Intervention Models and Forecasting
Transfer Functions with Binary Inputs

Jamie Monogan
University of Georgia

January 23, 2018
Objectives

By the end of this meeting, participants should be able to:

- Forecast future values of a time series with an estimated ARIMA model.
- Perform a static time series regression with ARIMA errors.
- Explain the logic behind intervention analysis.
- Use backshift algebra to make proofs about difference equations.
- Describe the substantive meaning of onset, effects, and dynamics (growth/decay).
- Estimate and graph the results of an impulse response function.
- Explain the purpose of additional numerator terms in intervention analyses.
- Interpret compound models. (Time permitting.)
Goals of Data Analysis

- Inference
- Prediction

Prediction in OLS

\[ E(Y|X) = \hat{Y} = X\hat{\beta} \]

Given a certain exogenous situation, what will we expect our endogenous variable to be?

Forecasting with ARIMA

- Similar to OLS, except projecting with estimated difference equations.
- All future projections are a function of observed values and parameter estimates.
- Performs well, often better than more complex models.
- Short-term outperforms long-term.
The AR(1) Case
Higher-Order AR is a Simple Extension

- We simply proceed with forward iteration.
- Suppose \( y_t = a_0 + a_1 y_{t-1} + \epsilon_t \)
- One period ahead, should simply be: \( y_{t+1} = a_0 + a_1 y_t + \epsilon_{t+1} \)
- Just like OLS, though, we use expectation:
  \( \hat{y}_{t+1} = E(y_{t+1}|y_t) = a_0 + a_1 y_t \)
- Now that we have one period ahead, let’s forecast two ahead:
  \( E(y_{t+2}|y_t) = a_0 + a_1 E(y_{t+1}|y_t) \)
- Which is: \( E(y_{t+2}|y_t) = a_0 + a_1(a_0 + a_1 y_t) \)
- Repeat the process \textit{ad infinitum}. Any concerns?
Error Variance

- We expect error in any prediction. This error aggregates over time.
- Recall: The formula for the variance of a linear combination is
  \[ \text{Var}(aX + bY) = a^2 \text{Var}(X) + 2ab \text{Cov}(X, Y) + b^2 \text{Var}(Y). \]
- For AR(1), our predictive error variance one lead out:
  \[ \text{Var}(\hat{y}_{t+1}) = \text{Var}(a_0 + a_1y_t + \epsilon_{t+1}) = \text{Var}(\epsilon_{t+1}) = \sigma^2. \]
- For AR(1), our predictive error variance two leads out:
  \[ \text{Var}(\hat{y}_{t+2}) = \text{Var}(a_0 + a_1\hat{y}_{t+1} + \epsilon_{t+2}) = \text{Var}(a_1\hat{y}_{t+1} + \epsilon_{t+2}) = \]
  \[ a_1^2 \text{Var}(\hat{y}_{t+1}) + \text{Var}(\epsilon_{t+2}) = a_1^2 \sigma^2 + \sigma^2. \]
- In general, our AR(1) forecast error variance \( j \) leads into the future is:
  \[ \sigma^2 [1 + a_1^2 + a_1^4 + a_1^6 + \ldots + a_1^{2(j-1)}]. \]
- This converges to: \( \frac{\sigma^2}{1 - a_1^2} \).

Confidence Intervals

- forecast \( \pm (\text{critical value} \times \sqrt{\text{root of forecast error variance}}) \)
- one ahead: \( a_0 + a_1y_t \pm 1.96\sigma \)
- two ahead: \( a_0(1 + a_1) + a_1^2y_t \pm 1.96\sigma(1 + a_1^2)^{1/2} \).
The General Case

- We can extend the logic of forward iteration to a general ARMA model if we assume $E(\epsilon_{t+j}|y_t) = 0$ for any $j > 0$.
- For example, an ARMA(2,1): $y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t + \beta_1 \epsilon_{t-1}$
- One period ahead: $y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \epsilon_{t+1} + \beta_1 \epsilon_t$
- Expectation: $E(y_{t+1}|y_t) = a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \epsilon_t$
- Two periods ahead: $y_{t+2} = a_0 + a_1 y_{t+1} + a_2 y_t + \epsilon_{t+2} + \beta_1 \epsilon_{t+1}$
- Expectation: $E(y_{t+2}|y_t) = a_0 + a_1 E(y_{t+1}) + a_2 y_t$
- Break to software for an example.
Inference: Estimating a Regression with ARIMA Errors

- We can now consider a case where we have a static time series regression with ARIMA errors (Hibbs 1974).
- \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon_t \]
- This is static because only \textit{current} \( x \) values affect current \( y \) values.

Software

- The \texttt{arima} commands in R and Stata can handle any number of right-hand-side regressors, and ARIMA errors.
- \begin{verbatim}
R: mod.1 <- arima(data$y, order=c(1,0,0),
                 xreg=cbind(data$x1, data$x2))

Stata: arima y x1 x2, AR(1)
\end{verbatim}
- Both of these are for the AR(1) case and will produce a ML regression with ML estimated \( \phi_1 \) simultaneously.
- This is still a very limited extension, because static models are not often appropriate specifications for longitudinal causality.
- If they were, the course could end now.
## Extending ARIMA to Static Regressions

### Percent Identifying as “Liberal” Over Time

A Real Example of a Static ARIMA Regression

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>MLE for AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E.</td>
</tr>
<tr>
<td>Great Society intervention</td>
<td>-5.92</td>
<td>0.65</td>
</tr>
<tr>
<td>Party control duration</td>
<td>-0.18</td>
<td>0.05</td>
</tr>
<tr>
<td>Post-intervention trend</td>
<td>-0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>Intercept</td>
<td>43.65</td>
<td>0.32</td>
</tr>
<tr>
<td>$\hat{\phi}_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2_{adj}$</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>N=70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Citation:

Two Issues in Time Series Regressions

1. Correlated errors in the residuals violate OLS assumptions, producing inefficient $\beta$ and biased $\sigma^2$, $t$, and $p$.
   - That’s the one we’ve solved.

2. Dynamics: $y$ is likely to be caused by previous values of $x$ and $y$.
   - This is the big one, producing biased and inconsistent $\beta$.
   - This is a violation of the Gauss-Markov assumption of proper functional form.
   - That is the focus of much of the rest of the course.
Thinking About Causality

- Social scientists have an odd attitude about causality:
  - We don’t talk about it in public and
  - we are obsessed by it.
- Virtually all theory is causal. Explanations are about cause, not about mere covariation.
- So saying that we are not going to talk about causation because there are thorny philosophical issues in it is, in a word, Stupid!
Causation in Cross-Sectional Data
James S. Coleman (1986)

- When we say \( x \) causes \( y \), we mean that at some time before the time of observation some dynamic process occurred which aligned \( y \) with \( x \) and furthermore, that process had reached an equilibrium level when we observed \( x \) and \( y \).
- The exciting difference in time series is that we get to observe \( x \) acting on \( y \).

Back to the Static Time Series Regression

- “The more the \( x \), the more the \( y \)” is not a satisfactory model of causality (as Coleman knew).
- And just because we have time series observations does not make it one.
- We need a sense of process, of delay, of decay, of cumulation, in our model.
The Interrupted Time-Series Design
A Type of Quasi-Experimental Design

- We have long known that the “interrupted time series” design has superb properties for causal inference.
- We make several observations ($O_t$) on a series and then introduce (or nature does it for us) $I$, an intervention:
  
  $$O_1 \ O_2 \ O_3 \ O_4 \ I \ O_5 \ O_6 \ O_7 \ O_8$$

- Then we observe whether the level of the series changes after introduction of a presumed causal effect.
- Three requisites of causal inference:
  1. Covariation between $x$ (or $I$) and $y$.
  2. Time order: $I$ precedes the change in $y$.
  3. Rule out all alternative explanations.
- The interrupted time series quasi-experiment handles #1 and #2 without problems. It doesn’t quite do #3.
  - It requires a *ceteris paribus* assumption between $O_1$-$O_4$ and $O_5$-$O_8$
  - And because time series have persistence issues (i.e. autocorrelation and equilibration), we don’t quite nail that *ceteris paribus*. 
### Why Begin with Binary Inputs?

- A *regular* transfer function, $y = f(x)$, has two random variables, each with stochastic error and (generally) systematic error aggregation.
- Systematic error aggregation processes in $x$ and $y$ generally are highly correlated. If this is the case, we cannot see causation between $x$ & $y$.
- $I_t$ on the other hand is a fixed variable, consisting of 0’s and 1’s reflecting researcher judgments.
  - It has no stochastic error.
  - Hence it has no error aggregation.
  - Hence $y$’s error aggregation cannot correlate with error aggregation in $I$.
  - Thus the binary case is easier & a good spot to start transfer functions.

**Box & Tiao 1975**

- Intervention models ($y = f(I) + N_t$) combine an ARIMA noise model with a causal specification for $f(I)$ to overcome the statistical limitations of the design.
- With ARIMA we can forecast $O_5-O_8$ and then have confidence that a divergence from that forecast represents causality.
The Setup for Intervention Models

- Let $I_t$ be a series of 0's and 1's, representing the on and off states of a presumed intervention.
  - $I_t$ may be either a “pulse” (... 0001000...) or a “step” (... 0001111...).
  - A step can be seen to be an integrated pulse.
- Then $y = f(I_t) + N_t$ is a model of structure, $f(I)$, and error, $N_t$.
- For a time series the appropriate function, $f$, will be a transfer function.

Transfer Functions

- Transfer functions are models of how the impulses in $x$ (or $I$ in the intervention case) are translated into future values of $y$.
- Transfer functions are classified by the order of their dynamics.
  - E.g. a regression $\beta$ is a zero-order (static) transfer function, denoted $\omega$. 
Dynamics

- Dynamics in a Transfer Function model the decay of effects—or additional effects for the integrated case.
- Only 1\textsuperscript{st} order dynamics appear in our literatures.
- The first order model has two parameters, $\omega$ in the numerator and $\delta$ in the denominator.
  - E.g., $y_t = \left[ \frac{\omega_0}{(1-\delta_1 B)} \right] I_{t-k} + N_t$.

Good News!

- You already know the zero order transfer function:
  - It is regression.
  - Just change notation: $y = \beta_0 + \omega_0 I_t + N_t$, substituting $\omega_0$ for the familiar $\beta_1$.
- And R or Stata’s ARIMA command can estimate this transfer function, because it is linear.
Notation

The Generic First Order Transfer Function

\[ y_t = \frac{(\omega_0 + \omega_1 + \cdots + \omega_k)I_{t-k}}{(1 - \delta_1 B)} + N_t \]

The One Input Parameter Case

\[ y_t = \left[ \frac{\omega_0}{(1 - \delta_1 B)} \right]I_{t-k} + N_t, \text{ where:} \]

- \( \omega_0 \) translates the current (0 lag) I into y,
- \( \delta_1 \) specifies 1st order dynamic decay, &
- \( N_t \) is an ARIMA noise model.
- We can use this information to craft the Implied Impulse Response Function.
The Advantage of the Decay Term

- Our non-linear model specification indicates that the effect of the input decays as time elapses.
- Leaving this out means the effect **abruptly disappears**.
- For a step input, though, $\delta$ is a **growth rate**.
- Where does the change in effect stop?

Equilibrium Effects for Two Cases

- **Pulse Input and stationary series:**
  \[ \lim_{t \to \infty} y_t = 0. \]

- **Step input or integrated series:**
  \[ \lim_{t \to \infty} y_t = \frac{\omega_0}{1 - \delta_1}. \]
Percent Liberal Divided by Sum(Liberal + Conservative)
Identification with the symbol “liberal” was damaged by the Great Society in three ways:

1. The excesses of LBJ and the 89th Congress (“Maximum Feasible Participation”).
3. The new clientele of liberalism.
   - From FDR’s “common man” to LBJ’s “poor person.”

Thus we entertain a first order dynamic intervention occurring in 1966, the first year in which a reaction to the Great Society could occur.
Modeling Steps for the Intervention Components

1. Identify tentative ARIMA error model.
2. Estimate tentative ARIMA model.
3. Specify Interventions (now theory-driven).
5. Diagnose Residuals.
6. Graph the estimated intervention models.

Software

- Stata’s ARIMA estimation routine can handle linear interventions, but not dynamic (nonlinear) ones like this example.
- Thus we need the “arimax” function in R
- Alternatively, RATS (Regression Analysis of Time Series) has a “BOXJENCK” procedure that will model ARIMA errors and dynamic inputs.
Example Intervention Graph

![Intervention Analysis Example](image-url)
Intervention Modeling Strategy

- In general, you begin the task with priors, unlike ARIMA, but weak ones.
- You know when to expect the effect, whether it should be positive or negative, etc.
- But you probably do not know whether or not it is permanent, whether or not it is dynamic, etc.
- This strategy will guide your search so that you don’t abort the modeling before you have entertained the right specification.
Suggestions on Specifying Interventions

1. Always begin with pulse input to model temporary effects for stationary series.
   - If it is not temporary, you will see evidence in the form of estimated $\delta_1 \geq 1$ which means precisely that the effect does not decay.

2. Always entertain first-order dynamics.
   - Same reason, if it is not a first order process we will see evidence that it is not, $\delta_1 \neq 0$.

3. Always graph your intervention model against the data.
   - More often than not, this exercise causes laughter, when you realize that the model in your head isn’t actually what you are imposing on the data!

4. Confirm specification by overfitting.
   - Add one extra step of complication. If your specification is right, it will be nonsignificant.
Interpretation: The Implied Impulse Response Function

- The IRF is a map of causal flow.
- It tracks how $y$ responds to an input $I_t$ for periods 0–$k$, where 0 is the onset of the effect, (as in $\omega_0$).
- The *implied IRF* is a deterministic function of parameter estimates.
  - It is what a transfer function models, but without the noise component.
- It has no parallel in the cross-sectional world because linear contemporaneous effects don’t require a map.

### Wave-by-Wave

- The Implied Impulse Response Function for $y_t = \left[ \frac{\omega_0}{1-\delta_1 B} \right] I_t + N_t$ at lag:

  0 $\omega_0$
  1 $\omega_0 \delta_1$
  2 $\omega_0 \delta_1^2$
  
  
  $\vdots$

  $T$ $\omega_0 \delta_1^T$

- For “large” $T$, $\delta_1^T \approx 0.0$ thus $\omega_0 \delta_1^T \approx 0.0$
Inputs

The equivalence: $(1-B)\text{step} = \text{pulse}$ and $\text{pulse} = \text{step}/(1-B)$

Equivalence

- $(1-B)\text{step} = \text{pulse}$. So pulse is a 1\textsuperscript{st} differenced step!
- $\text{Step} = \text{pulse}/(1-B)$. So step is a cumulated pulse!
- Therefore, a step input for a stationary series produces an identical impulse response to a pulse input for an integrated I(1) series.
Zero Order Transfer Functions

\[ y_t = \omega_0 I_t + N_t \]
First Order Transfer Functions
Pulse Input on Left, Step Input on Right

\[ y_t = \left[ \frac{\omega_0}{1 - \delta_1 B} \right] I_t + N_t \]
Integrated Series

- We must difference an integrated series before doing intervention analysis.
- So we proceed in steps:
  - First we create $w_t = y_t - y_{t-1}$
  - Then we fit a model to the stationary series $w_t$.
- This is equivalent to modeling $y_t$ itself and including the differencing operator $(1-B)$ in the denominator of the transfer function.
Differencing in the Transfer Function Denominator

Pulse Input on Left, Step on Right

\[ y_t = \left[ \frac{\omega_0}{1 - B} \right] l_t + N_t \]

The step input on the right goes to infinity and should not be entertained in a model of stable processes.
Adding Complications to the Numerator

- Our standard first order transfer function attains its maximum effect at the point of onset \((t+k)\) and is incapable of growing larger.

\[
y_t = \left[ \frac{\omega_0}{1 - \delta_1 B} \right] I_{t-k} + N_t
\]

- To model an effect that grows after onset we need additional \(\omega\) parameters in the numerator.
- To grow for one period after onset, for example:

\[
y_t = \left[ \frac{\omega_0 + \omega_1}{1 - \delta_1 B} \right] I_{t-k} + N_t
\]

- And that can be extended to any desired level of complexity by adding more \(\omega\) parameters, each of which will extend growth for one more period beyond onset, e.g.,

\[
y_t = \left[ \frac{\omega_0 + \omega_1 + \omega_2 + \omega_3}{1 - \delta_1 B} \right] I_{t-k} + N_t
\]

- Break to software.
For Next Time:

- Complete questions #2.d & 2.e from page 185 of *Political Analysis Using R*. (In brief, report what you concluded from the noise-modeling part of this question before.)

**Reading:**
- Carmines & Stimson (1986). Read the appendix, too.
- Gurnell & Fenn (1984). Figure 6 is a gem. Don’t dwell too much on sediment science.
Additional Material
More Complexity?

- Zero and first order transfer functions don’t quite handle the interesting dynamics we might hypothesize.
- The limitation comes from the choice of asymptotics for the 1\textsuperscript{st} order case:
  - Stationary: zero.
  - Integrated: $\omega_0/(1-\delta_1)$.
  - What about a third case, decay, but not all the way to zero?
Example: 9-11

- We might ask what was the effect of 9-11 on some political series, e.g., approval, partisanship, etc.
- But those effects are so huge that we would be posing a pretty trivial question, not much in need of a refined answer.
- But consider a more interesting question: *did the 9-11 events produce permanent change in American politics?*
- A 1\textsuperscript{st} order model with a pulse input can’t answer the question.
  - The model requires a zero asymptote.
  - If it happens to be a good fit to the data, it may mislead because it didn’t allow the possibility of permanent effects.
  - What we need is a model that allows that possibility, so that we might estimate a parameter for it.
A Compound Model

- A model can’t be both temporary and permanent.
  - Those contradict one another.
- The solution is to estimate two effects:
  - A big but temporary impulse-decay process, and
  - A smaller permanent residue.
  - That requires two transfer functions:
    - A 1st order model of impulse-decay (pulse input), and
    - a zero order model of permanent change (step input).
A Compound Model of Decaying and Permanent Effects

Decaying Temporary Effects on Top of Permanent Change

\[ y_t = \left[ \frac{\omega_{01}}{1 - \delta_1 B} \right] I_t + \left[ \frac{\omega_{02}}{1 - B} \right] I_t + N_t \]
9-11 Example

- For our example case, $\omega_{02}$ represents the permanent effect of 9-11 and is the key test of the permanent effect claim.
- A test of significance on the $\omega$ parameter is a test of the permanent effect claim.
For a stationary series, a constant term is just a mean.

Since real series rarely have a mean of zero, you should nearly always include a constant in an estimated model.

- R and Stata estimate constants by default.
- On the other hand, RATS does not. Leaving out a constant where one belongs is an estimation disaster.

For integrated series, you still have a choice, but the interpretation is different.

A constant in differenced data represents the mean gain/loss per time period—which is exactly the definition of “trend” in the Box-Jenkins tradition.

- Thus you include a constant when you postulate a trend or wish to test for one.
- A test of significance for $\theta_0$ is a test for linear trend.
Stationarity, Integration, and Constants II

- But beware: the empirically common random walk in a short time-series can produce false evidence of trend.
- The difference is crucial:
  - A real trend tends indefinitely in the same direction, because some underlying phenomenon produces it.
    - e.g., anything that is a function of population trends upward.
  - A random walk “trends” up and down for spells and eventually reverses itself.
- Theory helps. If we think something trends, we should know why.
- Spurious regressions: a major hazard.
Regular Transfer Functions

- What if we just substitute $x_t$ for $I_t$ in the transfer function model?
- Our ability to “see” the transfer function representation in the Box-Tiao intervention case depended upon having an $I_t$ measured without error.
- $x_t$ will have errors and (probably) systematic error aggregation processes.
- Thus the evidence of causality between $x$ and $y$ will have two components:
  1. True causal relationships, and
  2. Spurious relationships produced by covariance between the two error aggregation processes.
- Figuring out how to observe #1 all by itself is where we go next.