Time-series–Cross Section Data: Time-Series Issues

Nathaniel Beck (with a little help from his friends, particularly Jonathan Katz

Department of Politics, NYU, New York, NY 10012, nathaniel.beck@nyu.edu

Prepared for Workshop, UT Dallas, Nov. 11, 2006

Outline

Focus on dynamics

- Looking today at dynamic issues
- modeling issues related to time
- Assume from yesterday have dealt with cross-sectional issues correctly
- Dynamics in TSCS data is very similar to dynamics in time series data

Correlated errors

- "Nuisance" assumption - errors are serially correlated
- for yearly data usually AR, could easily generalize/test)

$$\epsilon_{i,t} = \rho \epsilon_{i,t-1} + \mu_{i,t} \quad (1)$$

  where the μ are independently distributed across time.
- (Aside: Old fashioned silliness - assumed that the temporal dependence of the errors could be modeled as a *unit-specific* AR1) process

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \mu_{i,t} \quad (2)$$

- Simulations show that it is better to assume $\rho_i = \rho$

GLS estimation

- Correct for this serial correlation in the usual (GLS) manner
- First run OLS
- Compute the serial correlation of the residuals
- (that is, regress the residuals on the lagged residuals and take the coefficient on the lagged residual as $\hat{\rho}$.)
- Then transform by subtracting $\hat{\rho}$ of the prior observation from the current one
- and run OLS on the transformed observations.
- Can do better by transforming rather than dropping first obs (Prais-Winsten)

Testing

- Test for serially correlated errors (with or without a lagged dependent variable)
- via the TSCS analogue of the standard Lagrange multiplier test
- Run OLS
- compute residuals
- regress the residuals on all the independent variables (including the lagged dependent variable if present) and the lagged residual
- If the coefficient on the lagged residual is significant (with the usual t-test), we can reject the null of independent errors.
- or TR^2 is χ^2_1
- Can do similar for higher order process, but with yearly data may not be needed (but why not test?)

Lagged Dependent Variable

- Just as with any time series, we could also model dynamics with a lagged dependent variable
- they make the dynamics part of the model, not just a nuisance
- there is seldom any reason to prefer serially correlated errors to a lagged dependent variable
- the LDV model assumes that the effects of all variables, measured and unmeasured (errors?) have impacts that die out exponentially
- whereas the AR1 error model assumes that the measured variables have only immediate impact
- but the unmeasured variables have impacts which die out exponentially
- lagged dependent variables usually simple to estimate and interpret (even if testing indicates SMALL remaining serial correlation)
- Can test for remaining serial correlation easily using LM test from previous slide
- If do not reject independence, need more complex model

LDV vs AR1 errors

The LDV model is (using ν to denote iid errors)

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \phi y_{i,t} + \nu_{i,t} \tag{3}$$

while the AR1 error model is

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \nu_{i,t} + \rho\epsilon_{i,t-1} \tag{4}$$

since

$$\epsilon_{i,t} = \nu_{i,t} + \rho\epsilon_{i,t-1} \tag{5}$$

and hence

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \rho y_{i,t-1} - \mathbf{x}_{i,t-1}\boldsymbol{\beta}\rho + \nu_{i,t} \tag{6}$$

so both LDV and AR1 errors are special cases of ADL model

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \rho y_{i,t-1} - \mathbf{x}_{i,t-1}\gamma + \nu_{i,t} \tag{7}$$

Choosing which spec

- Thus, following Hendry's advice to test from general to specific
- start with the ADL setup
- and then test the null that either $\gamma = 0$ or $\gamma = -\beta\rho$
- ADL equivalent to the single equation DHSY "error correction model" model

$$\Delta y_{i,t} = \Delta x_{i,t}\boldsymbol{\beta} - \phi(y_{t-1} - x_{t-1}\gamma) + \nu_t. \tag{8}$$

- with coefficients suitably interpreted

Back to specification

- To see how the specifications differ, it is easiest to look at the impact of a permanent one unit level change in x.
- AR1 model: y instantaneously adjusts, increasing by β_{ar1}.
- LDV model: y adjusts to the change in x geometrically; the initial impact of the change is β_{ldv}, with steady-state impact $\frac{\beta_{ldv}}{1-\phi}$
- The ADL model is more complicated; initially y responds to the level shift in x by increasing β_{adl} units, with the long run change in y being $\frac{\beta_{adl}+\gamma_{adl}}{1-\phi}$.
- If x is similar to the variables that make up the error process, one might expect the LDV model to be suitable.
- If x represents a change in regime that we expect to have an immediate one time impact, the AR1 formulation seems plausible.
- If the data are willing to speak, the ADL or DHSY model is a good compromise between the two
- and allows for testing the specializations

When does all this matter?

- If ϕ is relatively small, then shocks quickly die out in AR1 error model
- In a model with AR1 errors, the impact of any of the independent variables is felt only immediately.
- In the LDV model, where the long run impact of any given x is $\frac{\beta_x}{(1-\phi)}$, most of the long run effect is seen quickly if ϕ is small
- In this situation, the LDV and AR1 error specification will appear quite similar.
- As ϕ gets larger, the difference between the two models get larger.
- As ϕ gets towards one neither approach is correct
- we move into the world of non-stationary time series.
- One example with a small ϕ.

Garrett model of economic growth in 14 OECD nations, 1966–1990 (with fixed effects)

	LDV		AR1 Errors		AR1	
Variable	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE
GDP_{-1}	.16	.07			.15	.08
DEMAND	.72	.16	.70	.18	.70	.17
CORP	−.72	.60	−.78	.70	−.92	1.16
LEFT	−.77	.34	−.88	.38	−.63	.53
LEFTxCORP	.27	.14	.31	.15	.19	.20
PER6673	1.64	.37	1.98	.41	1.65	.37
CONSTANT	2.76	1.77	3.42	2.08	2.45	1.82
$DEMAND_{-1}$					.07	.18
$CORP_{-1}$					.23	1.09
$LEFT_{-1}$					−.23	.53
$LEFTxCORP_{-1}$					.14	.20
ϕ			.15	.08		
N	336		336		336	
BIC	4.3846		4.3945		4.4504	
SSE	1116.131		1127.226		1112.210	

Modeling non-stationary TSCS data

- While estimation of TSCS models with unit root data is just beginning to be studied
- Experience from single time series analysis tells us that we cannot simply use stationary methods
- Example: Huber and Stephens analysis of the determinants of social security spending
- 16 OECD countries in the post-World War II period (26 years)
- *SSBEN* is very smooth:

$$SSBEN_t = 1.003 SSBEN_{t-1} + \nu_t$$

  (se = .008)
- In the absence of co-integration, we can only explain short run changes in *SSBEN* by short run changes in the independent variable.
- Short run changes in SSBEN NOT explained by short run changes in political variables
- Conclusion very different from Huber and Stephens
- Try error correction if series appear to be cointegrated

Summing up

- Choose dyn spec for TSCS as you would for TS
- But remember the TSCS data have (usually) only low frequency (yearly) data
- Testing as in TS - LM test
- LDV's simplify, can test if appropriate, not atheoretical
- Worry about non-stationarity - less clear what to do

Models with Binary Dependent Variables

- So far have assumed continuous dependent variable.
- Now assume binary dependent variable (BTSCS)
- In IR much data is dyad year, that is, did country A and B fight in 1972
- or did A attack B in 1972 (directed dyad)
- But could also be did a country have an independent central bank in 1982
- May extend to more complicated limited DVs, but not clear
- DO NOT ASSUME WHAT IS SAID HERE EXTENDS!
- Many just ignore TS issues and just do ORDINARY logit
- (Logit and Probit pretty interchangeable here)
- Because N is easier to notate, much notation uses probit
- Estimation with logit more common item Remember that BTSCS data is not binary panel data
- Lots of issues for binary panel data (and possibilities) do not apply here

What goes wrong with BTSCS data

- Just as in standard TSCS, observations from same unit are not independent
- Often easiest to use a standard latent variables setup to think about issues.

$$y^*_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} \tag{9a}$$

$$y_{i,t} = 1 \text{ if } y_{i,t} > 0 \tag{9b}$$

where

$$x_{i,t} = \rho_x x_{i,t-1} + \nu_{i,t} \tag{10}$$

$$\epsilon_{i,t} = \rho_\epsilon \epsilon_{i,t-1} + \mu_{i,t} \tag{11}$$

- Ordinary probit is consistent here (Poirier and Ruud)
- se's not accurate; Simulation showed that with very high ρs errors may be off by 50%
- Huber grouped standard errors appeared quite accurate

Event History

- Much BTSCS data has long strings on 0's with few 1's
- E.g. conflict data (perhaps 3% of obs are 1's, with generous def of 1
- looks like event history data
- with each 1 marking a failure
- and the time between 1's, that is, the number of 0's, being the time until or between failures
- Make sure the number of 1's is not large, so we have a good distribution of failure times. item We could thus simply convert the binary TSCS data into event history data
- and use standard (Cox or parametric)

Helpful event history ideas

- Thinking about binary TSCS data as event history data helps
- The simple probit/logit approach is equivalent to the assumption of no duration dependence in event history analysis.
- normally we test for duration dependence,
- While it looks like we have $N \times T$ binary TSCS observations
- this is the same at N duration observations
- While we think of probit/logit as having troubles with rare events
- such rare events are the lifeblood of event history analysis
- Binary TSCS data in event form may have more than one event per unit.

Onset vs. Incidence, repeated events

- In the event history approach, we model strings of zeros which end with a 1
- that is, the probability of a transition from 0 to 1,
- or what is known in the medical world as ONSET (of a disease).
- We are NOT modelling the length of strings of 1's.
- The total proportion of 1's is called INCIDENCE. (Proportion of all people having the disease.)
- We could model length of time of string of 1's (spells of disease, war)
- THIS TURNS OUT TO BE A CRITICAL ISSUE (return to below)
- In the probit/logit setup, we assume that second and subsequent events can be modeled just like first events
- events history modelers realize that life is more complex
- Solution???? Kludge??? Counter for how many prior events

Discrete Time Duration Models

- Sticking with one event per unit for now
- Can model the time until an event in binary TSCS data by a discrete time duration model.
- Assume time measured in equal discrete intervals, $0, 1, \ldots, t, \ldots$ (years)
- We only observe whether someone dies in the interval $(t-1, t]$, (open on the left, closed on the right)
- and will assume this is a death at t
- Then we need discrete time analogues of the survivor and hazard function.
- The survivor function will simply be a step function, with steps at $1, 2, \ldots, t, \ldots$.
- Let y_i be the duration for the i'th unit;
- y_i is a discrete random variable, with support at the positive integers.

Discrete time maths

- Using standard notation we have

$$\begin{aligned} S(t) &= P(y > t) \\ &= P(y > t | y > t-1) P(y > t-1) \\ &= \prod_{i=0}^{t-1} P(y > t-i | y > t-i-1) \end{aligned} \tag{12}$$

where $S(0) = P(y > 0) = 1$. We still have

$$F(t) = 1 - S(t). \tag{13}$$

- Since F is discrete, we have an associated discrete density with support on the positive integers,

$$f(t) = F(t) - F(t-1). \tag{14}$$

Here the density is a probability; it is the unconditional probability of dying in the interval $(t-1, t]$.

Discrete time hazard

- Define the discrete hazard analogously to the continuous time hazard
- though simpler, since is now just a conditional probability
- $h(t)$ is the hazard of dying at time t (or in the interval $(t-1, t]$ given that one survived until $t-1$
- that is, probability of death in that interval given alive at start of interval

$$h(t) = \frac{f(t)}{S(t-1)}. \tag{15}$$

Since $1 - h(t)$ is the conditional probability of surviving at t given survival through $t-1$, substituting in Equation 12, we get

$$S(t) = \prod_{i=0}^{t-1} [1 - h(t-i)] \tag{16}$$

Estimation of discrete duration models via logit

- Estimating a BTSCS with dependent variable $y_{i,t}$ being whether unit i failed in the interval $(t-1, t]$
- (by probit, logit or any other binary model, will use "logit" as generic)
- as a function of covariates $\mathbf{x}_{i,t}$
- we are estimating a model for $h(t)$
- If the dependent variable is scored as 1 for non-failure, then we have a model for $1 - h(t)$
- Estimating via ordinary logit is assuming the hazard rate is time invariant (that is,

$$h_{i,t} = h(\mathbf{x}_{i,t}) \tag{17}$$

- To allow for duration dependence estimate we would need to estimate a binary model (with $y_{i,t}$ being one for the failure of unit i in the interval $(t-1, t]$ which has

$$h_{i,t} = h_t(\mathbf{x}_{i,t}). \tag{18}$$

Separate time counter

- Compromise to allow for different intercepts at each time point

$$h_{i,t} = a_t + h(\mathbf{x}_{i,t}) \tag{19}$$

- which would be estimated by putting in a period dummy in the logit.
- Or could use any $s(t)$ one liked, if flexible
- Remember, *the time variable is time since the last "event," not the particular period of the observation*.
- Sometimes the time dummies indicate that we don't need to include time in the specification (using standard tests on the coefficients of all the time variables).
- At that point we can assume no duration dependence and use ordinary logit.

A more formal derivation

- Start with a continuous time Cox proportional hazards model, so

$$h_i(t) = h_0(t)e^{\mathbf{x}_{i,t}\beta}. \tag{20}$$

- Letting $S(t)$ be the probability of surviving beyond t, by the math of hazard rates we have

$$S(t) = \exp(-\int_0^t h(\tau)d\tau). \tag{21}$$

- Only observe whether or not an event occurred between time $t_k - 1$ and t_k

More maths

- so model $P(y_{i,t_k} = 1)$

$$\begin{aligned} P(y_{i,t_k} = 1) &= 1 - \exp(-\int_{t_k-1}^{t_k} h_i(\tau)d\tau) \\ &= 1 - \exp(-\int_{t_k-1}^{t_k} e^{\mathbf{x}_{i,t_k}\beta} h_0(\tau)d\tau) \\ &= 1 - \exp(-e^{\mathbf{x}_{i,t_k}\beta}\int_{t_k-1}^{t_k} h_0(\tau)d\tau) \end{aligned}$$

Even more maths

- Since the baseline hazard is unspecified
- Can just treat the integral of the baseline hazard as an unknown constant
- Defining

$$\alpha_{t_k} = \int_{t_k-1}^{t_k} h_0(\tau)d\tau \text{ and}$$
$$\kappa_{t_k} = \log(\alpha_{t_k})$$

- we then have

$$P(y_{i,t_k} = 1) = 1 - \exp(-e^{\mathbf{x}_{i,t_k}\beta}\alpha_{t_k})$$
$$= 1 - \exp(-e^{\mathbf{x}_{i,t_k}\beta+\kappa_{t_k}})$$

- This is exactly a binary dependent variable model with a cloglog link and the κ (time dummy) terms added.
- There is almost no difference in practice between estimated a logit model and a cloglog model

Further Complications

- Thinking about the war data as event history data leads to thinking about other issues.
- Dyads can fight a number of wars.
- Durations of second events may follow different process than for first events
- This is difficult to model
- One solution is to add a variable to the hazard function which counts the number of previous failures
- Another issue is modeling onset vs. incidence
- To understand we need a detour

Markov Transition Matrices

- A Markov process (first order) assumes that whether or not you are 0 or 1 at time t is a function only of where you were at time $t-1$ and covariates.
- Thus would estimate two different probits (or logits) depending on prior state

$$P(y_{i,t} = 1 | y_{i,t-1} = 0) = \text{Probit}(\mathbf{x}_{i,t}\boldsymbol{\beta}) \tag{24}$$
$$P(y_{i,t} = 1 | y_{i,t-1} = 1) = \text{Probit}(\mathbf{x}_{i,t}\boldsymbol{\alpha}) \tag{25}$$

which can be writen more compactly as

$$P(y_{i,t} = 1) = \text{Probit}(\mathbf{x}_{i,t}\boldsymbol{\beta} + y_{i,t-1}\mathbf{x}_{i,t}\boldsymbol{\gamma}) \tag{26}$$

where

$$\boldsymbol{\gamma} = \boldsymbol{\alpha} - \boldsymbol{\beta}. \tag{27}$$

- Can test the hypothesis that prior state does not matter by testing $\boldsymbol{\gamma} = \mathbf{0}$

Comparison of Markov and LDV model

- An alternative that some have suggested is to use a lagged dependent variable

$$Pr(y_{i,t} = 1 | y_{i,t-1} = 1) = \text{logit}(\mathbf{x}_{i,t}\boldsymbol{\beta} + \rho)$$

whereas

$$P(y_{i,t} = 1 | y_{i,t-1} = 0) = \text{logit}(\mathbf{x}_{i,t}\boldsymbol{\beta})$$

so the two logit equations are parallel (in the latent space).

- Thus the only thing that differs by prior state is the intercept, the effect of all the iv's on the dv is the same regardless of whether past state was 0 or 1.
- Note this is just a strong restriction on the Markov transition model
- and could be tested by first estimating the full Markov model and then testing the null that all coefficients (other than the intercept) are the same in both probits

The transition model and event history

- Note that Equation 24 is just the duration *independent* form of our event history methods.
- The event history methods as generalizing this subset logit.
- Note in event history we ONLY model transitions from 0 to 1
- ignoring 1 to 0 (or, the same, 1 staying 1)
- But if you believe the Markov story, a different model for lengths of spells of 0s and 1s
- Note: because dv is binary, so no measure of variance in the latent (scaled to be one)
- Running two probits separately literally identical to one interaction model
- Thus the model for duration of spells of 0 independent of lengths of spells of 1
- Model onset differently from incidence
- Likely different models for lengths of spells of peace and war
- And if not, should first test, not assume

Summing up

- BTSCS and event history data are same thing
- Do not do ordinary logit (unless tests indicate okay)
- Do not lump together transitions form 0 and 1
- Lagged dv not enough
- For each spell, ordinary logit with the time counters is fine
- Markov transition model not quite as good

Data Analysis

- All results reported are probits
- Missing data (not much) dropped
- Huber standard errors, clustered on dyad, used for MID analysis
- Other analysis uses ordinary se's

MID's 1951–1992 - Oneal and Russett

- Spells of disputes
  - 2048 dyad-years of disputes (spells of dispute with first year of peace)
  - Dyads may have multiple spells
  - 307 dyads
  - 636 different spells of dispute not right censored
  - 52 spells of disputes right censored
  - Typical spell length of disputes is short
  - Mean length of disputes is 3.7, median is 2,
- Spells of peace
  - 32376 dyad-years of peace
  - (new dispute number is new dispute, not second year of previous dispute)
  - 1094 dyads
  - 1570 spells of peace right censored
  - 1061 spells not right censored (ends in dispute)

MIDS

Variable	ALL b	ALL SE	PEACE b	PEACE SE	MID b	MID SE	BKT b	BKT SE
DEM	−.06	.01	−.06	.01	−.02	.02	−.05	.01
TRADE	−93.83	31.82	−25.68	14.3	−115.62	42.9	−24.72	14.00
MAJPOW	.74	.25	.62	.22	.00	.27	.59	.17
LCAPRAT	−.29	.06	−.14	.05	−.23	.07	−.27	.05
LDIST	−.38	.09	−.31	.07	−.12	.10	−.21	.07
CONTIG	1.59	.24	1.61	.22	−.08	.30	1.19	.19
ALLIES	−.85	0.19	−.51	.15	−.52	.21	−.44	.14
C	−.86	.69	−2.53	.58	1.55	.81	PY	
N	32727		29745		1360		31174	

Transitions from Dem to Aut and Vice-versa 1951-1990 - Przeworski, et al.

- 135 countries
- Spells of democracy
  - 1683 country years
  - 72 spells of democ
  - 38 spells of democ end in autoc
  - 34 spells of democ right censored
- Spells of autocracy
  - 2530 country yeas item 101 spells of autoc
  - 49 spells end in democ
  - 52 spells right censored

Democracy/Autocracy

	ALL		DEMLAG		From AUT		From DEM	
Variable	**b**	**SE**	**b**	**SE**	**b**	**SE**	**b**	**SE**
GDPLAG	.33	.01	.16	.02	.12	.03	.22	.05
GDPLAG%	−.57	.35	−.18	.69	−1.97	.85	3.96	1.38
DEMLAG			3.75	.10				
C	−1.32	.04	−2.41	.08	−2.30	.10	1.12	.14
N	4126		3991		2407		1584	
hline								

NOTE: While significant durdep in both transition equations, coefficients and se's change by under 10%