Forecasting Sales: A Model and Some Evidence from the Retail Industry

Russell Lundholm
Sarah McVay
Taylor Randall

MICHIGAN BUSINESS SCHOOL

Why forecast financial statements?

- Seems obvious, but two common criticisms:
  - “Who cares, can’t we can look at the analyst forecasts?”
  - “Surely in the ‘real world’ analysts use much more sophisticated models than we could ever imagine.”

Why forecast sales?

- the sales forecast is the most important forecast in most financial forecasting applications
  - combine with margin to get income forecast
  - combine with turnover to get asset forecast

- A 1% change in the sales growth estimate can change the equity value estimate by about 15%.

- Our goal is to
  - develop a reasonably general sales-forecasting model
  - show how to estimate it on a retail firm
  - test its predictive ability in-sample and out-of-sample
  - compare it to analyst forecasts

What needs to be modeled?

- how do we mix together beliefs about
  - the number of new sales-generating units, and
  - the growth in sales from existing units?
- Half of the puzzle - # of stores - is commonly estimated/disclosed by the firm.
  - what does this imply about the sales forecast?
- Comp growth rates are more complicated than you might have thought

The Model – sales rates / store

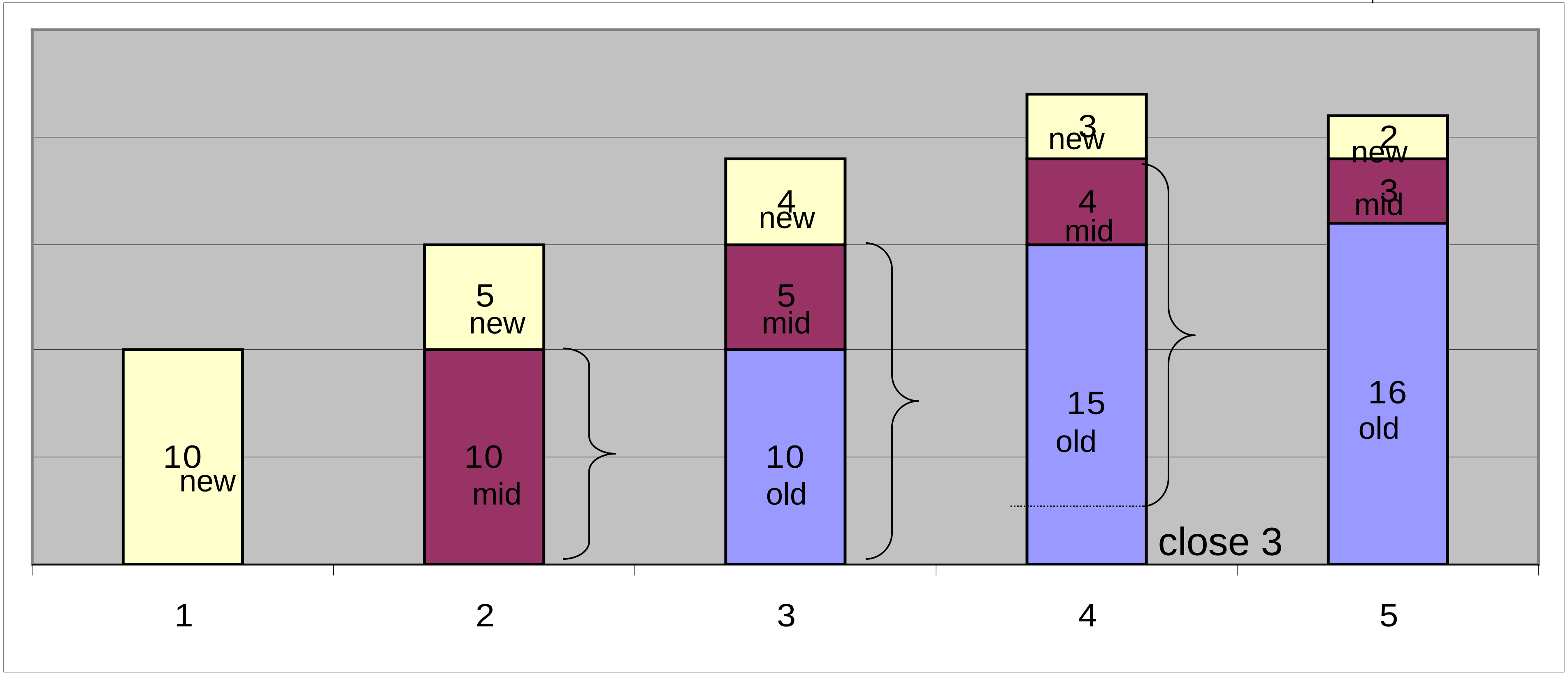
R_t^N = the average sales/store rate for the new stores in year t, (annualized in results)

R_t^M = the average sales/store rate for the mid stores in year t, and

R_t^O = the average sales/store rate for old stores in year t.

- $Sales_t = O_t R_t^O + M_t R_t^M + N_t R_t^N$

the number of stores and comps



Comparable store sales growth is the growth in sales from stores that where open at the beginning of the prior year and are currently still open.

The algebra of comps

$$1 + C_t = \frac{O_t R_t^O}{(O_{t-1} - D_t) R_{t-1}^O + M_{t-1} R_{t-1}^M}$$

- recall that $O_t = O_{t-1} + M_{t-1} - D_t$ so same stores are in numerator and denominator
- note that comps are sensitive to the movement of stores from mid to old (if the rates are different)
- if $R_t^M = R_t^O$ each period then $1 + C_t = \frac{R_t^O}{R_{t-1}^O}$

Using the comp growth to "clean up" the estimation

we know the # of stores of each type

$$Sales_t = O_t R_t^O + M_t R_t^M + N_t R_t^N$$

but the rates are **unknown.** However, we know something about the evolution of the rates over time.

The trick in the estimation is to account for the changes in past rates as part of the independent variable so that we can estimate the most recent rates as parameters.

The Model – something we can estimate?

$$\Delta Sales_t = (O_t R_t^O - O_{t-1} R_{t-1}^O) + (M_t R_t^M - M_{t-1} R_{t-1}^M) + (N_t R_t^N - N_{t-1} R_{t-1}^N)$$

- observe all the # of stores and control for the changing rates over time using the comps (plus some assumptions).

- Estimate $R_T^N \, R_T^M \, R_T^O$ where T is most recent year in data.

- Estimate by firm over time. Estimate in changes to control for non-stationarity.

Model 1

$$know \text{ that } R_t^O = R_{t-1}^O(1+C_t)$$

$$assume \text{ that } R_t^M = R_t^O$$

$$\text{and } R_t^N = R_{t-1}^N(1+C_t)$$

- after the end of first year, store is fully mature
- all rates evolve at the rate (1+C_t)

$$\Delta Sales_T = (O_T + M_T)R_T^O - (O_{T-1} + M_{T-1})R_{T-1}^O + N_T R_T^N - N_{T-1}R_{T-1}^N$$

change in stores and rates on old+mid

change in stores and rates on new

$$\Delta Sales_T = \left[(O_T + M_T) - \frac{(O_{T-1} + M_{T-1})}{(1+C_T)}\right]R_T^O + \left[N_T - \frac{N_{T-1}}{(1+C_T)}\right]R_T^N$$

adjusting the number of beginning stores down

Model 1 – its nicer than it looks!

$$\Delta Sales_{T-\tau} = \left[\frac{(1+C_{T-\tau})(O_{T-\tau} + M_{T-\tau}) - (O_{T-\tau-1} + M_{T-\tau-1})}{\prod_{i=T-\tau}^{T} (1+C_i)} \right] R_T^O +$$

$$\left[\frac{(1+C_{T-\tau}) N_{T-\tau} - N_{T-\tau-1}}{\prod_{i=T-\tau}^{T} (1+C_i)} \right] R_T^N$$

change in # of stores plus comp effect in current year

bring prior year data to year T basis by deflating with historical comps

- everything in brackets is known so can estimate the rates.

Model 2

$$R_t^M = kR_t^O$$

$$R_t^N = R_{t-1}^N(1+C_t)Q_t \text{ , where}$$

$$Q_t = \frac{O_{t-1} + kM_{t-1} - D_t}{O_t}$$

- mid rates and new rates differ by proportion k
- all rates evolve at the rate $(1+C_t)Q_t = \frac{R_t^O}{R_{t-1}^O}$
- if k > 1 (so Q>1) then prolonged new store “honeymoon period”.
- if k < 1 (so Q<1) then prolonged time to maturity.

models 3-6

- model 3: $\%\Delta Sales_{T-\tau} = (1+C_{T-\tau})(1+G_{T-\tau})-1$
  - assumes $R_t^M = R_t^O = R_t^N$

- model 4: $\Delta Sales_{T-\tau} = Sales_{T-\tau-1} SG_j$ where SG_j is historical sales growth for that decile (Nissim/Penman 2001)

- model 5: $\Delta Sales_{T-\tau} = \beta(\Delta \text{ total \# stores}) + \varepsilon$
- model 6: $\Delta Sales_{T-\tau} = \gamma_1 \Delta \text{ existing stores} + \gamma_2 \Delta \text{ new stores} + \varepsilon$

The Sample

- 87 retail firms (1036 firm-years) with at least 6 years of data on
  - sales
  - number of stores at year end
  - stores opened during the year
  - stores closed during the year
  - comparable store growth rate for the year
  - expected number of store openings/closings for the following year

Summary Regression Results (estimated firm-by-firm)

Models	Median p values for Estimated Sales Rates		Percent Positive Sales Rates		Mean(Median)	Median Adjusted R^2
	Old Rate	NewRate	Old	New	$k \in (0.8, 1.2)$	
Model 1: New/Old, k = 1 (Full Sample)	(0.001)	(0.014)	98.9%	92.0%	na	91.7%
Model 2: New/Old, $k \in (0.8, 1.2)$	(0.001)	(0.021)	98.9%	92.0%	1.0(1.0)	92.8%

In-Sample comparison of models

	Model 1 k=1	Model 2 k∈(.8, 1.2)	Model 3 (1+G)(1+C)–1
Median of Median [|Residual|/Sales]	2.49%	1.99%	3.24%

	Model 4 Mean Reversion	Model 5 ChgStores	Model 6 ChgNew/ Old Stores
Median of Median [|Residual|/Sales]	4.50%	4.61%	4.04%

examples

	old rate	new rate	k	R^2
Starbucks	1.18	0.66	k=1	98.9
Whole Foods	29.85	43.33	k=1	97.5
Walmart	111.65	322.74	k=1.2	94.0

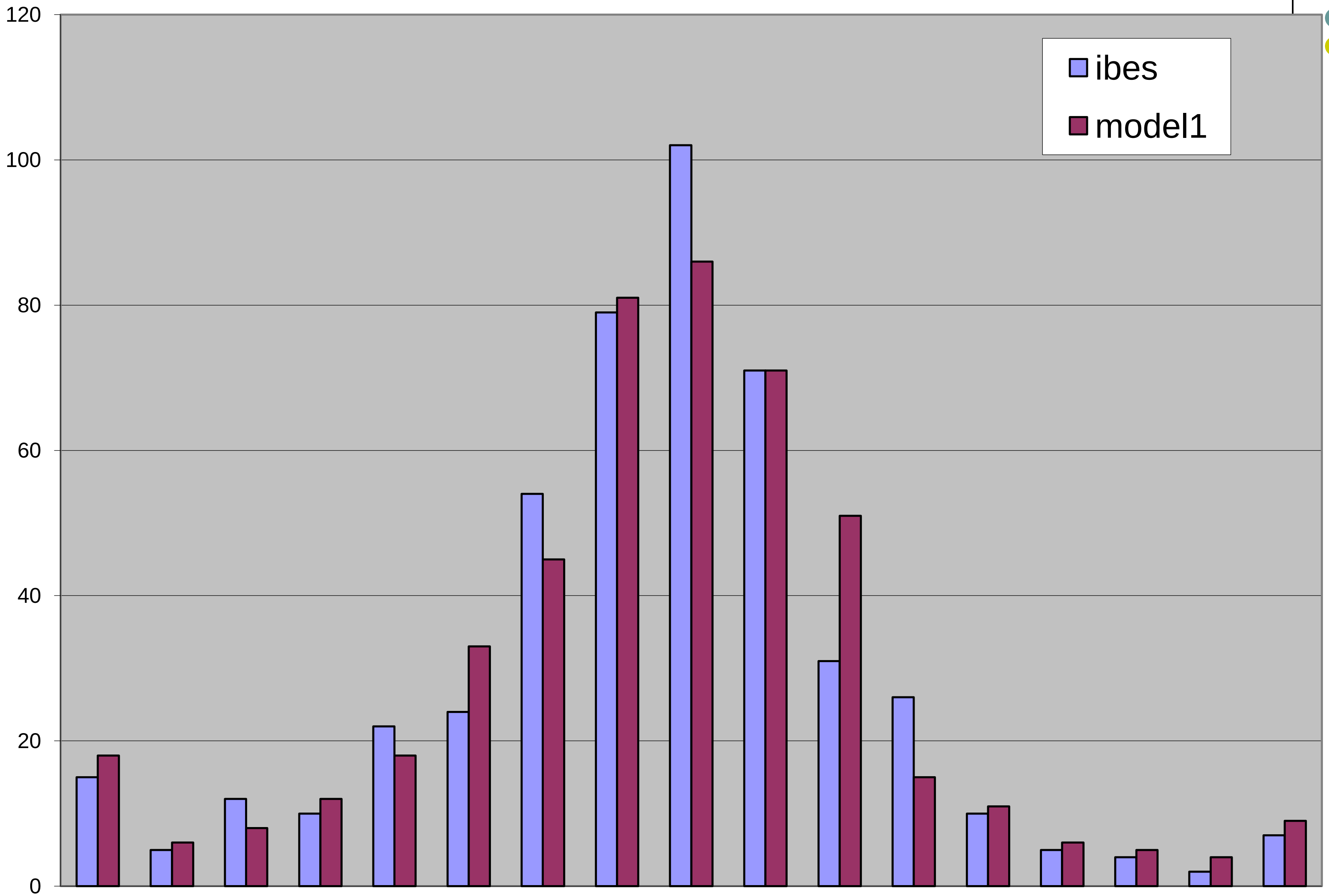
building an out-of-sample forecast

- need the company forecast of next year's # of new stores (already know old and mid)
- need an estimate of the old rate and new rate based on past data (use past 5 years)
- need an estimate of next year's comp
  - after much searching, best model is simply
  - $C_t - .02 = b^*(C_{t-1} - .02)$.

out-of-sample results

	Model 1: $k = 1$	Model 2: $k \in (.8, 1.2)$	Model 3: $(1+g)(1+c)-1$
Foresight Assumption	Median Error	Median Error	Median Error
Perfect Foresight for both Comp Growth and Change in Stores	2.39%	2.80%	2.95%
Company Forecasts of Change in Stores and Comp Growth $= b^*(C_{t-1} - .02)$	3.79%	3.88%	3.81%

model 1 versus analyst forecasts



median IBES analyst absolute forecast error is 3.20% of sales

median model 1 absolute forecast error is 3.79% of sales

incremental value relative to analyst forecasts

Variables	Coefficient (t-statistic)
Intercept	68.02 (1.50)
Model 1 Change in Sales Estimate	0.069 (6.09)
I/B/E/S Change in Sales Estimate	0.743 (45.34)
Adjusted R^2	94.34%
Number of firm-year observations	439

Review of Accounting Studies

Financial Statement Analysis and Valuation: Forecasting Firm and Industry Fundamentals

October 22-23, 2010

Mendoza College of Business The University of Notre Dame

Submission Deadline: May 5, 2010