

The background of the slide features a large, faint watermark of the Rutgers University seal. The seal is circular with a sunburst in the center and the words "RUTGERS UNIVERSITY" around the perimeter. The text "RUTGERS" is prominently displayed in red at the top left of the slide.

**RUTGERS**

THE STATE UNIVERSITY  
OF NEW JERSEY

# **Privacy-preserving Information Sharing within an Audit Firm**

Alexander Kogan & Cheng Yin

Accounting and Information Systems, Rutgers Business School

# Motivation

- **Big Data** – Demands of innovative audit data analysis techniques.
- **Information Sharing** – Information from companies facing similar financial environment (macro-economic cycle, market conditions, tax policies etc.).
- **Industry Specialization** - The same audit firm serves multiple clients competing in the same industry.
- **Privacy** – Avoid violating clients' confidentiality but gain benefits from sharing modified information with other engagement teams.

# Data Availability

- Previous studies only used company specific current data plus publicly available data but did not use contemporaneous data from peer companies
- The current regime requires auditors to protect clients' data confidentiality. Specifically, Rule 301 of the AICPA Code of Professional Conduct (1996) states that “a member in a public practice shall not disclose any confidential client information without specific consent of the client.”
- **However**, Guy and Carmichael (2002) interpreted the Statement of Auditing Standard (SAS) No. 56 and stated that “In circumstances where the auditor specializes in a specific industry, the auditor may use clients' data to develop plausible expectations.

# Findings

- **A generic sharing scheme** – Auditors can benefit from sharing contemporaneous information from peer companies in the same industry compared to original model without any sharing.
- **A residual-based sharing scheme** – Auditors can benefit comparably from sharing prediction residuals (Low-level) as from sharing predicted value (Medium-level) and true numbers (High-level).
- **A categorical sharing scheme** – To satisfy more conservative privacy concerns, we propose a sharing scheme based on categorical information derived from residuals. With a fine-tuned threshold, the sharing scheme achieves a comparable analytical performance.

## Extant Literature 1/3

- The effectiveness and **usefulness of using peer firms** as a benchmark.
  - e.g. Financial analysts use peer firms to support their valuation multiples, earnings forecasts and overall stock recommendations (De Franco et al. 2011).
  - e.g. Investment managers use peer firms in structuring their portfolios (Chan et al. 2007).
  - e.g. Peer firms are used by compensation committees in setting executive compensation (Albuquerque 2009; Albuquerque et al. 2013), in determining valuation multiples (Bhojraj et. al), as well as by auditors in conducting analytical procedures (Hoitash et al. 2006; Minutti-Meza 2013).

## Extant Literature 2/3

- Based on the usefulness of peer firms, previous papers also investigated the way of **choosing peers** (economically-comparable firms). (Hoitash 2006; De Franco et al. 2015; Minutti-Meza, M. 2011).
- A number of both financial accounting and auditing studies examined the importance of information **transfer** and industry expertise in providing high-quality audits.(Balsam et al. 2003; Krishnan 2003; Reichelt and Wang 2010, Owhoso et al. 2002; Solomon et al. 1999; Low 2004; Taylor 2000.)

## Extant Literature 3/3

- Since 1980s, there have been numerous studies concentrating on **improving the performance** of analytical procedures by means of using different data frequency (Wild 1987; Dzeng 1994), by applying more sophisticated statistical models (Dugan et al. 1985; Pany 1990; Leitch and Chen 2003), and by considering multiple companies in similar industries (Allen 1992; AICPA 1988; Lev 1980; Wheeler and Pany) as well as multi-location data (Allen et al. 1999).
- Other studies have examined the error detection performance of analytical procedures in the case of coordinated errors (Leitch and Cheng 2003) and account relations (Vandervelde et al. 2008).

## Data 1/4

- Sample preparation
  - **20** industries that contained the largest number of firms and experienced various sales growth rates from 1991–2015 were selected through **4** digits SIC codes.
  - **Quarterly data** for the total revenues, cost of revenues, accounts receivable, and accounts payable was downloaded from the Compustat fundamentals quarterly database for the period **1991 – 2015**.



# Data 2/4

## Descriptive Statistics Sample Companies from 1991–2015

SIC	Number of Firms	Account Payable	Cost of Goods Sold	Account Receivable	Revenue	Growth Rate
7372	320.00	28.82	35.71	109.56	166.50	14.05%
6798	236.00	68.56	65.47	204.38	105.89	11.22%
1311	216.00	205.40	307.19	202.43	438.93	23.63%
7370	180.00	209.38	245.57	470.39	468.90	19.30%
2834	152.00	160.39	156.02	403.59	597.10	25.22%
3674	140.00	85.14	120.34	142.48	292.63	12.98%
4911	134.00	280.97	516.87	350.11	734.13	4.86%
5812	124.00	48.97	209.97	37.91	287.86	9.52%
7373	120.00	42.22	79.94	112.21	133.65	13.57%
2836	111.00	64.21	47.92	107.46	158.05	26.51%
3845	100.00	14.80	21.99	49.34	62.58	17.13%
4813	99.00	359.07	438.57	662.59	873.85	14.88%
3663	82.00	187.46	292.45	269.43	461.72	11.05%
3841	68.00	34.87	38.31	65.74	92.55	17.46%
9995	67.00	94.34	45.49	154.13	60.18	5.92%
7990	65.00	28.08	100.50	41.79	171.47	14.63%
3714	63.00	244.00	403.37	338.75	497.17	9.75%
6331	62.00	1807.07	975.98	3666.59	1143.98	10.15%
6211	60.00	9534.34	379.00	12763.90	741.76	12.74%
3576	58.00	35.81	96.10	151.49	267.99	7.86%
3661	54.00	15.84	27.89	39.87	54.83	12.98%

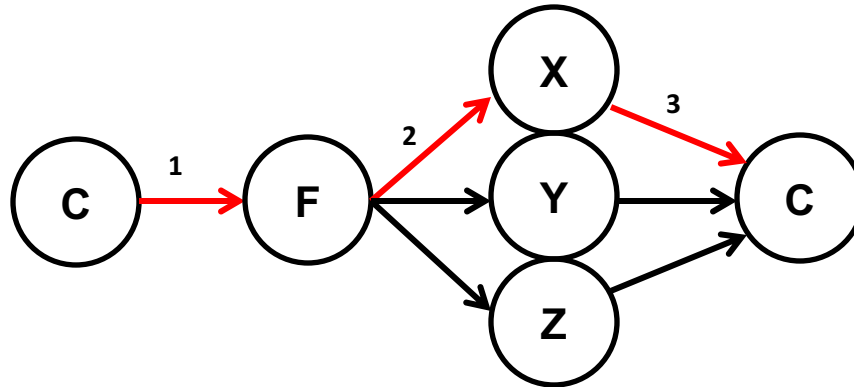
## Data 3/4

- Peer Selection
  - A company is assigned specific peers only if their **size** and **growth rankings** both fall into predefined ranking intervals.
  - If the firm does not have any peers starting in the fourth year to the end, this firm will be removed from the dataset.
  - The peer selection method is relatively crude in comparison to the analysis that can be performed by practicing auditors.

## Data 4/4

- Interpolation
  - Without loss of generality, we follow the **cubic splines interpolation method** introduced in the auditing literature by Chen and Leitch (1998) and Leitch and Chen (1999), and implemented by Hoitash et al. (2006).
- Standardization
  - A standard score is calculated for each company as follows  $z = \frac{y - \mu(y)}{\sigma(y)}$ , where  $y$  represents an interpolated monthly account balance.

# A Generic Sharing Scheme



In this example, C represents an engagement team of client C, F is the national office of C, and X, Y and Z are peer companies of C.

**Step 1** - C passes last year's revenue multiplied by a large number to F.

**Step 2** - F passes a random split of large number received from C to X, Y and Z separately.

**Step 3** - C receives the sum of private contaminated information from X, Y and Z and reduces the known self-generated large number to get the aggregation of private information from X, Y and Z.

# A Residual-Based Sharing Scheme: 3 levels

- Low-level sharing – standardized errors from peer companies
- Medium-level sharing – standardized predicted value from peer companies.
- High-level sharing – standardized true value from peer companies.

Sharing Schemes	Peer Sharing Model
Low Level	$\overline{\varepsilon_{i \neq j}}$ <span style="float: right;"><math>\varepsilon</math></span>
Medium Level	$\hat{y}_p$ <span style="float: right;"><math>\hat{y}</math></span>
High Level	$y_i, i \neq j$ <span style="float: right;"><math>y</math></span>

# Model Specification

$$IND_{XXX_t} = \frac{\sum_1^i Z_i}{i}$$

$$SALE_t = \alpha + \beta_1 SALE_{t-12} + \beta_2 AR_t + \varepsilon_t \quad \mathbf{(1)}$$

$$COGS_t = \alpha + \beta_1 COGS_{t-12} + \beta_2 AP_t + \varepsilon_t \quad \mathbf{(2)}$$

$$SALE_t = \alpha + \beta_1 SALE_{t-12} + \beta_2 AR_t + IND\_ERROR_t + \varepsilon_t \quad \mathbf{(3)}$$

$$COGS_t = \alpha + \beta_1 COGS_{t-12} + \beta_2 AP_t + IND\_ERROR_t + \varepsilon_t \quad \mathbf{(4)}$$

$$SALE_t = \alpha + \beta_1 SALE_{t-12} + \beta_2 AR_t + IND\_PREDICT_t + \varepsilon_t \quad \mathbf{(5)}$$

$$COGS_t = \alpha + \beta_1 COGS_{t-12} + \beta_2 AP_t + IND\_PREDICT_t + \varepsilon_t \quad \mathbf{(6)}$$

$$SALE_t = \alpha + \beta_1 SALE_{t-12} + \beta_2 AR_t + IND\_ACTUAL_t + \varepsilon_t \quad \mathbf{(7)}$$

$$COGS_t = \alpha + \beta_1 COGS_{t-12} + \beta_2 AP_t + IND\_ACTUAL_t + \varepsilon_t \quad \mathbf{(8)}$$

# Model Validation

- We use a simple auto regression as an illustration, but in reality, auditors can choose more **sophisticated** analytics.
- With predictive modeling, the omitted variables would not be a big issue based on the objective of better predictions not reliable coefficients.
- Similarly to “abnormal accruals” and “abnormal audit fees”, it is reasonable to utilize **residuals** as a information channel.
- In evaluation phase, we utilize **cross-validation** method and do not focus on examining collinearity or the fitness of model.

## Tests (Estimation Accuracy)

- Each regression model is estimated over 36 months or three years and is tested over the subsequent 12 months. (rolling regression with **fixed 36 months width moving window**)
- Every model is **estimated separately** for each company based on its unique set of peer companies
- Prediction performance is evaluated based on examining the mean absolute percentage error (**MAPE**) for each account-model.



## Tests (Error Detection) 1/2

- A simulation approach known as “**seeding errors**” is used to compare the anomaly detection capabilities of different models: we add artificial errors (based on the value of original value) into initial value to investigate whether the model will identify the polluted data.
- In the context of this research, the detection capability of models is measured using **cost of errors** determined by three metrics: the number of false negatives and false positives and the magnitude of cost ratio between two types of errors.
- We also test the **impact of magnitude of contaminated errors** ranging from 5% , 2% to 1%.

## Tests (Error Detection) 2/2

- The statistical investigation rule in this paper is to test whether the audit numbers provided by firms fall outside of **prediction interval** (rather than confidence interval). If the value of the prediction exceeds either the upper or lower limits of the PI, then the observation is flagged as an anomaly.
- The size of the prediction interval is determined by the value of the significance level  $\alpha$ . In our research, we use  $z$  to represent the z-values determined by value of  $\alpha$ .
- The entire error seeding procedure is repeated **ten** times to to reduce the variability of results.

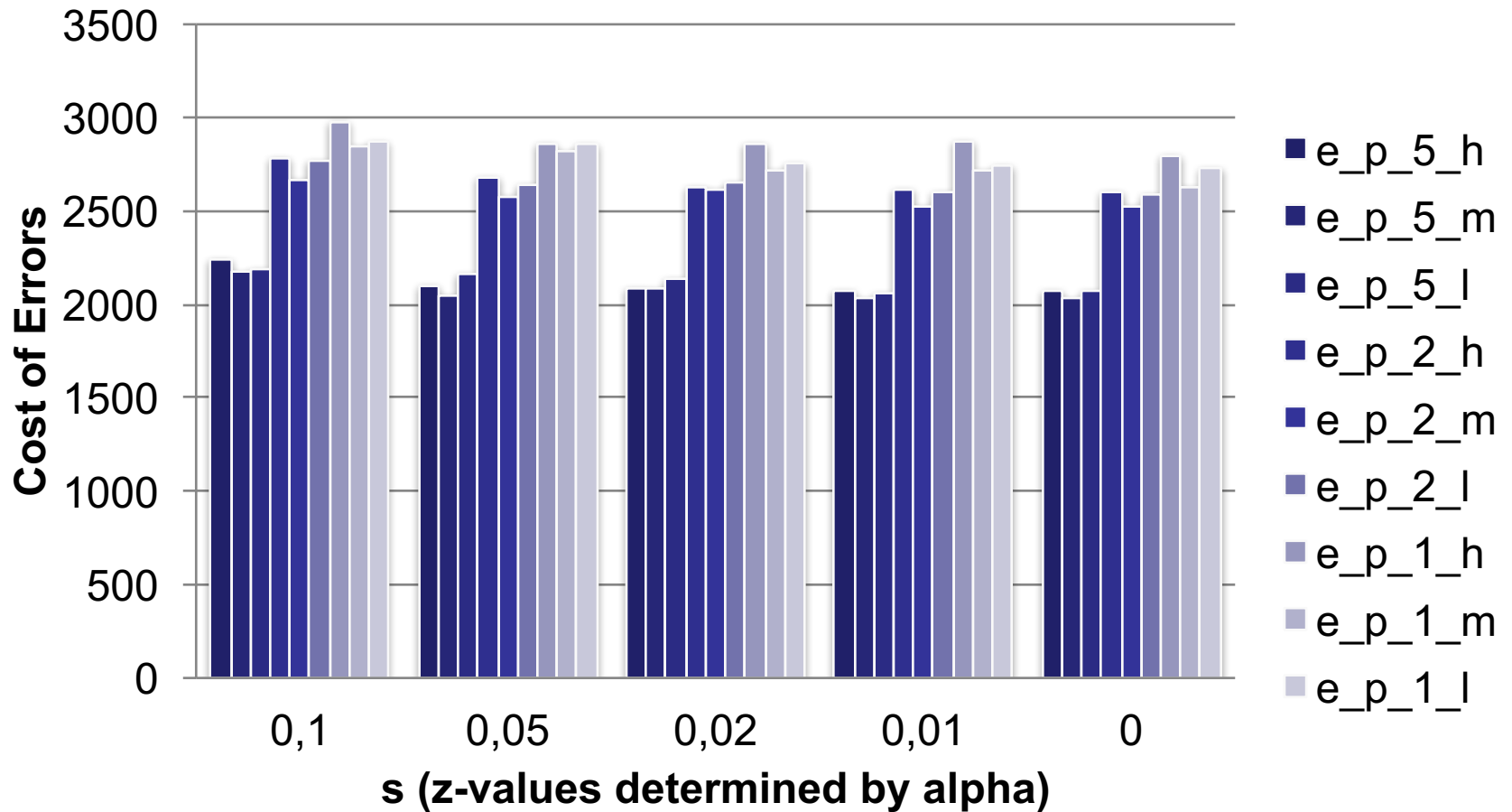
# Comparison of Median of MAPE for Revenue Account

SIC	Number of Firms	Median of Mape from Original Model	Median of Mape from Error Sharing Model	Median of Mape from Prediction Sharing Model	Median of Mape from Actual Sharing Model
7372	316	0.10	0.07	0.07	0.07
1311	212	0.18	0.12	0.13	0.12
7370	180	0.09	0.06	0.06	0.06
2834	150	0.12	0.08	0.08	0.08
3674	140	0.11	0.08	0.08	0.07
4911	126	0.09	0.06	0.07	0.06
5812	121	0.06	0.04	0.04	0.04
7373	120	0.10	0.07	0.07	0.07
2836	111	0.15	0.11	0.12	0.12
3845	100	0.10	0.07	0.07	0.07
4813	98	0.05	0.04	0.04	0.03
3663	82	0.12	0.09	0.09	0.09
4931	73	0.08	0.06	0.07	0.06
3841	68	0.06	0.04	0.04	0.04
9995	67	0.25	0.18	0.20	0.19
7990	65	0.09	0.07	0.07	0.06
3714	63	0.07	0.05	0.06	0.06
6331	62	0.07	0.05	0.05	0.05
6211	59	0.11	0.08	0.08	0.08
3576	58	0.10	0.08	0.08	0.09

***Error Detection Performance Comparison between Peer Models and Original (Benchmark) Model (Overestimating Revenue)***

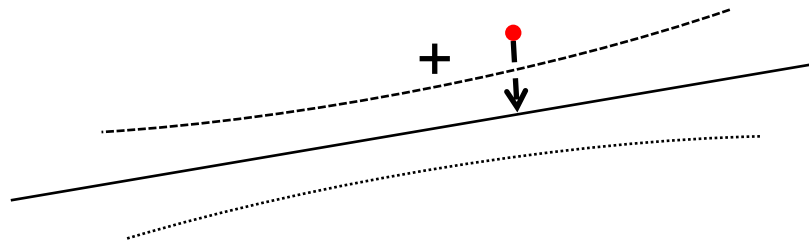
SIC:6211		Original Model		Low-level Sharing Scheme		Medium-level Sharing Scheme		High-level Sharing Scheme	
		FN_o	FP_o	FN_e	FP_e	FN_p	FP_p	FN_a	FP_a
s=0.1	e=0.05	20.41%	24.36%	<b>17.97%</b>	<b>24.51%</b>	<b>17.23%</b>	<b>24.50%</b>	<b>17.98%</b>	<b>24.46%</b>
	e=0.02	23.92%	23.73%	<b>22.82%</b>	<b>24.28%</b>	<b>22.04%</b>	<b>24.48%</b>	<b>22.45%</b>	<b>24.24%</b>
	e=0.01	25.00%	23.37%	<b>24.08%</b>	<b>23.69%</b>	<b>23.83%</b>	<b>24.34%</b>	<b>24.00%</b>	<b>23.87%</b>
s=0.05	e=0.05	19.92%	24.54%	<b>17.39%</b>	<b>25.48%</b>	<b>16.47%</b>	<b>25.75%</b>	<b>16.93%</b>	<b>24.88%</b>
	e=0.02	23.51%	23.96%	<b>21.99%</b>	<b>24.57%</b>	<b>21.10%</b>	<b>25.16%</b>	<b>21.60%</b>	<b>24.37%</b>
	e=0.01	24.52%	24.33%	<b>23.67%</b>	<b>24.72%</b>	<b>22.54%</b>	<b>25.25%</b>	<b>23.40%</b>	<b>24.54%</b>
s=0.02	e=0.05	19.78%	25.35%	<b>16.81%</b>	<b>25.39%</b>	<b>16.43%</b>	<b>26.77%</b>	<b>17.12%</b>	<b>25.64%</b>
	e=0.02	23.26%	24.30%	<b>22.26%</b>	<b>25.06%</b>	<b>21.28%</b>	<b>26.04%</b>	<b>21.49%</b>	<b>24.89%</b>
	e=0.01	23.99%	25.44%	<b>23.04%</b>	<b>26.09%</b>	<b>22.22%</b>	<b>27.04%</b>	<b>22.98%</b>	<b>25.96%</b>
s=0.01	e=0.05	18.83%	24.90%	<b>16.22%</b>	<b>25.16%</b>	<b>15.97%</b>	<b>26.28%</b>	<b>16.27%</b>	<b>25.11%</b>
	e=0.02	22.25%	24.96%	<b>20.85%</b>	<b>25.44%</b>	<b>20.49%</b>	<b>26.78%</b>	<b>20.62%</b>	<b>25.36%</b>
	e=0.01	24.54%	24.73%	<b>23.46%</b>	<b>25.18%</b>	<b>22.28%</b>	<b>26.06%</b>	<b>23.11%</b>	<b>24.91%</b>
s=0	e=0.05	19.23%	25.40%	<b>16.41%</b>	<b>25.93%</b>	<b>15.95%</b>	<b>26.60%</b>	<b>16.29%</b>	<b>25.56%</b>
	e=0.02	22.86%	25.31%	<b>21.33%</b>	<b>25.77%</b>	<b>20.40%</b>	<b>26.58%</b>	<b>20.85%</b>	<b>25.67%</b>
	e=0.01	23.14%	24.84%	<b>21.96%</b>	<b>25.29%</b>	<b>21.41%</b>	<b>26.68%</b>	<b>21.97%</b>	<b>25.24%</b>

# Comparison of Cost of Errors from Different Sharing Schemes - SIC 6211 (Cost Ratio = 1:10)



# A Categorical Sharing Scheme: Additional Test

- To cater for a more conservative demand of privacy, we propose a sharing scheme based on categorical information derived from residuals.
- We empirically test the estimation performance by using **the sign of prediction errors and the level of deviations**.
  - The sign of prediction errors (Over & Underestimate)
  - The level of deviation (Prediction Interval, Sigma)
  - The threshold (sensitive & insensitive)



Comparison of the mean of MAPEs from sharing categorical information with different threshold in revenue account											
SIC	Mean of MAPE from sharing the sign of prediction	P-value of t-test	Mean of MAPE from sharing the level of deviation ( $\delta=3$ )	P-value of t test	Mean of MAPE from sharing both categorical information ( $\delta=3$ )	P-value of t test	Mean of MAPE from sharing the level of deviation ( $\delta=1$ )	P-value of t test	Mean of MAPE from sharing both categorical information ( $\delta=1$ )	P-value of t test	
7372	0.53	0.8289	0.51	0.6812	0.64	0.9977	0.44	0.1959	0.42	0.0898	
1311	1.02	0.9330	0.76	0.0065	0.90	0.4661	0.85	0.0927	0.63	0.0000	
7370	0.28	0.0198	0.29	0.0589	0.45	0.8538	0.40	0.1391	0.38	0.0761	
2834	2.20	0.9227	1.32	0.8847	2.18	0.8863	1.57	0.8731	1.28	0.8458	
3674	0.22	0.0404	0.24	0.8131	0.26	0.9978	0.26	0.9875	0.17	0.0000	
4911	0.18	0.6991	0.19	1.0000	0.20	1.0000	0.17	0.3741	0.15	0.0000	
5812	0.10	0.8171	0.10	0.7552	0.10	0.8911	0.10	0.9397	0.08	0.0000	
7373	0.20	0.2096	0.20	0.3480	0.20	0.3658	0.31	0.9683	0.23	0.6526	
2836	1.68	0.9148	2.16	0.9909	1.18	0.0250	1.79	0.9742	1.11	0.0014	
3845	0.23	0.0001	0.25	0.5566	0.24	0.3347	0.26	0.7873	0.20	0.0000	
4813	0.10	0.5033	0.10	0.3938	0.11	0.9679	0.08	0.0000	0.08	0.0000	
3663	0.23	0.2414	0.26	0.9806	0.22	0.0463	0.22	0.0263	0.18	0.0000	
4931	0.15	0.4779	0.18	1.0000	0.18	1.0000	0.15	0.2029	0.13	0.0000	
3841	0.10	0.0000	0.10	0.0000	0.10	0.0000	0.10	0.0000	0.09	0.0000	
9995	0.63	0.0005	0.71	0.0222	0.75	0.1050	0.60	0.0001	0.49	0.0000	
7990	0.21	0.0174	0.23	0.0338	0.21	0.0125	0.20	0.0060	0.18	0.0002	
3714	0.12	0.0000	0.14	0.8461	0.13	0.1190	0.12	0.0003	0.11	0.0000	
6331	0.13	0.0012	0.17	0.9999	0.15	0.6741	0.13	0.0001	0.12	0.0000	
6211	0.56	0.0281	0.53	0.0176	0.42	0.0227	0.34	0.0137	0.30	0.0064	
3576	0.20	0.9814	0.25	1.0000	0.20	0.9949	0.21	0.9983	0.18	0.2327	

The error detection performance of categorical sharing scheme compared with original model and low-level sharing scheme in revenue account												
s	e	obs	fno_r	fpo_r	fns_r	fps_r	fnl_r	fpl_r	fne_r	fpe_r	fnm_r	fpm_r
<b>0.1</b>	0.05	10920	20.71%	23.32%	18.49%	24.36%	18.85%	22.10%	17.81%	23.45%	18.62%	22.77%
	0.02	10920	23.96%	23.56%	22.19%	24.62%	24.29%	22.79%	23.36%	24.06%	23.71%	23.25%
	0.01	10920	25.23%	23.03%	24.15%	24.16%	25.56%	22.03%	24.46%	23.32%	25.27%	22.56%
<b>0.05</b>	0.05	10920	19.79%	24.67%	17.66%	25.66%	18.57%	24.08%	17.11%	25.27%	18.21%	24.52%
	0.02	10920	23.81%	24.55%	22.10%	25.33%	22.64%	23.58%	22.01%	24.55%	22.96%	24.01%
	0.01	10920	24.13%	24.32%	23.07%	25.49%	24.29%	23.75%	23.09%	24.75%	24.16%	24.22%
<b>0.02</b>	0.05	10920	24.12%	24.24%	22.70%	25.33%	23.74%	23.85%	23.06%	24.99%	23.76%	24.29%
	0.02	10920	22.85%	24.88%	21.46%	25.95%	22.46%	24.72%	22.07%	26.03%	22.27%	25.01%
	0.01	10920	24.24%	24.66%	22.86%	25.66%	23.94%	23.99%	23.21%	25.20%	23.79%	24.50%
<b>0.01</b>	0.05	10920	19.15%	25.20%	17.05%	26.15%	17.65%	25.01%	16.41%	25.90%	17.70%	25.27%
	0.01	10920	23.00%	25.04%	21.67%	25.83%	21.64%	24.38%	21.52%	25.24%	21.85%	24.60%
	0.01	10920	23.74%	24.85%	22.52%	26.09%	23.40%	24.48%	22.54%	25.41%	23.50%	24.95%
<b>0</b>	0.05	10920	19.52%	24.72%	17.13%	25.69%	17.83%	24.24%	16.83%	25.47%	17.89%	24.74%
	0.02	10920	23.30%	25.70%	21.58%	26.24%	21.92%	24.83%	21.50%	25.87%	21.97%	25.29%
	0.01	10920	23.69%	26.10%	22.25%	26.88%	23.42%	25.75%	22.50%	26.64%	22.90%	25.91%



# Potential Risks

- Sample Selection Risk
  - Companies with no peers after the first three years are dropped from the sample but still need to be audited. The firms that do not have uninterrupted peers error are also removed from our sample.
- Simulated Data Risk
  - The results in this study should be carefully applied. The findings in this study are based on interpolated data points and not on real data. Thus there may be some abnormal data in our data set causing serious problems (outlier values of MAPE). While these results support our conjectures of general prediction improvements and the superior performance of error detections.

A large, faint watermark of the Rutgers University seal is visible in the background. The seal features a central sunburst design surrounded by a circular border containing the text "RUTGERS THE STATE UNIVERSITY OF NEW JERSEY".

# RUTGERS

THE STATE UNIVERSITY  
OF NEW JERSEY

Thank you !

A large, faint watermark of the Rutgers University seal is visible in the background. The seal features a central sunburst design surrounded by a circular border containing the text "RUTGERS THE STATE UNIVERSITY OF NEW JERSEY".

RUTGERS

THE STATE UNIVERSITY  
OF NEW JERSEY

## Appendix

# Research Questions

- The first research question examines whether models with aggregated peer information lead to different mean absolute percentage error (MAPE) in comparison to models that do not incorporate peer data. (Estimation Accuracy)
- The purpose of the second research question is to test the error detection performance regarding different information sharing schemes. (Error Detection Performance)

# Hypothesis Development 1/2

- Estimation Accuracy
- **Hypothesis 1:** The MAPEs generated from peer sharing model is less than those came from self-prediction model.
- **Hypothesis 2:** The MAPEs generated from peer sharing models with high level sharing scheme will result in a comparable fashion as peer sharing models with lower sharing and medium sharing schemes

## Hypothesis Development 2/2

- Error Detection Performance
- **Hypothesis 3:** The peer sharing models will be superior over original models without any sharing information.
- **Hypothesis 4:** The peer sharing models among different levels sharing schemes will perform comparable outputs.

# Data

- Sample preparation 2/2
  - Firms should have complete data without missing and zero values.
  - Firms should have uninterrupted quarterly data for five years.
  - Firms should have year-to-year sales growth of no more than 500 percent.
  - Firms should be eliminated from our sample if there is an acquisition or merger.
  - SIC 6798 (REIT) was removed from the sample due to its idiosyncratic characteristics.

Our final sample includes **7,516** quarterly observations.

# Data

- Peer Selection 2/2

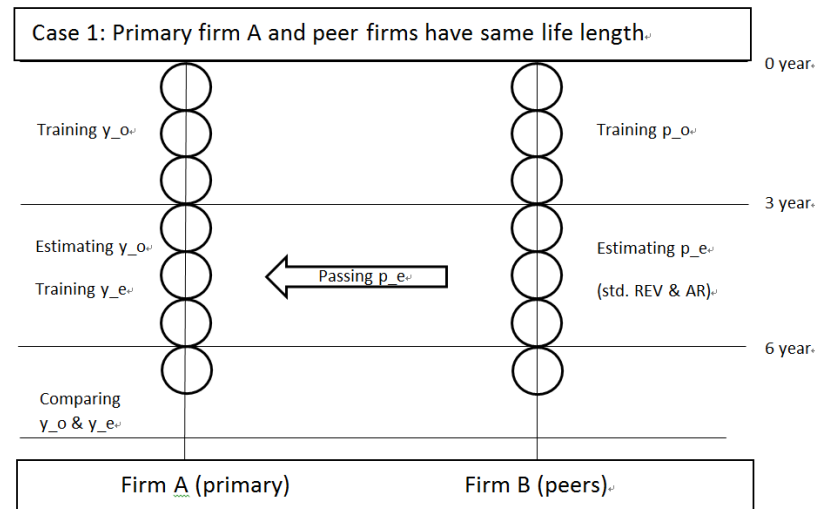
Table 2. An illustration of the Peer Selection Criteria  
(1996, SIC=2821, 17 firms, S: n/5, G: n/4)

Year	Company ID	rank_r(revenue)	rank_g (growth rate)	Selected Peers
1996	A	6	4	/
1996	B	1	5	G
1996	C	5	1	/
1996	D	3	3	/
1996	E	7	2	/
1996	F	4	7	/
1996	G	2	6	B

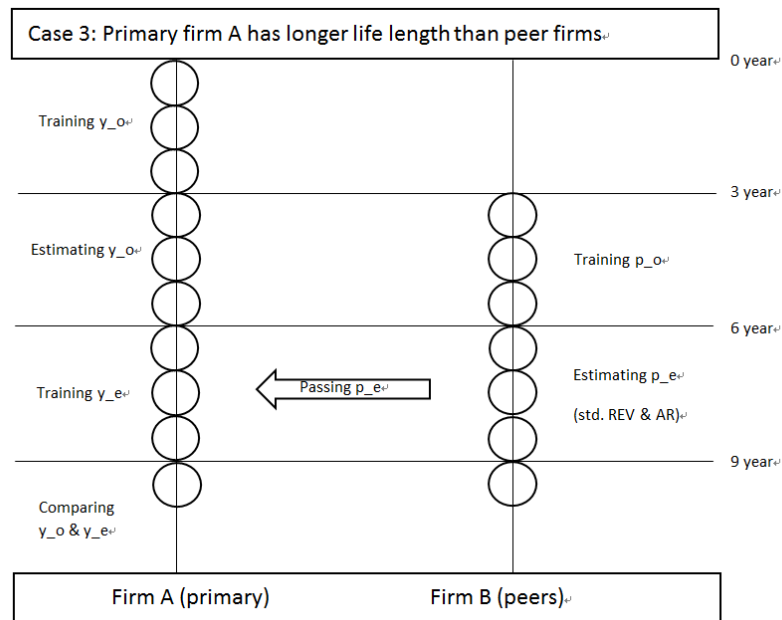
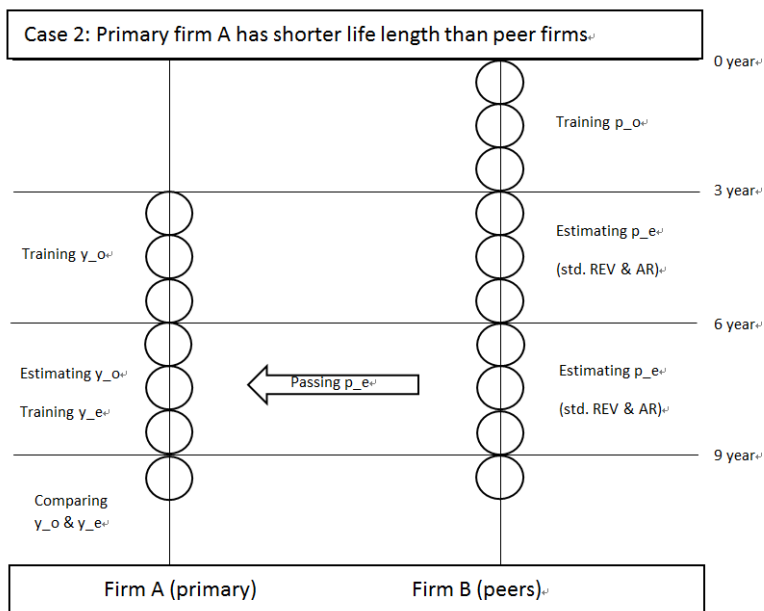


## Tests (Estimation Accuracy) 2/3

- Considering different life length of primary firm and peer firms, the estimation cases will be different and the detail illustration is presented in the following.



# Tests (Estimation Accuracy) 3/3



# Results Interpretations (Estimation Accuracy)

Table 6. The Comparison of Mean of MAPE for COGS Account

SIC	Number of Firms	Mean of Mape from Original Model	Mean of Mape from Error Sharing Model	Mean of Mape from Prediction Sharing Model	Mean of Mape from Actual Sharing Model	Min	Actual	Prediction	Error	Original
7372	316	1.57	0.73	0.73	0.73	0.73	1	1	1	0
1311	212	4.54	2.00	1.85	2.06	1.85	0	1	0	0
7370	180	0.77	0.45	0.41	0.34	0.34	1	0	0	0
2834	150	0.58	0.42	0.34	0.38	0.34	0	1	0	0
3674	140	0.39	0.25	0.23	0.24	0.23	0	1	0	0
4911	126	0.31	0.18	0.19	0.18	0.18	1	0	1	0
5812	121	0.11	0.08	0.08	0.08	0.08	1	1	1	0
7373	120	0.33	0.20	0.21	0.20	0.20	1	0	1	0
2836	111	1.92	0.85	0.85	0.82	0.82	1	0	0	0
3845	100	0.53	0.31	0.29	0.31	0.29	0	1	0	0
4813	98	0.31	0.21	0.22	0.21	0.21	1	0	1	0
3663	82	0.28	0.21	0.20	0.20	0.20	1	1	0	0
4931	73	0.25	0.14	0.15	0.14	0.14	1	0	1	0
3841	68	0.44	0.29	0.24	0.28	0.24	0	1	0	0
9995	67	0.90	0.60	0.60	0.54	0.54	1	0	0	0
7990	65	0.30	0.19	0.16	0.17	0.16	0	1	0	0
3714	63	0.20	0.13	0.13	0.13	0.13	1	1	0	0
6331	62	0.30	0.21	0.20	0.21	0.20	0	1	0	0
6211	59	0.26	0.17	0.16	0.17	0.16	0	1	0	0
3576	58	0.27	0.18	0.18	0.18	0.18	1	1	1	0
							12	13	7	0

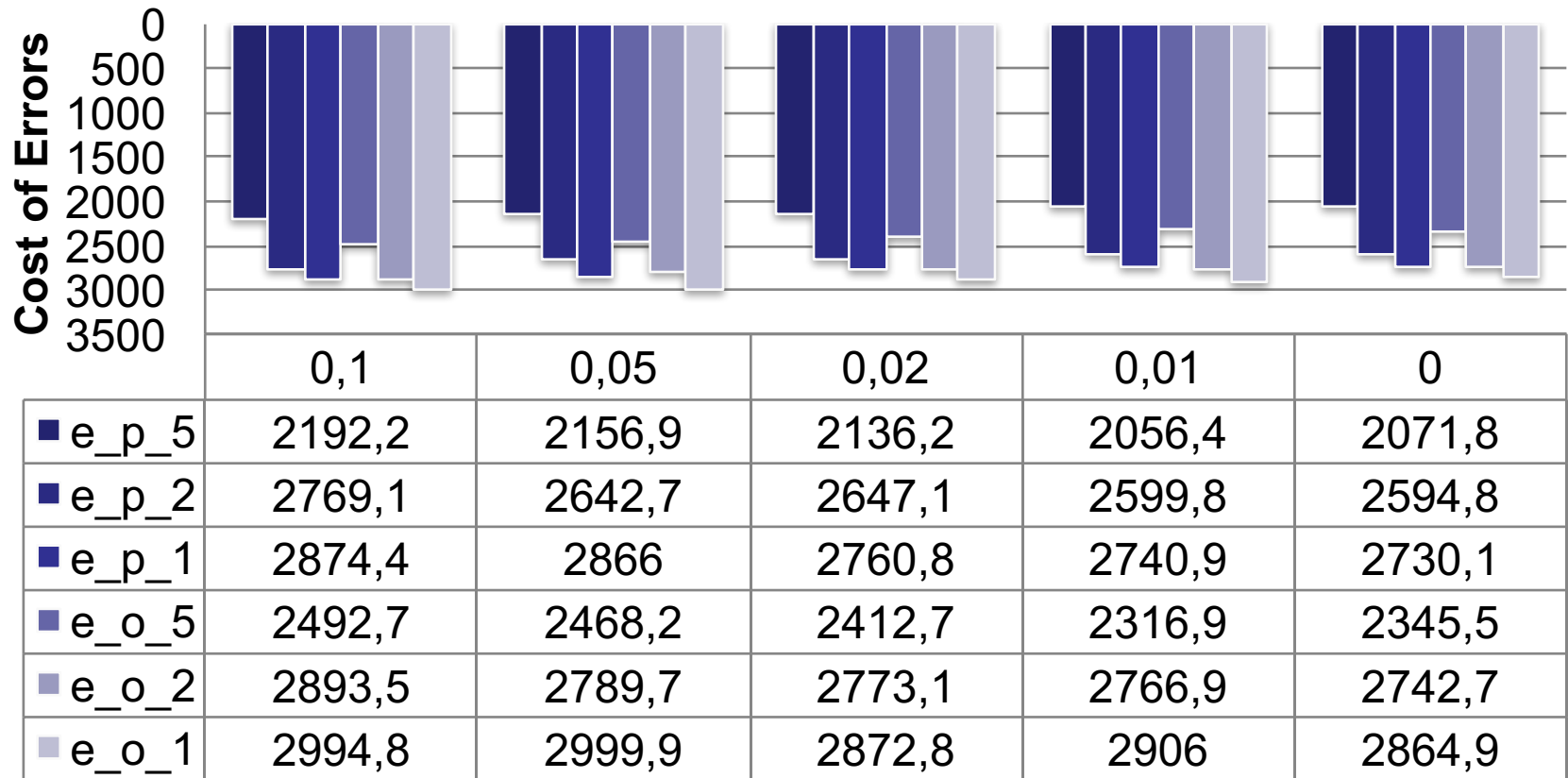
# Results Interpretations (Estimation Accuracy)

Table 8. The Comparison of Median of MAPE for COGS Account

SIC	Number of Firms	Median of Mape from Original Model	Median of Mape from Error Sharing Model	Median of Mape from Prediction Sharing Model	Median of Mape from Actual Sharing Model	Min	Actual	Prediction	Error	Original
7372	316	0.18	0.12	0.12	0.12	0.12	1	1	1	0
1311	212	0.28	0.20	0.20	0.20	0.20	1	1	1	0
7370	180	0.14	0.09	0.09	0.09	0.09	1	1	1	0
2834	150	0.16	0.11	0.11	0.11	0.11	1	1	1	0
3674	140	0.14	0.10	0.10	0.09	0.09	1	0	0	0
4911	126	0.11	0.08	0.09	0.08	0.08	1	0	1	0
5812	121	0.06	0.04	0.04	0.04	0.04	1	1	1	0
7373	120	0.11	0.08	0.08	0.08	0.08	1	1	1	0
2836	111	0.18	0.12	0.12	0.12	0.12	1	1	1	0
3845	100	0.14	0.10	0.10	0.10	0.10	1	1	1	0
4813	98	0.11	0.07	0.08	0.08	0.07	0	0	1	0
3663	82	0.15	0.11	0.10	0.10	0.10	1	1	0	0
4931	73	0.10	0.08	0.08	0.07	0.07	1	0	0	0
3841	68	0.11	0.07	0.07	0.07	0.07	1	1	1	0
9995	67	0.31	0.20	0.20	0.20	0.20	1	1	1	0
7990	65	0.09	0.07	0.07	0.06	0.06	1	0	0	0
3714	63	0.08	0.06	0.06	0.06	0.06	1	1	1	0
6331	62	0.10	0.07	0.07	0.07	0.07	1	1	1	0
6211	59	0.13	0.09	0.09	0.09	0.09	1	1	1	0
3576	58	0.14	0.11	0.09	0.10	0.09	0	1	0	0
							18	15	15	0

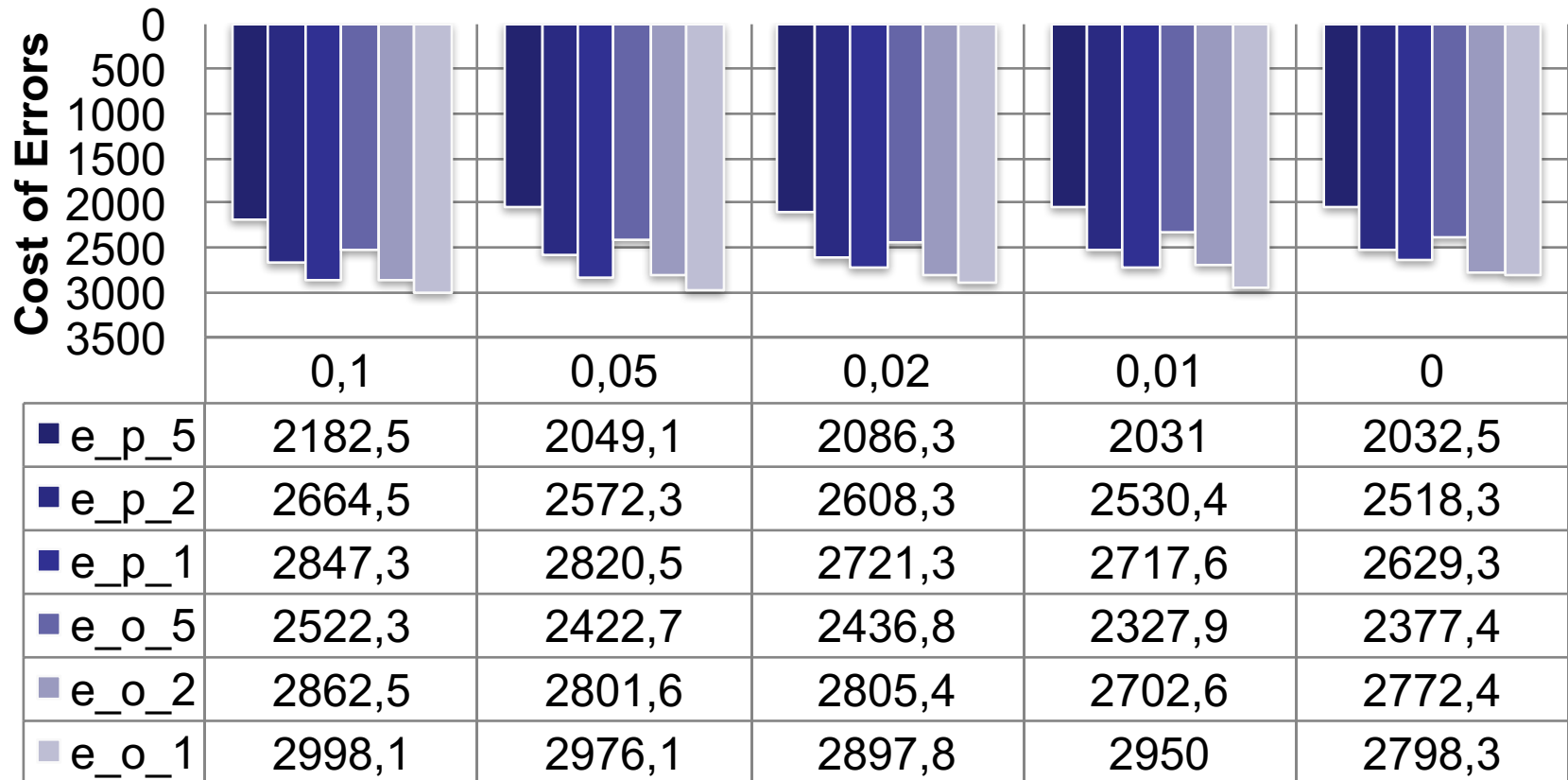
# Results Interpretations (Error Detection)

Comparison of Cost of Errors by Low-Level Sharing Scheme - SIC 6211  
Cost Ratio = 1: 10



# Results Interpretations (Error Detection)

Comparison of Cost of Errors by Medium-Level Sharing Scheme - SIC 6211  
Cost Ratio = 1: 10



# Results Interpretations (Error Detection)

Comparison of Cost of Errors by High-Level Sharing Scheme - SIC 6211  
Cost Ratio = 1: 10

