POLS571 - Longitudinal Data Analysis

September 27, 2001

Intervention Analysis

The simplest form of multivariate time-series analysis involves investigating the impact of some event (or "intervention") on the level of the series. Some people refer to these as "transfer function" models (which is a term from engineering), or "impact assessments".

There are two key things to remember before beginning an intervention analysis:

1. The researcher needs to know, *a priori*, when and how the intervention occurred;

2. The nature of the event needs to be reflected as closely as possible.

Condition (1) just means that intervention analysis proceeds by assessing the influence of a known change on a time series. It is *not*, therefore, a "fishing expedition" in which changes in the time series are uncovered and causal mechanisms are then found for those changes.

Condition (2) means that, to the extent possible, the operationalization of the intervention should match its reality. For example, in an early study of the impact of gun enforcement laws, Zimring (1975) had to deal with the fact that, while the law was passed at a particular point in time, its enforcement was implemented more slowly. Thus, rather than a single step-change (i.e., $X = 0$ before the law's passage and $X = 1$ after) he has the intervention take place over six months: $X = 1/6$ in the first month, $X = 2/6$ in the second, etc. until $X = 1$ in the sixth month and thereafter. In other words, *use your theory, and use your head.*

Consider a conventional stationary ARMA(p,q) time series:

$$\phi_p L^p Y_t = \theta_q L^q u_t \tag{1}$$

Both long- and short-term changes in Y_t may occur purely as a result of "shocks" (extreme values of u). Call the collective effect of short- and long-term chances in Y_t due to the "shocks" u_t N_t. If we consider the influence of some intervention X_t on the time series, we can write a very general model as:

$$Y_t = f(X_t) + N_t \tag{2}$$

The challenge in intervention analysis is to distinguish changes in Y due to the "noise" N_t from those due to the intervention X_t.

1 Zero-Order Models

The simplest way in which an event can impact a series is as a step function. Recalling that AR processes can be rewritten as sums of moving averages of the disturbances, we can consider the ARIMA aspects of the model as the "noise" component. A simple version of a zero-order intervention model can then be written as:

$$Y_t = \omega X_t + N_t \tag{3}$$

where N_t summarizes the "noise" (ARMA) components of the model, X_t is the intervention variable (e.g., coded $X = 0$ prior to the event and $X = 1$ after) and ω is the impact of the intervention on Y. Note that the effect of X on Y is *instantaneous*, *constant* and *permanent*. Estimating $\hat{\omega}$ provides an estimate of the difference between the pre- and post-intervention levels of the process.

Intervention modeling begins with establishing the ARIMA properties of the series (i.e., the N_t component). This is done in the standard way. Once the properties of the series are determined, it becomes possible to estimate the model in (3) in a straightforward way. So, for example, if Y was initially determined to follow an ARMA(1,0) process:

$$\begin{aligned} Y_t &= \phi Y_{t-1} + u_t \\ &= \sum_{j=0}^{T} \phi^j u_{t-j} \end{aligned} \tag{4}$$

then the model with the intervention is simply:

$$Y_t = \omega X_t + \sum_{j=0}^{T} \phi^j u_{t-j} \tag{5}$$

In Stata, we can estimate this model using the `-arima-` command; we'll do an example in a bit...

2 First-Order Models

The assumption that an event's impact is immediate and discrete is often untenable; an alternative approach would allow for a gradual effect of an intervention over time. A first-order intervention model does this. Consider the model in equation (2), but write it as:

$$Y_t^* = Y_t - N_t \tag{6}$$

That is, consider only the deterministic part of the equation $f(X_t)$. In (3), we used a simple linear function to define Y_t^* in the zero-order model. For a first-order model, we require an additional parameter; define $f(X_t)$ as:

$$Y_t^* \equiv f(X_t) = \frac{\omega}{1 - \delta L} X_t \tag{7}$$

where we require that $-1 < \delta < 1$ (we'll see why in a minute). From this specification of $f(X_t)$, we get the following model:

$$\begin{aligned} Y_t^* &= \frac{\omega}{1 - \delta L} X_t \\ (1 - \delta L) Y_t^* &= \omega X_t \\ Y_t^* &= \delta Y_{t-1}^* + \omega X_t \end{aligned} \tag{8}$$

Note further that, because $Y_{t-1}^* = \delta Y_{t-2}^* + \omega X_{t-1}$, and $|\delta| < 1$ (so that $\delta^j Y_{t-j}^* \to 0$ as $T \to \infty$), we can substitute back into (8) to get:

$$Y_t^* = \omega \sum_{j=0}^{T} \delta^j X_t \tag{9}$$

This means that, for all observations prior to the intervention (i.e., where $X_t = 0$), $Y_t^* = 0$ as well. In the first period $t + 1$ in which $X_t = 1$, we get:

$$
\begin{aligned}
Y^*_{t+1} &= \delta Y^*_t + \omega X_{t+1} \\
&= \delta(0) + \omega(1) \\
&= \omega
\end{aligned}
$$

In the second period after the intervention,

$$
\begin{aligned}
Y^*_{t+2} &= \delta Y^*_{t+1} + \omega X_{t+2} \\
&= \delta\omega + \omega
\end{aligned}
$$

and the third period is:

$$
\begin{aligned}
Y^*_{t+3} &= \delta Y^*_{t+2} + \omega X_{t+3} \\
&= \delta(\delta + \omega) + \omega \\
&= \delta^2\omega + \delta\omega + \omega
\end{aligned}
$$

More generally, for the kth postintervention period in which $X_t = 1$:

$$
\begin{aligned}
Y^*_{t+k} &= \delta(\delta^{k-1}\omega + ... + \delta\omega + \omega) \\
&= \sum_{j=0}^{k} \delta^k \omega \qquad (10)
\end{aligned}
$$

Because $|\delta| < 1$, the terms being summed are decreasing over time. If we consider the limiting cases, note that $\delta = 0$ corresponds to the zero-order model discussed above

$$
Y^*_t = \omega X_t
$$

Similarly, $\delta = 1$ corresponds to a model in which the intervention yields a deterministic trend with slope ω:

$$
Y^*_t = \omega \sum_{j=0}^{T} X_t
$$

(its unlikely that this latter formulation will be very useful...). For $|\delta| < 1$ and $\neq 0$, δ determines the size of the effect of X on Y, as well as the rate at which the shock approaches its asymptotic limit. Using (**??**), the latter is simply:

$$\begin{aligned} Y^*_{t\to\infty} &= \sum_{j=0}^{\infty} \delta^j \omega \\ &= \frac{\omega}{1-\delta} \end{aligned} \tag{11}$$

The effects of a shock on a model with $\omega = 1$ and with varying values of δ are presented graphically in Figure 1.

3 Shifts and Pulses

So far, we've discussed interventions as a permanent change in X (that is, a *shift*). Alternatively, an intervention's effect may only be temporary; it may fade out over time. In that case, X may equal 1 only for a given period or periods, after which it reverts back to a zero value.[1] Both zero-and first-order models of pulse functions exist.

Mathematically, models for these two types of interventions are the same; what differs is the substance of their interpretation. When X is a pulse, a zero-order model remains:

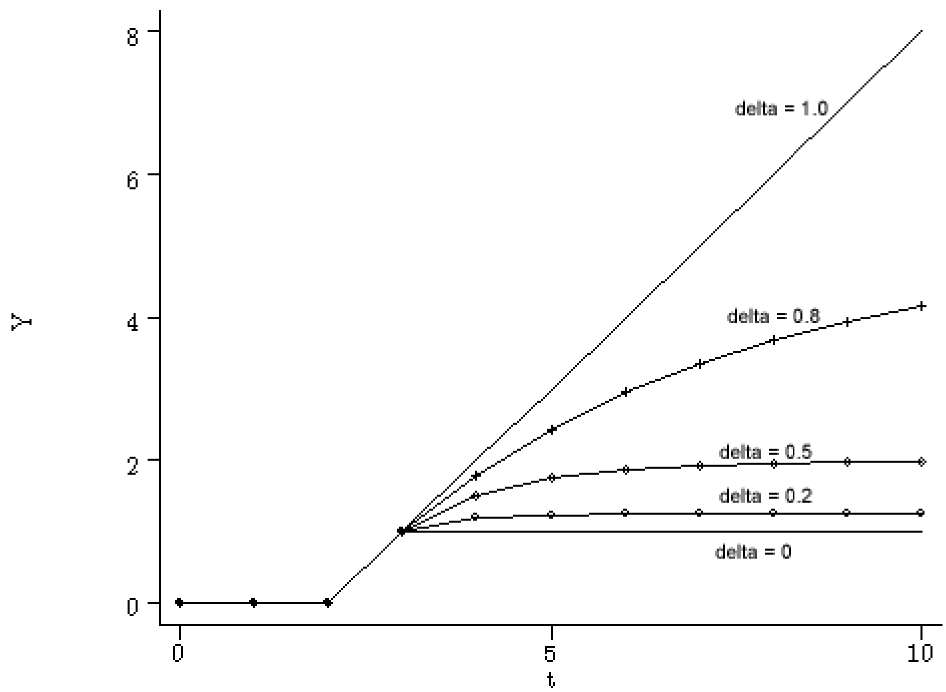
$$Y_t = \omega X_t + N_t \tag{12}$$

such that the value of Y^*_t is simply ω when $X_t = 1$ and 0 otherwise. In the first-order model, following equation (9), we have:

$$Y^*_t = \omega \sum_{j=0}^{T} \delta^j X_t \tag{13}$$

[1]Note that one way to think of a pulse function is as a differenced shift.

Figure 1: First-Order Impulse Response, $\omega = 1$, for Varying δs



but the response of this function to the pulse is quite different. At time $t+1$, the result is the same (that is, $Y^*_{t+1} = \omega$). At time $T + 2$, however, we have:

$$\begin{aligned} Y^*_{t+2} &= \delta Y^*_{t+1} + \omega X_{t+2} \\ &= \delta\omega + \omega(0) \\ &= \delta\omega \end{aligned}$$

and at time $t + 3$:

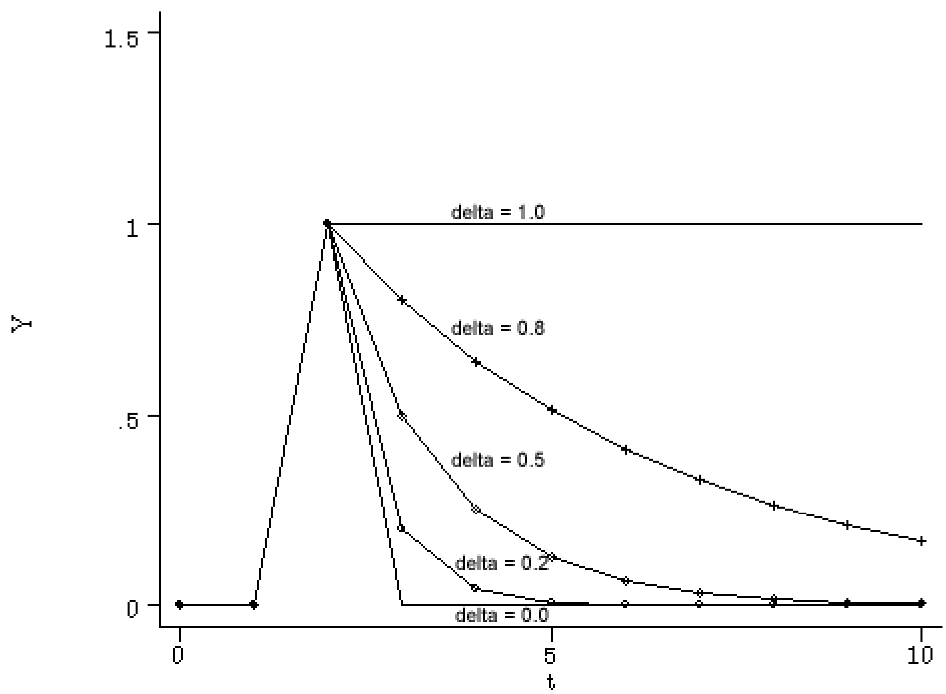
$$\begin{aligned} Y^*_{t+3} &= \delta Y^*_{t+2} + \omega X_{t+3} \\ &= \delta(\delta\omega) + \omega(0) \\ &= \delta^2\omega \end{aligned}$$

In general, for the "pulse" model:

$$Y^*_{t+k} = \delta^k \omega$$

This means that the effect of the "pulse" is now dying out at a geometric rate determined by the value of δ. Not surprisingly, $\delta = 1$ corresponds to a model in which the effect of the pulse persists permanently (i.e., it is equivalent in effect to a zero-order shift model). Conversely, $\delta = 0$ yields a model in which the effect of the shock is only felt in the period in which it occurs. Figure 2 displays the values if Y^*_t over time for a model in which $\omega = 1$ and a single pulse occurs at $t = 3$, for differing values of δ.

Figure 2: First-Order Impulse Response, $\omega = 1$, for Varying δs, Single Pulse

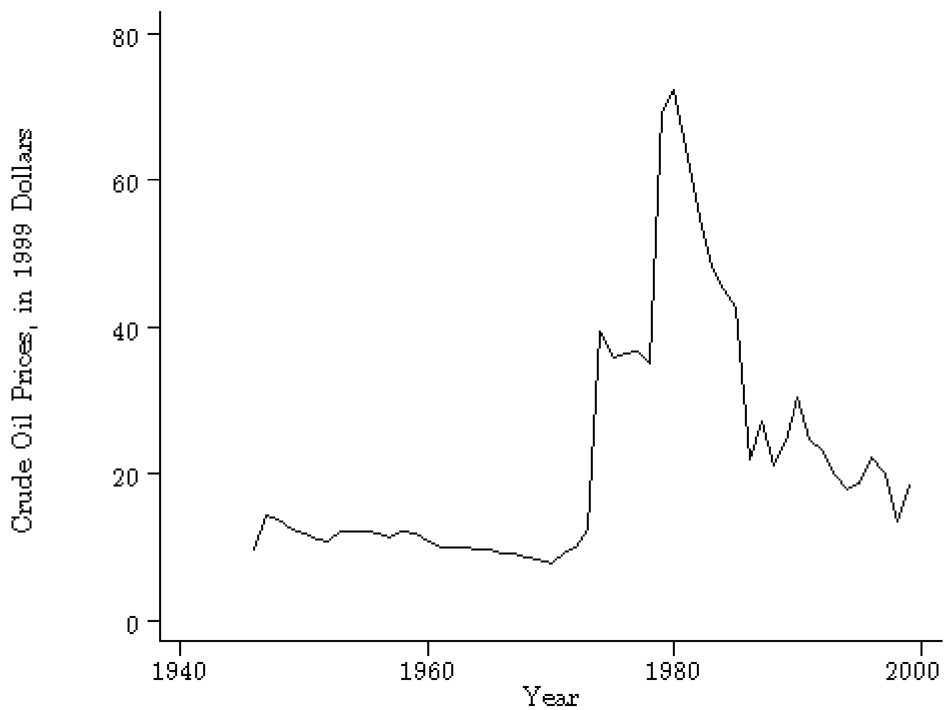


4 An Example: Oil Prices

Please Note: I'm still trying to figure out how to estimate transfer function models in Stata, so this part of the notes may change at any time.

As an example, we'll consider average annual U.S. crude oil prices, between 1946 and 1999 inclusive ($T = 54$). The plot of these prices is in Figure 3.

Figure 3: Average Annual U.S. Crude Oil Prices, in 1999 Dollars



I did some ARIMA diagnostics, and determined that the logged series is I(1), and that its first difference is a stationary ARMA(1,0) process (just trust me on that one, OK?...). So, we'll be working with the first difference of logged oil prices. The idea is to examine the effect of the 1973 OPEC oil price shock on prices overall.

One thing we can do is estimate different models which reflect different ideas about how the effect of the shock occurred. For example, a zero-order model with a shift variable implies an *abrupt, permanent* change in prices, while a first-order model implies a gradual, permanent change. Similarly, a zero-order model with X as a "pulse" indicates an abrupt shift, followed by an equally-abrupt dissipation of the effect, while a first-order pulse model allows for an abrupt shift followed by a gradual decline in effect. This suggests that one strategy is to estimate several such models, and see which one best reflects the data.

As noted above, the zero-order model posits an abrupt shift; here, we can write that model as:

$$\Delta PRICE_t = \alpha + \omega SHIFT_t + u_{ARMA(1,0)} \tag{14}$$

for the model with a shift in X and

$$\Delta PRICE_t = \alpha + \omega PULSE_t + u_{ARMA(1,0)} \tag{15}$$

for the model with X as a one-period "pulse". In Stata, this model is estimated using the -arima- command:

```
. arima logprix shift73, arima(1,1,0) hessian

. arima logprix pulse73, arima(1,1,0) hessian
```

For the first-order models, recall that the systematic component of the model is:

$$Y_t^* = \alpha + \delta Y_{t-1}^* + \omega X_t$$

equation
that is,

$$\Delta PRICE_t = \alpha + \delta \Delta PRICE_{t-1} + \omega SHIFT_t + u_{ARMA(1,0)} \tag{17}$$

for the shift model and

$$\Delta PRICE_t = \alpha + \delta \Delta PRICE_{t-1} + \omega PULSE_t + u_{ARMA(1,0)} \tag{18}$$

for the pulse model. The latter two are estimated in Stata by:

```
. arima logprix LD_logpr shift73, arima(1,1,0) hessian

. arima logprix LD_logpr pulse73, arima(1,1,0) hessian
```

The results of these models are presented in the table below:

Parameter	Zero-order Shift	Zero-order Pulse	First-order Shift	First-order Pulse
$\hat{\alpha}$	-0.01	0.01	-0.02	0.01
(s.e.)	(0.02)	(0.03)	(0.02)	(0.03)
$\hat{\omega}_{Shift}$	1.19	-	1.20	-
(s.e.)	(0.19)		(0.18)	
$\hat{\omega}_{Pulse}$	-	0.61	-	0.64
(s.e.)		(0.14)		(0.15)
$\hat{\delta}$	-	-	-0.02	0.05
(s.e.)			(0.10)	(0.10)
$\hat{\rho}$	-0.14	0.10	-0.12	0.06
(s.e.)	(0.15)	(0.14)	(0.17)	(0.17)

Several things are interesting about these results:

- Both he shift and the pulse results are statistically significant and positive, indicating that the 1973 shock led directly to a rise in oil prices.
- Neither δ parameter is statistically distinguishable from zero; this indicates that the effects of the shock are more-or-less instantaneous, and die off quickly as well. That is, the effect neither occurs gradually, as would be the case if $\delta > 0$ in the shift model, nor does it die off slowly, as would be the case for $\delta > 0$ in the pulse model.

Next time: Distributed Lag Models...