Higher-Dimension Markov Models

David Epstein   Sharyn O'Halloran

Columbia University

July 26, 2005

Iron Laws of the Political Science Profession

Tolstoy's Law of Journal Reviews

All good reviews are good in the same way; all bad reviews are bad in different ways.

Corollary

Never revise a paper for a new journal submission in response to comments by a bad reviewer at the previous journal.

The James Bond Law of Previous Literature

Never say never; you are sure to be wrong.

Iron Laws of the Political Science Profession

Tolstoy's Law of Journal Reviews

All good reviews are good in the same way; all bad reviews are bad in different ways.

Corollary

Never revise a paper for a new journal submission in response to comments by a bad reviewer at the previous journal.

The James Bond Law of Previous Literature

Never say never; you are sure to be wrong.

Iron Laws of the Political Science Profession

Tolstoy's Law of Journal Reviews

All good reviews are good in the same way; all bad reviews are bad in different ways.

Corollary

Never revise a paper for a new journal submission in response to comments by a bad reviewer at the previous journal.

The James Bond Law of Previous Literature

Never say never; you are sure to be wrong.

Basics of Markov Models

- Given: a system that can exist in any of a finite number of states in each period.
- Markov models estimate the probabilities π_{ab} of transitions from state a at time $t-1$ to state b at time t.

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_{N1} & \pi_{N2} & \cdots & \pi_{NN} \end{pmatrix}$$

Basics of Markov Models

- Given: a system that can exist in any of a finite number of states in each period.
- Markov models estimate the probabilities π_{ab} of transitions from state a at time $t-1$ to state b at time t.

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_{N1} & \pi_{N2} & \cdots & \pi_{NN} \end{pmatrix}$$

Basics of Markov Models

- Given: a system that can exist in any of a finite number of states in each period.
- Markov models estimate the probabilities π_{ab} of transitions from state a at time $t-1$ to state b at time t.

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_{N1} & \pi_{N2} & \dots & \pi_{NN} \end{pmatrix}$$

Typical Uses

1. International Relations: Friendly ↔ War
2. Election Challenger: Unchallenged ↔ Challenger
3. Transitions: Autocracy ↔ Democracy

- This is the simplest 2x2 case
- But one can imagine higher dimension categories...

Typical Uses

1. International Relations: Friendly $\leftrightarrow$ War
2. Election Challenger: Unchallenged $\leftrightarrow$ Challenger
3. Transitions: Autocracy $\leftrightarrow$ Democracy

- This is the simplest 2x2 case
- But one can imagine higher dimension categories...

Typical Uses

1. International Relations: Friendly $\leftrightarrow$ War
2. Election Challenger: Unchallenged $\leftrightarrow$ Challenger
3. Transitions: Autocracy $\leftrightarrow$ Democracy

- This is the simplest 2x2 case
- But one can imagine higher dimension categories...

Typical Uses

1. International Relations: Friendly $\leftrightarrow$ Tense $\leftrightarrow$ War
2. Election Challenger: None $\leftrightarrow$ Weak $\leftrightarrow$ Strong
3. Transitions: Autocracy $\leftrightarrow$ Partial Dem. $\leftrightarrow$ Democracy

- These higher dimension models have never been used in political science.

Typical Uses

1. International Relations: Friendly $\leftrightarrow$ Tense $\leftrightarrow$ War
2. Election Challenger: None $\leftrightarrow$ Weak $\leftrightarrow$ Strong
3. Transitions: Autocracy $\leftrightarrow$ Partial Dem. $\leftrightarrow$ Democracy

- These higher dimension models have never been used in political science.

Typical Uses

1. International Relations: Friendly $\leftrightarrow$ Tense $\leftrightarrow$ War
2. Election Challenger: None $\leftrightarrow$ Weak $\leftrightarrow$ Strong
3. Transitions: Autocracy $\leftrightarrow$ Partial Dem. $\leftrightarrow$ Democracy

- These higher dimension models have never been used in political science.

Typical Uses

1. International Relations: Friendly $\leftrightarrow$ Tense $\leftrightarrow$ War
2. Election Challenger: None $\leftrightarrow$ Weak $\leftrightarrow$ Strong
3. Transitions: Autocracy $\leftrightarrow$ Partial Dem. $\leftrightarrow$ Democracy

- Previous implementations in political science include:
  1. Dean and Moran (1977)
  2. Jones, Kim and Starz (2005)
  3. Walker (2005)
  4. Others??

Models of Transition Dynamics

Given C categories for the dependent variable, let π_{ab} be the transition probabilities from state a to state b be, where $0 \leq a, b \leq C-1$ and $\sum_b \pi_{ab} = 1$.

The simplest Markov process consists of a two-state system:

$$\begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix} \end{array}.$$

Using a logit link, this two-state case could be estimated by a single regression:

$$\Pr(Y_t = 1) = \text{logit}(X_{t-1}\beta).$$

Problems With the Single-Equation Approach

- This formulation implicitly assumes that the factors moving the state from 0 to 1 are equal and opposite from those that move it from 1 to 0.
- In many substantive applications, we would not wish to assume this *a priori*:
  - Religious factionalization starts ethnic wars; international intervention stops them
  - A bad economy makes people want to elect Democrats; war makes them want to elect Republicans
  - Good economic conditions foster transitions out of autocracy; group-based politics and violence trigger reversals to autocracy

Problems With the Single-Equation Approach

- This formulation implicitly assumes that the factors moving the state from 0 to 1 are equal and opposite from those that move it from 1 to 0.
- In many substantive applications, we would not wish to assume this *a priori*:
  - Religious factionalization starts ethnic wars; international intervention stops them
  - A bad economy makes people want to elect Democrats; war makes them want to elect Republicans
  - Good economic conditions foster transitions out of autocracy; group-based politics and violence trigger reversals to autocracy

Problems With the Single-Equation Approach

- This formulation implicitly assumes that the factors moving the state from 0 to 1 are equal and opposite from those that move it from 1 to 0.
- In many substantive applications, we would not wish to assume this *a priori*:
  - Religious factionalization starts ethnic wars; international intervention stops them
  - A bad economy makes people want to elect Democrats; war makes them want to elect Republicans
  - Good economic conditions foster transitions out of autocracy; group-based politics and violence trigger reversals to autocracy

Problems With the Single-Equation Approach

- This formulation implicitly assumes that the factors moving the state from 0 to 1 are equal and opposite from those that move it from 1 to 0.
- In many substantive applications, we would not wish to assume this *a priori*:
  - Religious factionalization starts ethnic wars; international intervention stops them
  - A bad economy makes people want to elect Democrats; war makes them want to elect Republicans
  - Good economic conditions foster transitions out of autocracy; group-based politics and violence trigger reversals to autocracy

Problems With the Single-Equation Approach

- This formulation implicitly assumes that the factors moving the state from 0 to 1 are equal and opposite from those that move it from 1 to 0.
- In many substantive applications, we would not wish to assume this *a priori*:
  - Religious factionalization starts ethnic wars; international intervention stops them
  - A bad economy makes people want to elect Democrats; war makes them want to elect Republicans
  - Good economic conditions foster transitions out of autocracy; group-based politics and violence trigger reversals to autocracy

The Markov Approach to Transition Dynamics

The Markov approach is to estimate the 2x2 system by a pair of logit regressions, each depending explicitly on the prior state of the system:

$$\begin{aligned} \Pr(Y_t = 1 | Y_{t-1} = 0) &= \text{logit}(X_{t-1}\beta) \\ \Pr(Y_t = 1 | Y_{t-1} = 1) &= \text{logit}(X_{t-1}\alpha) \end{aligned}$$

which can be written more compactly as

$$\Pr(Y_t = 1) = \text{logit}(X_{t-1}\beta + Y_{t-1}X_{t-1}\gamma)$$

where

$$\gamma = \alpha - \beta.$$

The Markov Approach to Transition Dynamics

The Markov approach is to estimate the 2x2 system by a pair of logit regressions, each depending explicitly on the prior state of the system:

$$\begin{aligned} \Pr(Y_t = 1 | Y_{t-1} = 0) &= \text{logit}(X_{t-1}\beta) \\ \Pr(Y_t = 1 | Y_{t-1} = 1) &= \text{logit}(X_{t-1}\alpha) \end{aligned}$$

which can be written more compactly as

$$\Pr(Y_t = 1) = \text{logit}(X_{t-1}\beta + Y_{t-1}X_{t-1}\gamma)$$

where

$$\gamma = \alpha - \beta.$$

The Markov Approach to Transition Dynamics

The Markov approach is to estimate the 2x2 system by a pair of logit regressions, each depending explicitly on the prior state of the system:

$$\begin{aligned} \Pr(Y_t = 1 | Y_{t-1} = 0) &= \text{logit}(X_{t-1}\beta) \\ \Pr(Y_t = 1 | Y_{t-1} = 1) &= \text{logit}(X_{t-1}\alpha) \end{aligned}$$

which can be written more compactly as

$$\Pr(Y_t = 1) = \text{logit}(X_{t-1}\beta + Y_{t-1}X_{t-1}\gamma)$$

where

$$\gamma = \alpha - \beta.$$

The Markov Approach to Transition Dynamics

The Markov approach is to estimate the 2x2 system by a pair of logit regressions, each depending explicitly on the prior state of the system:

$$\begin{aligned} \Pr(Y_t = 1 | Y_{t-1} = 0) &= \text{logit}(X_{t-1}\beta) \\ \Pr(Y_t = 1 | Y_{t-1} = 1) &= \text{logit}(X_{t-1}\alpha) \end{aligned}$$

which can be written more compactly as

$$\Pr(Y_t = 1) = \text{logit}(X_{t-1}\beta + Y_{t-1}X_{t-1}\gamma)$$

where

$$\gamma = \alpha - \beta.$$

Higher-Dimension Processes

For the 3-state case we want to estimate the probabilities π_{ab} in the transition matrix

$$\begin{array}{c} \\ Y_0 \\ Y_1 \\ Y_2 \end{array} \begin{array}{c} \begin{array}{ccc} Y_0 & Y_1 & Y_2 \end{array} \\ \begin{pmatrix} \pi_{00} & \pi_{01} & \pi_{02} \\ \pi_{10} & \pi_{11} & \pi_{12} \\ \pi_{20} & \pi_{21} & \pi_{22} \end{pmatrix} \end{array}$$

We could run nine regular logits for each entry in the matrix; this is known as the "fully saturated" model.

(Note nine vs. two in the 2x2 case.)

But there are some improvements we can make.

Higher-Dimension Processes

For the 3-state case we want to estimate the probabilities π_{ab} in the transition matrix

$$\begin{array}{c} \\ Y_0 \\ Y_1 \\ Y_2 \end{array} \begin{array}{c} \begin{array}{ccc} Y_0 & Y_1 & Y_2 \end{array} \\ \begin{pmatrix} \pi_{00} & \pi_{01} & \pi_{02} \\ \pi_{10} & \pi_{11} & \pi_{12} \\ \pi_{20} & \pi_{21} & \pi_{22} \end{pmatrix} \end{array}$$

We could run nine regular logits for each entry in the matrix; this is known as the "fully saturated" model.

(Note nine vs. two in the 2x2 case.)

But there are some improvements we can make.

Cumulative Probabilities

First, it is easier to work with *cumulative* transition probabilities: $Y_a^* = 1$ if $Y \leq a$.

Given the cumulative probabilities, we can recover the cell probabilities since $\Pr(Y \leq a) = \Pr(Y \leq a-1) + \Pr(Y = a)$.

In the 3-state case the translation from Y to Y^* is:

Y:	0	1	2
Y_0^*:	1	0	0
Y_1^*:	1	1	0

Note that $Y_2^* = 1$.

Cumulative Probabilities

First, it is easier to work with *cumulative* transition probabilities: $Y_a^* = 1$ if $Y \leq a$.

Given the cumulative probabilities, we can recover the cell probabilities since $\Pr(Y \leq a) = \Pr(Y \leq a - 1) + \Pr(Y = a)$.

In the 3-state case the translation from Y to Y^* is:

Y :	0	1	2
Y_0^*:	1	0	0
Y_1^*:	1	1	0

Note that $Y_2^* = 1$.

Example

As a simple example, the log-odds model of cumulative probabilities is:

$$\text{logit}\,\Pr(Y \leq a) = \log \frac{\Pr(Y \leq a)}{\Pr(Y > a)} = \theta_a + X\beta.$$

If $X = 0$, then $\Pr(Y \leq a) = e^{\theta_a}/(1 + e^{\theta_a})$, which is non-decreasing in a, so $\theta_0 \leq \theta_1 \leq \ldots \leq \theta_{C-2}$.

If $\theta_a = \theta_{a+1}$, then $\Pr(Y \leq a) = \Pr(Y \leq a + 1)$, and categories a and $a + 1$ can therefore be collapsed.

Combining Equations

Second, we can run each column of the matrix as a single estimation equation, as in the 2x2 case.

Assume that for any given a, the model to be estimated is

$$\Pr(Y_t = b | Y_{t-1} = a) = \text{logit}\left(\theta_{ab} + X\beta_a\right)$$

Then we can write:

$$\Pr(Y_t = b) = \text{logit}\left(X_{t-1}\beta + \sum_a Y^*_{at-1} X_{t-1}\gamma_a\right)$$

Combining Equations

Second, we can run each column of the matrix as a single estimation equation, as in the 2x2 case.

Assume that for any given a, the model to be estimated is

$$\Pr(Y_t = b | Y_{t-1} = a) = \text{logit}\left(\theta_{ab} + X\beta_a\right)$$

Then we can write:

$$\Pr(Y_t = b) = \text{logit}\left(X_{t-1}\beta + \sum_a Y^*_{at-1} X_{t-1}\gamma_a\right)$$

Combining Equations

Second, we can run each column of the matrix as a single estimation equation, as in the 2x2 case.

Assume that for any given a, the model to be estimated is

$$\Pr(Y_t = b|Y_{t-1} = a) = \text{logit}\left(\theta_{ab} + X\beta_a\right)$$

Then we can write:

$$\Pr(Y_t = b) = \text{logit}\left(X_{t-1}\beta + \sum_a Y^*_{at-1}X_{t-1}\gamma_a\right)$$

Combining Equations

Second, we can run each column of the matrix as a single estimation equation, as in the 2x2 case.

Assume that for any given a, the model to be estimated is

$$\Pr(Y_t = b | Y_{t-1} = a) = \text{logit}\left(\theta_{ab} + X\beta_a\right)$$

Then we can write:

$$\Pr(Y_t = b) = \text{logit}\left(X_{t-1}\beta + \sum_a Y^*_{at-1} X_{t-1}\gamma_a\right)$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Now if $X = 0$, then

$$\begin{aligned} \Pr(Y_t = b|Y_{t-1} = 2) &= \beta_0 \\ \Pr(Y_t = b|Y_{t-1} = 1) &= \beta_0 + \beta_2 \\ \Pr(Y_t = b|Y_{t-1} = 0) &= \beta_0 + \beta_2 + \beta_1 \end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Now if $X = 0$, then

$$\begin{aligned} \Pr(Y_t = b | Y_{t-1} = 2) &= \beta_0 \\ \Pr(Y_t = b | Y_{t-1} = 1) &= \beta_0 + \beta_2 \\ \Pr(Y_t = b | Y_{t-1} = 0) &= \beta_0 + \beta_2 + \beta_1 \end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Now if $X = 0$, then

$$\begin{aligned}
\Pr(Y_t = b | Y_{t-1} = 2) &= \beta_0 \\
\Pr(Y_t = b | Y_{t-1} = 1) &= \beta_0 + \beta_2 \\
\Pr(Y_t = b | Y_{t-1} = 0) &= \beta_0 + \beta_2 + \beta_1
\end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Now if $X = 0$, then

$$\begin{aligned} \Pr(Y_t = b | Y_{t-1} = 2) &= \beta_0 \\ \Pr(Y_t = b | Y_{t-1} = 1) &= \beta_0 + \beta_2 \\ \Pr(Y_t = b | Y_{t-1} = 0) &= \beta_0 + \beta_2 + \beta_1 \end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Now if $X = 0$, then

$$\begin{aligned}\Pr(Y_t = b | Y_{t-1} = 2) &= \beta_0 \\ \Pr(Y_t = b | Y_{t-1} = 1) &= \beta_0 + \beta_2 \\ \Pr(Y_t = b | Y_{t-1} = 0) &= \beta_0 + \beta_2 + \beta_1\end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Similarly, for general values of X

$$\begin{aligned}
\Pr(Y_t = b | Y_{t-1} = 2, X) &= \gamma_0 \\
\Pr(Y_t = b | Y_{t-1} = 1, X) &= \gamma_0 + \gamma_2 \\
\Pr(Y_t = b | Y_{t-1} = 0, X) &= \gamma_0 + \gamma_2 + \gamma_1
\end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Similarly, for general values of X

$$\begin{aligned} \Pr(Y_t = b | Y_{t-1} = 2, X) &= \gamma_0 \\ \Pr(Y_t = b | Y_{t-1} = 1, X) &= \gamma_0 + \gamma_2 \\ \Pr(Y_t = b | Y_{t-1} = 0, X) &= \gamma_0 + \gamma_2 + \gamma_1 \end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Similarly, for general values of X

$$\begin{aligned} \Pr(Y_t = b | Y_{t-1} = 2, X) &= \gamma_0 \\ \Pr(Y_t = b | Y_{t-1} = 1, X) &= \gamma_0 + \gamma_2 \\ \Pr(Y_t = b | Y_{t-1} = 0, X) &= \gamma_0 + \gamma_2 + \gamma_1 \end{aligned}$$

Example

Say we have three states and one independent variable. Then we estimate:

$$\Pr(Y_{it} = b) = \beta_0 + \beta_1 y_0^* + \beta_2 y_1^* + \gamma_0 X + \gamma_1 X y_0^* + \gamma_2 X y_1^*$$

Similarly, for general values of X

$$\begin{aligned}
\Pr(Y_t = b|Y_{t-1} = 2, X) &= \gamma_0 \\
\Pr(Y_t = b|Y_{t-1} = 1, X) &= \gamma_0 + \gamma_2 \\
\Pr(Y_t = b|Y_{t-1} = 0, X) &= \gamma_0 + \gamma_2 + \gamma_1
\end{aligned}$$

Interpretation of Interaction Terms

To summarize, for any given independent variable X:

- The γ_0 coefficient on the un-interacted X term gives the impact of X_{t-1} on Y_t when $Y_{t-1} = Y_{C-1}$, the "last" category of Y.
- The γ_a coefficients on the interaction terms Xy_a^* give the *differential* impact of X_{t-1} on Y_t between $Y_{t-1} = Y_{a+1}$ and $Y_{t-1} = Y_a$.
- The cell probabilities giving the impact of X on $\Pr(Y_t = b|Y_{t-1} = a)$ when $a < C - 1$ can be recovered as the sums of the γ coefficients, in the order $\gamma_0 + \gamma_{C-1} + \gamma_{C-2} + \ldots + \gamma_{a+1}$.

Interpretation of Interaction Terms

To summarize, for any given independent variable X:

- The γ_0 coefficient on the un-interacted X term gives the impact of X_{t-1} on Y_t when $Y_{t-1} = Y_{C-1}$, the "last" category of Y.
- The γ_a coefficients on the interaction terms Xy_a^* give the *differential* impact of X_{t-1} on Y_t between $Y_{t-1} = Y_{a+1}$ and $Y_{t-1} = Y_a$.
- The cell probabilities giving the impact of X on $\Pr(Y_t = b|Y_{t-1} = a)$ when $a < C - 1$ can be recovered as the sums of the γ coefficients, in the order $\gamma_0 + \gamma_{C-1} + \gamma_{C-2} + \ldots + \gamma_{a+1}$.

Interpretation of Interaction Terms

To summarize, for any given independent variable X:

- The γ_0 coefficient on the un-interacted X term gives the impact of X_{t-1} on Y_t when $Y_{t-1} = Y_{C-1}$, the "last" category of Y.
- The γ_a coefficients on the interaction terms Xy_a^* give the *differential* impact of X_{t-1} on Y_t between $Y_{t-1} = Y_{a+1}$ and $Y_{t-1} = Y_a$.
- The cell probabilities giving the impact of X on $\Pr(Y_t = b|Y_{t-1} = a)$ when $a < C - 1$ can be recovered as the sums of the γ coefficients, in the order $\gamma_0 + \gamma_{C-1} + \gamma_{C-2} + \ldots + \gamma_{a+1}$.

Interpretation of Interaction Terms

$$Y_{t-1} = 0$$

$$Y_0^* \quad \rightarrow \qquad\qquad \leftarrow \quad X \cdot Y_0^*$$

$$Y_{t-1} = 1$$

$$Y_1^* \quad \rightarrow \qquad\qquad \leftarrow \quad X \cdot Y_1^*$$

$$Y_{t-1} = 2$$

Insignificant values of coefficients on interactions of X with Y_a^* mean that X has a similar impact on Y for categories a and $a+1$, so we can collapse those categories in the analysis.

We can then test down to a more parsimonious model, eliminating some of the interactive terms.

Interpretation of Interaction Terms

$$Y_{t-1} = 0$$

$$Y_0^* \quad \longrightarrow \qquad \qquad \longleftarrow \quad X \cdot Y_0^*$$

$$Y_{t-1} = 1$$

$$Y_1^* \quad \longrightarrow \qquad \qquad \longleftarrow \quad X \cdot Y_1^*$$

$$Y_{t-1} = 2$$

Insignificant values of coefficients on interactions of X with Y_a^* mean that X has a similar impact on Y for categories a and $a+1$, so we can collapse those categories in the analysis.

We can then test down to a more parsimonious model, eliminating some of the interactive terms.

Interpretation of Interaction Terms

$$\begin{array}{ccccc} & & Y_{t-1} = 0 & & \\ Y_0^* & \rightarrow & & \leftarrow & X \cdot Y_0^* \\ & & Y_{t-1} = 1 & & \\ Y_1^* & \rightarrow & & \leftarrow & X \cdot Y_1^* \\ & & Y_{t-1} = 2 & & \end{array}$$

Insignificant values of coefficients on interactions of X with Y_a^* mean that X has a similar impact on Y for categories a and $a+1$, so we can collapse those categories in the analysis.

We can then test down to a more parsimonious model, eliminating some of the interactive terms.

Interpretation of Interaction Terms

$$
\begin{array}{ccccc}
 & & Y_{t-1} = 0 & & \\
Y_0^* & \rightarrow & & \leftarrow & X \cdot Y_0^* \\
 & & Y_{t-1} = 1 & & \\
Y_1^* & \rightarrow & & \leftarrow & X \cdot Y_1^* \\
 & & Y_{t-1} = 2 & &
\end{array}
$$

Insignificant values of coefficients on interactions of X with Y_a^* mean that X has a similar impact on Y for categories a and $a+1$, so we can collapse those categories in the analysis.

We can then test down to a more parsimonious model, eliminating some of the interactive terms.

Interpretation of Interaction Terms

$$\begin{array}{ccccc} & & Y_{t-1} = 0 & & \\ Y_0^* & \rightarrow & & \leftarrow & X \cdot Y_0^* \\ & & Y_{t-1} = 1 & & \\ Y_1^* & \rightarrow & & \leftarrow & X \cdot Y_1^* \\ & & Y_{t-1} = 2 & & \end{array}$$

Insignificant values of coefficients on interactions of X with Y_a^* mean that X has a similar impact on Y for categories a and $a + 1$, so we can collapse those categories in the analysis.

We can then test down to a more parsimonious model, eliminating some of the interactive terms.

Summary

Key points in higher-dimension Markov models:

- Work with cumulative probabilities.
- Combine cases for transitions to state b using interactions with the y^* terms.
- Start with the saturated model with all interactive terms and test down.

To overcome small-N cell problems, you can also combine the analysis for all values of b into a single ordered probit/logit.

Summary

Key points in higher-dimension Markov models:

- Work with cumulative probabilities.
- Combine cases for transitions to state b using interactions with the y^* terms.
- Start with the saturated model with all interactive terms and test down.

To overcome small-N cell problems, you can also combine the analysis for all values of b into a single ordered probit/logit.

Summary

Key points in higher-dimension Markov models:

- Work with cumulative probabilities.
- Combine cases for transitions to state b using interactions with the y^* terms.
- Start with the saturated model with all interactive terms and test down.

To overcome small-N cell problems, you can also combine the analysis for all values of b into a single ordered probit/logit.

Summary

Key points in higher-dimension Markov models:

- Work with cumulative probabilities.
- Combine cases for transitions to state b using interactions with the y^* terms.
- Start with the saturated model with all interactive terms and test down.

To overcome small-N cell problems, you can also combine the analysis for all values of b into a single ordered probit/logit.

Application: Democratic Transitions

- Modernization theory (Lipset 1959) says that as countries get richer, they get democratic.
- This was always thought of as a causal relationship, although there has always been a dispute about the mechanics.
- Przeworski, et. al. ("PACL" 2000) challenge this, saying that the process could be:
  1. Countries become democratic randomly.
     Once there, higher GDP per capita helps keep them there.
- So a GDP-democracy relationship could develop, even though modernization doesn't *cause* democracy.

Application: Democratic Transitions

- Modernization theory (Lipset 1959) says that as countries get richer, they get democratic.
- This was always thought of as a causal relationship, although there has always been a dispute about the mechanics.
- Przeworski, et. al. ("PACL" 2000) challenge this, saying that the process could be:
  1. Countries become democratic randomly.
  2. Once there, higher GDP per capita helps keep them there.
- So a GDP-democracy relationship could develop, even though modernization doesn't *cause* democracy.

Application: Democratic Transitions

- Modernization theory (Lipset 1959) says that as countries get richer, they get democratic.
- This was always thought of as a causal relationship, although there has always been a dispute about the mechanics.
- Przeworski, et. al. ("PACL" 2000) challenge this, saying that the process could be:
  1. Countries become democratic randomly.
  2. Once there, higher GDP per capita helps keep them there.
- So a GDP-democracy relationship could develop, even though modernization doesn't *cause* democracy.

Application: Democratic Transitions

- Modernization theory (Lipset 1959) says that as countries get richer, they get democratic.
- This was always thought of as a causal relationship, although there has always been a dispute about the mechanics.
- Przeworski, et. al. ("PACL" 2000) challenge this, saying that the process could be:
  1. Countries become democratic randomly.
  2. Once there, higher GDP per capita helps keep them there.
- So a GDP-democracy relationship could develop, even though modernization doesn't *cause* democracy.

Application: Democratic Transitions

- Modernization theory (Lipset 1959) says that as countries get richer, they get democratic.
- This was always thought of as a causal relationship, although there has always been a dispute about the mechanics.
- Przeworski, et. al. ("PACL" 2000) challenge this, saying that the process could be:
  1. Countries become democratic randomly.
  2. Once there, higher GDP per capita helps keep them there.
- So a GDP-democracy relationship could develop, even though modernization doesn't *cause* democracy.

Application: Democratic Transitions

- Modernization theory (Lipset 1959) says that as countries get richer, they get democratic.
- This was always thought of as a causal relationship, although there has always been a dispute about the mechanics.
- Przeworski, et. al. ("PACL" 2000) challenge this, saying that the process could be:
  1. Countries become democratic randomly.
  2. Once there, higher GDP per capita helps keep them there.
- So a GDP-democracy relationship could develop, even though modernization doesn't *cause* democracy.

Results from PACL Table 2.12

Indep. Var.	$D \rightarrow A$	$A \rightarrow D$ (Original)	$A \rightarrow D$ (Corrected)
Constant	-1.144** (0.000)	-2.524** (0.000)	-2.524** (0.000)
GDP	-0.201 (0.162)	0.329 (0.484)	0.329** (0.004)
GDP^2	-0.003 (0.874)	-0.029 (0.191)	-0.029 (0.069)
GDP Growth	-0.042** (0.003)	-0.021** (0.000)	-0.021* (0.015)

PACL's Extended Transition Model

- PACL also run a regression adding a number of covariates, but without GDP^2.
- Actual results here are more favorable to their hypothesis:
  - GDP helps keep democracies from backsliding, but has no effect on autocracy.
  - These results are somewhat fragile to specification.
- Leaves open the central question of modernization and democracy.

PACL's Extended Transition Model

- PACL also run a regression adding a number of covariates, but without GDP^2.
- Actual results here are more favorable to their hypothesis:
  - GDP helps keep democracies from backsliding, but has no effect on autocracy.
  - These results are somewhat fragile to specification.
- Leaves open the central question of modernization and democracy.

PACL's Extended Transition Model

- PACL also run a regression adding a number of covariates, but without GDP^2.
- Actual results here are more favorable to their hypothesis:
  - GDP helps keep democracies from backsliding, but has no effect on autocracy.
  - These results are somewhat fragile to specification.
- Leaves open the central question of modernization and democracy.

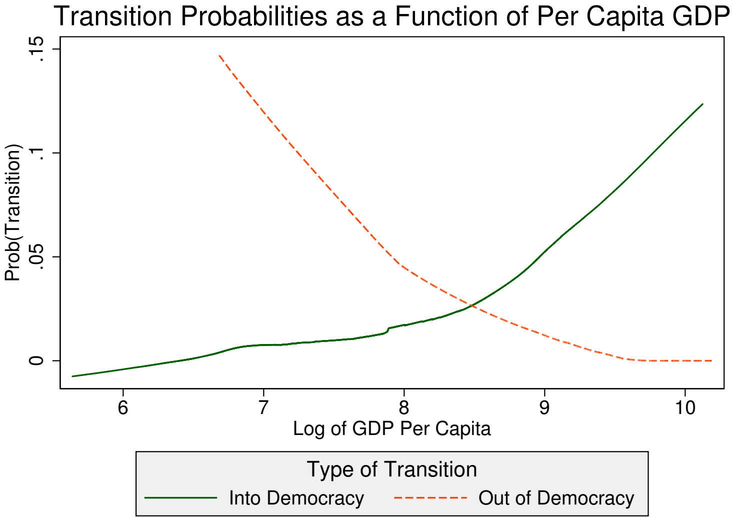
PACL's Extended Transition Model

- PACL also run a regression adding a number of covariates, but without GDP^2.
- Actual results here are more favorable to their hypothesis:
  - GDP helps keep democracies from backsliding, but has no effect on autocracy.
  - These results are somewhat fragile to specification.
- Leaves open the central question of modernization and democracy.

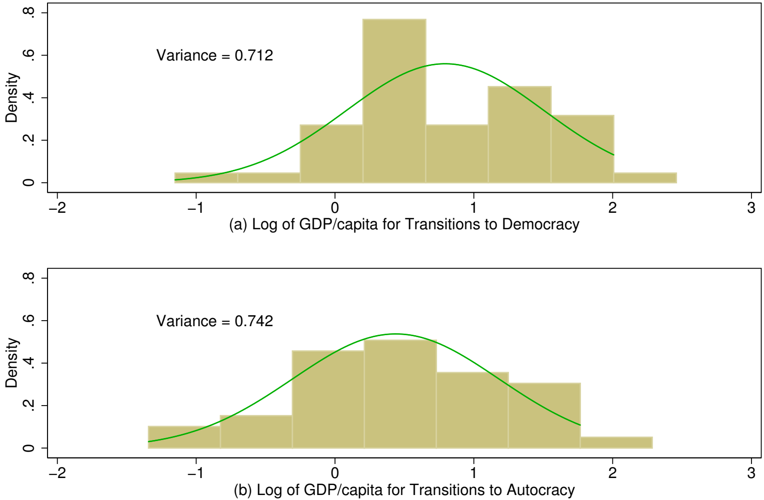
PACL's Extended Transition Model

- PACL also run a regression adding a number of covariates, but without GDP^2.
- Actual results here are more favorable to their hypothesis:
  - GDP helps keep democracies from backsliding, but has no effect on autocracy.
  - These results are somewhat fragile to specification.
- Leaves open the central question of modernization and democracy.

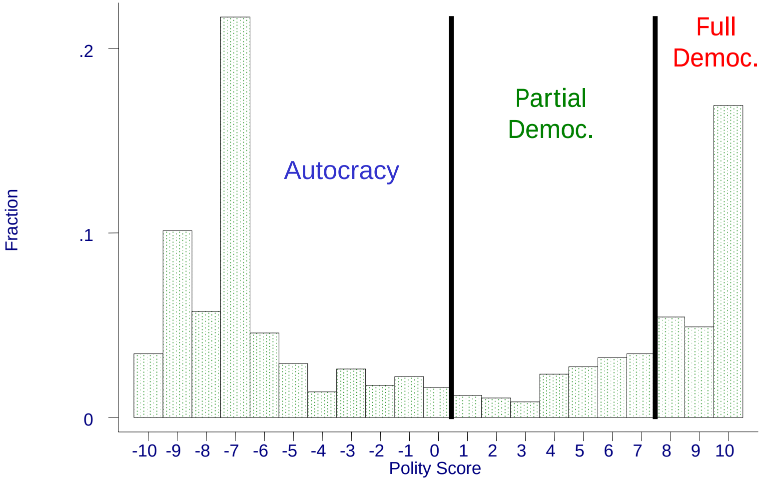
Impact of GDP on Transition Probabilities



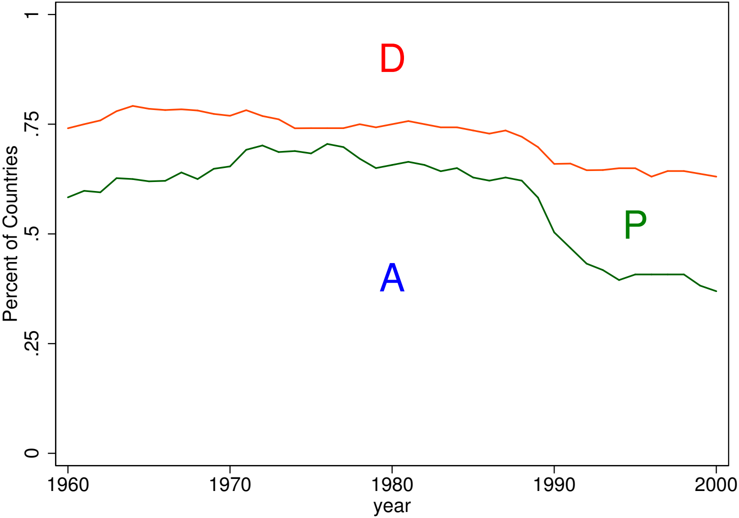
Distribution of GDP for Transition Years



Definition of Partial Democracy



World Democratization Trends, 1960-2000



Transition Marginals

	Current Year		
Previous Year	Autocracy	Partial Democracy	Democracy
Autocracy	97.3%	2.1%	0.7%
	(3,121)	(66)	(22)
Partial Democracy	6.4%	90.4%	3.3%
	(49)	(695)	(25)
Democracy	1.1%	0.8%	98.2%
	(16)	(12)	(1,496)
Total	3,186	773	1,543

Note: Numbers in parentheses are cell counts.

Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Polity Score	-0.45	7.58	-10	10	5671
Regime Category	0.70	0.88	0	2	5671
Log of Per Capita GDP	8.14	1.04	5.64	10.21	4417
Percent Change in GDP	0.02	0.06	-0.52	1.01	4475
Percent Urban Pop.	44.94	24.29	2.3	100	5245
Log of Population Density	3.61	1.46	-0.49	8.77	5600
Log of Trade Openness	3.98	0.62	0.43	6.16	4902
Previous Transitions	3.96	6.41	0	31	5671
Resource Curse	0.23	0.42	0	1	5671

Regression Results

Adding Partial Autocracies

	Polity Range
Autocracy	(-10,-7)
Partial Aut.	(-6,0)
Partial Dem.	(1,7)
Democracy	(8,10)

Check to see if we should split the autocracies as well.

Adding Partial Autocracies

			Polity Range
		Autocracy	(-10,-7)
Y_0^*	$\longrightarrow$		
		Partial Aut.	(-6,0)
Y_1^*	$\longrightarrow$		
		Partial Dem.	(1,7)
Y_2^*	$\longrightarrow$		
		Democracy	(8,10)

Use the Y^* variables to test for collapsing adjacent categories.

Adding Partial Autocracies

			Polity Range
		Autocracy	(-10,-7)
		$\updownarrow$	
		Partial Aut.	(-6,0)
Y_1^*	$\longrightarrow$		
		Partial Dem.	(1,7)
Y_2^*	$\longrightarrow$		
		Democracy	(8,10)

Only Y_0^* is insignificant, lending support to our three-way classification vs. four-way classification with partial autocracies.

Comparison with PACL

Autocracy

Partial Dem. a

Partial Dem. b

Democracy

Split our partial democracies into PACL autocracies (a) and PACL democracies (b).

Comparison with PACL

Autocracy
$\updownarrow$ PACL
Partial Dem. a

Partial Dem. b
$\updownarrow$ PACL
Democracy

PACL collapse the partial democracies into the full autocracies and full democracies.

Comparison with PACL

Autocracy

Partial Dem. a

$\updownarrow$ O'Epstein

Partial Dem. b

Democracy

We collapse the partial democracies into a single category.

Comparison with PACL

Autocracy

$Y_0^* \longrightarrow$

Partial Dem. a

$Y_1^* \longrightarrow$

Partial Dem. b

$Y_2^* \longrightarrow$

Democracy

Again use the Y^* variables to discriminate.

Comparison with PACL

Autocracy

$Y_0^* \quad \longrightarrow$

Partial Dem. a

$\updownarrow$

Partial Dem. b

$Y_2^* \quad \longrightarrow$

Democracy

Only Y_1^* is insignificant, lending support to our three-way classification vs. PACL's dichotomous classification.