Update on RA Project 3/30/2015

Summary of Holdings Data for 1908, 1911, and 1913

I. Matching Methodology

1) Examined data to manually find/fix common mistakes (ex. incorrectly iterating manual year) and cleaned those common mistakes (most of the time required on this update). This included cleaning company names, coupons, maturity, par value of invested capital, and investor state. Investor city and investor name were not cleaned.

2) Company names were matched within a given year based on what was entered, since naming was consistent within a year.

3) Company names were not precise across years, so matched across years in excel using Microsoft's proprietary fuzzy logic matching algorithm. The best match across years was entered as well as the level of confidence for the score. It appears that scores of 90% and higher tend to be very accurate, so this is used as a cut-off for a viable match. If two company names have a match of 90% or higher (say two companies from 1911 match a company name from 1908), then the higher match is used and the other is designated as a new firm. One concern with this methodology is that similar company names are likely parents/subsidiaries which we would be missing. To do this very accurately at some point we need to create a map of all firm relationships.

4) Within each company information on security are used to match security identifiers across years. The key variable used are the coupon range and maturity range. Oftentimes one entry will cover multiple securities so these were converted from text (ex. "4s to 6.5s") to minimum and maximum values (Ex. low = 4 and high=6.5) and then any bonds in that range in subsequent years are matched.

5) Investor names were matched within a given year and across years using the same fuzzy logic algorithm where all investor names with an 85% match with each other are labeled as one investor. They still retain listed investor location (some investors have multiple locations/branches).

II. Across Year Matching Results

Firm-Level Matches (Weighted by $)			
	1908	1911	1913
1908	1	87%	80%
1911		1	79%
1913			1

Firm-Level Matches (Equal weight)			
	1908	1911	1913
1908	1	58%	44%
1911		1	47%
1913			1

Issue-Level Matches (Weighted by $)			
	1908	1911	1913
1908	1	61%	55%
1911		1	62%
1913			1

Issue-Level Matches (Equal weight)			
	1908	1911	1913
1908	1	35%	26%
1911		1	38%
1913			1

Discussion:

- Using a 90% cut-off the matching results overall look fairly reasonable.
- Weighting by invested capital there is a 87% match for firm names between 1908 and 1911, which falls to 80% between 1908 and 1913. This is exactly what we would expect if some firms with bonds expire/delist/merging/renaming.
- The match is only 58% and 44% respectively equally weighted which is consistent with larger firms surviving more years (ex. Smaller firms could also be bought out by large parent companies which could confuse this analysis).
- Results are qualitatively similar looking at security-level matches which is comforting.

III. Explaining missing matches

This is to address Eric Hilt's excellent suggestion:

For the issue-level matches of securities you did over time (that's in the first table), does that low level of matches between 1908 and 1913 make sense, given the maturities of the securities? In other words, is the proportion of issues that disappear equal to the proportion that is of 5 years maturity or less?

# Bonds w/ Maturity in 1908	13,342
# Missing in 1913	9,624
% Missing in 1913	72%
% of Missing w/ Mat Year <=1913	50%
% of Missing w/ Mat Year <=1913 or # Investors <5	95%

As you can see from the above table while a large number of bonds in 1908 are missing in 1913 (~9.6k) of those 95% are either bonds who have matured or bonds where we have less than 5 investors in our data. For bonds with less than 5 investors it is fairly reasonable that investors not included in the holdings manuals could have taken those positions, which is why they disappear. We can get an even clearer picture once we match these securities to the amount outstanding so we can see how much of the bonds ownership we see.

IV. Summary Statistics

	1908	1911	1913
# Obs	68,640	70,270	75,368
# Firms	10,905	9,363	9,087
# Securities	22,144	21,090	22,231
$ Assets (bil)	$3.65	$4.54	$4.96
# Investors	6,851	11,022	8,844
$/Investor	532,769	411,903	560,832
$/Firm	334,709	484,887	545,835
# Investors/Firm	0.63	1.18	0.97

Discussion:

- Total value of assets increased, but more securities weren't listed, the value just rose.
- Investors didn't own more, instead the number of investors per issue increased (almost doubled) and they actually took smaller positions. This would be consistent with smaller investors entering the market post-ratings, which would be consistent with a reduction in information asymmetry.
- The reduced concentration of ownership could also be an interesting area to explore more
- One minor concern with this analysis is the investor identifiers are based on exact matches. Company names are consistent within a manual year, but it isn't obvious that investor name will be. This means changes in the number of investors could be driven by time trends in consistency of investor naming.

V. Regional Breakdown over time

Rank	Region	1908 $ Invested	1908 % of Total Assets	1911 $ Invested	1911 % of Total Assets	1913 $ Invested	1913 % of Total Assets
1	ny	1,705,227,392	47%	2,262,639,104	50%	2,406,004,992	49%
2	pa	461,659,360	59%	522,360,384	61%	569,328,768	60%
3	ma	361,147,392	69%	429,317,888	71%	458,551,456	69%
4	cn	316,700,320	78%	329,199,808	78%	425,537,248	78%
5	nj	245,996,592	85%	283,619,040	84%	346,437,216	85%
6	ri	87,410,440	87%	91,001,144	86%	97,484,696	87%
7	me	83,610,864	89%	94,692,864	88%	96,485,216	89%
8	wi	73,991,272	91%	80,333,904	90%	91,462,144	91%
9	ca	72,019,536	93%	64,777,472	92%	32,174,330	91%
10	canada	53,090,840	95%	137,138,800	95%	107,313,792	93%
Total		3,646,282,116		4,540,264,874		4,956,837,973	

Rank	Region	# Investors 1908	# Investors 1911	# Investors 1913	% of Total 1908	% of Total 1911	% of Total 1913	Cum % of Total 1908	Cum % of Total 1911	Cum % of Total 1913
1	ny	510	616	600	17%	18%	20%	17%	18%	20%
2	pa	506	576	572	17%	17%	19%	34%	35%	39%
3	ma	340	383	355	11%	11%	12%	45%	46%	51%
4	ca	282	284	92	9%	8%	3%	55%	55%	54%
5	nj	197	234	197	7%	7%	7%	61%	62%	61%
6	cn	179	203	213	6%	6%	7%	67%	67%	68%
7	nh	166	115	111	6%	3%	4%	73%	71%	72%
8	me	142	151	96	5%	4%	3%	78%	75%	75%
9	ri	110	88	74	4%	3%	2%	81%	78%	78%
10	canada	80	122	119	3%	4%	4%	84%	81%	82%
11	vt	73	71	60	2%	2%	2%	86%	84%	84%
12	il	43	71	62	1%	2%	2%	88%	86%	86%
13	england	36	41	50	1%	1%	2%	89%	87%	87%
14	mi	36	39	24	1%	1%	1%	90%	88%	88%
15	va	35	31	22	1%	1%	1%	91%	89%	89%
16	oh	33	48	41	1%	1%	1%	93%	90%	90%
17	md	26	62	46	1%	2%	2%	93%	92%	92%
18	in	25	37	26	1%	1%	1%	94%	93%	93%
19	la	15	14	13	1%	0%	0%	95%	94%	93%
20	de	13	12	10	0%	0%	0%	95%	94%	94%
21	mn	12	17	9	0%	0%	0%	96%	95%	94%
22	mo	12	39	24	0%	1%	1%	96%	96%	95%
23	ne	11	15	10	0%	0%	0%	96%	96%	95%
24	wi	11	11	9	0%	0%	0%	97%	96%	95%
25	nc	10	11	10	0%	0%	0%	97%	97%	96%

Discussion:

- Top 7 are in New England are around 90% of all invested capital and 70% of investors in all years.
- As expected New York has about 4 times the invested capital of any other location and represent about half of all invested capital. This is despite the fact that Pennsylvania has a larger # of investors, but with average positions less than 4-6 times smaller throughout the years.
- In this extremely simple description there isn't much change in the relative amount of investors in NY (or even New England). This doesn't adjust for changes in firms/securities listed or their headquarter locations.

- As with the summary statistics we see that the total investors increase overall and the $/investor falls in 1911 (the notable exception is Canada which has increases in #, $, and $/investor).

5. 2 Addressing the following questions from Eric:

- Which states' institutions are actually included in the securities holdings volumes? I assume it's not all 48. Which ones are excluded? And, equally important, does the coverage change over time?

- Within each state, which types of institutions are included (eg savings banks, insurance companies, state banks, trust companies, etc?) I know that many states will have idiosyncratic laws that might lead to hybrid institutions (savings banks + trust company in same firm) and some states might not even have insurance companies (who knows) but it would be good to know what's actually covered in that volume. I assume it's a compilation of data published in state reports and there may have been states that simply did not publish such data for some or all of their institutions.

	# Investors			Chg # Investors	
	1908	**1911**	**1913**	**1911-1908**	**1913-1911**
All States	3147	3514	3020	367	-494
All States (w/ State IDs)	2992	3402	2973	410	-429
Top 10 States as of 1908	2512	2772	2429	260	-343
These don't match aggregates precisely since investors can be in multiple regions.					

Response:

1) It looks like the top 10 are as informative as the whole sample which suggests to me it isn't about some types included that weren't before, but we can look at this in more depth by examining the investors themselves:

VI. Investors

6.1 Matching (using fuzzy logic):

1) Start with 214,650 observations. Lose 404 which didn't have investor name entered (I'm removing these. Obviously given tiny size out of the 214,650 observations they shouldn't matter).

2) This becomes 19,367 investors when look at exact matches.

3) Becomes 12,510 investors with manual fixes (commas, spaces, abbreviations, common mistakes, etc...).

4) Fuzzy logic matching based on 85% match takes this down to 3,190 investors (a bit more complex than before since creating optimal matching "groups", but similar idea). Manual work could probably put this just below 3k, but it would be a decent amount of work and would be hard to be sure they weren't actually different without a "master list".

6.2 Revised Summary Statistics:

	1908	1911	1913
# Obs	68,640	70,270	75,368
# Firms	10,905	9,363	9,087
# Securities	22,144	21,090	22,231
$ Assets (bil)	$3.65	$4.54	$4.96
# Investors	1,583	1,860	1,781
$/Investor	2,305,749	2,440,860	2,784,952
$/Firm	334,709	484,887	545,835
# Investors/Firm	0.15	0.20	0.20

The summary statistics change since the # of investors falls, but the earlier interpretation qualitatively still holds.

6.3 Investor Type:

There are 6 investor types:

1) Trusts
2) savings banks
3) state banks
4) Insurance companies
5) other banks
6) other investors

Investor type rules:

1) If includes word "trust" it is a trust.

2) If includes "savings" and not the word "trust" then it is a "savings bank".
3) If it includes "state" and "bank" and not words in 1)-2) then it is a "state bank".
4) If includes "insurance", "indemnity", or "assurance" and not key words from 1-3 it is an insurance company.
5) If doesn't include any key words in 1)-4), but does include "bank" then it is "other banks".
6) If doesn't fit any of the criteria for 1-5 it is "other investors".

of Investors by Investor Type each Year

	# of Investors		
	1908	1911	1913
savings bank	485	613	588
insurance	467	636	644
trust	365	467	457
other bank	232	255	177
other investor	196	298	300
state bank	12	15	8

Based on the breakdown it looks like the majority of investors are banks, trusts, or insurance companies (as expected). The # of trusts, insurance companies, savings banks, and other investors increase over this period, while "other banks" stay relatively flat.

6.4 Investor Type by Region

of Investors by Investor Type and State each Year

	# Investors														
	Insurance			Other Bank			Other Investor			Savings Bank			Trusts		
	1908	1911	1913	1908	1911	1913	1908	1911	1913	1908	1911	1913	1908	1911	1913
All States	901	1044	990	273	287	192	293	391	357	810	953	781	709	710	645
All States (w/ State IDs)	898	1044	990	273	287	192	291	391	357	809	953	781	708	710	645
Top 10 States	638	701	652	264	272	185	218	285	267	743	858	739	636	639	578

		Insurance			Other Bank			Other Investor			Savings Bank			Trusts		
Rank	Region	1908	1911	1913	1908	1911	1913	1908	1911	1913	1908	1911	1913	1908	1911	1913
1	ny	175	211	208	15	8	13	95	104	115	85	164	149	140	129	115
2	pa	114	99	111	57	74	86	36	59	52	69	70	63	227	269	254
3	ma	88	102	94	14	10	8	16	30	26	200	219	214	22	22	13
4	ca	15	17	10	149	157	52	11	8	1	65	64	17	33	27	10
5	nj	39	37	37	3	3	3	17	19	10	47	64	33	91	111	114
6	cn	49	71	64	11	8	9	9	14	16	87	87	103	23	22	21
7	nh	43	21	15	6	6	6	10	6	3	65	60	67	42	22	20
8	me	24	24	9	3	1	2	4	8	3	92	105	73	19	13	9
9	canada	52	81	75				17	33	39	7	5	2	4	3	3
10	ri	39	38	29	6	5	6	3	4	2	26	20	18	35	21	19
11	vt	9	12	5	3	2	2	3	1	2	31	25	17	27	31	34
12	il	28	38	34		2		9	14	15	2	7	5	4	10	8
13	md	11	32	28	1			9	17	10	1	10	3	4	3	5
14	england	26	33	40				7	7	8	1	1	2	2		
15	oh	22	33	31	1	2	1	4	9	6	4	3	2	2	1	1
16	mi	16	16	12		1		10	12	8	6	4	2	4	6	2
17	in	10	17	18		1	1	9	12	6	1	4		5	3	1
18	va	21	24	16				2	3	2	4	1	1	8	3	3
19	mo	5	13	12				2	9	8	3	16	3	2	1	1
20	la	13	9	8	1	1	1			2	1	3	2		1	
21	mn	9	15	8					1	1	2	1		1		
22	ne	7	5	4		1	1	3	7	5	1	2				
23	de	12	11	9				1							1	1
24	nc	5	5	5		1	1	4	3	2		1		1	1	2
25	wi	8	8	9				1	2		1	1		1		

This might be difficult to read, but there isn't any obvious sudden increase/decrease of investor types in a given region, which leads me to believe coverage was pretty similar throughout the period (though it could have slowly gotten better coverage, which would be difficult to see).

This is consistent with no reference in the actual manuals to missing data in 1908, 1911, or 1913 (just 1912) so hopefully converge should be similar throughout the period.

From 1908 handbook:

Poor's Handbook of Investors Holdings shows ownership of securities by savings banks trust companies insurance companies and other financial institutions The data has been compiled from state banking and insurance reports published in the United States and Canada

From 1913 handbook:

SHOWING OWNERSHIP OF SECURITIES BY SAVINGS BANKS TRUST COMPANIES INSURANCE COMPANIES AND OTHER INSTITUTIONS COMPILED FROM STATE AND OTHER DOCUMENTS PUBLISHED IN THE UNITED STATES AND CANADA

POOR'S HANDBOOK OF INVESTORS HOLDINGS shows ownership of securities by financial institutions It is compiled from reports of savings banks trust companies insurance companies and benevolent orders In several of the larger States and also in Canada these reports are published by the State authorities in other States they are gathered by a special canvass for publication in this book In New York where the detailed reports of holdings of savings banks and trust companies for 1912 were omitted from the annual report of the Superintendent of Banks a special canvass resulted in the return of the holdings of more than 50 per cent of these institutions The dates of these reports were August and September 1913 Holdings of the institutions that did not respond to this special canvass also appear in this number but under date of their last reports to the Superintendent of Banks

VII. Merging Files

7.1 List of Datasets

1) weekly_data_cleaned_v1.dta
All weekly closing NYSE bond market data with winsorized data at 99th percentile and perpetuity yield, returns, price impact and $ volume computed.

2) intraday_data_cleaned_v1.dta
All intra-day NYSE bond market data with winsorized data at 99th percentile and perpetuity yield, returns, price impact and $ volume computed.

3) HoldingsData_Final.dta
All data on investors holdings merged together for 1908, 1910, and 1911.

7.2 Dataset Summary Statistics

1) weekly_data_cleaned_v1.dta

61,787 weekly bond perpetuity yield observations

260 unique dates from 1905-1915 (implies average of 237 bonds/day)

2,112 unique bonds listed (# obs/bond; mean=29; median=4; 25[th]=1; 75[th] = 29)

Summary statistics for all observations:

	#	Mean	Median	Stdev	25th	75th
Perp yield	61,787	4.98%	4.72%	1.27%	4.26%	5.26%
Price Impact	60,904	0.050%	0.009%	0.11%	0.002%	0.041%
Abs return	61,013	0.172	0.105	0.243	0.0399	0.212

2) intraday_data_cleaned_v1.dta

51,431 weekly bond perpetuity yield observations

190 unique dates from 1908-1915 (implies average of 271 transactions/day)

320 unique bonds listed (# obs/bond; mean=161; median=1; 25[th]=0; 75[th] = 9)

Summary statistics for all observations:

	#	Mean	Median	Stdev	25th	75th
Perp yield	51,431	5.04%	4.80%	1.22%	4.23%	5.43%
Price Impact	50,695	0.094%	0.024%	0.202%	0.005%	0.084%
Abs return	50,707	19.48%	11.79%	26.74%	4.12%	24.81%
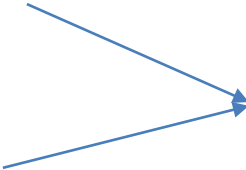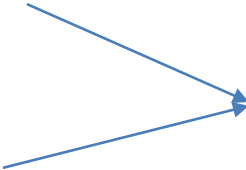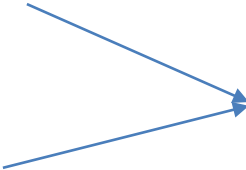
3) HoldingsData_Final.dta

	1908	1911	1913
# Obs	68,640	70,270	75,368
# Firms	10,905	9,363	9,087
# Securities	22,144	21,090	22,231
$ Assets (bil)	$3.65	$4.54	$4.96
# Investors	1,583	1,860	1,781
$/Investor	2,305,749	2,440,860	2,784,952
$/Firm	334,709	484,887	545,835
# Investors/Firm	0.15	0.20	0.20

7.3 Merging Files Statistics

7.3.1 Holdings and End-of-Day NYSE Bond Data

7.3.1.1 Matching by Issuer (Company)

Holdings Data	
# Obs	214,583
# Issuers	18,221
# Securities*	48,938

NYSE EOD Bond Data	
# Obs	63,923
# Issuers	539
# Securities	1,718

NYSE EOD Bond - Holdings Data	
# Obs	81,669
# Issuers	402
# Securities	1,559

* Stocks, corporate bonds, municipal bonds, and sovereign bonds
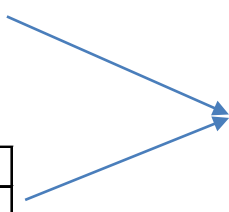
7.3.1.2 Matching by Security

Security-level identifiers are internally consistent in both files, but don't connect to each other well yet. Ideally an algorithm of some kind would be optimal, but otherwise will just have to do it manually within each matched issuer identifier.

7.3.2 Holdings and Intra-Day NYSE Bond Data

7.3.2.1 Matching by Issuer (Company)

Holdings Data	
# Obs	214,583
# Issuers	18,221
# Securities*	48,938

NYSE Intraday Bond - Holdings Data	
# Obs	81,469
# Issuers	402
# Securities	1,147

NYSE Intraday Bond Data	
# Obs	55,198
# Issuers	402
# Securities	1,147

 * Stocks, corporate bonds, municipal bonds, and sovereign bonds

7.3.2.2 Matching by Security

Security-level identifiers are internally consistent in both files, but don't connect to each other well yet. Ideally an algorithm of some kind would be optimal, but otherwise will just have to do it manually within each matched issuer identifier.

7.4 End-of-Day NYSE Stock Data Update

1) I spoke with Marc and he is comfortable with us using his NYSE data from this period. It is end-of-month and includes closing price, bid-ask (where available), and trading volume.

2) I have a record in my notes that Conrad matched the firm names from both databases already, just in case we were allowed to use it, but I have not found the file yet. I'll continue to look and if I can't find it for some reason I'll try to follow-up with Conrad.

VIII. Additional Holdings Data Geographic Concentration Analysis

From Table 8.1 it doesn't appear that there was necessarily a sudden "influx" of capital from investors not in NY. In contrast actually it appears that % of investors in NY rises from 1908-1911.

Table 8.1 $ Investment by Geographic Region

Rank	Region	1908		1911		1913	
		$ Invested	% of Total Assets	$ Invested	% of Total Assets	$ Invested	% of Total Assets
1	ny	1,705,227,392	47%	2,262,639,104	50%	2,406,004,992	49%
2	pa	461,659,360	59%	522,360,384	61%	569,328,768	60%
3	ma	361,147,392	69%	429,317,888	71%	458,551,456	69%
4	cn	316,700,320	78%	329,199,808	78%	425,537,248	78%
5	nj	245,996,592	85%	283,619,040	84%	346,437,216	85%
6	ri	87,410,440	87%	91,001,144	86%	97,484,696	87%
7	me	83,610,864	89%	94,692,864	88%	96,485,216	89%
8	wi	73,991,272	91%	80,333,904	90%	91,462,144	91%
9	ca	72,019,536	93%	64,777,472	92%	32,174,330	91%
10	canada	53,090,840	95%	137,138,800	95%	107,313,792	93%
Total		3,646,282,116		4,540,264,874		4,956,837,973	

It also appears when looking at the concentration of investor portfolios in table 8.2 that investors were less diversified in their $ distribution of investments across securities (though of course in this simple analysis we can't look at the factor structure of these allocations).

Table 8.2 Herfindahl Index of Investors' Portfolios

Herfindahl		
	Mean	Median
1908	4.91%	2.08%
1911	5.52%	2.07%
1913	5.06%	1.98%

From table 8.3 though it does look that for a given security the ownership is a bit more distributed after 1908.

Table 8.3 Herfindahl Index of Security Owners Distribution

Ownership Herfindahl		
	Mean	Median
1908	42.1%	27.1%
1911	39.1%	23.8%
1913	38.5%	22.5%

This becomes even more striking looking at the median geographic distribution of investors in a security. The median Herfindahl index of $ ownership by region falls from 90.6% in 1908 to 65.8% in 1911.

Table 8.4 Herfindahl Index of Security Owners Regional Distribution

Ownership Herfindahl		
	Mean	Median
1908	60.5%	90.6%
1911	57.6%	65.8%
1913	57.5%	66.0%

Combining this with the total investment from earlier this would be consistent with more NY investors taking concentrated positions in companies which do not predominantly have NY investors (ex. Firms headquartered outside of NY). So rating agencies may have reduced the information acquisition costs so that NY investors could invest in companies in say MO, which previously only MO-based firms felt comfortable investing in.

I also examined if this relationship differs by the type of security (stock vs. bond), but in results not shown here they are qualitatively the same.

Potential Next Steps:

1) Try to get firm headquarters to see if this is true
2) Compare this analysis for RRs vs. Industrials
3) Compare this analysis depending on rating a firm gets

Note:

There are 18,000+ issuers in the holdings data so assigning headquarters and industry might not be that straightforward. Of course the scope falls substantially when we focus only on firms with secondary market data. What should we do?

IX. "Other Banks" & "Other Investors"

Some Examples of "Other Banks"

investor_name_hold	investor_type	Investor region
Bank of Alameda	other bank	ca
Bank of Alvarado	other bank	ca
Bank of Amador County	other bank	ca
Bank of Antioch	other bank	ca
Bank of Areata	other bank	ca
Bank of Arroyo Grande	other bank	ca
Bank of Auburn	other bank	ca
Bank of Bakersfield	other bank	ca
Bank of British North America	other bank	ca
Bank of Campbell	other bank	ca
Bank of Centerville	other bank	ca
Bank of Charlerol	other bank	pa
….		
Sonoma Valley Bank.	other bank	ca
South Berkeley Bank	other bank	ca
South Norwalk Sav. Bank	other bank	cn
South Pasadena Bank	other bank	ca
South Side Bank	other bank	pa

Some Examples of "Other Investors"

investor_name_hold	investor_type	Investor region
Afodera Woodmen of America	other investor	il
American Exchange Underwriters	other investor	ny
American Guild	other investor	va
Catholic Knights of America (Sup. Coun.)	other investor	mo
Commercial Travelers Mutual Aid Association	other investor	ny
Equitable Security Company	other investor	ny
Fire Association of Philadelphia	other investor	pa
Order of Scottish Clans	other investor	ma
Order of the Colder Seal	other investor	nj
Order of the Golden Seal	other investor	ny
United States Fidelity ft Guaranty Company	other investor	md
United States Lloyds	other investor	ny
Woodmen Circle	other investor	ne

X. Merging EOD and Intraday NYSE Bond Data

10.1 Matching by Issuer (Company) 1905-1915

NYSE EOD Bond Data	
# Obs	63,923
# Issuers	539
# Securities	1,718

NYSE Intraday Bond Data	
# Obs	55,198
# Issuers	402
# Securities	1,147

EOD-Intraday Bond Data	
# Obs	77,885
# Issuers	402
# Securities	1,559
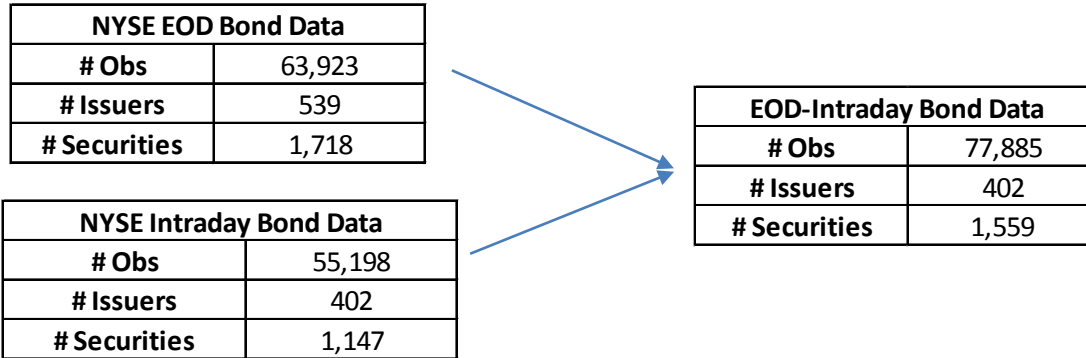
 * Stocks, corporate bonds, municipal bonds, and sovereign bonds

EOD Bond Data Coverage
- Collection day-of-the-week is Sundays (constitutes all trades during the week)
- Monthly 1/14/05-4/11/08; Weekly 4/11/08-4/23/10; Monthly 5/14/10-1/4/13; Weekly 1/4/13-7/11/14; No data 7/11/14-12/18/14; Weekly 12/19/14-1/2/15

Intraday Bond Data Coverage
- Collection day-of-the-week is Thursday paper (constitutes all trades on Wednesday)
- Weekly 4/29/08-7/29/14; No data 8/1/14-12/22/14; Weekly 12/23/14-1/6/14

10.2 Matching by Issuer (Company) 1908-1915

NYSE EOD Bond Data	
# Obs	56,532
# Issuers	518
# Securities	1,620

EOD-Intraday Bond Data	
# Obs	72,569
# Issuers	401
# Securities	1,488

NYSE Intraday Bond Data	
# Obs	55,198
# Issuers	402
# Securities	1,147
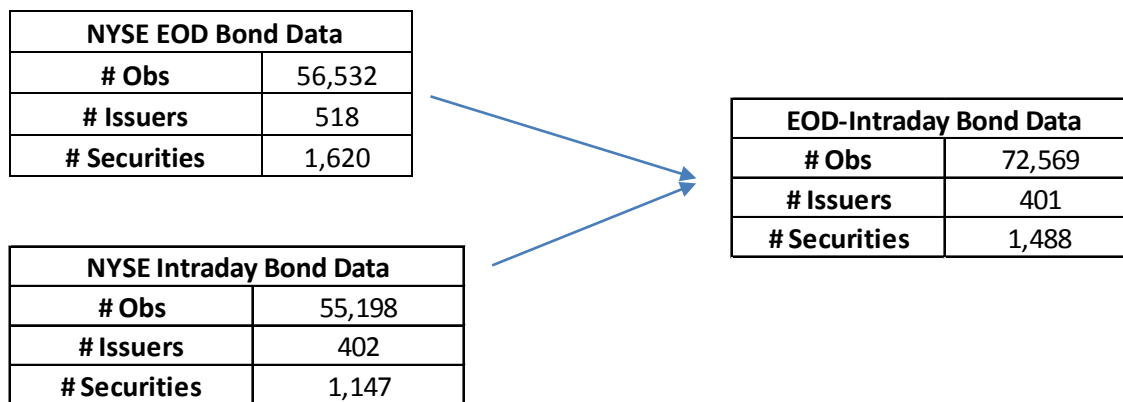
 * Stocks, corporate bonds, municipal bonds, and sovereign bonds

10.3 Matching by Issue (Security) 1908-1915

NYSE EOD Bond Data	
# Obs	56,532
# Issuers	518
# Securities	1,620

EOD-Intraday Bond Data	
# Obs	69,018
# Issuers	395
# Securities	823

NYSE Intraday Bond Data	
# Obs	55,198
# Issuers	402
# Securities	1,147

* Stocks, corporate bonds, municipal bonds, and sovereign bonds

Since we can see that EOD tends to be a superset of intraday trades during this period losing 7 companies and especially ~300 securities should definitely be able to be improved.

XI. Exploring Difference in #Firms for EOD vs. Intraday

EOD Excerpt:

Complete Bond Transactions

BONDS.

Sales W'k Ended Aug. 1, $15,319,000	First.	High.	Low.	Last.	Sales.
Adams Express 4s	91¼	92	91⅛	92	19
Albany & Susquenanna 3½s	95⅞	95½	95½	95½	8
American Cotton Oil 4½s	93	93	93	93	1
American Hide & Leather 6s	89½	89¾	89½	89¾	15

Week Ended Aug. 1, 1908.

Intraday Excerpt:

TRANSACTIONS IN BONDS.

NEW YORK STOCK EXCHANGE.
Tuesday, July 28, 1908.

Imp Jap Govt 4½s
sterling loan
1,000.........89½
500.........89¾
Rep of Cuba 5s
1,000.........105
New Yk City 4½s
1957
20,000.........108¾
New
2,000.........108⅞
49,000.........108¾
Adams Expr 4s
1,000.........91¾
Alb & Susq 3½s
8,000.........95½
Am Cot Oil 4½s
1,000.........93

Laclede Gas Lgt of
St L 1st 5s
1,000.........102½
L S & M S 4s
1931
1,000.........93¾
1,000.........98⅝
L S & M Sou 4s
2,000.........94¼
5,000.........94½
15,000.........94¾
L S & M So 3½s
5,000.........92
5,000.........93
Lex Av & P F 1st
Lo, N A & C 1st
1,000.........102⅜
Louisv & Nashv

As you can see American Cotton Oil traded once on Tuesday 7/28/1908, but it still shows up in the EOD file on Sunday 8/1/1908 since this is a summary of all weekly activity. However we only collect intraday data for Wednesday so this is an example where this firm and bond could be excluded since it is very illiquid.

XII. Bid-Ask Data from Eric vs. Intraday Data
- Important NOTE***** The images Eric collected say Oct 14, 1908, but refer to the closing information for Oct 13th, 1908. Our files all reference the closing date, not the newspaper date. So our intra-day data for Oct 14th, 1908 is actually for the next day.

- Uploaded the PDF from Oct 13th, 1908's intra-day data for comparison in the dropbox folder Eric provided. It looks like most of the bid-asks from the CFC data would not be reproducible via the intra-day data.

XIII. Fixing Table 8.4

Table 13.1 *FIXED* Herfindahl Index of Security Owners Regional Distribution

	Ownership Herfindahl	
	Mean	**Median**
1908	82.5%	100%
1911	80.4%	100%
1913	78.9%	100%

Originally the % share was being computed at the firm-level when it should have been at the region-level which led to the unusual previous results. That mistake has been fixed and the table above is now correct.

XIV. Adding Industries to Holdings Data

As Eric pointed out we can actually get industry from the original PDF files since the companies are in order by section. Having these would let us do some preliminary analysis in a diff-in-diff framework without relying on merging with any other files at all. Examining which of the railroads actually showed up in the manual would also be interesting. It would also improve the within holdings data matches as well as those from the holdings data to other data files. In short, it is very important.

Current status:
1) Industries have been assigned to each of the raw entries
2) Matches to the final holdings database works for 23,595 names, but I'm not able to match raw data to final data for 2,827 names. The transition from raw names to final names is documented throughout my files (ex. removing extra spaces, etc…) and I'm going through it now, but it will take a little while longer to get the full matches.

Important Note*
Looks like we did not enter the "Bank and Trust Co. Securities" for 1911 or 1913 (though we did for 1908).

Proposed Next Steps:
1) Get all industry identifiers matched to the final holdings database.
2) Redo analysis done in previous sections except broken out by the different industries

3) Redo internal holdings data identifiers to be consistent with new industry information
4) Redo merging of holding with other files now accounting for new industry information
5) Complete Stock IDs and merge analysis
6) Complete RA TOC and Holdings data merge analysis
7) Redo distribution analysis looking at changes within each investor or within each issuer (basically investor and issuer fixed effects)

XV. Eric's Explanation of Bid-Ask Data vs. Market Data

The quotes are definitely for the date that is printed at the top. So the Wednesday quotes are the correct ones. But the reason they may not always correspond to some of the intra-day trading very well is that they are quotes as of 11 AM - on the 4th page of every issue, it says "Quotations At 11 AM."

So this might explain why the quotes seem to correspond to the Tuesday trading - everything that happens after 11 AM on Tuesday doesn't get reflected in the printed quotations until the Wednesday morning issue.

In the original transactions data sheets (which are what the NYTimes compiles and reports the following day, and which are available at the NYSE), the transactions are divided into 3 periods - those up to 11 AM, and then I think 2 PM, and then 4 PM - or something like that. So any change in the quotes between two days should reflect trading that occurred prior to 11 AM on the second day, and also trading that occurred after 11 AM on the first day.

I think this probably means we can't really do anything super clever with the intra-day data and the bid/asked quotes together, unless we were to have the original NYSE intra day sales sheets entered, so that we could separate out the pre-11 AM trading vs. the post-11 AM trading. That's actually do-able, but it would be expensive and difficult and time-consuming, so we should only consider it if we have some super brilliant, game-changing idea to pursue with it.

But I doubt that will be the case, so the bid-ask quotes, if we want to use them, are going to have to be treated as a kind of separate measure of market liquidity - we can compile data such as: is there a bid price, is there an asked price, are there both bid and asked prices, what is the spread, etc., and treat those as separate outcome measures that we can't necessarily link to our trading volume data directly.

XVI. Holdings Data Analysis by Industry Group

Table 16.1 Securities Per Industry 1908-1913

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	1,349	13,533	3,783	2,339	1,227
1911	8	13,988	3,494	2,467	1,259
1913	8	14,867	3,559	2,687	1,276

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	6%	61%	17%	11%	6%
1911	0%	66%	16%	12%	6%
1913	0%	66%	16%	12%	6%

As you can see in Table 16.1 Bank & Trust companies basically disappear under the new classification (the 8 here are cases of firms that were labeled as Bank & Trust in 1908 and show up mostly under "Misc" in other years). There isn't an obvious trend in the total securities except it appears that "industrial & misc" # of securities was the only one (besides Bank & Trust) to see a decline after 1908. Not sure what to make of it, but does suggest that all of Bank & Trust didn't just enter "Misc".

Another general takeway is that most of the securities are government securities and steam railorads is the 2nd smallest number of issuances.

Table 16.2 Issuers Per Industry 1908-1913

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	1,307	4,784	3,108	909	872
1911	7	4,880	2,715	954	899
1913	4	4,732	2,758	882	874

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	12%	44%	28%	8%	8%
1911	0%	52%	29%	10%	10%
1913	0%	51%	30%	10%	9%

I don't explicitly compute those in a table here, but ti should be clear that the government securities in all years tended to have the most securuties per issuer, while industrial tended to have very few. Difference is dramatic. It tends to be almost 3 securities/issuer for governments, but just slightly more than 1 per firm.

Table 16.3 $ Invested (millions) Per Industry 1908-1913

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	155.6	888.2	373.0	1,964.9	264.8
1911	0.4	1,397.8	442.2	2,382.8	318.0
1913	0.2	1,462.2	475.1	2,690.8	357.0

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	4%	24%	10%	54%	7%
1911	0%	31%	10%	52%	7%
1913	0%	29%	10%	54%	7%

When we look at actual $ invested, as we might expect, now Railroads constitute the majority of the assets. From this though again there isn't an obvious differential change from 1908-1911. If anything government securities overtook railroads slightly.

Table 16.4 $ Invested/Security (000s) Per Industry 1908-1913

	Bank & Trust	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	115	66	99	840	216
1911	50	100	127	966	253
1913	30	98	133	1,001	280

Table 16.4 gets at the crux of why railroads are so important during this period. The average investments in our holdings data per security is about 10x larger for railroads than it is for government securities. If anything this differential did appear to grow over period.

Table 16.5 Mean Herfindahl Index of Security Owners Regional Distribution

	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	76.6%	64.6%	23.8%	39.0%
1911	72.6%	59.1%	24.9%	37.2%
1913	73.7%	54.9%	25.2%	35.0%

Table 16.6 Median Herfindahl Index of Security Owners Regional Distribution

	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	100.0%	100.0%	5.7%	21.1%
1911	100.0%	79.0%	5.8%	18.1%
1913	100.0%	55.6%	4.8%	13.5%

If you look at mean or median it is clear that that majority of government securities are held locally. Though this is also true of industrial securities and since those are the two smallest industries by security size (proxied by $ investment/security), the distribution of investors could be driven by size of issuance rather than (or more likely in addition to) restrictions placed on local bank investments (as suggested by Eric).

I think it is also worth making a brief note that Railroads are the only industry that see a rise in concentration from 1908-1911. That isn't obviously consistent with a narrative of home bias falling after the introduction of ratings, but I thought it was worth noting.

XVII. Holdings and 1909 TOC Matches

Table 17.1 Steam Railroads in 1909 Moody's Manual TOC (just parents names)

	$ Value (mils)	# Issues	# Issuers
1908	1,020	848	98
1911	1,277	964	131
1913	1,382	961	98

	$ Value (mils)	# Issues	# Issuers
1908	52%	36%	11%
1911	54%	39%	14%
1913	51%	36%	11%

Looking at matches between our holdings data and the 1909 TOC it looks like only 10% of railroads in our holdings data in 1908 show up in the TOC of 1909. This group though represents more than 1/3rd of all securities and more than ½ of all invested assets. This seems likely since Moody's would have rated the most important (aka largest firms) first, but there is another confounding factor. The 1909 TOC only lists 101 parent company names (or railroad system names). Without knowing the parent-subsidiary relationship we are probably missing a huge number of railroads. The 1911 TOC has all subsidiaries as well, so we could use that to help define our parent-subsidiary relationship database, but then we would need to also be careful about changes. Alternatively we could have a student flip through the whole manual and enter all subsidiaries as well...

Proposed Next Steps:
1) Details about improved holdings ID matches using industry classifications
2) Initial attempt at matching issue ids from holdings data to other data sources using fuzzy logic
3) Create template and explanation for RA to manually check fuzzy logic matches and fill-in missing values
4) Complete NYSE Equities matching/IDs
5) Initial analysis of NYSE secondary market data on equities

XVIII. Methodology to Collect Bid/Ask Data for a Few Dates

Best strategy to enter these for only a small number of dates (to see if there is anything interesting we can do with them)?

Eric: Enter all names and then only enter new names if need otherwise just enter new price information… Following up on 5/1 with Eric.

XIX. Manually Checking Company Name IDs

19.1. Checking Internal Consistency of Company IDs in Holdings Data Across Years

The file "All_Holdings_CompanyNamesByYear.xlsx" has all unique company identifiers for each book year (1908, 1911, or 1913), the original name, the cleaned name, and the industry associated with it. The file has been sorted by cleaned company name so that it should be clear if there are issues (in most cases). Here is an example of what it looks like:

firm_id_holdings	cname_hold_orig	cname_hold_clean	book_year_hold	industry	firm_id_holdings_new	notes
1009433	\YLMA	\ylma	1911	Government		
2010	A SPIN WALL, PA	a spin wall pa	1908	Government		
2010	A SPIN WALL, PA	a spin wall pa	1911	Government		
1824	ABBEVILLE, ALA ,	abbeville ala	1911	Government		
1824	ABBEVILLE, ALA	abbeville ala	1913	Government		
1824	ABBEVILLE, GA	abbeville ga	1908	Government		
6928	ABBEVILLE, S, C ,	abbeville s c	1911	Government		
6928	ABBEVILLE, S C	abbeville s c	1913	Government		
2000010	ABERDEEN AND ASHEBORO	aberdeen and asheboro	1913	Steam Railroad		
1825	ABERDEEN, MISS	aberdeen miss	1908	Government		
1825	ABERDEEN, MISS	aberdeen miss	1911	Government		
1825	ABERDEEN, MISS	aberdeen miss	1913	Government		

I think for the most part the matches are pretty good, but it might be worth somebody taking another glance since this is the basis for all the company ID matches used in other files. There are over 28K entries so a real manual check of everything would probably be challenging.

Carola: Let's call them issuers/investors. What if only include only one of cleaned names which are identical. Check the residual could be much smaller… (assume cleaning was fine). Only print them if more than 1 cleaned name within firm ID. This could be much more reasonable. Have

"cleaned" be more basic though. Just remove commas and spaces and make lower case. Way fewer things (don't make corporation vs. company gone).

Make list of all things use an how much it reduces the list…

19.2. Checking Company IDs NYSE Bond -> Holdings

The file "BondNamesAndIDs.xlsx" has the list of all company names from our NYSE bond data with their NYSE bond data company unique identifiers as well as the current matches to the holdings data. There are a few missing, so it is definitely worth going through and it should be fairly quick (only 539 entries). Here is an example of what it looks like:

nyse_bond_issuer_id	companyname	industry_codebook_companyname	firm_id_holdings	cname_hold_orig	cname_hold_clean	industry	notes
1	Adams Express Co.	Adams Express	6605	ADAMS EXPRESS	adams express	Industrial and Misc	done
359	Al. S. & L.	Al. S. & L.					
800	Alabama Central	Alabama Central	4	ALABAMA CENTRA	alabama central	Steam Railroad	

To make this easier I have saved down a file called "HoldingsDataNames_withBond.xlsx" which has only the companies from the holdings data that have bonds and are not governments (since we only collected non-government nyse bond data). Here is an example of what it looks like:

firm_id_holdings	cname_hold_orig	cname_hold_clean	industry
2000010	ABERDEEN AND ASHEBORO	aberdeen and asheboro	Steam Railroad
1828	ABILENE LIGHT AND WATER COMPANY	abilene light and water	Industrial and Misc
9727	ABINGTON NATIONAL, BANK	abington national bank	Bank and Trust

19.3. Checking Company IDs NYSE Stock -> Holdings

The file "StockPriceDataNames.xlsx" has the list of all company names from Marc's NYSE equity data with company unique identifiers. I filled in a few as an example, but somebody needs to go through and enter all of these manually. Note: The majority will not have matches since this covers a large range of dates during our period and many have ambiguous names. This should be noted in the "notes" section though. Also anybody entering it should give new "stock_issuer_id_new" ids if there are issues with the originals which were not cleaned that intensely (database was never actually used yet). This should take a little while, but hopefully not more than a week (~1,700 entries). Here is an example of the file:

stock_is	CleanedName	stock_issuer_id_new	firm_id_holdings	cname_hold_orig	cname_hold_clear	industry	notes
1	A						ambiguous
2	AAO						ambiguous
3	ABA						ambiguous
4	AbitibiPowerPaper						missing
5	AbranceRumely	5					missing
6	ACD						ambiguous
7	ACF						ambiguous
8	AcmeTea						missing
9	AdamsExpress		6605	ADAMS EXPRESS	adams express	Industrial and Misc	done

To help make this easier I have saved an excel file called "HoldingsDataNames_withStock.xlsx" which has only those holdings data companies with stock who are not governments (since those also were not in the NYSE data Marc collected). Here is an example from that file:

firm_id_holdings	cname_hold_orig	cname_hold_clean	industry
1000018	ABSEQUAN HOTEL COMPANY	absequan hotel	Industrial
6600	ABSTRACT ANI) TITLE COMPANY	abstract ani) title	Industrial
6601	ACADEMY REALTY COMPANY	academy realty	Industrial
6602	ACME HARVESTER COMPANY	acme harvester	Industrial

19.4. Checking Company IDs Collected Ratings -> Holdings

TBD

XX. 1909 Moody's RR Manual Parent-Subsidiary Relationship and All Rated Securities

Figure 20.1 1909_Moodys_Abbreviated_TOC.pdf:

Figure 20.2 From 1909_Moodys_Full_TOC.pdf:

Figure 20.3 pg. 275 of 1909 Moody's Manual

TABLE D.—Bond Record and Ratings (Based on 10-Year Results, Per Mile of Road).

NAME OF ISSUE.	Interest Payable.	Maturity.	Lien on Miles.		Average Income Available.	Interest R'q'r'd per Mile of System.	Factor of Safety.	BASIS FOR RATING.		Net Rating.	1908 PRC. RANGE.		
								Security.	Sal'bility.		Low.	High.	Last.
Prior Mortgages.													
1. Milwaukee & L. Winnebago first 6s.................	J&J	Jl 1912	(1st)	66	(Pr) $2,357	$37	97%	High.	High.	**Aa.**	....		102*b*
2. Chic. Wis. & Minn. first 6s....	M&S	Mr 1916	(1st)	115	(Pr) 2,357	47	97%	"	"	**Aa.**	....		106
Main Company Issues.													
3. Wisconsin Cent. first gen. 4s..	J&J	Jl 1949	(1st & con)		(Jt) 2,273		40%	Good.	"	**Baa.**	80 –92¼		92¼
4. Wisconsin Cent. Sup. & Dul. term. first 4s............	M&N	My1936	(1st)	161	(Jt) 2,273		40%	"	"	**Baa.**	85¼–87¼		87½*b*
5. Wisconsin Cent. Marsh. & S. E. Div. 4s..............	M&N	My1951	(1st)	33	(Jt) 2,273	1,310	40%	"	"	**Baa.**	85 –93		88*b*
6. Wisconsin Cent. Minn. Term. P. M. 3½s................	J&J	Jl 1950	(1st) *Ter.*		(Jt) 2,273		40%	"	"	**A.**	....		.....

"Pr" signifies Prior, and "Jt" Joint Claim on Total Net Income.

Figure 20.4 pg. 206 of 1909 Moody's Manual

both preferred issues have received 4%.					
Wisconsin Central Railway (preferred)........................	....	**Ba.**	33	72½	72¼
Outstanding, $11,267,104. Par $100.					
Wisconsin Central Railway (common)........................	....	**C.**	13½	38¼	38¼
Outstanding, $24,543,000. Par $100.					

I've uploaded the file 1909_Moodys_Full_TOC.pdf, which it turns out is the full table of contents that lists every security in the 1909 manual. Unfortunately the previous TOC entered was the abbreviated version and not what we wanted. In addition to getting a complete list of companies and securities in the 1909 manual, we can also establish the parent subsidiary relationship, which we would want anyway.

Notice in figure 20.3 that "Milwaukee & L. Winnebage first 6s" is included in the same table as all the bonds for "Wisconsin Central". In the abbreviated TOC in figure 20.1 we can't see this, but in the full TOC we can see that "Milwaukee & L. Winnebage first 6s" appears on the same page as the "Wisconsin Central" bonds. Since there is only 1 bond record per page in the full manual, any bonds that appear on the same page, must have some kind of parent subsidiary relationship.

Issue #1: Sometimes bond records appear on more than 1 page. Could have difficulty distinguishing bond records that appear on multiple pages from separate bond records on consecutive pages.

Issue #2: Need to clearly look at bonds not stocks, since stocks all appear on a list together in the middle of the manual with their ratings.

Steps for 1909 Ratings Manual:
1) Enter all entries for "1909_Moodys_Full_TOC.pdf" into spreadsheet "1909_Moodys_Full_TOC.xlsx "

2) Match company and security names to "HoldingsData_Final_wIndustry.dta" firm_id_holdings
3) Establish parent-subsidiary relationship using page numbers and code these into dataset

Eric and Asaf will talk about it more Friday. Have firm enter data from Bond Records.

XXI. Random Idea

A recent RFS paper by Crotty and Back called "The Informational Role of Stock and Bond Volume" (in dropbox under literature called "Info_Role_StockBond_Vol.pdf") points out that 1) The sign of cross-market (bonds to stocks) of Kyle's lambda (price impact) is same as correlation between firm's debt and equity returns and 2) sign of Kyle's lambda is positive (negative) if private information concerns the mean (risk) of the firm's assets. Since "surprise" ratings alter the mean, but any ratings alter the risk I think this gives us nice variation in micro-structure predictions for Kyle's lambda (which we can do with the data we have) and predictions for bond/stock market correlations (which we can also do). I think we can also relate these of course to the holdings distribution and real effects, but requires a bit more thought.

Next Steps:
1) Enter 1909 Full TOC
2) Check all company names (as detailed above)
3) Create template and strategy for matching/double checking security-level matches
4) Enter rest of steps in "Steps for 1909 Ratings Manuals" (should come after checking company/security names/IDs)
5) Analysis of matching success at company and security-level
6) Redo preliminary holding analysis
7) Initial analysis of NYSE equity data

XXII. Manually Checking Issuer Name IDs (redux)

22.1. Checking Internal Consistency of Issuer IDs in Holdings Data Across Years

Step	Description	# to Check
1	Original issuer name	13,994
2	Lower case	13,933
3	Remove "company"	13,707
4	Remove "corporation"	13,698
5	Remove "()" and everything inside of them	13,326
6	Remove apostrophes	13,291
7	Replace "5" w/ "S", "8" w/ "S", "1" w/ "t", then leading numbers with blank	13,286
8	Remove trailing and leading spaces	10,263
9	Remove commas	4,798
10	Use industry codes	4,183

In table 22.1 (above) I look at the number of issuer name entries that still exist after removing those that only have one "cleaned" issuer name per issuer ID # assigned via fuzzy logic. As you can see after all the basic cleanings I did there are around 4,183 issuer names that have multiple matches (within a given industry, where industry is defined from the holdings manual).

22.2 Example of First 12 Entries of Result

firm_id_holdings	book_year_hold	industry	cname_hold_clean_code
14	1913	Steam Railroad	ann arbor railroad
14	1913	Steam Railroad	ann arbor
18	1911	Steam Railroad	ashley riverl
18	1908	Steam Railroad	ashley river
22	1911	Steam Railroad	atchison topeka & santa fe
22	1913	Steam Railroad	atchison topeka and santa fe
27	1911	Steam Railroad	atlanta knonville and northern
27	1913	Steam Railroad	atlanta knoxville and northern
29	1908	Steam Railroad	atlantic and birmingham construction
29	1911	Steam Railroad	atlantic and birmingham construction co
36	1913	Steam Railroad	atlantic coast line rr of south carolina
36	1911	Steam Railroad	atlantic coast line of south carolina

Table 22.2 is a sample of the first 12 entries that result after running the simple cleaning methods above. As you can see the majority of what the fuzzy logic was able to do was eliminate a lot of typos

22.3 Example of Mistakes

firm_id_holdings	book_year_hold	industry	cname_hold_clean_code
79	1913	Steam Railroad	york and peach bottom
79	1908	Steam Railroad	boston revere beach and lynn
80	1913	Steam Railroad	boston terminal
80	1913	Steam Railroad	houston belt and terminal

While at first glance the majority look like logical things for the fuzzy logic to fix, it didn't take long to find some unusual mistakes. Here are the 1st two examples I could find. For the first I think that "peach" looks very similar to "beach" and maybe "boston" looks similar to "bottom", but other than that I'm not sure why it decided those were a match, but clearly they are not. For the 2nd I imagine "boston" looks similar to "Houston", both had the word "terminal", and filler words like "and" typically don't get much weight.

I have saved the full file of these in "HoldingIssuerNamesToCheck.xlsx" in the dropbox under "Data_Programs\Data\20150506". As you can see above and from what I've seen it is a pretty easy gut check on most of these and I imagine one of us could do it in just a few hours. I'm not sure how long it would take a good undergrad RA, but I imagine similar…

XXIII. Collecting Bond Ratings for 1909 for All Bonds

Eric took the pictures so that it is parent name and then the bond record information (see below) from Moody's.

NAME OF ISSUE.	Inter-est Pay-able.	Matur-ity.	Lien on Miles.	Average Income Available.	Interest R'q'r'd per Mile of System.	Factor of Safety.	BASIS FOR RATING.		Net Rat-ing.	1908 PRC. RANGE.		
							Security.	Sal'bility.		Low.	High.	Last.
1. New Orleans & N.-E. prior-lien 6s.	A&O	N 1915	(1st) 196	$3,599	404	89%	Very high.	High.	Aaa.	....		103b
				3,195	30	99	" "	"	Aaa.	....		.....
2. New Orleans & N.-E. first 6s	J&J	Ja 1911	(2d) 196	3,065	1,261	60	High.	"	Aa.	....		.....
3. New Orleans & N.-E. gen.4½s.	J&J	Ja 1952	(3d) 196	1,804	344	75	"	"	Aa.	....		.....
4. New Orleans & N.-E. inc. 4½s.	Dec.1	Jl 1952	(4th) 196									

The plan is to have Eric's data collection company enter the parent name and then everything in the bond record table pictured above except the 1908 price range, low, high, and last information. We should get those anyway from our data and sal'bility rating is probably a better proxy for liquidity anyway.
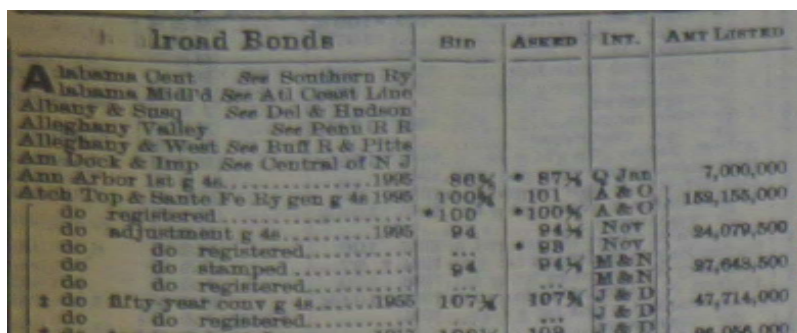
XXIV. Collecting Bid-Ask Data

I will talk with Samasource on Monday about getting estimates for the data collection project. Our current strategy will be to collect the following:

industry_heading	issuer_name	coupon	maturity	class	bid	asked	bid_asterisk	asked_asterisk	int	amt_listed

Carola: Change column headings so they appear in the order in the image…

Eric: Remember that amount listed applies to multiple versions…

From this source (full sample from Eric is in the dropbox as "bid_ask_19090407.pdf":

Where "int" is the interest payment dates and "amt_listed" in the amount listed and for these we will only include those entries for one date (ex. April 7, 1909), but will be skipped for other dates. One outstanding question is if we want to enter only rows where there are a bid and ask or if there are either?

Next Steps:
1) Manually check the 4,183 issuer ids in the holdings data
2) Start data collection for 1909 bond record information
3) Get quotes from Samasource for bid-ask entry information
4) Initial analysis of NYSE equity data (by itself) to understand what we have
5) Work on ideas that could be completed in time for the conference with data we already have

XXV. Manually Checking Issuer Name IDs (redux #2)

25.1. Checking Internal Consistency of Issuer IDs in Holdings Data Across Years

Step	Description	# to Check
1	Original issuer name	13,994
2	Lower case	13,933
3	Remove "company"	13,707
4	Remove "corporation"	13,698
5	Remove "()" and everything inside of them	13,326
6	Remove apostrophes	13,291
7	Replace "5" w/ "S", "8" w/ "S", "1" w/ "t", then leading numbers with blank	13,286
8	Remove trailing and leading spaces	10,263
9	Remove commas	4,798
10	Use industry codes	4,183
11	Replace "&" w/ "and", multiple blanks, and "-"	3,828
12	Replace "ry" and "rr"	3,824
13	Differ only by 1 letter	2,579
14	Differ by only 2 letters	2,168
15	Differ by only 3 letters	1,658

I don't really like "differ by only 3 letters", since that is actually quite a bit. For example "sweetwater tenn" and "sweetwater texas" differ by only 3 letters but are definitely different municipalities. For all analysis below then I only include through steps 14 so size is 2,168.
Carola: Maybe only do 1 letter difference.

25.2. Industry Breakdown of Duplicates

Industry	#	%
Government	1,457	67
Industrial and Misc	448	21
Steam Railroad	138	6
Street Railway	129	6

Most of the duplicates are in the "government" securities sections, so if we are not interested in those right away that cuts out about 2/3rds of the work…

25.3. Effective Fuzzy Logic Examples

firm_id_holdings	cname_hold_orig	book_year_hold	industry
847	TENNESSEE COAL, IRON AND RAILROAD,	1911	Industrial and Misc
847	TENNESSEE COAL, IRON AND RAILROAD Birmingham Division	1913	Industrial and Misc
6665	AMERICAN GRAPHOPHONE COMPANY	1908	Industrial and Misc
6665	AMERICAN GRAPHOPHONF COMPANY Notes	1911	Industrial and Misc
9448	UNITED ELECTRIC, GAS AND POWER COMPANY	1911	Industrial and Misc
9448	UNITED GAS, ELECTRIC AND POWER COMPANY	1911	Industrial and Misc

25.4. Ineffective (false positive) Fuzzy Logic Examples

firm_id_holdings	cname_hold_orig	book_year_hold	industry
6634	AMADOR COUNTY ROLLER FLOUR MILLS (Ione, Cal)	1908	Industrial and Misc
6634	KEEWATIN FLOUR MILL	1913	Industrial and Misc
6644	AMERICAN CAR AND FOUNDRY,	1913	Industrial and Misc
6644	AMERICAN PIPE AND FOUNDRY	1911	Industrial and Misc

As you can see above the fuzzy logic algorithm is a bit cleverer at times than the cleaning methods used above. For example it doesn't get "tricked" by the order of words. The problem is of course that it allows for a lot of false positives. My gut feeling from looking through the files now is that the correct matches only slightly outnumber the false positives in the remaining 2,168 entries to check. We could use only issuer ids generated using the cleaning method above as a robustness check instead of having an RA go through the whole thing…

Carola: Let's use one (or both) of these. She might have an RA to help, but don't wait for those. Name cleaning, name cleaning w/ 1 letter difference, and fuzzy logic as 3 methods.

Eric: Concentration increases RRs, but not for other things, so maybe look at this more…

Plan: 1) Look at within holdings data analysis. 2) Next week should get bond record info from Eric. Try to quickly match it and see what can produce.

XXVI. Collecting Bid-Ask Data (update)

I had a meeting with them this Weds (5/13) and was told I would get pricing estimates (under different assumptions for how we want to do the project) by today (5/15). Here are the highlights from the Weds call:

1) They have a $5k minimum, so we would need the job to be at least that size (an accurate estimate of the number of dates is critical)
2) The more photos we can send them (Eric some help there?) the better. There only real concern was image quality.

3) If we do the project I will send even more detailed collection instructions.
4) There is a fee for onboarding images so if we can "host" them ourselves and share them via a link could save money. Daniele (the pricing person) will give me more details on 5/15.
5) They could use some more info on our budget and our timeline (I said collected by the fall).

Update from email sent on 5/15/2015:

"OK-- so Daniele and I worked with our project management teams to scope and price your project and it's definitely looking more intensive than what we've done in the past. Please let Daniele and me know if you think any of these assumptions are particularly out of line--- we based this on the example pdf provided and the numbers we discussed on Wednesday.

Here are the numbers we calculated:

480 lines per page
3 pages per pdf
150 pdfs
= 216,000 total lines

180 skipped lines per page (due to Bid and Ask being blank)
* 3 pages per pdf * 150 pdfs
= 81,000 total skips

216000-81000 = **135,000 lines that need to be transcribed**

Is this correct by your calculations?

Pricing:

$0.21 - $0.40 per line to be transcribed
= a pretty large range total cost here of anywhere from $28,500 - $54,000 with volume assumptions made above (300 lines to transcribe per page, 3 pages per pdf, 150 pdfs.)

There is such a big range because we need to consult further with our team on the project workflow, associated time for each task, concerns about quality of image, necessary review process to ensure quality, etc....

There is also a setup fee which covers recruitment of agents, creation of training materials, training of agents, creation of the project on our platform, and engineering requirements and this is typically about $1500, which would apply here since this is a new project.

Daniele and/or I are more than happy to jump on a call to discuss the pricing if you want to chat about this. I realize there are some pretty large ranges presented, and that these numbers are pretty high, but I hope this gives you a good general idea. We could also set some time to walk through the actual workflow with you to eliminate certain parts of the task that are causing the time and complexity (and price) to increase."

Eric: Should be 54 dates for 1908-1909. For 1913-1914 would be another 50+. So would be $10k-$20k.
Asaf: call back to work on pricing (Closer to $10k).

XXVII. Collecting Bond Ratings for 1909 for All Bonds (update)

From Eric (5/13/15):

"So the Moody's bond rating tables actually include stuff that is more like short-term notes than regular coupon bonds. (The data entry company keeps coming back to me with questions.) In a case where it's "ctf. of indeb. (4%)," or "3-year 5% notes" I'm having them enter the interest rate as the "coupon," and other stuff as class.

Also there's a few cases of jointly issued bonds, eg, "Atlantic Coast Line - L. & N." and I think the Northern Pacific and Great Northern had at least one joint issue as well. I'm having them enter that full name under issuer. In any case I think there's very few of those."

That sounds like a reasonable way to handle it, but I just wanted to include it in the update document so we have a record of it.

XXVIII. June Conference Presentation

28.1 Original Regression Table from 5/20/14

Dependent Variable	(1) Yield (log)	(2) Yield (log)	(3) Yield (log)	(4) Yield	(5) Sales (log)
Above-median rating x post April 1909	-0.0307* (0.0134)		-0.0412* (0.0134)	-0.00267* (0.00100)	0.1031 (0.1146)
Below-median rating x post April 1909		0.0134 (0.0103)	0.0353* (0.0117)	0.00189* (0.00064)	0.2047+ (0.1168)
Rating Aa x post April 1909			-0.0197 (0.0128)	-0.00134 (0.0090)	0.1662 (0.1398)
Rating A x post April 1909			-0.0442* (0.0136)	-0.00276* (0.0085)	0.2758* (0.1073)
Rating Baa or Lower x post April 1909			-0.0697* (0.0159)	-0.00381* (0.00096)	0.1457 (0.1236)
Constant	-3.17* (0.002)	-3.17* (0.002)	-3.16* (0.006)	0.0426* (0.00039)	7.38* (0.0468)
Observations	20,001	20,001	20,001	20,001	20,000
Clusters	174	174	174	174	174
R-squared	0.550	0.548	0.553	0.499	0.356
Firm FE	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES

Standard errors adjusted for clustering on firms

These are the regression results from last summer from using the secondary market NYSE bond data. We have made changes and unfortunately I don't have the exact same code and/or underlying data, but I did my best to replicate the results in STATA do file 20140716_PerpYieldResponse_Approx_Same_As_Slide31.do in the dropbox.

28.2 Approximate Replicating Table

Dependent Variable	(1) Yield (log)	(2) Yield (log)	(3) Yield (log)	(4) Yield	(5) Sales (log)	(6) Price Impact (bps)
Above-median rating x post April 1909	-0.0293* (0.0161)		-0.0461* (0.0180)	-0.00327* (0.00150)	0.3085* (0.1226)	-1.03+ (0.59)
Below-median rating x post April 1909		0.0075 (0.0103)	0.0314* (0.0111)	0.00191* (0.00067)	0.1592 (0.1147)	-0.636 (0.416)
Rating Aa x post April 1909			-0.0271+ (0.0164)	-0.00209 (0.00135)	0.2130 (0.1400)	-0.00 (0.563)
Rating A x post April 1909			-0.0523* (0.0158)	-0.00351* (0.0011)	0.0442 (0.1319)	0.164 (0.44)
Rating Baa or Lower x post April 1909			-0.0729* (0.0189)	-0.00437* (0.0013)	0.0055 (0.1710)	0.039 (0.069)
Constant	-3.03* (0.001)	-3.03* (0.004)	-3.02* (0.005)	0.0493* (0.00033)	5.064* (0.0648)	8.80* (0.237)
Observations	13,225	13,225	13,225	13,225	13,263	12,984
Clusters	172	172	172	172	172	172
R-squared	0.5233	0.5217	0.5264	0.4844	0.3539	0.1202
Firm FE	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES

Standard errors adjusted for clustering on firms

Generally speaking the results look similar. There were some "fake" observations in the original table (missing values counted as observations) which are fixed here which is why the observations fall a lot. The only other big difference that jumped out at me was column (5), which I think might be related to how I computed sales. I'm still looking into that.

I added one new piece of analysis in column 6 which is the price impact (Absolute Value of change in Price (weekly or daily)/ total sales (daily or weekly)). It looks like the price impact falls substantially for those firms with surprise ratings (not sure why that would be). Even though both up and down aren't significant, they are quantitatively large, and a combined test certainly would be significant.

Eric: Run this intra-day and EOD of separately.
Asaf: Look at more clever intra-day stuff…
Eric: Is there anything systematic of surprise rated bonds… Most senior bonds of bad RRs or worst bonds of good RRs. Look more at who the surprises (up or down) is… How about the liquidity? Maturity?

Next Steps:
1) Work on ideas that could be completed in time for the conference with data we already have

2) Look especially at intraday data

XXIX. Industry Assignment Problem

For Part 2 of the 1911 entries it appears (I didn't notice till now) they transcription company altered the order of a small subset of the company entries in the middle of the transcription. Example:

BERGEN COUNTY. N. J	1911	N/A	4	1911-24		Provident Institution for Savings
UNION RAILWAY COMPANY OF CHESTER, PA	1911	N/A	5	1913		North Philadelphia Trust Company
UNION RAILWAY COMPANY OF CHESTER, PA	1911	N/A	5	1913		Savings Fund Society of Germantown & Its Vic
UNION RAILWAY COMPANY OF NEW YORK	1911	N/A	5	1942		Albany Insurance Company.

The problem is that union railway in 1911, for example, is now assigned to "government securities". I need to go through the part 2 file in full manually, fix the order, and the reassign industries to those issuers.

XXX. Moody's 1909 Ratings

In order to make the data from the Moody's 1909 ratings manual usable we need to match in the parent name and issuer name to the holdings data. The issue is that even fuzzy logic might have a tough time because of all the abbreviations in the issuer_name. It could probably handle matching "parent" names to "cname_hold_orig", but there are only 92 unique parent names so that would be very easy to do by hand.

Example:

parent	issuer_name	parent_firm	parent_cnam_hold_orig	firm_id_holdings	cname_hold_orig
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Chicago & St. Louis Ry.	22	ATCHISON, TOPEKA & SANTA FE	156	CHICAGO AND ST LOUIS
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Chic. S.F. & Cal.	22	ATCHISON, TOPEKA & SANTA FE		
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	A. T. & S. F.	22	ATCHISON, TOPEKA & SANTA FE	22	ATCHISON, TOPEKA & SANTA FE
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Hutchinson & So. Ry.	22	ATCHISON, TOPEKA & SANTA FE	368	HUTCHINSON AND SOUTHERN
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	San Fran. & San Joaquin V.	22	ATCHISON, TOPEKA & SANTA FE	783	SAN FRANCISCO AND SAN JOAQUIN VALLEY
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Sante Fe, Prescott & P.	22	ATCHISON, TOPEKA & SANTA FE	785	SANTA FE PRESCOTT AND PHOENIX
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Cane Belt R. R.	22	ATCHISON, TOPEKA & SANTA FE		
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Gulf, Beaumont & K. C.	22	ATCHISON, TOPEKA & SANTA FE	349	GULF BEAUMONT AND KANSAS CITY
(THE) ATCHISON TOPEKA & SANTA FE RAILWAY COMPANY	Prescott & Eastern	22	ATCHISON, TOPEKA & SANTA FE		
ALABAMA & VICKSBURG RAILWAY	Vicksburg & Meridian R.R.	3	ALABAMA AND VICKSBURG	886	VICKSBURG AND MERIDIAN
ALABAMA & VICKSBURG RAILWAY	Ala. & Vicks. Ry	3	ALABAMA AND VICKSBURG	3	ALABAMA AND VICKSBURG

The challenge is the 1,021 issuer names which are abbreviated. Doing 100 names/hr is definitely feasible manually so I think we should just have somebody spend 10 hours to do this (could even be Shrey).

XXXI. New RA: Shrey

I spoke with Shrey about the task of manually checking holdings names were it looks like there might be a discrepancy. Some of them are going to be obvious to Shrey (like "boston terminal" vs. "Houston belt and terminal"), but others are not ("ann arbor company" vs. "ann arbor railroad"), so I asked him to make two columns, one with clear problems and ones with uncertainty. He is also going to make a word document where he notes common problems so

perhaps we can code/automate them. It sounds like he will probably be able to start helping us starting next Tuesday (6/2).

Carola: How long is he willing to work?

Carola: Have Shrey do the Moody's matching rather than the holdings data…

Eric: Have Shrey use full name in TOC if need to.

XXXII. Holdings Analysis Redux

Table 32.1 Mean Herfindahl Index of Security Owners Regional Distribution

	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	76.6%	64.1%	23.7%	39.2%
1911	72.7%	59.9%	24.9%	36.8%
1913	73.8%	55.2%	25.5%	34.9%
1908-1911	-3.9%	-4.2%	+1.2%	-2.4%
1911-1913	+1.1%	-4.7%	+0.6%	-1.9%

Table 32.2 Median Herfindahl Index of Security Owners Regional Distribution

	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	100.0%	100.0%	5.4%	20.7%
1911	100.0%	80.4%	5.9%	16.0%
1913	100.0%	55.6%	5.3%	14.6%
1908-1911	-	-19.6%	+0.5%	-4.7%
1911-1913	-	-24.8%	-0.6%	-1.4%

As we noted previously, steam railroads are the only industry to see an increase in investor regional concentration from 1908-1911. This is true looking at the mean or the median Herfindahl.

Eric: Have we looked at changes in the stock (amount) of bonds per year? Are firms issuing more of the same bond?

Table 32.3 1908-1911 Chg in Mean Herfindahl Index Security Owners Regional Distribution

1908 Herf	Government	Industrial & Misc.	Steam Railroad	Street Railway
0-10%	+18%	+14%	+6%	+7%

10-25%	+12%	+10%	-2%	+7%
25-50%	+4%	-1%	-9%	-10%
50-75%	-10%	-10%	-15%	-19%
75-100%	-18%	-28%	-57%	-47%

Table 32.4 1908-1911 Chg in Median Herfindahl Index Security Owners Regional Distribution

1908 Herf	Government	Industrial & Misc.	Steam Railroad	Street Railway
0-10%	+3%	+2%	0%	+1%
10-25%	+1%	-3%	-5%	+3%
25-50%	-6%	-13%	-13%	-12%
50-75%	-7%	-7%	-12%	-10%
75-100%	-	-	-80%	-50%

It looks like now once we control for initial concentration the railroads actually see the largest decline.

Table 32.5 Mean Herfindahl Index of Security Owners Regional Distribution (<$10k issuance)

	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	96%	92%	91%	91%
1911	78%	85%	50%	67%
1913	79%	86%	43%	64%
1908-1911	-17%	-6%	-41%	-25%
1911-1913	1%	1%	-8%	-3%

Table 32.6 Mean Herfindahl Index of Security Owners Regional Distribution ($0.5-1Mil)

	Government	Industrial & Misc.	Steam Railroad	Street Railway
1908	26%	19%	13%	13%
1911	20%	19%	14%	7%
1913	14%	30%	12%	3%
1908-1911	-6%	0%	1%	-6%
1911-1913	-6%	11%	3%	-4%

As we might expect size issuance (as proxied by total $ value listed in holdings books) is also very related to concentration. The smaller the issue in 1908 the higher the initial concentration and the more negative the change in concentration from 1908 to 1911.

Table 32.6 Regression Herfindahl Index of Security Owners Regional Distribution

```
      Source │       SS          df       MS              Number of obs =    7681
─────────────┼────────────────────────────────           F(  9,  7671) =  227.07
       Model │ 210.815475        9   23.4239417           Prob > F      =  0.0000
    Residual │ 791.329797     7671   .103158623           R-squared     =  0.2104
─────────────┼────────────────────────────────           Adj R-squared =  0.2094
       Total │ 1002.14527     7680   .130487666           Root MSE      =  .32118


chg_conc~1908 │     Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
──────────────┼───────────────────────────────────────────────────────────────
 _Iconcentra_2 │ -.0804878   .0154072     -5.22   0.000    -.1106901   -.0502855
 _Iconcentra_3 │ -.2171034   .0157001    -13.83   0.000      -.24788   -.1863268
 _Iconcentra_4 │ -.2937301   .0158937    -18.48   0.000     -.324886   -.2625742
 _Iconcentra_5 │ -.4603748   .0113075    -40.71   0.000    -.4825406    -.438209
  _Iissue_siz_2 │ -.0277645   .0094914     -2.93   0.003    -.0463702   -.0091588
  _Iissue_siz_3 │ -.1417393   .0134265    -10.56   0.000    -.1680589   -.1154197
  _Iissue_siz_4 │ -.1525715   .0223433     -6.83   0.000    -.1963705   -.1087726
  _Iissue_siz_5 │ -.1991203   .0196075    -10.16   0.000    -.2375563   -.1606844
       SteamRR │ -.0699949   .0113442     -6.17   0.000    -.0922327   -.0477571
         _cons │  .3287162   .0129964     25.29   0.000     .3032398    .3541926
```

The interpretation of the above simple OLS regression is that controlling for concentration and size fixed effects for each quintile the change in concentration from 1908 to 1911 for Steam Railroad securities is -7.0 percentage points, with is consistent with interpretation before. We can also see that, as expected, the higher the initial concentration and the larger the size the larger the decline in concentration from 1908 to 1911.

On the other hand if I run a "placebo test" (sort of) by running the above regression from 1911 to 1913 using 1911 size and concentration buckets I get a decline of 8 percentage points for Steam Railroad Securities. So either the effect takes more than 2 years to alter holdings or more likely we are picking up some omitted variables that cause a time trend and drive both…

Carola: Maybe look at this by institution type…

XXXIII. Microstructure Response to Ratings Introduction

TBD

XXXIII. Samasource Update

Call today (5/27/2015) at 3pm EST.

Eric: For presentation:
2) Redo summary statistics on slide 18.
3) Take out 1913 and 1914 (only high frequency data for period around 1909 ratings release). Try 1906-1911. After 1911 RRs start to get in trouble in a way completely unrelated to Ratings… Remove those from plot in slide 19.
4) Would the above or below median rating change that Eric used?
5) Make sure presentation is up-to-snuff…

2 Major Comments
6) Slide 26: Should be 2-dimensional plot with duration and ratings (include either duration or maturity buckets). Have buckets or surprises within buckets… Remaining life of the bond quartiles… For every security do we have maturity?
7) In regression on slide 31 maybe control for years to maturity, duration, or something else (maybe buckets of those)
8) There should be an effect on firm value (equity value)…

9) Maybe look at what % is held outside of NY or NY & Chicago… We can try stuff like that easily in a regression… Maybe clear with more controls. Changes are interesting and that might give different picture…

XXXIV. Samasource Update

They will send an SOW by this Friday. Here are big picture details:

10) Price per line is $0.14/line.
11) Duration of project: 3 months
12) Project total cost estimate: $10,000
13) Assumption is that we have 68,100 lines (300 lines per page, 227 pages).
14) This project has a one-time set-up fee of $500 to cover for training and workflow design
- They probably won't be fully done till sometime in October, but they will send data as it is completed so RAs can still do stuff.
- Meeting next week to go over guide and SOW. Can kick off as early as next week if all is good.
- I need to be reachable for two weeks during training.
- I need to send them additional explanation materials by this Friday (7/10/15) if we hope to start by next week

XXXV. Literature Review

EMPIRICAL

35.1 Shocks to Asymmetric Information via Exogenous Equity Analyst Coverage Declines

Instrument: Shut-down of sector level analyst coverage increases asymmetric information which they argue are primarily driven by broker-specific factors not performance of the sector (some papers use broker mergers).

Exclusion restriction evidence: Sectors not covered have not recently had worse earnings or other related factors and don't cluster for multiple brokerage firms so usually related to areas where brokers have made poor predictions or other broker-specific effects.

1) **Kelly and Ljungqvist (RFS 2012)** **"Testing Asymmetric-Information Asset Pricing Models"**
   a. Proxies for asymmetric information rose (bid-ask spreads for example)
   b. Stock prices fell
   c. Retail investor demand fell (more institutional owners of the securities)
   d. Liquidity betas rise
   e. Market beta falls (SMB and HML are unclear)
"Our results confirm that information asymmetry is priced and imply that a primary channel that links asymmetry to prices is liquidity."

Eric: Alternative could be marketing story… They are "selling" the security.

2) **Irani and Oesch (JFE 2013)** **"Monitoring and corporate disclosure: Evidence from a natural experiment"**
   a. Firms disclose less in financial reporting

b. Concentrated among firms with bad governance

3) **Derrien and Kecskes (JF 2012)** [**"The Real Effects of Financial Shocks: Evidence from Exogenous Changes in Analyst Coverage"**](#)

    a. Investing and external financing drop 2% (as percent of assets)

    b. Use less equity and long-term debt, but no change in short-term debt and increase use of cash and retained earnings.

4) **Wei (2013)** [**"Information Asymmetry and Insider Trading: Evidence from Exogenous Changes in Analyst Coverage"**](#)

    a. Insiders earn higher abnormal returns after asymmetric information increases, but do not trade more frequently

Carola:

1) Most shocks to asymmetric information are equity analyst coverage and here is why ours is better
2) Explain in intro how different

35.2 Asymmetric Information Shocks and Debt Ratings

1) **Sufi (RFS 2009)** [**"The Real Effects of Debt Certification: Evidence from the Introduction of Bank Loan Ratings"**](#)

Methodology: Introduction of syndicated bank loan ratings by Moody's and Poors in 1995 as shock to debt certification and look at firms that obtain a rating and compare to firm's that don't.

Findings for firm's that obtain a rating:

    a. Increased use of debt

    b. Higher asset growth

    c. More cash acquisitions

    d. Increased investment by less informed investors (still pretty sophisticated though since they are banks and other large institutions)

    e. Effects larger for lower credit quality firms

Eric: Does he look at better vs. worse? Do ratings help on average?

2) **Tang (JFE 2009)** [**"Information Asymmetry and Firms' Credit Market Access: Evidence from Moody's credit rating format refinement"**](#)

Methodology: In 1982 Moody's increased the granularity of their ratings to include +/- which he uses as a shocked reduction to the level of asymmetric information and informational shock on debt ratings.

    a. Firms with refinement upgrades experience and additional decrease in their ex-post borrowing cost, issuer more debt, rely more on debt (over equity) financing, invest more, have less cash accumulation, and faster asset growth.

Eric: 3rd party security certification.

THEORY

1) **Boot, Milbourn, and Schmeits (RFS 2006)** "Credit Ratings as Coordination Mechanism"
   a. There are multiple equilibria since firm's optimal investment depends on investor priors about the firm's credit quality
   b. If enough investor priors are that the firm's credit is high quality the optimal project is undertaken
   c. CRA (credit rating agency) monitors firms and provides information to market that co-ordinates priors and allows for optimal equilibrium

Comment: Key to mechanism here is that CRA can actively monitor and influence firm behavior via "warnings" that they will alter the equilibrium if inappropriate action is taken.

2) **Fulghieri, Garcia, and Hackbarth (WP 2015)** "Asymmetric Information, security design, and the pecking (dis)order"
   a. "shape of the optimal security crucially hinges on the 'location' of asymmetric information in the firm-value distribution"
   b. Introduce Information Costs in the Right Tail (ICRT) and show when asymmetric information has small effect on right tail of firm value risky debt is optimal for raising low levels of capital, when raising large levels of capital equity is optimal
   c. Assuming insiders have private information on 1st moment (ex. assets in place), but not 2nd moment (ex. growth opportunities), show that pecking order can be reversed and firms more likely to issue equity

Comment: Paper still needs some work, but drives home the point that equity vs. debt in pecking order depends on what part of firm value distribution has asymmetric information (ex. 1st vs. 2nd moment).

3) **Strebulaev, Zhu, and Zryumov (RR JF 2014) "Dynamic Information Asymmetry, Financing, and Investment Decisions"**
   a. Identical mechanism to (FGH 2015 above), but in a dynamic setting. Again asymmetric information about losses to asset in place (1st moment) rather than 2nd moment flips pecking order result

4) **Lambert, Leu, and Verrecchia (RF 2012)** "Information Asymmetry, Information Precision, and the Cost of Capital"
   a. If markets are perfectly competitive then an increase in asymmetric information between the firm and the average investor ("information precision") raises the cost of capital, while an increase in asymmetric information between investors (holding precision constant) does not.
   b. In perfectly competitive markets if "insiders" become more informed (or own more of the capital), but non-insiders have the same amount of information (or same amount of capital), asymmetric information between investors would rise, but the cost of capital would actually fall because information precision rose.
   c. With imperfect competition, an increase in asymmetric information between investors reduces liquidity which can raise the cost of capital

35.3 Differences between Introduction of Rating Agencies and Previous Literature

1) Exogenous analyst coverage literature

a. We look at asymmetric information in the debt market rather than the equity market.

b. Analysts provide on-going qualitative coverage, while we are specifically focusing on ratings which arrive less frequently and provide more quantitative evaluations of risk.

2) Debt certification literature

a. We look at bonds not loans which seems like a more natural place for debt certification to matter since buyers are the public not sophisticated institutions.

b. We look at secondary market prices and (if collected) bid-ask spreads so can tie (hopefully) market channel more clearly to "real channel"

c) In our period ratings or investor paid not issuer paid (so who gets ratings is endogenous but on the part of Moody's based on general interest rather than by the firm which as shown by Sufi is related to many variables likely to relate to outcomes). Thus using controls for unrated firms is likely to be problematic (though still a bit) than it is for modern literature.

d) There are no regulatory implications of ratings in our period so results are all about asymmetric information/certification and not about regulatory-induced demand making interpretation more cleanly identified.

e) In our period ratings are given after issuance while now it is prior to issuance.

XXXVI. Next Steps

1) Send Samasource materials (ex. acceptable fractions) so they can begin data collection next week (Asaf)
2) Review SOW Samasource sends Friday so we can respond quickly (All)
3) Meet with Samasource to review materials and SOW next Thursday (7/16) at 3:30pm EST (Asaf)
4) Next week discuss literature review, novel empirical predictions, etc in more depth. Papers added if needed (All)
5) Document and write-up tasks Shrey was doing and establish plan to complete them or assign them (Asaf)
   a. oDesk for 1st page of CFC
   b. Holdings id matches manually checking
   c. Matching 1909 manual ratings and other data to holdings data
   d. Merge in and summarize equity data
   e. 1-linear for Kellogg professional RA
6) Look at different specifications of regressions in current draf (Asaf)

     a.   Rating-level or 1908-yield-level interactions
     b.   With and without rating level controls
     c.   Smaller window around rating release
     d.   Issue vs. issuer level response
     e.   Duration and/or maturity controls/interactions

Eric: Peter Koudjis's papers do well. Think about how he writes and presents to get people engaged...

https://people.stanford.edu/koudijs/research

https://people.stanford.edu/koudijs/sites/default/files/18thcentury.pdf

XXXVII. Samasource Update

15) Bid-ask data collection has started and they have sent a 1-pg sample which generally looked good and I went over it with them to fix small issues.

XXXVIII. List of Datasets

1) NYSE Bond Market Data

a) weekly_data_cleaned_v1.dta
All weekly closing NYSE bond market data with winsorized data at 99th percentile and perpetuity yield, returns, price impact and $ volume computed.
- Created using "RawFileSetup.do" and raw input file "20140427_AllCleanMktData_v5.dta".
More details on process can be seen section XXXX below
- Collection day-of-the-week is Sundays (constitutes all trades during the week)
- Monthly 1/14/05-4/11/08; Weekly 4/11/08-4/23/10; Monthly 5/14/10-1/4/13; Weekly 1/4/13-7/11/14; No data 7/11/14-12/18/14; Weekly 12/19/14-1/2/15
- All variables well labeled. Key: issuer_id & issue_id & trade_id

b) intraday_data_cleaned_v1.dta
All intra-day NYSE bond market data with winsorized data at 99th percentile and perpetuity yield, returns, price impact and $ volume computed.
- Created using "RawFileSetup.do" and raw input file "20140427_AllCleanMktData_v5.dta".
More details on process can be seen section XXXX below
- Collection day-of-the-week is Thursday paper (constitutes all trades on Wednesday)
- Weekly 4/29/08-7/29/14; No data 8/1/14-12/22/14; Weekly 12/23/14-1/6/14
- All variables well labeled. Key: issuer_id & issue_id & trade_id

c) merged_intra_wkly.dta
Merged combination of intra-day and weekly data where intra-day data are aggregated to a daily level using closing prices or sum of total sales.
- Created using "RawFileSetup.do" and raw input file **weekly_data_cleaned_v1.dta**
 & **intraday_data_cleaned_v1.dta**. More details on process can be seen section XXXX below
- All variables well labeled. Key: issuer_id & issue_id & trade_id

d) BidAskNYSEBondData.dta
Currently being collected by Samasource. Expected finish date: Early Oct.

2) Security Holdings Data

a) HoldingsData_Final_wIndustry.dta.dta

All data on investors holdings merged together for 1908, 1910, and 1911.

- All variables well labeled. Key: firm_id_holdings, issue_id_holdings, investor_id

- Note: nyse_bond_issuer_id matches "issuer_id" from 1C and was the first attempt at matching across those datasets. Still needs to be checked more. Also no issue-level identifier match across those datasets has happened yet.

3) Ratings Data

a) Firm_Ratings_v2.dta

All ratings (overall, risk, and saleability) aggregated (median, mean, max, min, outstanding-weighted) to a firm level for 1909 & 1914. *Only collected for firms initially in the bond NYSE market data.*

- Created using "RawFileSetup.do" and raw input file "201404171636_BondRatingRecordInformation_cmb.csv". More details on process can be seen section XX below

- All variables well labeled. Key: issuer_id

- Still needed: issue_id key/matches

- Could drop this file in favor of 3B and 3C and later on add 1914 ratings in full if we decide to do that…

b) Moodys_1909_Bond_Ratings.xlsx

All ratings (overall, risk and saleability), maturity, lien, miles, income for the bond, interest per mile, and factor of safety *for all bonds in 1909 manual.*

- Need to match issuer and issue names to identifiers in other files

- Currently a spreadsheet. Need to code it up so usable in Stata (ex. Convert "4 1/2s" under coupond to 4.5)

c) Moodys_1909_Stock_Ratings.xlsx

Stock rating, dividend, outstanding, priority, and other info *for all stocks in 1909 manual.*

- Initial attempt has been made to match company names to "issuer_id" key from **merged_intra_wkly.dta**, but this needs to checked again and probably fixed. Also need to match identifiers in holdings data.

- Currently a spreadsheet. Need to code it up so usable in Stata

4) NYSE Stock Price Data

a) MonthlyNYSEStockPriceData.dta

All NYSE closing prices, bid, ask, and trading volume at a monthly frequency from 1893-1925.

- No issuer key yet. Needs to be created based on company names (which are only partially cleaned).

- Needs to be matched to other datasets
- Needs to be complemented with weekly/daily data around the introduction of ratings?

5) Accounting Data (Eric Hilt has details)

XXXIX. Completed/Outstanding Tasks

Completed

1) 1A & 1B have matched internal issuer, issue, and trade IDs and across each other which appear in 1C
2) 2A within each year issuer and issue IDs are good.
3) 2A issuer level identifiers have been matched to 1C, but this needs to checked again.
4) 3A issuer-level identifiers have been merged/matched successfully to 1C

Outstanding Tasks

1) 2A need to double check issuer and issue level IDs across years
2) For 2A need to create issue-level identifier matches from 2A->1C as well as the other datasets. My guess is holdings data (2A) is a superset of all securities so can really use this as the master that everything merges to, rather than doing a huge number of bilateral matches…
3) Once data collected for 1D issuer, issue, and trade IDs will need to created and matched to 1C
4) 3A needs issue-level ids and need to merge 3A -> 1C (unless we want to drop 3A for now…)
5) 3B need to generate issuer and issue-level ids, code up all variables, and match 3B->1C and 3B->2A
6) 3C need to generate issuer -level ids, code up all variables, and match 3C->1C and 3C->2A, 3C->4A
7) 4A need to generate issuer -level ids, code up all variables, and match 4A->1C and 4A->2A, 4A->3C.

One way to do this is match everything to holdings data as first step AND have IDs that are consistent internally so that can avoid that issue!!!! **** Rather than having all outstanding tasks this is the clear way to do it.

Asaf should write what simplifications fuzzy logic is doing so I can follow it…

Asaf tasks before JMP overtakes him completely:

1) Send along code and data that includes only small subset of holdings data (based on "simple cleaning techniques") that need to be checked.

2) Summarize files/code that produce presentation results

XXXX. RawFileSetup.do

I. Initial Set-up

Input:

20140427_AllCleanMktData_v5.dta – All raw weekly and intraday bond market data.

Output:

intraday_data.dta – All intra-day bond market data with variable labels.

weekly_data.dta – All weekly bond market data with variable labels.

Purpose:

Split raw data into separate data sets for intra-day and weekly bond market data and provide variable labels.

II. Matt's Automated False Negatives Code

Input:

weekly_data.dta - All weekly bond market data with variable labels.

Output:

weekly_data_MattCleaned.dta – All weekly bond market data with false negatives automatically fixed.

Purpose:

Use an automated process to find bonds which are the same but were given different identifiers

III. Just Weekly Data

Input:

weekly_data_MattCleaned.dta - All weekly bond market data with variable labels.

Output:

weekly_data_cleaned_v1.dta – All weekly bond market data with winsorized data at 99th percentile and perpetuity yield, returns, price impact and $ volume computed.

Purpose:

Generate summary statistics for weekly bond market data, winsorize data, and generate new variables perpetuity yield, returns, price impact and $ volume.

Results:

61,787 weekly bond perpetuity yield observations

260 unique dates from 1905-1915 (implies average of 237 bonds/day)

2,112 unique bonds listed (# obs/bond; mean=29; median=4; 25[th]=1; 75[th] = 29)

Summary statistics for all observations:

	#	Mean	Median	Stdev	25th	75th
Perp yield	61,787	4.98%	4.72%	1.27%	4.26%	5.26%
Price Impact	60,904	0.050%	0.009%	0.11%	0.002%	0.041%
Abs return	61,013	0.172	0.105	0.243	0.0399	0.212

III. Just Intra-day Data

Input:

intraday_data.dta - All intra-day bond market data with variable labels.

Output:

intraday_data_cleaned_v1.dta – All intra-day bond market data with winsorized data at 99th percentile and perpetuity yield, returns, price impact and $ volume computed.

Purpose:

Generate summary statistics for intra-day bond market data, winsorize data, and generate new variables perpetuity yield, returns, price impact and $ volume.

Results:

51,431 weekly bond perpetuity yield observations

190 unique dates from 1908-1915 (implies average of 271 transactions/day)

320 unique bonds listed (# obs/bond; mean=161; median=1; 25th=0; 75th = 9)

Summary statistics for all observations:

	#	Mean	Median	Stdev	25th	75th
Perp yield	51,431	5.04%	4.80%	1.22%	4.23%	5.43%
Price Impact	50,695	0.094%	0.024%	0.202%	0.005%	0.084%
Abs return	50,707	19.48%	11.79%	26.74%	4.12%	24.81%

V. Firm-level Ratings and Characteristics

Input:

201404171636_BondRatingRecordInformation_cmb.csv – All data on bond ratings and characteristics taken from Moody's manuals.

Output:

Firm_Ratings_v2.dta – All ratings (overall, risk, and saleability) aggregated (median, mean, max, min, outstanding-weighted) to a firm level.

Purpose:

Generate firm-level ratings with variable labels and combine company name into one variable called "company name".

XXXXXI. Samasource Update

- Agents have completed through 4/14/1909, next week's delivery will likely cover through 5/25/1909.
- 33,198 lines transcribed in 33 dates - about 1,006 lines/date.
- They've submitted 33 of 57 dates (58%)

XXXXXII. Quick Explanation of Fuzzy Logic Algorithm

Fuzzy Lookup uses Jaccard similarity, which is defined as the size of the set intersection divided by the size of the set union for two sets of objects. For example, the sets {a, b, c} and {a, c, d} have a Jaccard similarity of 2/4 = 0.5 because the intersection is {a, c} and the union is {a, b, c, d}. The more that the two sets have in common, the closer the Jaccard similarity will be to 1.0. The method I used also assigns weights based on the frequency that records occur. Frequent words, like say "Corporation", receive less weight in the algorithm than infrequent words. I also allow for transformations so that the program automatically considers common corrections of typos and after trying all corrections computes the minimum weighted Jaccard similarity. For more details see this link:

https://atidan.files.wordpress.com/2013/08/fuzzy-lookup-add-in-for-excel.pdf

Add-in for Microsoft excel is available here:

http://www.microsoft.com/en-us/download/details.aspx?id=15011

XXXXXIII. Eric Questions

In particular, I'm wondering about the analysis part of the programs. That doesn't seem to be in the current RawFilesSetup program - or am I missing it? I'm particularly interested to see the code that identifies suprises in response to the ratings. I found some stuff in some of the archive folders but I am not sure whether that's 'current.'

Answer: The code to run the analysis in the paper so far and produce surprises is called "201506_AnalysisForPresentation.do" and can now be found in "C:\Users\asafb\Dropbox (MIT)\RatingAgencies\Data_Programs\Code\MostRecent". The associated data to run this is described in the previous update and can be found here:
"C:\Users\asafb\Dropbox (MIT)\RatingAgencies\Data_Programs\Data\MostRecent"
The only additional data file is "rating_1909_surprise_v5_WithHoldingsIDMatches.csv", where I manually entered the holdings data IDs at an issuer-level for all surprises to make sure they were accurate for the presentation.

Also, a broad question about the surprises: we have firms where we observe prices on multiple bonds. In a case like that, how is the firm-level surprise variable defined? Is it ever the case that within the same firm there could be suprises in opposite directions, or some bonds with surprises and some without surprises?

Answer: The firm-level surprise is based on the median yield in 1908 for all bonds (weighted by # of observations) of the firm and the median rating across all bonds (if the median is halfway between two rating grades I round up). At this point we haven't done any analysis of surprises at an issue-level. This is at least partially because we would need to create issue-level IDs in the ratings data we are comfortable matching to the NYSE bond data.