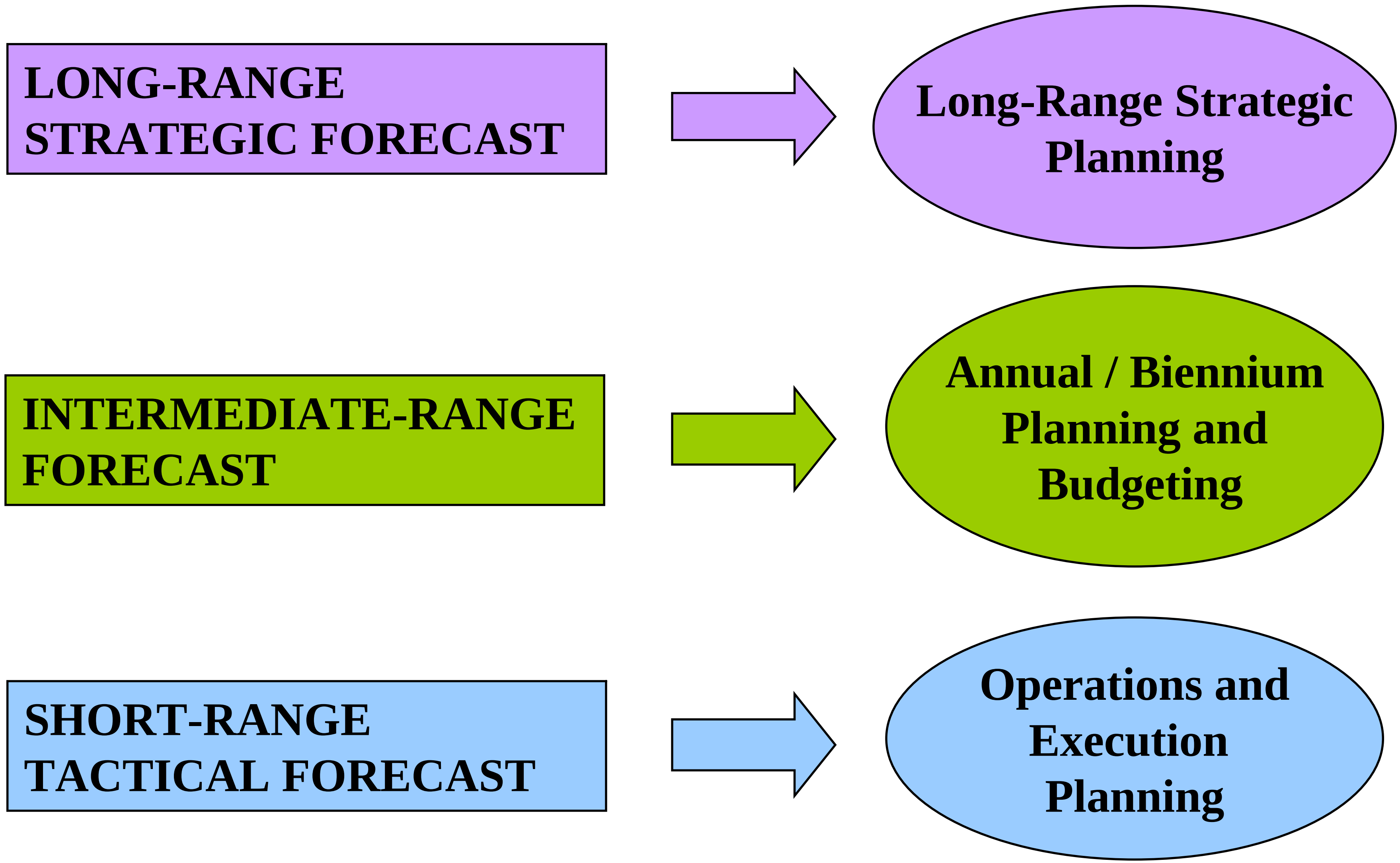
A Comparison of Time Horizon Models to Forecast Enrollment

Lawrence J. Redlinger and Stanley L. Gordon
The University of Texas at Dallas

Presented at AIR 2004 Forum, Boston

June 2, 2004

FORECAST PROCESS with DIFFERENT TIME HORIZONS



Forecasting Supports Planning

FORECAST PROCESS with DIFFERENT TIME HORIZONS

LONG-RANGE STRATEGIC FORECAST (10-15 YEARS OUT)

Enrollment, SCH, Revenues, Expenditures, FTE Faculty, Federal R&D Expenditures, Endowment, Physical Plant / Campus Master Plan, Degree Programs

INTERMEDIATE-RANGE FORECAST (1-2 YEARS OUT)

Enrollment, SCH, Revenues, Expenditures, Faculty and Staff Hiring, Student Financial Aid, Tuition / Fees, Administrative Computer Systems, Course Inventory

SHORT-RANGE TACTICAL FORECAST (1-3 SEMESTERS OUT)

Enrollment, SCH, Revenue, Expenditures, Classroom Scheduling

Forecasts with Interlocking Time Horizons

ENROLLMENT DRIVES EVERYTHING

Enrollment = f (X1, X2, … , Xn)

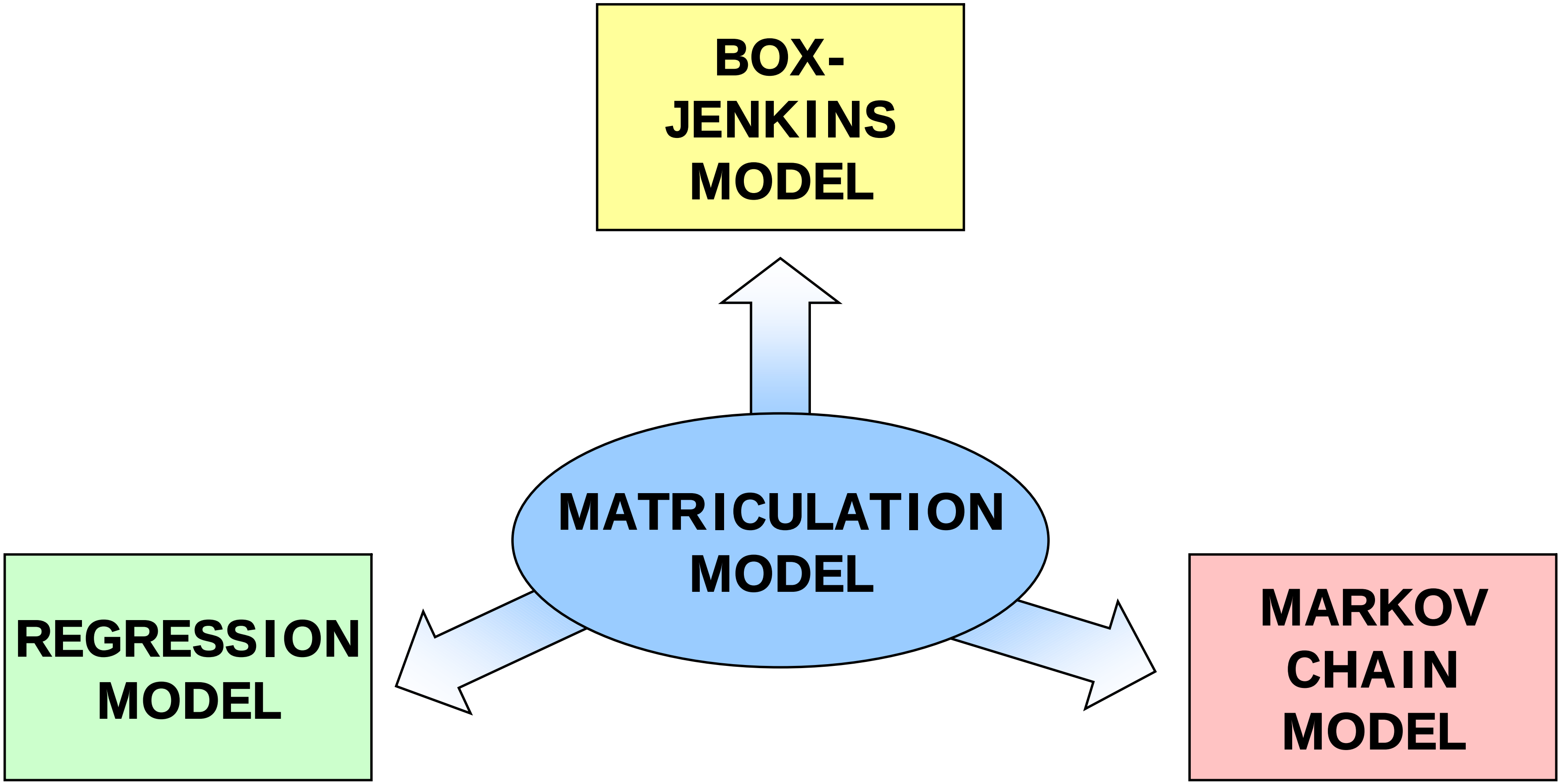
SCH = g(Enrollment)

Revenue ($) = h(SCH)

These models are recursive; i.e., the output of one model becomes the input of the next model.

HUB AND SPOKES

Goal: Accuracy &Timing



Short-Range Tactical Forecast System Structure

REGRESSION MODEL

- Good for both aggregated and disaggregated data
- Handles unusual events using dummy variables
- Prediction variables must themselves be forecasted for future semesters (this is usually done using compound growth formula, Box-Jenkins, business judgment / scenario forecasting)
- Sample size can vary from a few observations to many
- Can establish “causal” relationships among variables
- Appropriate for capturing turning points in data
- Generally requires that random error term be distributed as bell curve (hypothesis testing)
- Useful for goal-seeking (run model in reverse to assess inputs that yield targeted output)
- SOFTWARE TOOLS: SAS, Stata, RATS, Eviews, S+, SPSS

BOX-JENKINS MODEL

- Requires a sample of 40-60 observations
- Extremely simple model with no prediction variables, only time-series history
- Requires intervention of analyst to identify model structure using graphs of sample autocorrelation and partial autocorrelation functions; thus, can be time-consuming to construct many models
- Usually works best on more aggregated data with discernable trend and seasonal patterns
- Seasonal models much more difficult to formulate
- Requires stationary data; i.e., de-trended (no linear or curvilinear trend) and constant variance across sample
- SOFTWARE TOOLS: SAS, Stata, RATS, Eviews, S+

MARKOV CHAIN MODEL

- Sample size of three semester-to-semester changes is adequate to estimate state transition probabilities, if stable
- Closely replicates the physical student pipeline
- Transition probabilities are required to be time invariant (constant over time)
- New students (first time and transfers-in) need to be forecasted outside the MC system using traditional methods
- Lends itself to finer granularity of data; e.g., baccalaureate degree model has classification states (freshmen, sophomores, juniors, seniors)
- No goodness of fit statistics to assess model adequacy a priori
- Gain in-depth of understanding of underlying data generating process (DGP)
- Transaction-oriented approach to modeling time-series
- SOFTWARE TOOLS: Customized programming in SAS, Excel

TREND-SEASONAL REGRESSION MODEL

Use stepwise selection of key prediction variables from Master Menu of 33 Time-Sensitive Trend, Seasonal, and Event History variables conjectured to affect enrollment by semester:

Enrollment(t) = B0 (Intercept) +

B1 * TIME +

B2 * Spring Intercept Dummy +

B3 * Spring Slope Dummy +

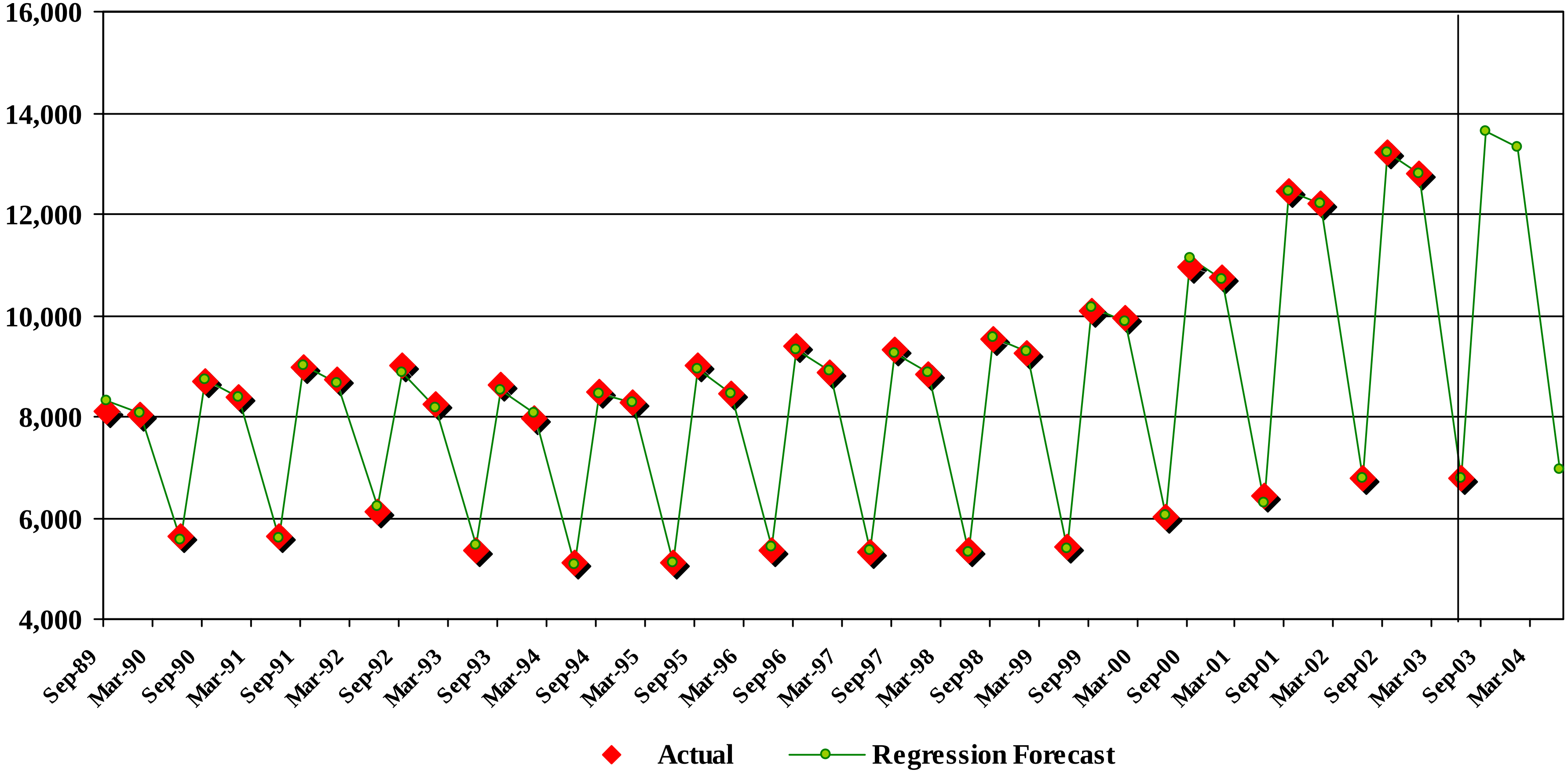
B4 * Summer Intercept Dummy +

B5 * Summer Slope Dummy +

(AY Dummies) + random error(t)

Correct for all violations of underlying statistical assumptions to Sharpen forecasts.

UTD ENROLLMENT FORCASTING – REGRESSION MODEL



TWO TECHNICALLY EQUIVALENT BOX-JENKINS MODELS

Notation: Y(t) = Fall Enrollment at time t

Z(t) = Y(t) – Y(t-1) = First Difference to de-trend the data

e(t) = random error (bell curve)

Example: Enrollment Fall 2001 = 12,455 = Y(t-1)

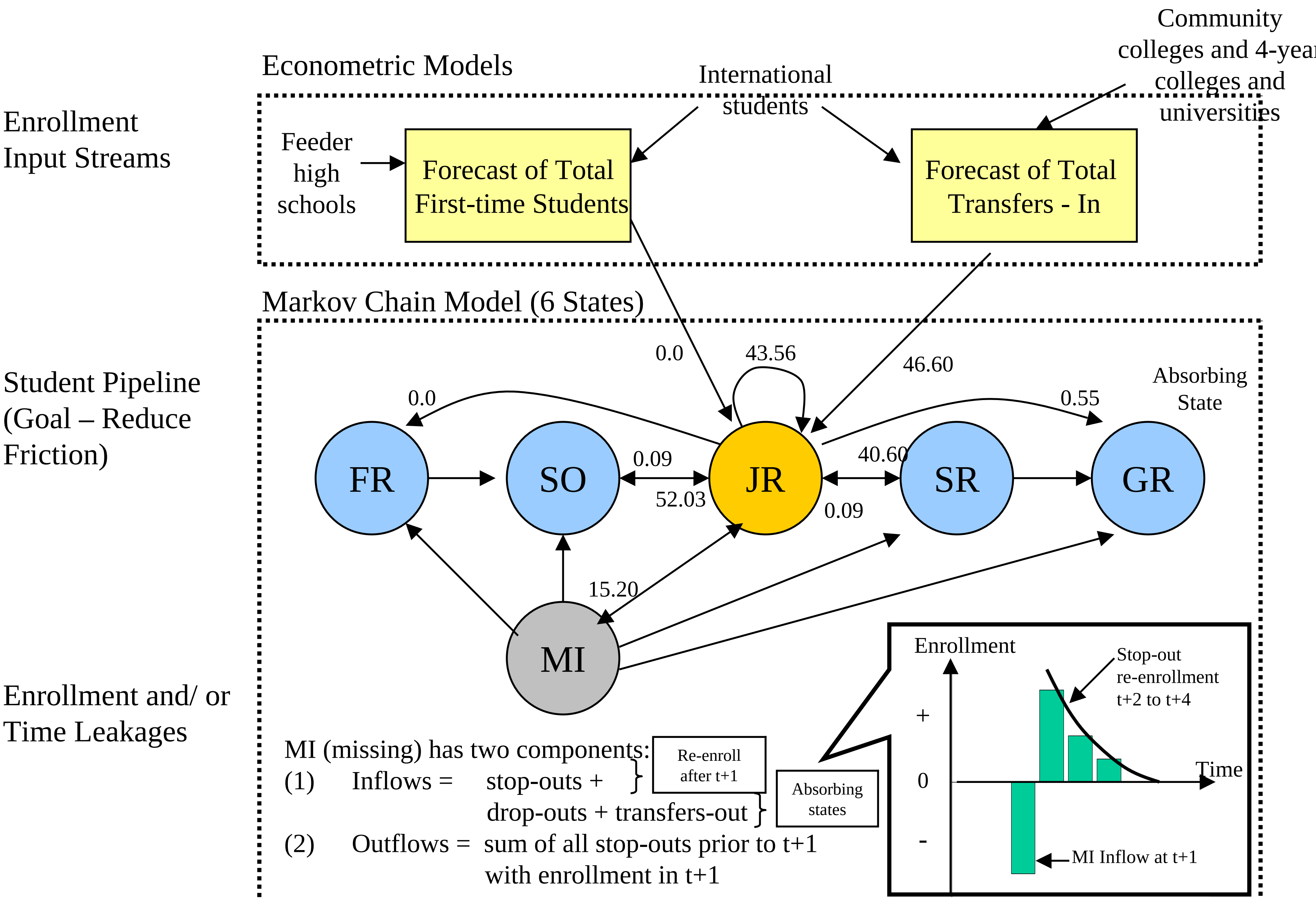
Enrollment Fall 2002 = 13,229 = Y(t)

First Difference = 13,229 – 12,455= 774 = Z(t)

MA(1) Model: Z(t) = B0 + e(t) – [B1 * e(t-1)]

AR(1) Model: Z(t) = A0 + e(t) + [A1 * Z(t-1)]

MARKOV CHAIN TRANSITION GRAPH (JUNIORS)



Source Data: Spring 2001 (t) to Fall 2001 (t + 1)

MC TRANSISTION PROBABILITIES (%)

P_ij = Probability of state j in fall 2003 given state i in Spring 2003.

	FR	SO	JR	SR	GR	MI	TOTAL
FR	37.26	40.95	1.00	0.00	0.05	20.74	100.00
SO	0.00	26.47	53.69	0.74	0.00	19.10	100.00
JR	0.00	0.04	45.83	38.59	0.26	15.28	100.00
SR	0.00	0.00	0.04	52.64	31.17	16.15	100.00
GR	0.00	0.00	0.00	0.00	100.00	0.00	100.00
MI	17.55	16.08	31.26	35.10	0.00	0.00	100.00

Fall 2003 Enrollment = Spring 2003 Enrollment – Graduates – Missing + First-Time Fall 2003 + Transfers-In Fall 2003 + Stop-Outs Returning Fall 2003

= 7,570 – 851 – 1,339 + 947 +1,325 + 578

= 8,230 undergraduate students

ENROLLMENT FORECASTS FOR FALL 2003

- Trend-Seasonal Regression Model = 13,656
- Box-Jenkins MA(1) Model = 13,570
- Box-Jenkins AR(1) Model = 13,803
- Average of two Box-Jenkins Models = 13,687
- Markov Chain Model = 13,717
- Average of three models = 13,687

For MC model, baccalaureate enrollment forecast was inflated using 60% - 40% split between undergraduate and graduate students)

THECB Certified Fall 2003 Headcount = 13,718

NEXT STEPS IN RESEARCH AND DEVELOPMENT

- Finish development of Undergraduate Markov Chain Model before tackling Masters and Doctoral models
- Refine estimation of missing (MI) or “out-of-system” state
- Model time-varying state transition probabilities using logistic regression, as required
- Develop robust econometric models for enrollment input streams (first-time students and transfers-in)
- Implement a working software system that includes simulation and what-if capabilities
- Disaggregate data by student classification and academic school
- Generate SCH forecasts driven by enrollment forecasts using distributions of semester credit hours for part-time and full-time students
- Repeat above for Revenue ($) using distribution of tuition and fees for in-state and out-of-state tuition status