When Lag Regressions Fail: A Tale of Two Techniques

Marie C. McGraw
Elizabeth A. Barnes

Atmospheric Science, Colorado State University, Fort Collins, CO, USA

March 2, 2017

Lagged regression–a popular analysis tool

Lagged regression: a great tool to establish causality

simple: just one equation: $Y(t) = B_\tau \cdot X(t - \tau) + \epsilon$

popular: "lagged regression" in 1500+ articles in *J. Clim* since 1990

effective: sense of spatial and temporal variations and patterns

... but lagged regression has weaknesses under certain conditions.

Lagged regression–a popular analysis tool

Lagged regression: a great tool to establish causality

simple: just one equation: $Y(t) = B_\tau \cdot X(t - \tau) + \epsilon$

popular: "lagged regression" in 1500+ articles in *J. Clim* since 1990

effective: sense of spatial and temporal variations and patterns

... but lagged regression has weaknesses under certain conditions.

Lagged regression–a popular analysis tool

Lagged regression: a great tool to establish causality

simple: just one equation: $Y(t) = B_\tau \cdot X(t - \tau) + \epsilon$

popular: "lagged regression" in $1500+$ articles in *J. Clim* since 1990

effective: sense of spatial and temporal variations and patterns

... but lagged regression has weaknesses under certain conditions.

Lagged regression–a popular analysis tool

Lagged regression: a great tool to establish causality

simple: just one equation: $Y(t) = B_\tau \cdot X(t - \tau) + \epsilon$

popular: "lagged regression" in 1500+ articles in *J. Clim* since 1990

effective: sense of spatial and temporal variations and patterns

... but lagged regression has weaknesses under certain conditions.

Lagged regression–a popular analysis tool

Lagged regression: a great tool to establish causality

simple: just one equation: $Y(t) = B_\tau \cdot X(t - \tau) + \epsilon$

popular: "lagged regression" in $1500+$ articles in *J. Clim* since 1990

effective: sense of spatial and temporal variations and patterns

. . . but lagged regression has weaknesses under certain conditions.

Another way–Granger causality

Figure: Sir Clive Granger, economist, Nobel laureate.

A tool for using one time series to forecast another, popular in:

- Economics[1]
- Neuroscience[2]
- Detection and attribution studies[3,4,5]

Granger causality–just a few extra steps

1) Lagged regression of dependent variable ($y_{t-\tau}$) on itself (y_t)

2) Multivariate lagged regression of independent variable ($x_{t-\tau}$) and $y_{t-\tau}$ on y_t

3) Evaluate additional variance explained by including x

Does adding information about x_t increase our ability to predict y_t beyond the information from y_t itself?

Granger causality–just a few extra steps

1) Lagged regression of dependent variable ($y_{t-\tau}$) on itself (y_t)

2) Multivariate lagged regression of independent variable ($x_{t-\tau}$) and $y_{t-\tau}$ on y_t

3) Evaluate additional variance explained by including x

Does adding information about x_t increase our ability to predict y_t beyond the information from y_t itself?

Granger causality–just a few extra steps

1) Lagged regression of dependent variable ($y_{t-\tau}$) on itself (y_t)

2) Multivariate lagged regression of independent variable ($x_{t-\tau}$) and $y_{t-\tau}$ on y_t

3) Evaluate additional variance explained by including x

Does adding information about x_t increase our ability to predict y_t beyond the information from y_t itself?

Granger causality–just a few extra steps

1) Lagged regression of dependent variable ($y_{t-\tau}$) on itself (y_t)

2) Multivariate lagged regression of independent variable ($x_{t-\tau}$) and $y_{t-\tau}$ on y_t

3) Evaluate additional variance explained by including x

Does adding information about x_t increase our ability to predict y_t beyond the information from y_t itself?

Granger causality–just a few extra steps

1) Lagged regression of dependent variable ($y_{t-\tau}$) on itself (y_t)

2) Multivariate lagged regression of independent variable ($x_{t-\tau}$) and $y_{t-\tau}$ on y_t

3) Evaluate additional variance explained by including x

Does adding information about x_t increase our ability to predict y_t beyond the information from y_t itself?

Granger causality has some limitations

Only tests X causes Y–could be something else (Z) causing both

Assumes **linearity**

Assumes **stationarity**

Granger causality has some limitations

Only tests X causes Y–could be something else (Z) causing both

Assumes **linearity**

Assumes **stationarity**

Granger causality has some limitations

Only tests X causes Y–could be something else (Z) causing both

Assumes **linearity**

Assumes **stationarity**

Statistical model

1. Create Y, a red-noise time series with some auto-correlation coefficient (α_y)

$$Y(t) = \alpha_y \cdot Y(t-1) + (1-\alpha_y^2)^{1/2}\epsilon(t) \tag{1}$$

2. Create X using Y: X is simply Y lagged by some number of steps (τ) with added noise, ϵ

$$X(t) = Y(t-\tau) + \epsilon(t) \tag{2}$$

3. Perform lagged regressions and Granger causality analysis in both the "correct" ($Y \rightarrow X$) and "incorrect" ($X \rightarrow Y$) directions
4. Repeat 50,000 times

McGraw and Barnes, in prep

Statistical model

1. Create Y, a red-noise time series with some auto-correlation coefficient (α_y)

$$Y(t) = \alpha_y \cdot Y(t-1) + (1-\alpha_y^2)^{1/2}\epsilon(t) \tag{1}$$

2. Create X using Y: X is simply Y lagged by some number of steps (τ) with added noise, ϵ

$$X(t) = Y(t-\tau) + \epsilon(t) \tag{2}$$

3. Perform lagged regressions and Granger causality analysis in both the "correct" ($Y \rightarrow X$) and "incorrect" ($X \rightarrow Y$) directions
4. Repeat 50,000 times

McGraw and Barnes, in prep

Statistical model

1. Create Y, a red-noise time series with some auto-correlation coefficient (α_y)

$$Y(t) = \alpha_y \cdot Y(t-1) + (1-\alpha_y^2)^{1/2}\epsilon(t) \quad (1)$$

2. Create X using Y: X is simply Y lagged by some number of steps (τ) with added noise, ϵ

$$X(t) = Y(t-\tau) + \epsilon(t) \quad (2)$$

3. Perform lagged regressions and Granger causality analysis in both the "correct" ($Y \rightarrow X$) and "incorrect" ($X \rightarrow Y$) directions
4. Repeat 50,000 times

McGraw and Barnes, in prep

Statistical model

1. Create Y, a red-noise time series with some auto-correlation coefficient (α_y)

$$Y(t) = \alpha_y \cdot Y(t-1) + (1-\alpha_y^2)^{1/2}\epsilon(t) \tag{1}$$

2. Create X using Y: X is simply Y lagged by some number of steps (τ) with added noise, ϵ

$$X(t) = Y(t-\tau) + \epsilon(t) \tag{2}$$

3. Perform lagged regressions and Granger causality analysis in both the "correct" ($Y \rightarrow X$) and "incorrect" ($X \rightarrow Y$) directions
4. Repeat 50,000 times

McGraw and Barnes, in prep

Statistical model

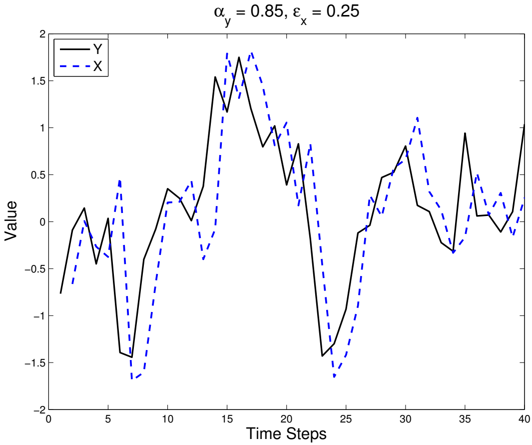

Figure: Example of X created by lagging Y one day.

McGraw and Barnes, in prep

Statistical model–The Right Direction

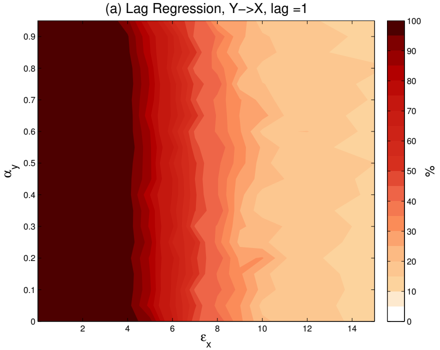

Figure: Testing hypothesis that $Y \rightarrow X$ with (left) lagged regression and (right) Granger causality at 95% confidence. Shading represents percentage of significant results.

McGraw and Barnes, in prep

Statistical model–The Right Direction

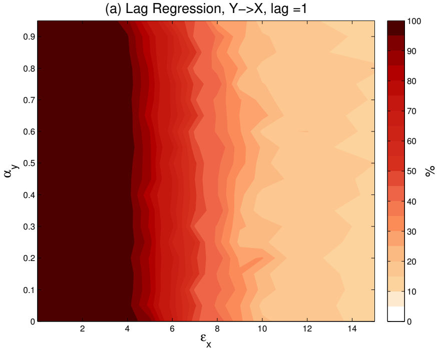

Figure: Testing hypothesis that $Y \rightarrow X$ with (left) lagged regression and (right) Granger causality at 95% confidence. Shading represents percentage of significant results.

McGraw and Barnes, in prep

Statistical model–The Wrong Direction

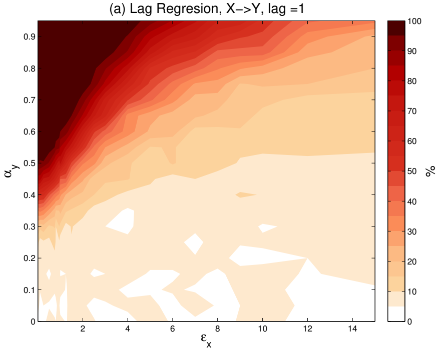

Figure: Testing hypothesis that $X \rightarrow Y$ with (left) lagged regression and (right) Granger causality at 95% confidence. Shading represents percentage of significant results.

McGraw and Barnes, in prep

Statistical model–The Wrong Direction

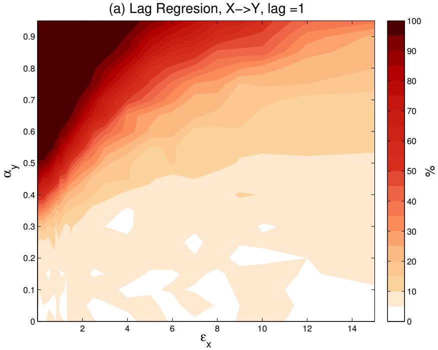

Figure: Testing hypothesis that $X \rightarrow Y$ with (left) lagged regression and (right) Granger causality at 95% confidence. Shading represents percentage of significant results.

McGraw and Barnes, in prep

A Real World Example

We think about ...

$$\text{ENSO} \rightarrow \text{T}$$

but what about

$$\text{T} \rightarrow \text{ENSO}$$

A Real World Example

We think about . . .

$$\text{ENSO} \rightarrow \text{T}$$

but what about

$$\text{T} \rightarrow \text{ENSO}$$

Observations–The Right Direction

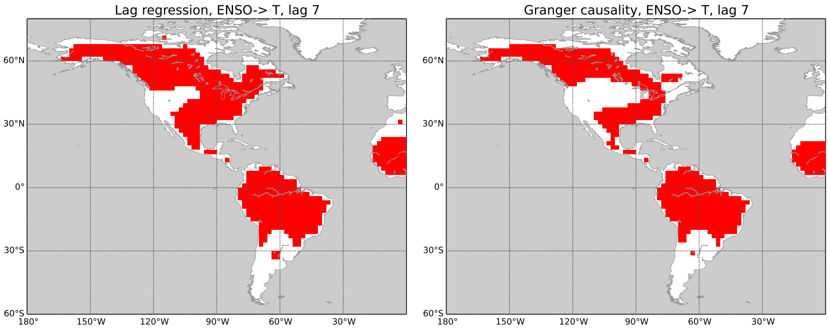

Figure: Testing the hypothesis that ENSO causes changes in surface temperature with (left) lagged regression and (right) Granger causality at 95% confidence. Red indicates a significant lagged relationship identified at up to 7 months.

McGraw and Barnes, in prep

Observations–The Wrong Direction

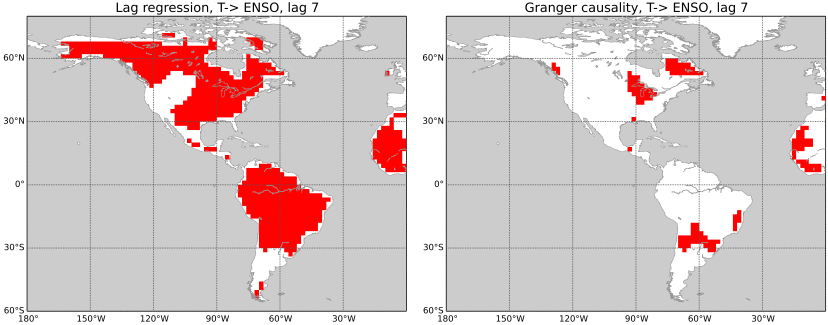

Figure: Testing the hypothesis that surface temperature causes changes in ENSO with (left) lagged regression and (right) Granger causality at 95% confidence. Red indicates a significant lagged relationship identified at up to 7 months.

McGraw and Barnes, in prep

Granger causality can improve your analysis:

Ensures significant results are not due to memory in data

Similar to lag regression in the "right" direction ...
but will fail in the "wrong" direction

Granger causality–more robust results for a little extra work!
Try it out!

Granger causality can improve your analysis:

Ensures significant results are not due to memory in data

Similar to lag regression in the “right” direction ...
but will fail in the “wrong” direction

Granger causality–more robust results for a little extra work! Try it out!

Granger causality can improve your analysis:

Ensures significant results are not due to memory in data

Similar to lag regression in the "right" direction ...
but will fail in the "wrong" direction

Granger causality–more robust results for a little extra work!
Try it out!

Granger causality can improve your analysis:

Ensures significant results are not due to memory in data

Similar to lag regression in the “right” direction ...
but will fail in the “wrong” direction

Granger causality–more robust results for a little extra work! Try it out!

References

1 Granger, C.W.J. (1969): "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica,* **37**, 424-438.

2 Seth, A.K., A.B. Barrett, and L. Barrett (2015): "Granger causality analysis in neuroscience and neuroimaging." *J. Neurosci.*, **35**, 3293-3297.

3 Attanasio, A., A. Pasini, and U. Triacca (2012): "A contribution to attribution of recent global warming by out-of-sample Granger causality analysis." *Atmos. Sci. Lett.,* **13**, 67-72.

4 Pasini, A., U. Triacca, and A. Attanasio (2012): "Evidence of recent causal decoupling between solar radiation and global temperature." *Environ. Res. Lett.,* **7**, 034020.

5 Attanasio, A., A. Pasini, and U. Triacca (2013): "Granger causality analyses for climatic attribution." *Atmos. Clim. Sci.,* **3**, 515-522.

BACKUP SLIDES

Statistical model–The Right Direction

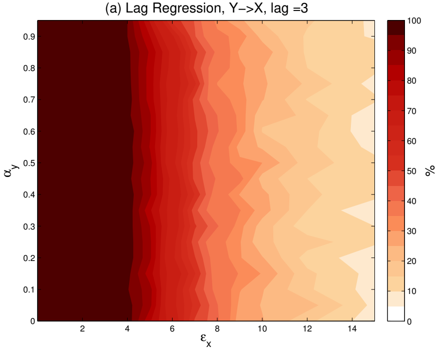

Figure: Testing hypothesis that $Y \rightarrow X$ with (left) lagged regression and (right) Granger causality at 95% confidence. Shading represents percentage of significant results (e.g., false positives).

Statistical model–The Wrong Direction

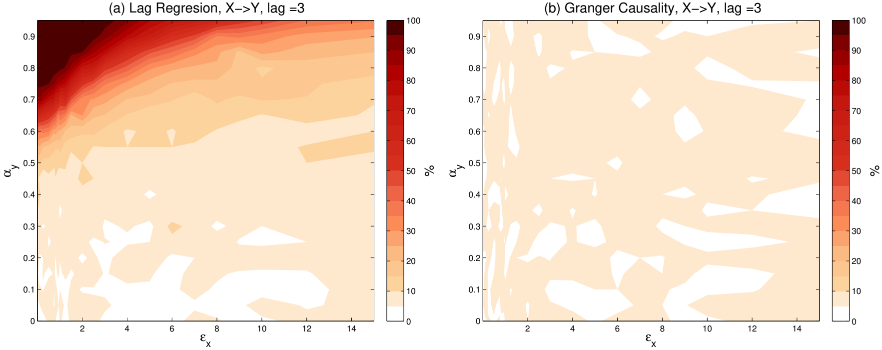

Figure: Testing hypothesis that $X \rightarrow Y$ with (left) lagged regression and (right) Granger causality at 95% confidence. Shading represents percentage of significant results.

Statistical model

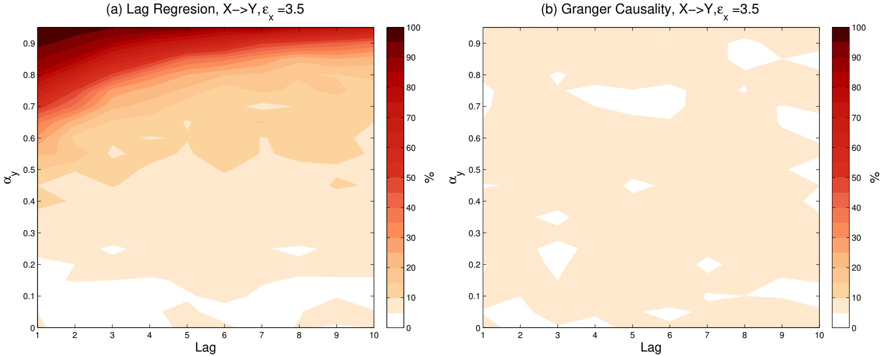

Figure: Percentage of significant results as a function of lag for (left) lagged regression and (right) Granger causality.

Statistical model

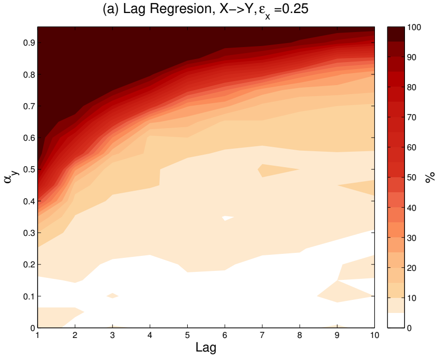

Figure: Percentage of significant results as a function of lag for (left) lagged regression and (right) Granger causality.

McGraw and Barnes, in prep