilon_t\varepsilon_{t-2} + \lambda\varepsilon_t\varepsilon_{t-3} + \lambda\varepsilon_{t-1}\varepsilon_{t-2} + \lambda^2\varepsilon_{t-1}\varepsilon_{t-2}]$
- $= 0 + \lambda 0 + \lambda 0 + \lambda^2 0$
- $= 0$

- This is related to the order:
- An MA series of order N has all autocovariances >N equal to zero.

MA(1) Autocovariance

- An MA series of order N has all autocovariances >N equal to zero.

- If we have an MA(1) residual parameterized by θ, then we know what the full covariance matrix looks like… What is it?

- Diagonals $= (1 + \lambda^2)\sigma^2$
- 1st off – diagonal $= \lambda\sigma^2$

- Zero everywhere else

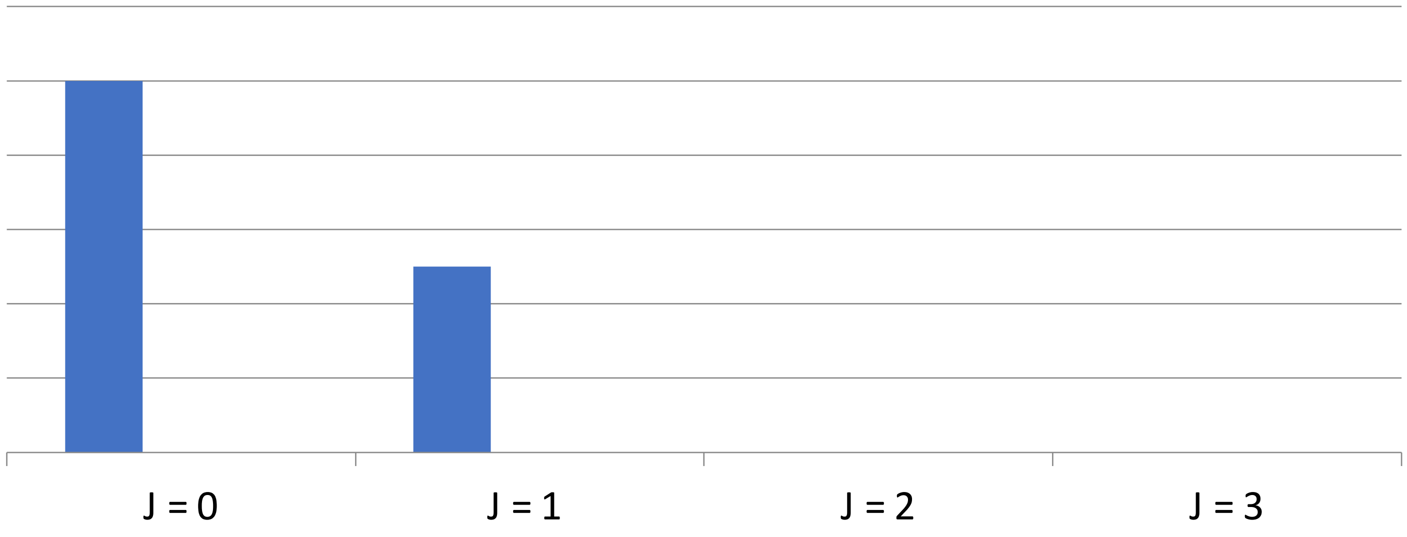
MA(1) Autocorrelation

- Define the first autocorrelation as the autocovariance of y_t scaled by the product of the standard deviations (just like regular correlation):

- First Autocorrelation $= \frac{First\ Autocovariance}{\sqrt{Var(y_t)}\sqrt{Var(y_{t-1})}}$
- $= \frac{\lambda\sigma^2}{\sqrt{(1+\lambda^2)\sigma^2}\sqrt{(1+\lambda^2)\sigma^2}}$
- $= \frac{\lambda\sigma^2}{(1+\lambda^2)\sigma^2}$
- $= \frac{\lambda}{(1+\lambda^2)}$

MA(1) Autocorrelation

- Can plot the autocorrelations as a function of the lag, this is called an “autocorrelgram”.



- Q: What is the height of each bar?

MATLAB: MA(1) Autocorrelation

Lecture 3 MA.m

Generate a random, first order moving average process, graph the time series and the autocorrelgram

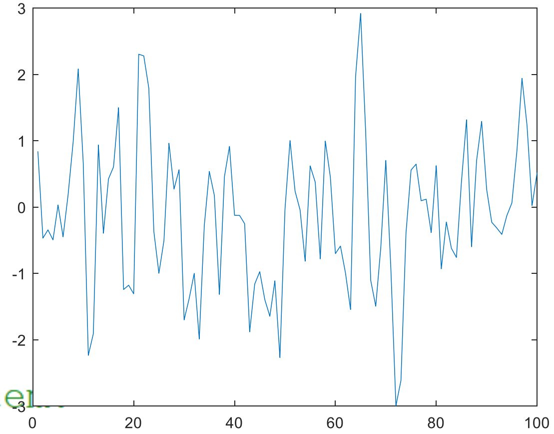
Will set $\lambda = 0.5$

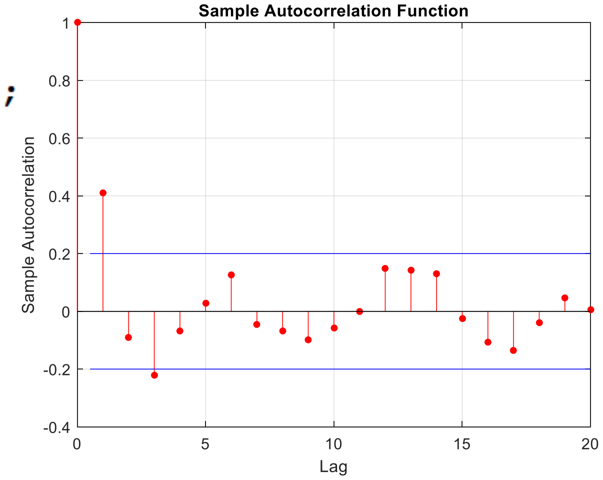
Q: What is first autocorrelation?

$$\frac{\lambda}{(1+\lambda^2)} = \frac{0.5}{1+0.25} = 0.4$$

MA(1) Autocorrelation

```
%generate a first order MA process, plot autocorrelgram

mu = 0;
lambda = 0.5;
T = [1:1:100]';
eps = mvnrnd(0,1,100);

y = zeros(100,1);
% assume eps(0) = 0 to initialize the system
y(1) = mu+eps(1);
for i = 2:100
    y(i) = mu + eps(i) + lambda*eps(i-1);
end

plot(T,y)
autocorr(y)
```





ACF and PACF

- k^{th} Autocorrelation $= r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{T}(y_t - \overline{y})^2}$
-
- The Partial Autocorrelation Function (PACF) is the ACF without the intervening lags:
-
- $r_k{}^* = \frac{\sum_{t=k+1}^{T}(y_t{}^* y_{t-k}{}^*)}{\sum_{t=k+1}^{T}(y_t{}^*)^2}$
- Where * indicates that we take the residual from a regression of y_t and y_{t-k} on $[1, y_{t-1}, y_{t-2}, \dots, y_{t-k+1}]$

ACF and PACF

- The Partial Autocorrelation Function (PACF) is the ACF without the intervening lags:
-
- $r_k{}^* = \frac{\sum_{t=k+1}^{T}(y_t{}^* y_{t-k}{}^*)}{\sum_{t=k+1}^{T}(y_t{}^*)^2}$
- Where * indicates that we take the residual from a regression of y_t and y_{t-k} on $[1, y_{t-1}, y_{t-2}, \dots, y_{t-k+1}]$

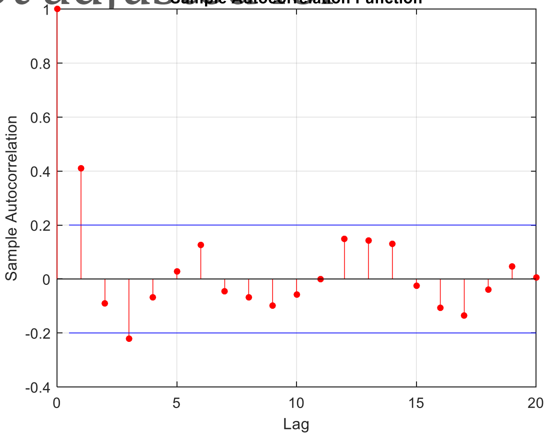
- This is the correlation between y_t and y_{t-k} conditional on the effect of the intervening values.

Inference on Autocorrelegrams

- $r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{T}(y_t - \overline{y})^2}$
-
- Numerator looks like the sum of squared residuals.. Looks like we are dividing by the variance….

Inference on Autocorrelegrams

- What if instead of time series, we had a standard, normal, white noise series, we could square the residuals and add them up, we should get a chi-squared distribution…this looks a lot like our ACF:
- $r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{T}(y_t - \overline{y})^2}$
- We can estimate the relationship between the variables, get the residuals, and then test the residuals to see if they look like this.... In particular, if $|\sqrt{T}\, r_k| > 1.96 \Rightarrow$ we can reject the null hypothesis of zero autocorrelation at lag k at the 5% level (not adjusted for multiple tests!).
- In our MATLAB example: $r_k > \frac{1.96}{\sqrt{T}} = \frac{1.96}{\sqrt{100}} = \frac{1.96}{10} = 0.196$



Autoregressive Process

- Let
- $y_t = c + \phi y_{t-1} + \varepsilon_t$
- Where c, ϕ are constants and ε_t is a white noise process.

- We call this a "First Order Autoregressive Process"
-

MA Representation of an AR(1) Process

- $y_t = c + \phi y_{t-1} + \varepsilon_t$

- $y_t = c + \phi(c + \phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$

- $y_t = c + \phi c + \phi^2(c + \phi y_{t-3} + \varepsilon_{t-2}) + \phi\varepsilon_{t-1} + \varepsilon_t$

- Combine terms by ϕ:

- $y_t = c + \varepsilon_t + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3 y_{t-3} \ldots$

MA Representation of an AR(1) Process

- $y_t = c + \varepsilon_t + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3 y_{t-3} \ldots$
- $y_t = \sum_{j=0}^{\infty} \phi^j c + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$

- This is an (infinite order) MA process!
- So autocovariances should never be zero…

- Can use this form to solve for the moments as well.

Mean of an AR(1) Process

- $E[y_t] = E[\sum_{j=0}^{\infty} \phi^j c] + E[\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}]$
- $E[y_t] = \sum_{j=0}^{\infty} \phi^j c + \sum_{j=0}^{\infty} \phi^j E[\varepsilon_{t-j}]$
- $E[y_t] = \sum_{j=0}^{\infty} \phi^j c + \phi E[\varepsilon_t] + \phi^2 E[\varepsilon_{t-2}] + \cdots$
- $E[y_t] = \sum_{j=0}^{\infty} \phi^j c + 0 + 0 + 0 \ldots$
- $E[y_t] = \sum_{j=0}^{\infty} \phi^j c$
- An example of a "Geometric Series"; with $|\phi| < 1$, this converges to:
  - $\mu = \frac{c}{1-|\phi|}$

Variance of an AR(1) Process

- $E[y_t - \mu]^2 = E[[\cancel{\sum_{j=0}^{\infty} \phi^j c}] + E[\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}] - [\cancel{\sum_{j=0}^{\infty} \phi^j c}]]^2$

- $E[\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}]^2 = E(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots)^2$

- Note only terms of the same lag have non zero expectations since ε_t is white noise
- so we have no ϕ on ε_t^2, ϕ^2 on ε_{t-1}^2 ...

Variance of an AR(1) Process

- Note only terms of the same lag have non zero expectations since ε_t is white noise
- $\sigma^2 + \phi^2\sigma^2 + \phi^4\sigma^2 + \cdots = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j}$
- Another geometric series… $|\phi^2| < 1 \leftrightarrow |\phi| < 1$, this converges to:

$$\frac{\sigma^2}{1-\phi^2}$$

Autocovariance of an AR(1) Process

- $E[(y_t - \mu)(y_{t-1} - \mu)] = E[\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots][\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots]$
- Only the same lag matter again, and those will give you sigma squared. We can pull a ϕ out too.
- $\phi\sigma^2[1 + \phi^2 + \phi^4 + \cdots] = \phi\sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\phi\sigma^2}{1-\phi^2}$

Autocovariance of an AR(1) Process

- $E[(y_t - \mu)(y_{t-1} - \mu)] = E[\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots][\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots]$
- Only the same lag matter again, and those will give you sigma squared. We can pull a ϕ out too.
- $\phi\sigma^2[1 + \phi^2 + \phi^4 + \cdots] = \phi\sigma^2\sum_{j=0}^{\infty}\phi^{2j} = \frac{\phi\sigma^2}{1-\phi^2}$
- Note:

$$\frac{\phi\sigma^2}{1-\phi^2} = \phi\frac{\sigma^2}{1-\phi^2} = \phi var(y_t) \Rightarrow$$

autocorrelation of an AR(1) process is ϕ

AR(1) in MATLAB

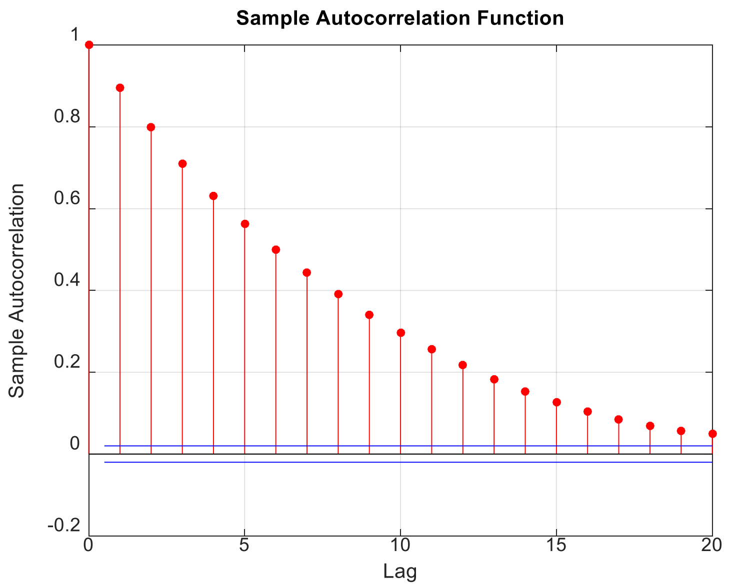
Lecture 3 AR1.m

Generate an AR(1) process,
Compare the empirical and analytical mean, variance
Plot autocorrelegram/autocorrelation function

AR(1) in MATLAB

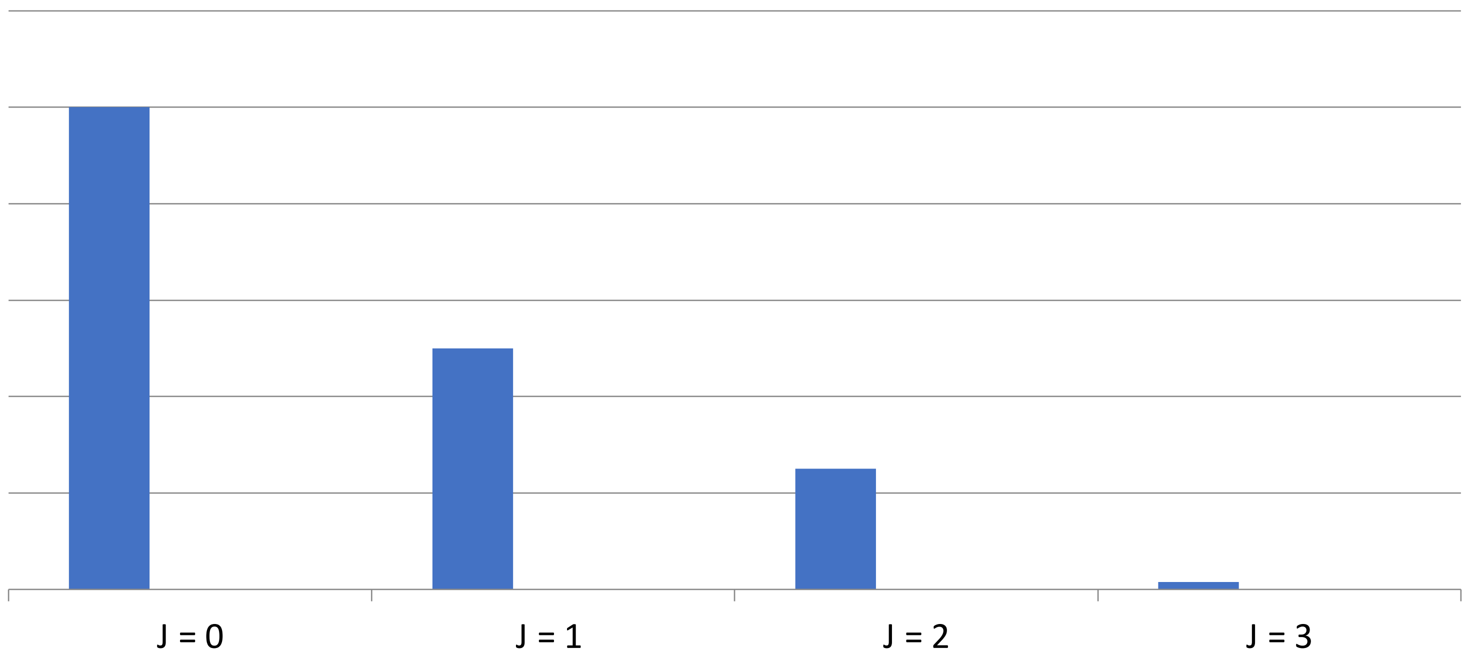
```
%generate a first order AR process, plot autocorrelgram, mean and variance

mu = 0;
phi = 0.9;
c = 0.1;
T = [1:1:10000]';
%epsilon is mean zero, variance one
eps = mvnrnd(0,1,10000);

y = zeros(10000,1);

for i = 2:10000
    y(i) = c + phi*y(i-1) +eps(i);
end

mean(y)
theor_mean = c/(1-phi)
var(y)
theor_var = 1/(1-phi^2)
autocorr(y)
```

```
ans =

    1.0317


theor_mean =

    1.0000


ans =

    5.0871


theor_var =

    5.2632
```



Autocorrelgram of an AR(1) Process

The Paul Merage School of Business



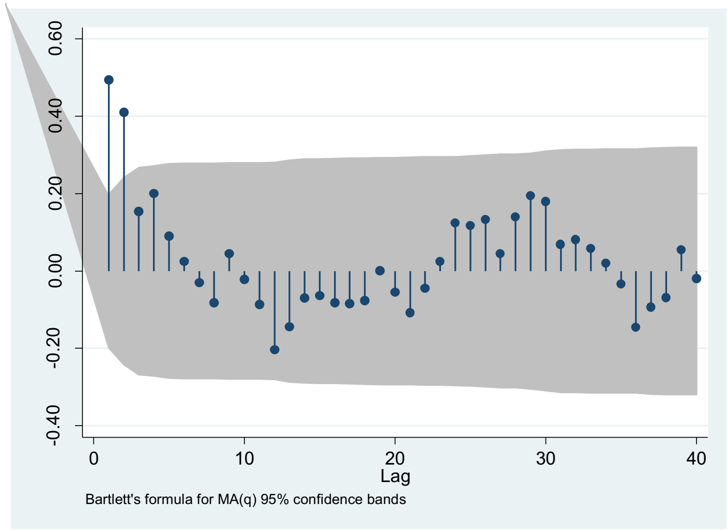
ACF/PACF for AR(1) and MA(1) Process

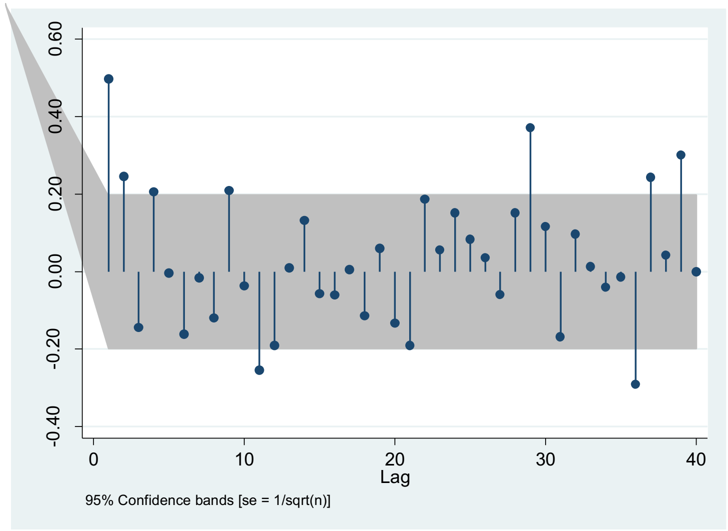
- For an AR specification, the ACF decays slowly, and the PACF drops to zero after the order of the model.
- For an MA specification, the ACF drops to zero after the order of the model and the PACF decays slowly.
- It is sometimes easy to identify MA(q) processes from the ACF and AR(p) processes from the PACF. You can sometimes detect the order this way too!

Estimating AR or MA models

- Using okun dataset – GDP growth from 1985Q2-2009Q3

```
. gen obsno = _n

. tsset obsno
        time variable:  obsno, 1 to 98
                delta:  1 unit

. ac g

. pac g
```



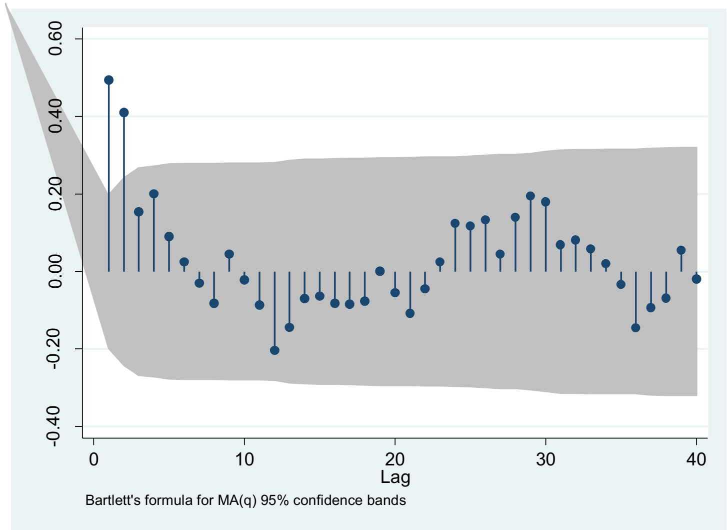


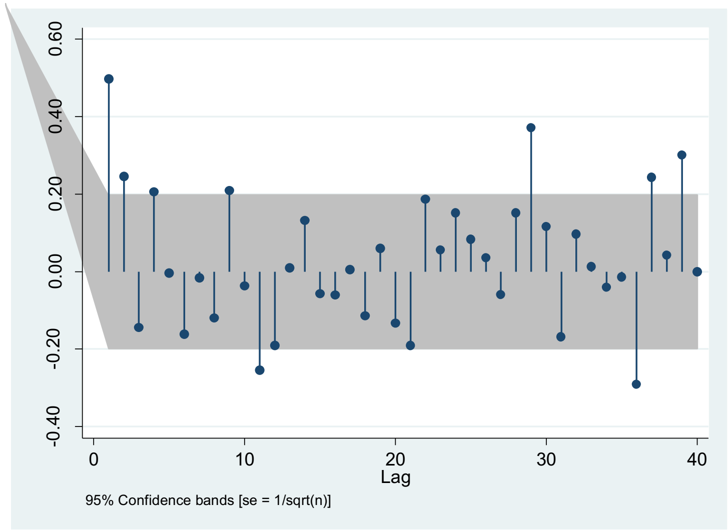
Give Stata a time series variable
Set dataset as a time series
Plot autocorrelations and partial autocorrelations

Estimating AR or MA models

- Using okun dataset – GDP growth from 1985Q2-2009Q3

```
. gen obsno = _n

. tsset obsno
        time variable:  obsno, 1 to 98
                delta:  1 unit

. ac g

. pac g
```





AC: AR(1) or MA(2)?
PAC: AR(1)?

Testing for Autocorrelation: LM Test

- While the AC and PAC plots are helpful to detect patterns, they are not formal tests (though they often include confidence bands).

- To check for residual autocorrelation, we need a consistent estimate of the residuals.

- That is, we start with the model
- $y = x\beta + \epsilon$

Testing for Autocorrelation: LM Test

- $y = x\beta + \epsilon$
- Can use OLS to estimate
- $\hat{\beta}$ and e for the coefficient and fitted residuals respectively.

- Regress
- $e_t = \rho e_{t-1} + u_t$

- And we can include as many lags as we want in this expression and jointly test for their significance (using an F-test or χ^2 test as in regular OLS).

Testing for Autocorrelation: LM Test

- Because of the FW Theorem, this is equivalent to testing for significance of the lagged fitted residuals in a regression of
- $y = x\beta + \rho \widehat{e_{t-1}} + v_t$

- This is referred to as the Lagrange multiplier test or the Breusch Godfrey test.

- Because we rely on consistent estimates of the residuals, it is an asymptotic test.

Testing for Autocorrelation: LM Test

Apply to same GDP dataset:

1) Regress GDP on a constant term (there are no independent variables)
2) Predict residuals
3) Regress GDP on lagged residuals

```
. reg g l.ehat

      Source |       SS           df       MS      Number of obs   =        97
-------------+----------------------------------   F(1, 95)        =     30.96
       Model |  9.97415839         1  9.97415839   Prob > F        =    0.0000
    Residual |  30.6064602        95  .322173265   R-squared       =    0.2458
-------------+----------------------------------   Adj R-squared   =    0.2378
       Total |  40.5806186        96  .422714777   Root MSE        =     .5676

------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ehat |
         L1. |   .4970809   .0893375     5.56   0.000     .3197236    .6744382
             |
       _cons |   1.272816    .057633    22.08   0.000       1.1584    1.387232
------------------------------------------------------------------------------
```

There are usually built in tests, but you should always read the help files!

This example helps show why…

Testing for Autocorrelation: LM Test

Description for estat bgodfrey

estat bgodfrey performs the Breusch-Godfrey test for higher-order serial correlation in the disturbance. This test does not require that all the regressors be strictly exogenous.

Options for estat bgodfrey

is it 1? or >1?

lags(*numlist*) specifies a list of numbers, indicating the lag orders to be tested. The test will be performed separately for each order. The default is order one.

nomiss0 specifies that Davidson and MacKinnon's approach (1993, 358), which replaces the missing values in the initial observations on the lagged residuals in the auxiliary regression with zeros, not be used.

small specifies that the p-values of the test statistics be obtained using the F or t distribution instead of the default chi-squared or normal distribution.

Testing for Autocorrelation: LM Test


```
. qui reg g

. predict ehat, resid

. qui reg g l.ehat

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	5.727	1	0.0167

```
H0: no serial correlation
```

These regressions are only possible for AR terms- identifying MA terms is a different exercise for later in the course..

Testing for Autocorrelation: LM Test

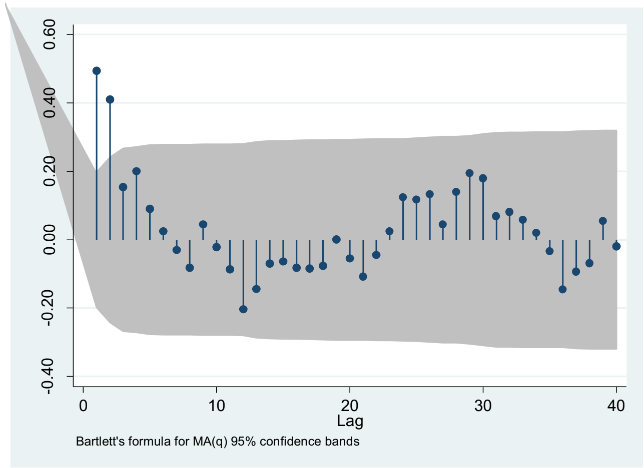

```
. reg g l(1/2).ehat

      Source |       SS           df       MS      Number of obs   =        96
-------------+----------------------------------   F(2, 93)        =     19.06
       Model |  11.6417916         2  5.82089582   Prob > F        =    0.0000
    Residual |  28.4081042        93  .305463486   R-squared       =    0.2907
-------------+----------------------------------   Adj R-squared   =    0.2754
       Total |  40.0498958        95  .421577851   Root MSE        =    .55269

------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ehat |
         L1. |   .3770015    .100021     3.77   0.000     .1783797    .5756233
         L2. |   .2462394   .1028688     2.39   0.019     .0419623    .4505165
             |
       _cons |   1.261312   .0564417    22.35   0.000      1.14923    1.373394
------------------------------------------------------------------------------

. estat bgodfrey, small

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          F                   df                 Prob > F
-------------+-------------------------------------------------------------
       1     |        5.727               (  1,    94 )         0.0187
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Should become comfortable reading through files to review implementation of tests and reconcile test cases to ensure code is working properly.

Always know what the default settings are and think through if they are appropriate to your problem.

Not a good mistake to make!

Intuition/Example



If an AR(1) was sufficient, the second term would be very near the first term squared.

$$0.4943^2 = 0.244 \text{ "} \neq \text{"} 0.411$$

```
. corrgram g, lag(5)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.4943   0.4971   24.681  0.0000
2        0.4107   0.2462     41.9  0.0000
3        0.1544  -0.1451    44.36  0.0000
4        0.2004   0.2052   48.549  0.0000
5        0.0904  -0.0035    49.41  0.0000
```

Choosing the "best" number of lags

- So how do we decide what the right number of lags/AR or MA terms to include is? We can get suggestive evidence by looking at the AC and PAC plots, though we can make this more precise.

- 1) Can check for residual serial correlation with the LM test
- 2) Can ensure that the signs and magnitudes of the variables we care about are as expected
- 3) Can review functions of model performance to determine appropriateness of the number of lags.

Choosing the "best" number of lags

- One such measure of performance is the Akaike Information Criteria
- $AIC = \ln\left(\frac{SSE}{T}\right) + \frac{2K}{T}$

- Where SSE is the sum of squared errors in the model, and K reflects the df in the two regressions.

- That is, $K = p + q + 2 =$ number of X variables + number of residual lags + 2 (intercept terms).

Choosing the “best” number of lags

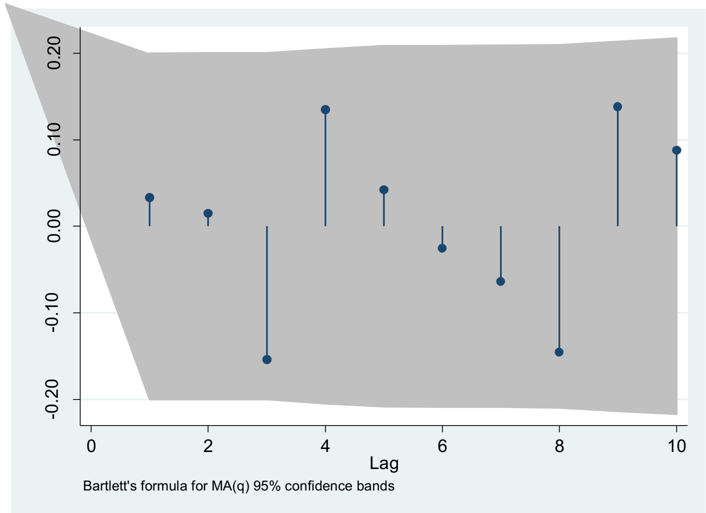
- One such measure of performance is the Bayesian Information Criteria (also called the Schwarz criterion)
- $BIC = \ln\left(\frac{SSE}{T}\right) + \frac{K ln(T)}{T}$

- Note: $\ln(T) > 2$ for $T > 8$, so the BIC penalizes additional terms more than the AIC.
- These are performance measures penalizing model complexity. Conceptually similar to Adjusted R^2
- These are often presented in terms of the “likelihood” (including in Stata) – we will return to this later in the course.

Example

```
. reg g l.ehat l2.ehat

      Source |       SS           df       MS      Number of obs   =        96
-------------+----------------------------------   F(2, 93)        =     19.06
       Model |  11.6417916         2  5.82089582   Prob > F        =    0.0000
    Residual |  28.4081042        93  .305463486   R-squared       =    0.2907
-------------+----------------------------------   Adj R-squared   =    0.2754
       Total |  40.0498958        95  .421577851   Root MSE        =    .55269

------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ehat |
         L1. |   .3770015    .100021     3.77   0.000     .1783797    .5756233
         L2. |   .2462394   .1028688     2.39   0.019     .0419623    .4505165
             |
       _cons |   1.261312   .0564417    22.35   0.000      1.14923    1.373394
------------------------------------------------------------------------------

. predict ehat_l2, resid
(2 missing values generated)
```

Example



Most measures suggest that two lags is sufficient, and with two lags, there does not appear to be any meaningful residual serial correlation…

```
. varsoc g

  Selection-order criteria
  Sample:  5 - 98                              Number of obs      =        94

 +---------------------------------------------------------------------------+
 |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
 |----+----------------------------------------------------------------------|
 |  0 | -93.1957                        .434433   2.00416   2.01509   2.03122  |
 |  1 | -79.8232  26.745     1  0.000  .333888   1.74092   1.76278   1.79503  |
 |  2 | -77.0929  5.4608*    1  0.019  .321823    1.7041   1.73689*  1.78527* |
 |  3 |  -76.265  1.6557     1  0.198  .323014   1.70777   1.75148   1.81599  |
 |  4 | -74.8172  2.8956     1  0.089  .319967*  1.69824*  1.75288   1.83352  |
 +---------------------------------------------------------------------------+
  Endogenous:  g
   Exogenous:  _cons
```

Note that Stata uses the Likelihood form of AIC and BIC (read the help files!):

$$AIC = -2\ln(L) + 2K$$
$$BIC = -2\ln(L) + \mathrm{K}\log(n)$$

Repeat Inference

A key component of statistical model building is reducing the (very large) space of possible drivers to a parsimonious set of final model candidates. The conceptual guidance from econometrics and common practices in the industry (including academia) are often quite different

- *"Econometricians sometimes teach what people never use in practice – they simply teach correct things to use, and it is up to you to decide if you want to do wrong or correct things in practice." – Victor Chernozhukov*

Repeat Inference

There are many common techniques for sparse model selection. Forward selection, backward selection, best subsets, etc. All have limitations that make themselves obvious when used.

Usually around repeat inference…

Why does this make sense?

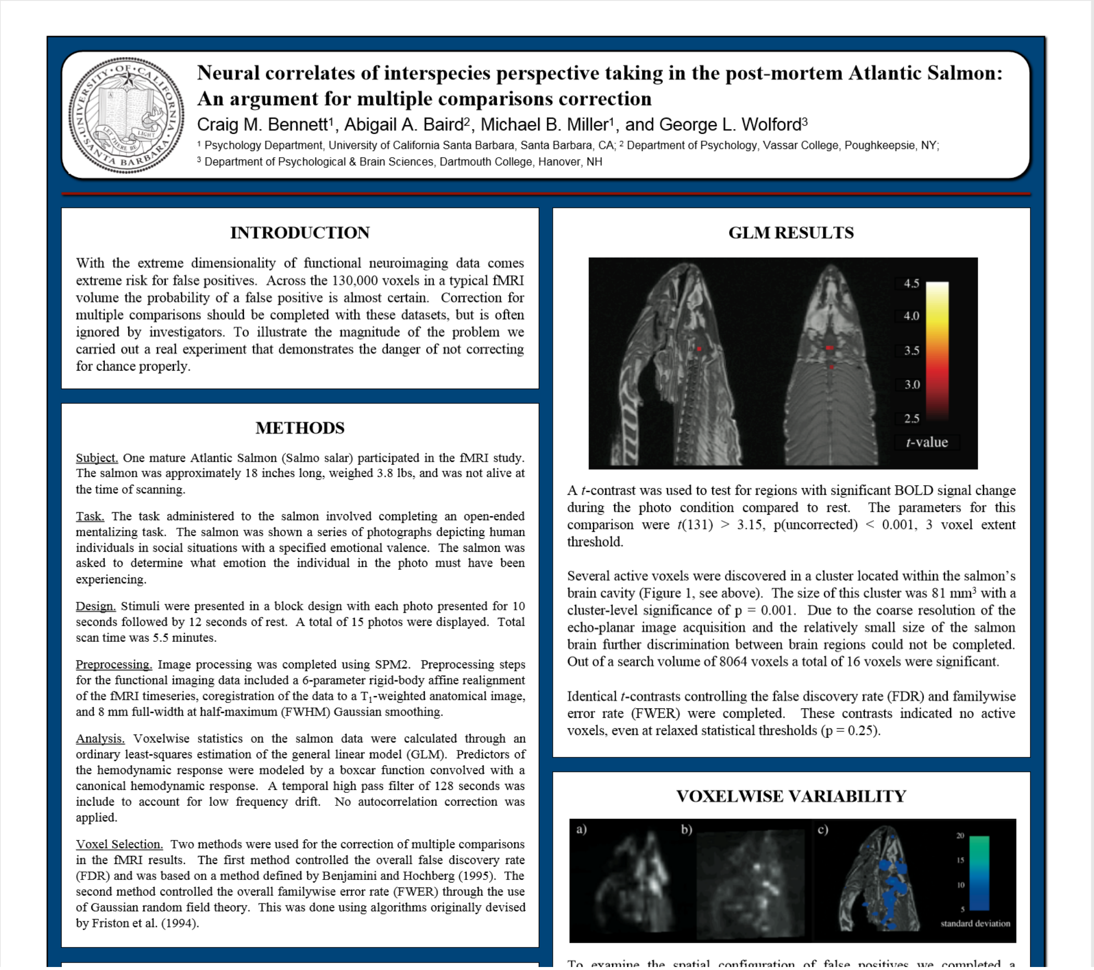
Repeat Inference

Repeat Inference:

Running multiple different (often correlated) hypothesis tests means that we expect to get false positives and false negatives with some positive probability.

This may be the most important thing you can take away from this class. If you are making sequential decisions based on p-values or a single decision based on multiple p values the standalone confidence level is not correct!

Repeat Inference



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY; [3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we carried out a real experiment that demonstrates the danger of not correcting for chance properly.

METHODS

Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

Design. Stimuli were presented in a block design with each photo presented for 10 seconds followed by 12 seconds of rest. A total of 15 photos were displayed. Total scan time was 5.5 minutes.

Preprocessing. Image processing was completed using SPM2. Preprocessing steps for the functional imaging data included a 6-parameter rigid-body affine realignment of the fMRI timeseries, coregistration of the data to a T_1-weighted anatomical image, and 8 mm full-width at half-maximum (FWHM) Gaussian smoothing.

Analysis. Voxelwise statistics on the salmon data were calculated through an ordinary least-squares estimation of the general linear model (GLM). Predictors of the hemodynamic response were modeled by a boxcar function convolved with a canonical hemodynamic response. A temporal high pass filter of 128 seconds was include to account for low frequency drift. No autocorrelation correction was applied.

Voxel Selection. Two methods were used for the correction of multiple comparisons in the fMRI results. The first method controlled the overall false discovery rate (FDR) and was based on a method defined by Benjamini and Hochberg (1995). The second method controlled the overall familywise error rate (FWER) through the use of Gaussian random field theory. This was done using algorithms originally devised by Friston et al. (1994).

GLM RESULTS

A t-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

Several active voxels were discovered in a cluster located within the salmon's brain cavity (Figure 1, see above). The size of this cluster was 81 mm³ with a cluster-level significance of $p = 0.001$. Due to the coarse resolution of the echo-planar image acquisition and the relatively small size of the salmon brain further discrimination between brain regions could not be completed. Out of a search volume of 8064 voxels a total of 16 voxels were significant.

Identical t-contrasts controlling the false discovery rate (FDR) and familywise error rate (FWER) were completed. These contrasts indicated no active voxels, even at relaxed statistical thresholds ($p = 0.25$).

VOXELWISE VARIABILITY

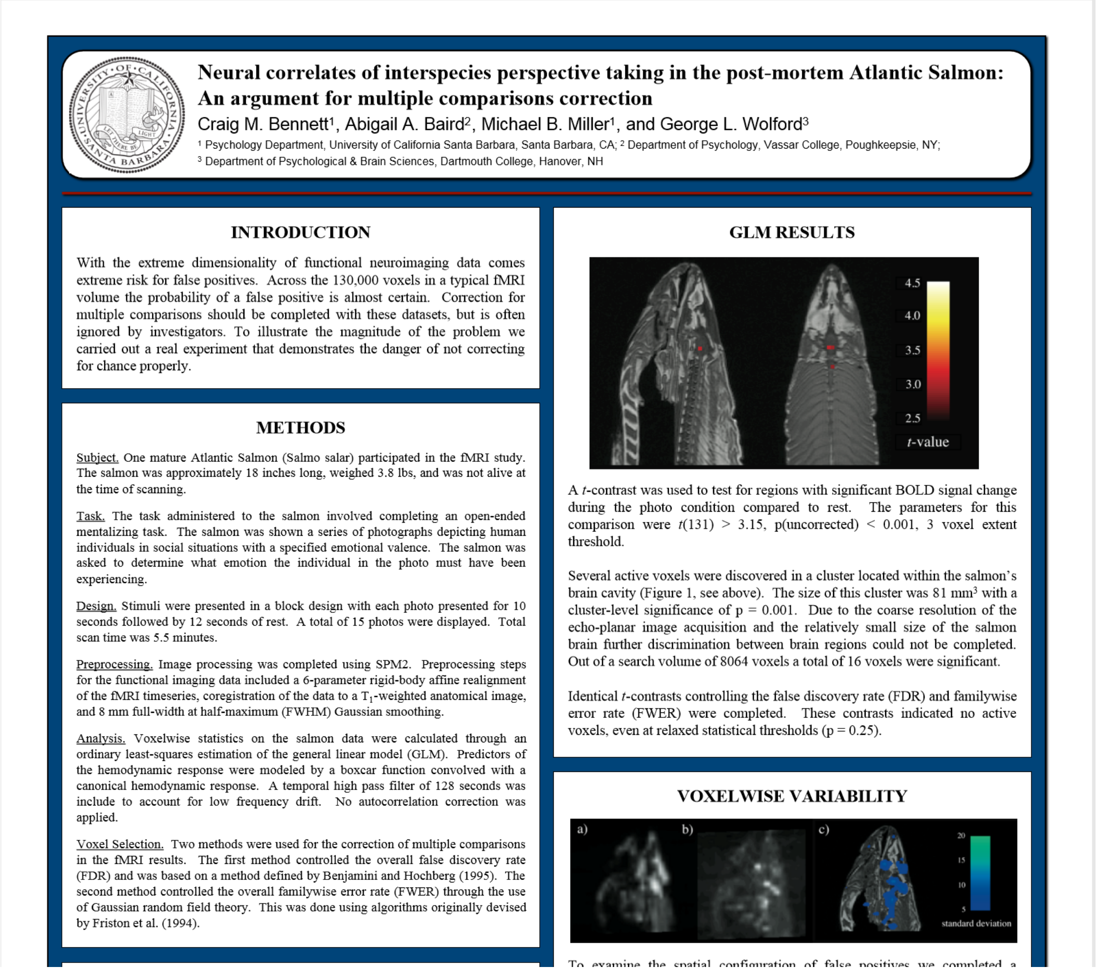
To examine the spatial configuration of false positives we completed a

For each 95% confidence test, there is a 1/20 chance we will get a false negative.

If we run 100 hypothesis tests, it's very likely that we will reject the null at least a few times, even if it is true!

I have kept this paper in my office throughout my career, no matter the firm!

Repeat Inference



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY;
[3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we carried out a real experiment that demonstrates the danger of not correcting for chance properly.

METHODS

Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

Design. Stimuli were presented in a block design with each photo presented for 10 seconds followed by 12 seconds of rest. A total of 15 photos were displayed. Total scan time was 5.5 minutes.

Preprocessing. Image processing was completed using SPM2. Preprocessing steps for the functional imaging data included a 6-parameter rigid-body affine realignment of the fMRI timeseries, coregistration of the data to a T_1-weighted anatomical image, and 8 mm full-width at half-maximum (FWHM) Gaussian smoothing.

Analysis. Voxelwise statistics on the salmon data were calculated through an ordinary least-squares estimation of the general linear model (GLM). Predictors of the hemodynamic response were modeled by a boxcar function convolved with a canonical hemodynamic response. A temporal high pass filter of 128 seconds was include to account for low frequency drift. No autocorrelation correction was applied.

Voxel Selection. Two methods were used for the correction of multiple comparisons in the fMRI results. The first method controlled the overall false discovery rate (FDR) and was based on a method defined by Benjamini and Hochberg (1995). The second method controlled the overall familywise error rate (FWER) through the use of Gaussian random field theory. This was done using algorithms originally devised by Friston et al. (1994).

GLM RESULTS

4.5
4.0
3.5
3.0
2.5
t-value

A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

Several active voxels were discovered in a cluster located within the salmon's brain cavity (Figure 1, see above). The size of this cluster was 81 mm³ with a cluster-level significance of $p = 0.001$. Due to the coarse resolution of the echo-planar image acquisition and the relatively small size of the salmon brain further discrimination between brain regions could not be completed. Out of a search volume of 8064 voxels a total of 16 voxels were significant.

Identical *t*-contrasts controlling the false discovery rate (FDR) and familywise error rate (FWER) were completed. These contrasts indicated no active voxels, even at relaxed statistical thresholds ($p = 0.25$).

VOXELWISE VARIABILITY

a) b) c)
20
15
10
5
standard deviation

To examine the spatial configuration of false positives we completed a

Who can think of examples of repeat inference concerns in finance?

Asset Management
Banking
Credit Modeling

Many issues with machine learning boil down to this point

Bonferroni Correction

Boole's Inequality:

For any finite sequence of events, the probability that at least one of the events happens is no greater than the sum of the probabilities of the individual events.

$$P(\bigcup_i A_i) \leq \sum_i P(A_i)$$

Bonferroni Correction

The Bonferroni correction: for $i = 1, \ldots, m$

$$P\left(\bigcup_i p_i \leq \frac{\alpha}{m}\right) \leq \sum_i P\left(p_i \leq \frac{\alpha}{m}\right) \leq \frac{m\alpha}{m} = \alpha$$

In repeated hypothesis tests, add up the p values to get the total level of confidence.

Bonferroni Correction

Example:

Overall confidence level α, run 20 hypothesis tests.

Can run each test at level $\frac{\alpha}{20}$, or with some fraction and sequence of $\{m_i\}$: $\sum_i \frac{\alpha}{m_i} = \alpha$

Run a 4% test and a 1% test => 5% overall.

Bonferroni Correction

Other fixes are possible, and there is some evidence this is overly punitive.

But it does highlight how little power there is after repeat inference!

Bootstrapping Correction (Chernozhukov)

Define the supremum distance measure as:

$$\|v\|_\infty := \max\{|v_i| : i \in \{1, \ldots, k\}\}.$$

Bootstrapping Correction

Consider a p_1-dimensional subvector β_1 of the coefficient vector β. Assume, without loss of generality, that these are the first p_1 components. Assume that

$$\hat{\beta}_1 - \beta_1 \overset{a}{\sim} N(0, V_{11}/n),$$

where V_{11} is the upper-left $p_1 \times p_1$ sub-block of V, in the sense that

$$\sup_{A \in \mathcal{A}} \left| \mathrm{P}\left(\sqrt{n}(\hat{\beta}_1 - \beta_1) \in A \right) - \mathrm{P}(N(0, V_{11}) \in A) \right| \to 0, \quad n \to \infty, \tag{6.1}$$

where $\mathcal{A}$ is a collection of rectangles in $\mathbb{R}^{p_1}$.

Translation: Take p_1 consistent estimates and their covariance matrix.

Bootstrapping Correction

The value of c can be determined as $(1-\alpha)$-quantile of

$$\|N(0,C)\|_\infty,$$

where C is the correlation matrix associated with V_{11}, that is,

$$C = S^{-1/2}V_{11}S^{-1/2}$$

where $S = \mathrm{diag}(V_{11})$ is a diagonal matrix with the diagonal of V_{11} in its diagonal and zeroes elsewhere. The constant c can be approximated by simulation.

This constant c is the right one by the following argument:

$$\begin{aligned}
\mathrm{P}(\beta_1 \in [\ell, u]) &= \mathrm{P}(\sqrt{n}(\hat{\beta}_1 - \beta_1) \in S^{1/2}[-c, c]) \\
&= \mathrm{P}(N(0, V_{11}) \in S^{1/2}[-c, c]) + o(1) \\
&= \mathrm{P}(S^{-1/2}N(0, V_{11}) \in [-c, c]) + o(1) \\
&= \mathrm{P}(\|N(0, S^{-1/2}V_{11}S^{-1/2})\|_\infty \leq c) + o(1) \\
&= 1 - \alpha + o(1),
\end{aligned}$$

Bootstrapping Correction

Translation:

Simulate p_1 correlated standard normals with the correlation matrix of your betas.

Take the upper envelope of values C such that no estimate has more than α values beyond C.