Causal Inference, Time Series and Economic History

3. Single-equation Models

Jason Lennard

Overview

- Distributed lag models
- Autoregressive distributed lag models
- Local projections
- *Class discussion paper:* Kelly, M. and Ó Gráda, C., 'Numerare est errare: Agricultural output and food supply in England before and during the Industrial Revolution', *Journal of Economic History*, 73 (2013), pp. 1132-63.

Dynamic Models

- The time series regression model is not always appropriate:

$$y_t = \alpha + \beta x_t + u_t \tag{1}$$

- One reason is that time series are sometimes persistent
- Another is that some independent variables affect the dependent variable with a lag

Lags in Economics

- The effect of an independent variable on a dependent variable may not be instantaneous but delayed
- Consider the response of inflation to a change in monetary policy
- One potential mechanism:

1. Central bank increases rate charged to commercial banks
2. Commercial banks increase deposit and borrowing rates
3. Saving more attractive, borrowing less attractive
4. Consumption and aggregate demand falls
5. Inflation falls

Lags in Economics

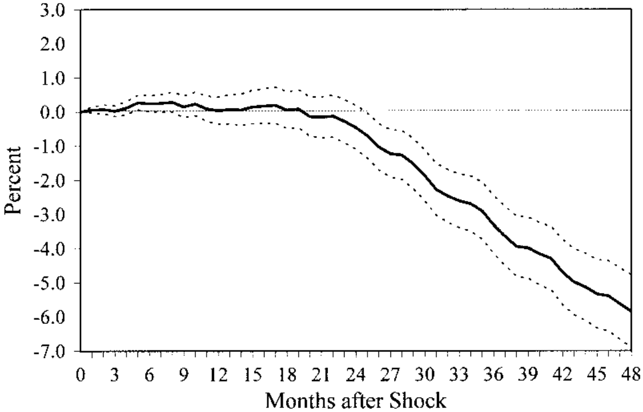

Figure 1. The Effect of Monetary Policy on the Price Level
Source: Romer and Romer (2004)

Dynamic Models

Three commonly-used types of dynamic models:

1. Finite distributed lag
2. Autoregressive distributed lag
3. Local projections

Finite Distributed Lag Model (FDL): Bivariate

- The dependent variable is modeled as a function of current and past values of the independent variables
- Generalized $FDL(Q)$:

$$y_t = \alpha + \sum_{q=0}^{Q} \beta_q x_{t-q} + u_t$$

- With 1 lag, $FDL(1)$:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

- With 2 lags, $FDL(2)$:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

Finite Distributed Lag Model (FDL): Multivariate

- Generalized $FDL(Q, R)$ with 2 independent variables:

$$y_t = \alpha + \sum_{q=0}^{Q} \beta_q x_{t-q} + \sum_{r=0}^{R} \gamma_r z_{t-r} + u_t$$

- Example with 2 lags each, $FDL(2, 2)$:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \gamma_0 z_t + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + u_t$$

Finite Distributed Lag Model (FDL): Interpretation

- Impact multiplier = β_0
- Intermediate multiplier at horizon $h = \beta_0 + \ldots + \beta_h$
- Total long-run effect = $\beta_0 + \ldots + \beta_Q$
    - In other words, the long-run effect of a change in x

Autoregressive Distributed Lag Model (ARDL):Bivariate

- The dependent variable is modeled as a function of its own past values and current and past values of the independent variables
- Generalized $ARDL(P, Q)$:
  - P = number of lags of y and Q = number of lags of x

$$y_t = \alpha + \sum_{p=1}^{P} \delta_p y_{t-p} + \sum_{q=0}^{Q} \beta_q x_{t-q} + u_t$$

- With 1 lag each, $ARDL(1, 1)$:

$$y_t = \alpha + \delta_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

- With 2 lags each $ARDL(2, 2)$:

$$y_t = \alpha + \delta_1 y_{t-1} + \delta_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

Autoregressive Distributed Lag Model (ARDL): Multivariate

- Generalized $ARDL(P, Q, R)$ with 3 independent variables:

$$y_t = \alpha + \sum_{p=1}^{P} \delta_p y_{t-p} + \sum_{q=0}^{Q} \beta_q x_{t-q} + \sum_{r=0}^{R} \gamma_r z_{t-r} + u_t$$

- Example with 2 lags each, $ARDL(2, 2, 2)$:

$$y_t = \alpha + \delta_1 y_{t-1} + \delta_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \gamma_0 z_t + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + u_t$$

Autoregressive Distributed Lag Model (ARDL): Interpretation

- Impact multiplier = β_0
- Intermediate multiplier at horizon h for $h < p$= $\frac{\beta_0+\ldots+\beta_h}{1-\delta_1-\ldots-\delta_h}$
- Intermediate multiplier at horizon h for $h \geq p$= $\frac{\beta_0+\ldots+\beta_h}{1-\delta_1-\ldots-\delta_P}$
- Total long-run effect = $\frac{\beta_0+\ldots+\beta_Q}{1-\delta_1-\ldots-\delta_P}$
  - In other words, the long-run effect of a change in x
- See the appendix for the derivation of these multipliers

How Many Lags to Include? 3 Approaches

1. Average economic regression (AER)
   - Start with simplest model and build up until you have a well-behaved model
   - i.e., correct signs, significant coefficients, a good fit and white-noise residuals
   - However, this approach potentially suffers from omitted variable bias from the start
2. The Hendry approach
   - Start with the most general model and work down
   - Difficult to know when to stop
3. Use information criteria

Information Criteria

- Information criteria (IC) are measures of a model's relative goodness of fit
- Two main ICs:

1. Akaike Information Criterion (AIC) by Akaike (1974):

$$AIC = \left(\frac{SSR}{n}\right) e^{2k/n}$$

2. Schwarz Bayesian Information Criterion (SBIC) by Schwarz (1978):

$$SBIC = \left(\frac{SSR}{n}\right) n^{k/n}$$

- where SSR is the sum of squared residuals, n is the number of observations, $e \approx 2.71828$ (base of the natural log) and k is the number of parameters (including the intercept)
- Trade off:
  - ICs decrease with the goodness of fit (higher R^2 and lower SSR)
  - ICs increase with the number of parameters (k)

Information Criteria

- Step 1: Estimate different models (lag length may vary by variable)
- Step 2: Calculate information criteria
- Step 3: Select model that *minimizes* a given information criteria

Notes:

- ICs may give conflicting results
- SBIC penalizes complexity more heavily
- Stata uses slightly different formulas to those in the previous slide, but the intuition is the same

Local Projections

- The local projections model was developed by Jordà (2005)
- It is something between an FDL/ARDL and VAR model
- A simple example with no controls is:

$$y_{t+h} = \alpha_h + \beta_h x_t + u_{t+h}$$

- The coefficient β_h gives the impulse response of y at time $t+h$ to a shock in the independent variable at time t. Thus, one constructs the impulse responses as a sequence of the β_h's estimated in a series of single regressions for each horizon

Local Projections

- For the impulse response at $h = 0$, we would estimate:

$$y_t = \alpha_0 + \beta_0 x_t + u_t$$

- For the impulse response at $h = 1$, we would estimate:

$$y_{t+1} = \alpha_1 + \beta_1 x_t + u_{t+1}$$

- For the impulse response at $h = 2$, we would estimate:

$$y_{t+2} = \alpha_2 + \beta_2 x_t + u_{t+2}$$

- And so on ...

Local Projections: Standard Errors

- The residuals will be serially correlated because they are a moving average of the forecast errors from t to $t + h$
- This violates the Gauss-Markov assumption of no serial correlation of errors
- Therefore, best practice is to include y_{t-1} as a control variable, which is known as lag-augmented local projections (Montiel Olea and Plagborg-Møller, 2021)
- Lag-augmented local projections are also valid for non-stationary time series

Local Projections: Pros and Cons

Strengths

- Impulse response functions calculated directly
- Standard errors calculated directly
- Robust to misspecification of the data generating process
- Very flexible

Weaknesses

- Can be less efficient relative to other models
- Can be more variable relative to other models
- Multiple (h) residuals to monitor

Local Projections: Flexibility

- Differential responses by regimes (expansions (E_t) and recessions (R_t), for example):

$$y_{t+h} = E_t(\alpha_h^E + \beta_h^E x_t) + R_t(\alpha_h^R + \beta_h^R x_t) + u_{t+h}$$

- Differential responses by sign of shocks (positive (+) and negative (−)):

$$y_{t+h} = \alpha_h + \beta_h^+ x_t^+ + \beta_h^- x_t^- + u_{t+h}$$

- Differential responses by size of shocks (small (s) and large (l)):

$$y_{t+h} = \alpha_h + \beta_h^s x_t + \beta_h^l x_t^2 + u_{t+h}$$

Dynamic Models: Excluding x_t

- As with the static time series regression model, dynamic models must satisfy the Gauss-Markov assumptions (including the zero conditional mean assumption) to be the Best Linear Unbiased Estimator (BLUE)
- But many relationships of interest suffer from reverse causality
- In order to overcome the issue of reverse causality, a common approach is to exclude the contemporaenous value of an independent variable (e.g. x_t)
- The idea is that whereas there may be reverse causality between y_t and x_t, only x_{t-1} can affect y_t but y_t can't affect x_{t-1}
- Reed (2015) shows that this is a mistake and that OLS is still biased

Class Discussion Paper: Kelly and Ó Gráda (2013)

Research Question

- To compare various estimates of agricultural output for England between 1270 and 1800
- To construct a compromise estimate for 1750/70 and 1800

Class Discussion Paper: Kelly and Ó Gráda (2013)

- Crude death rate (Wrigley and Schofield, 1981)
- Real wage (Clark, 2010)
- Real agricultural output per head (Broadberry et al., 2015)
- Real GDP per head (Broadberry et al., 2015)
- Sample: 1546-99, 1600-49, 1650-99, 1700-49, 1750-99

Class Discussion Paper: Kelly and Ó Gráda (2013)

Model

$$\Delta cdr_t = \alpha + \sum_{p=1}^{4} \delta_p \Delta cdr_{t-p} + \sum_{q=0}^{4} \beta_q \Delta i_{t-q} + u_t \tag{2}$$

where:
Δcdr_t is the change in the log crude death rate
Δi_t is the change in the log of a measure of real income

Class Discussion Paper: Kelly and Ó Gráda (2013)

Results

TABLE 6

SHORT-RUN RESPONSE OF THE DEATH RATE TO ANNUAL VARIATIONS IN REAL INCOME

Period and Lags	Wage	Agricultural Output per Head	GDP per Head
1546–1599			
0	–0.131	0.054	0.267
1	–0.273	–0.266 *	–0.518 *
2	–0.572 **	–0.383 *	–0.868 **
3	–0.197	–0.400 **	–0.999 **
4	–0.331 *	0.136	–0.132
1600–1649			
0	0.223	0.063	0.155
1	–0.518 **	0.065	0.165
2	–0.049	–0.189	–0.371
3	–0.085	–0.049	–0.119
4	–0.112	0.070	0.101
1650–1699			
0	–0.024	–0.020	0.038
1	–0.000	0.040	0.069
2	–0.110	0.087	0.194
3	0.206	0.045	0.123
4	0.194	0.041	0.202
1700–1749			
0	–0.017	0.176	–0.077
1	–0.522 **	–0.076	–0.196
2	–0.180	–0.305 **	–0.660 **
3	–0.199	–0.139	0.087
4	0.036	0.261	0.405
1750–1799			
0	0.151	0.145	0.468
1	0.214	–0.048	–0.004
2	0.187	0.289 **	0.619 *
3	0.351 **	0.198 **	0.259
4	0.534 **	0.110	0.130

* = Significant at 10 percent.
** = Significant at 5 percent.
Note: Estimated using robust regressions.

Next Class

- *Class discussion paper:* Nicolini, E. A., ’Was Malthus right? A VAR analysis of economic and demographic interactions in pre-industrial England’, *European Review of Economic History*, 11 (2007), pp. 99-121.

Further Reading

- Stock and Watson, *Introduction to Econometrics*, chapters 15 (pp. 578-82) and 16

- Koudijs, P., 'The boats that did not sail: Asset price volatility in a natural experiment', *Journal of Finance*, 71 (2016), pp. 1185-226
- Steinwender, C., 'Real effects of information frictions: When the States and the Kingdom became United', *American Economic Review*, 108 (2018), pp. 657-96
- Fochesato, M., 'Origins of Europe's north-south divide: Population changes, real wages and the "little divergence" in early modern Europe', *Explorations in Economic History*, 70 (2018), pp. 91-131

Appendix

- In order to derive the $ARDL(P, Q)$ multipliers, let's begin with the simplest form, an $ARDL(1, 1)$ model:

$$y_t = \alpha + \delta_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

- Take expected values (note that the expectation of a variable is the same at each point in time, therefore: $Ey_t = Ey_{t-1}$ and so on):

$$Ey_t = \alpha + \delta_1 Ey_t + \beta_0 Ex_t + \beta_1 Ex_t + u_t$$

- Subract $\delta_1 Ey_t$ from both sides:

$$Ey_t - \delta_1 Ey_t = \alpha + \beta_0 Ex_t + \beta_1 Ex_t + u_t$$

Appendix

- Collect terms:

$$(1 - \delta_1) \, Ey_t = \alpha + (\beta_0 + \beta_1) \, Ex_t + u_t$$

- Divide both sides by $1 - \delta_1$:

$$Ey_t = \frac{\alpha}{1 - \delta_1} + \frac{\beta_0 + \beta_1}{1 - \delta_1} Ex_t + u_t$$

- The partial derivative of the above with respect to the independent variable is:

$$\frac{\partial Ey_t}{\partial Ex_t} = \frac{\beta_0 + \beta_1}{1 - \delta_1}$$

Appendix

- Let's move on to a more complex form, an $ARDL(2,2)$ model:

$$y_t = \alpha + \delta_1 y_{t-1} + \delta_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

- Take expected values:

$$Ey_t = \alpha + \delta_1 Ey_t + \delta_2 Ey_t + \beta_0 Ex_t + \beta_1 Ex_t + \beta_2 Ex_t + u_t$$

- Subract $\delta_1 Ey_t + \delta_2 Ey_t$ from both sides:

$$Ey_t - \delta_1 Ey_t - \delta_2 Ey_t = \alpha + \beta_0 Ex_t + \beta_1 Ex_t + \beta_2 Ex_t + u_t$$

Appendix

- Collect terms:

$$(1 - \delta_1 - \delta_2)\, Ey_t = \alpha + (\beta_0 + \beta_1 + \beta_2)\, Ex_t + u_t$$

- Divide both sides by $1 - \delta_1 - \delta_2$:

$$Ey_t = \frac{\alpha}{1 - \delta_1 - \delta_2} + \frac{\beta_0 + \beta_1 + \beta_2}{1 - \delta_1 - \delta_2} Ex_t + u_t$$

- The partial derivative of the above with respect to the independent variable is:

$$\frac{\partial Ey_t}{\partial Ex_t} = \frac{\beta_0 + \beta_1 + \beta_2}{1 - \delta_1 - \delta_2}$$

- Use this approach to derive the mutliplier for any ARDL model