Gender Difference in Financial Confidence During the Great Recession: A Longitudinal Study

Ian Cero, Tenchi Smith, Zheran Wang
Department of Psychology, Department of Industrial & Systems Engineering, Department of Statistics
Tuesday, 4/25/17

Abstract

The dataset from the Interuniversity Consortium for Political and Social Research (ICPSR)'s research in public response of financial crisis was used to conduct a longitudinal analysis using a linear spline model and SAS' "PROC MIXED" statistical package. The main goal was to study how males' and females' financial confidence levels might differ during The Great Recession beginning in 2008. It was found that confidence varied significantly across time and had a significantly slower change after the third measurement occasion for both genders. Women's confidence began lower than men's and increased at a slower rate throughout the study period.

Table of Content

List of Tables and Figures

Introduction

From late September 2008, peak of the financial crisis, until August 2011, a nation-wide survey was conducted at 8 different points of time to collect public response and to understand the trajectory of risk perception in the midst of the economic crisis. (Burns, 2016)

Among many factors considered in the survey such as retirement and job status, anxiety toward the crisis, confidence that the crisis will be successfully managed, as well as personal information collected such as age and gender, we chose gender and days as covariates, and confidence across different time points as response variable. Researchers believe there is a significant gender difference in confidence; people's confidence would also vary significantly over time. Such information would help predict the general public's behavior in possible future financial crises and therefore is valuable to study.

All subjects were given the same survey at 8 different time points during the financial crisis. The measurement time points are displayed as follows. Note that the measurement intervals are unequal. Among the initial 937 respondents, those who did not respond at any of the 8 measurement time points were excluded. As a results, 556 remaining subjects have at least one recording of confidence at a measurement time point. We analyzed and compared the complete cases (n=90) versus all cases (n=556) which included complete and incomplete cases. Incomplete cases were analyzed using multiple imputation.

- Measurement Time Point 1: Sept 29 - Oct 1, 2008 (N=802)
- Measurement Time Point 2: Oct 8-10, 2008 (n=712)
- Measurement Time Point 3: Nov 5-7, 2008 (N=755)
- Measurement Time Point 4: Dec 6-18, 2008 (N=654)
- Measurement Time Point 5: March 21-26, 2009 (N=609)
- Measurement Time Point 6: June 30 - July7, 2009 (N=715)
- Measurement Time Point 7: Oct 3-9, 2009 (N=645)
- Measurement Time Point 8: Aug 9-15, 2011 (N=325).

ID	Subject number
Confidence	Observations in scale 1 to 100
logC	= log[(Confidence+1) / (100-confidence+1)] New dependent variable after transformation
Days	Number of days after the first observation.\
Days3	Knot at the third measurement point
G	Gender: 1=male, 0=female

Table 1 : List of Variables

Statement of Problem and Statistical Analysis Approach

We wanted to assess the difference, if any, between gender in financial confidence during The Great Recession.

We first examined the relationship between the mean transformed confidence versus days from baseline for each gender. Means of each group were examined using PROC MEANS with OUTPUT OUT statement. Mean confidence versus days was plotted using PROC SGPLOT (Figure 1). We observed in Figure 1 that at time point 3, there was likely a slope change in the prediction of confidence. We decide to use spline model to fit the data with one knot at measurement time point 3.

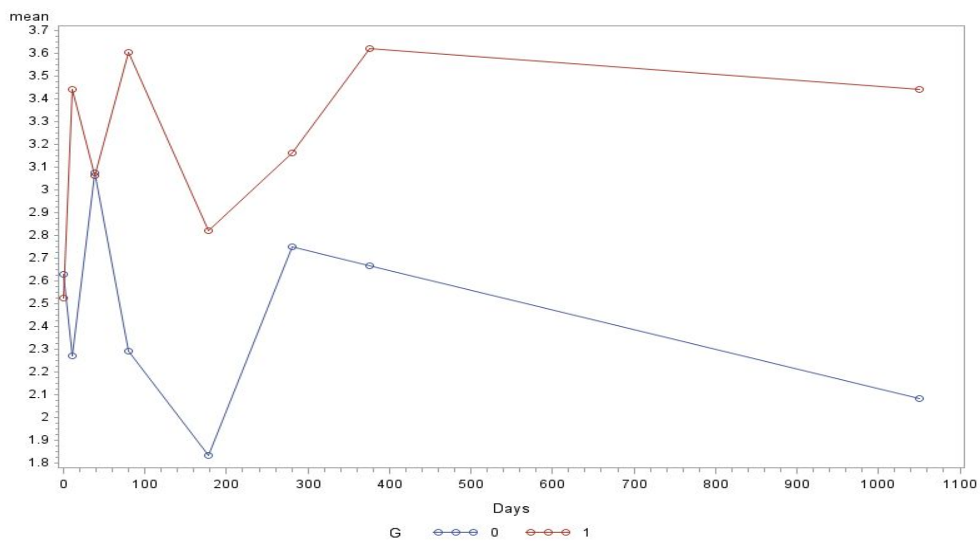

Figure 1. Plot Means Confidence Versus Days

Full Model

We proposed a full model as follows:

$$Log((Confidence +1) / (100\text{-}Confidence+1)) =$$
$$\beta_0 + \beta_1 * G_i + \beta_2 * Days_{ij} + \beta_3 * Days_{3ij+} + \beta_4 * (G_i{*}Days_{ij}) + \beta_5 * (G_i{*}Days_{3ij+}) + e_{ij}$$

Confidence score, originally ranged from 0 to 100, was transformed to Log((Confidence +1) / (100-Confidence+1)) to exand the range to negative to positive infinity. “Confidence + 1” was used to handle cases when respondents rated confidence as 0 while reserving the weight of their ratings.

Correlation and Covariance

First, we examined the correlation and covariance matrix. (Table 2 and 3)

Estimated R Correlation Matrix for ID 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.5933	0.3069	0.4997	0.3141	0.5394	0.3662	0.4370
2	0.5933	1.0000	0.1723	0.6279	0.2427	0.4125	0.4249	0.5961
3	0.3069	0.1723	1.0000	0.3499	0.4341	0.4499	0.4905	0.3639
4	0.4997	0.6279	0.3499	1.0000	0.5771	0.4160	0.4182	0.6299
5	0.3141	0.2427	0.4341	0.5771	1.0000	0.4303	0.2892	0.3714
6	0.5394	0.4125	0.4499	0.4160	0.4303	1.0000	0.4026	0.3860
7	0.3662	0.4249	0.4905	0.4182	0.2892	0.4026	1.0000	0.4847
8	0.4370	0.5961	0.3639	0.6299	0.3714	0.3860	0.4847	1.0000

Table 2: Correlation Matrix

Estimated R Matrix for ID 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	8.1233	4.2960	2.1901	3.4688	2.8940	3.8894	2.4243	3.2268
2	4.2960	6.4543	1.0959	3.8852	1.9936	2.6510	2.5073	3.9238
3	2.1901	1.0959	6.2709	2.1339	3.5149	2.8498	2.8533	2.3613
4	3.4688	3.8852	2.1339	5.9313	4.5437	2.5632	2.3657	3.9747
5	2.8940	1.9936	3.5149	4.5437	10.4523	3.5191	2.1717	3.1114
6	3.8894	2.6510	2.8498	2.5632	3.5191	6.3997	2.3658	2.5301
7	2.4243	2.5073	2.8533	2.3657	2.1717	2.3658	5.3955	2.9174
8	3.2268	3.9238	2.3613	3.9747	3.1114	2.5301	2.9174	6.7133

Table 3: Covariance Matrix

Based on the correlation matrix, we observed that the correlation for each pair of time points follows an alternating pattern of association (i.e., higher then lower, at every other lag). In covariance matrix, we observed that the variance were heterogeneous.

We decided to fit the data to an unstructured covariance model first and then, since our time interval are not equal, to an exponential covariance structure. Note, because our time intervals were unequal, this was among the only appropriate alternative structures available for comparison against our baseline unstructured model. We fit the exponential model by using PROC MIXED and type=exp(sp(exp)(Days)).

Unstructured Model

We fit a spline model with a knot at time point 3 (39 days from baseline), assuming an unstructured covariance model using PROC MIXED.

Fit Statistics	
-2 Res Log Likelihood	3159.0
AIC (Smaller is Better)	3231.0
AICC (Smaller is Better)	3234.9
BIC (Smaller is Better)	3321.0

Table 4: Fit Statistics for Unstructured Model 1

Solution for Fixed Effects						
Effect	G	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		2.7692	0.4010	88	6.91	<.0001
Days		0.01827	0.008760	88	2.09	0.0399
G	0	-0.2715	0.5226	88	-0.52	0.6047
G	1	0	.	.	.	.
Days*G	0	-0.01285	0.01141	88	-1.13	0.2632
Days*G	1	0	.	.	.	.
Days3		-0.01845	0.008893	88	-2.07	0.0410
Days3*G	0	0.01257	0.01159	88	1.08	0.2809
Days3*G	1	0	.	.	.	.

Table 5: Fixed Effects Result for Unstructured Model 1

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Days	1	88	4.31	4.31	0.0380	0.0409
G	1	88	0.27	0.27	0.6034	0.6047
Days*G	1	88	1.27	1.27	0.2602	0.2632
Days3	1	88	4.40	4.40	0.0359	0.0387
Days3*G	1	88	1.18	1.18	0.2779	0.2809

Table 6: Type 3 Fixed Effects Result for Unstructured Model 1

Based on the tables above, the gender and days effects were not significant. The knot effect days3 was significant. Therefore, there was no main gender and timing effect in our data. Since the knot was significant, there existed the turning point for two groups. Final model results were interpreted in detail in the Results section.

Exponential Model

Fit Statistics	
-2 Res Log Likelihood	3425.1
AIC (Smaller is Better)	3429.1
AICC (Smaller is Better)	3429.1
BIC (Smaller is Better)	3434.1

Table 7: Fit Statistics for Exponential Model 1

Solution for Fixed Effects						
Effect	G	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		2.7046	0.4255	88	6.36	<.0001
Days		0.01366	0.01194	626	1.14	0.2530
G	0	-0.2793	0.5545	88	-0.50	0.6158
G	1	0	.	.	.	.
Days*G	0	-0.01038	0.01556	626	-0.67	0.5052
Days*G	1	0	.	.	.	.
Days3		-0.01343	0.01211	626	-1.11	0.2676
Days3*G	0	0.009728	0.01578	626	0.62	0.5377
Days3*G	1	0	.	.	.	.

Table 8: Fixed Effects for Exponential Model 1

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Days	1	626	1.19	1.19	0.2761	0.2765
G	1	88	0.25	0.25	0.6145	0.6158
Days*G	1	626	0.44	0.44	0.5050	0.5052
Days3	1	626	1.18	1.18	0.2773	0.2777
Days3*G	1	626	0.38	0.38	0.5375	0.5377

Table 9: Type 3 Fixed Effects for Exponential Model 1

None of the variables were significant from the exponential model. Since the exponential structure is nested within the unstructured model. We used the log likelihood to test which structure is better. Null hypothesis was that the reduced model (exponential) was adequate. Alternative hypothesis was that unstructed covariance was required. (Table 10)

	-2 (REML)log-likelihood	AIC	df	log-likelihood diff	df diff	P
Unstructured	3159	3231	36			
Exponential	3425.1	3429.1	2	266.1	34	8.62748E-38

Table 10: Log Likelihood Ratio Test

The p-value is 8.6 * 10^(-38), less than 0.05. Therefore we rejected the H0, and concluded that the full model with unstructured covariance was preferred than the exponential structure. P-value was obtained from Microsoft Excel instead of SAS as the p-value was much too small.

Missing Value

The data has 937 observations initially. After deleted all the missing observations, the data only has 90 complete-case observation left. Deleting the observation might cause misleading assumption and skewness of the distribution. Figure2 shows the histogram of the data after deleting records that had missing value.

The histogram (Figure 2) indicated skewness of the distribution. In the following analysis, we filled in the missing observations by multiple imputation and fit the model we mentioned in previous section again.

To fill the missing observations, we first looked closely at the data again to decide the missing values' tupes. Several typical missingness of this dataset was shown in Table 11. Some of the missingness was due to dropout, while some were missing completely at random or missing at random. We therefore decided to use PROC MI statement in SAS to complete the observations as it is a general way of handling all above-mentioned types of missingness. (Maurice, 2012) The histogram of the data after filling the missing values are shown in Figure 3.

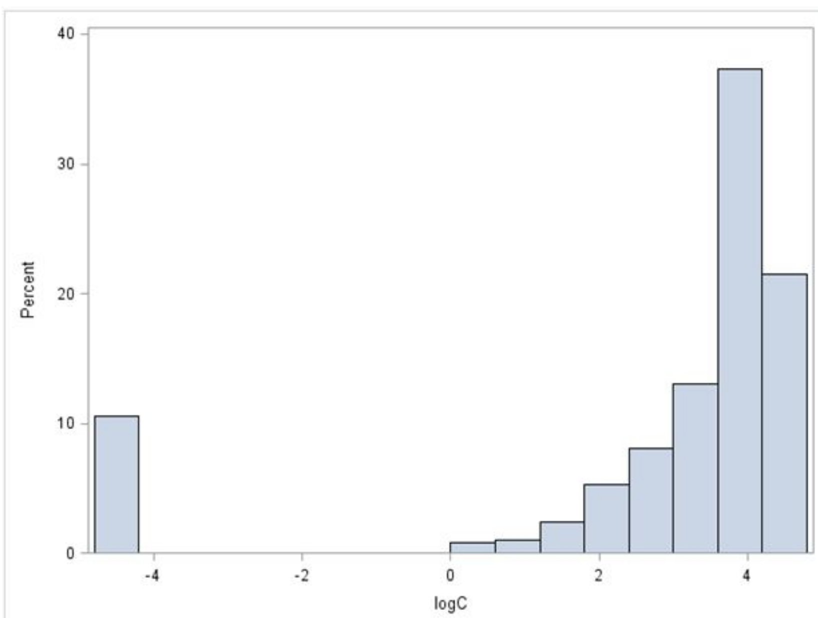

Figure 2: Histogram After Delete Missing Value

ID	CONFIDENCE1	CONFIDENCE2	CONFIDENCE3	CONFIDENCE4	CONFIDENCE5	CONFIDENCE6	CONFIDENCE7	CONFIDENCE8	GENDER
16	0	0				0			0
36	0	20	30	20	30	60	50	30	1
47	0	30	50	25	40	58	76	40	1
112	0	0	20	40	50				0
163	0	25	50	20					0
175	0	60	10	40	35				0
251	0	15	25	10	25	10	50	10	0
403	0		50						1
423	0		0	30				25	0
465	0	0	40	0	0	50	0	0	0
471	0	0	50	0	0	0	10	50	0
478	0					0	0		0
488	0	0	0	0					0
509	0	0	30	0	20	20	30	0	0
511	0	0	20	5		2	60		0
533	0	10	5	30		80			0
636	0	0						0	0
637	0	0	0	0	5	0	10		0
653	0	0							0
660	0	10	1	25	25	50			0
719	0	0	0	0	0	0			0
739	0	30	20	50	40				0
755	0		0	0	0		0	1	0

Table 11: Observations with Missing Value

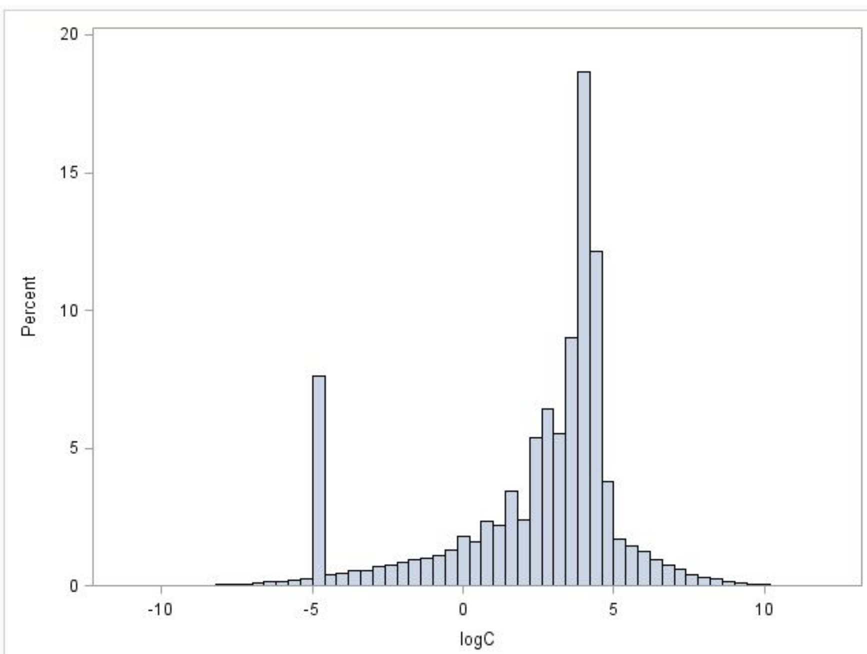

Figure 3: Histogram After Adding Missing Value

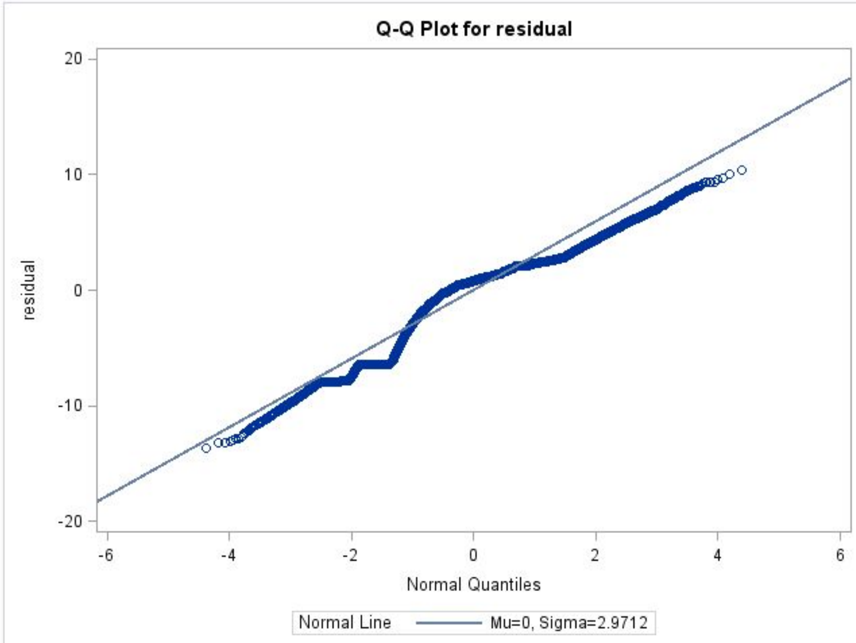

Figure 4: QQ-plot of the Full Observations

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.165544	Pr > D	<0.0100
Cramer-von Mises	W-Sq	875.967	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4947.817	Pr > A-Sq	<0.0050

Table 12: Test for Normality

After multiple imputation, the histogram (Figure 3) was more normal than the "complete case" scenario showed earlier. Based on the results of the normality test, as the QQ-plot and normality test showed above, we concluded that even though the data did not pass the normality test, it had still become more normal and less biased than before. Note that normality assumption was not a requirement for PROC MI.

We then examined the mean plot of confidence versus days, and the variance-covariance matrices again to spot any change. Figure 5, Table 13 and 14 showed the data with the missing value replaced almost followed the same trend as the data without the missing value replaced. Therefore, it was reasonable for us to conduct the same tests again.

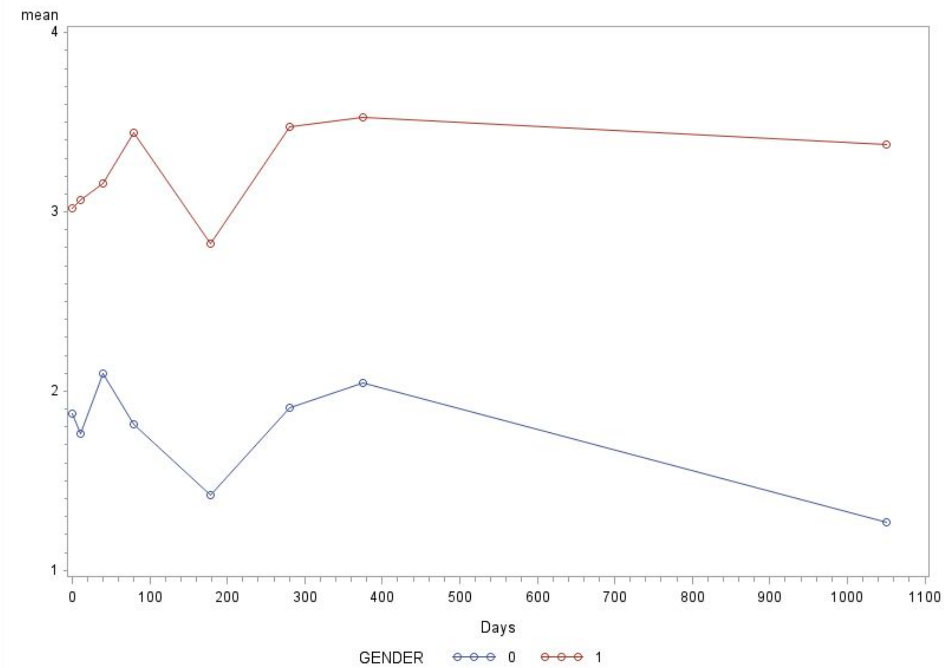

Figure 5: Plot Means Confidence Versus Days(2)

Estimated R Matrix for ID 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	9.6978	5.2218	5.1841	4.7096	5.4440	4.7791	5.8166	3.8842
2	5.2218	7.0405	4.5609	4.1346	4.1998	4.8761	4.4222	3.2327
3	5.1841	4.5609	8.1144	4.8616	4.7130	4.8217	5.7778	4.5144
4	4.7096	4.1346	4.8616	11.2546	4.7582	5.2988	4.0305	3.7964
5	5.4440	4.1998	4.7130	4.7582	7.7540	4.4389	5.4953	4.7537
6	4.7791	4.8761	4.8217	5.2988	4.4389	8.3746	5.0271	4.5517
7	5.8166	4.4222	5.7778	4.0305	5.4953	5.0271	9.1877	5.5683
8	3.8842	3.2327	4.5144	3.7964	4.7537	4.5517	5.5683	9.1073

Table 13: Covariance Matrix

Estimated R Correlation Matrix for ID 1								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.6320	0.5844	0.4508	0.6278	0.5303	0.6162	0.4133
2	0.6320	1.0000	0.6034	0.4645	0.5684	0.6350	0.5498	0.4037
3	0.5844	0.6034	1.0000	0.5087	0.5942	0.5849	0.6692	0.5251
4	0.4508	0.4645	0.5087	1.0000	0.5093	0.5458	0.3964	0.3750
5	0.6278	0.5684	0.5942	0.5093	1.0000	0.5508	0.6511	0.5657
6	0.5303	0.6350	0.5849	0.5458	0.5508	1.0000	0.5731	0.5212
7	0.6162	0.5498	0.6692	0.3964	0.6511	0.5731	1.0000	0.6087
8	0.4133	0.4037	0.5251	0.3750	0.5657	0.5212	0.6087	1.0000

Table 14: Correlation Matrix

Unstructured Model

We fit the unstructured model again with the missing value replaced again. All independent variables were significant.

Fit Statistics	
-2 Res Log Likelihood	495400.6
AIC (Smaller is Better)	495472.6
AICC (Smaller is Better)	495472.6
BIC (Smaller is Better)	495744.0

Table 14: Fit Statistics for Unstructured Model 2

Solution for Fixed Effects						
Effect	GENDER	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		2.9692	0.03770	14E3	78.75	<.0001
Days		0.01026	0.000730	14E3	14.06	<.0001
GENDER	0	-1.1331	0.04800	14E3	-23.60	<.0001
GENDER	1	0	.	.	.	.
Days*GENDER	0	-0.00525	0.000929	14E3	-5.65	<.0001
Days*GENDER	1	0	.	.	.	.
Days3		-0.01018	0.000738	14E3	-13.80	<.0001
Days3*GENDER	0	0.004557	0.000939	14E3	4.85	<.0001
Days3*GENDER	1	0	.	.	.	.

Table 15: Fixed Effects Result for Unstructured Model 2

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Days	1	14E3	270.10	270.10	<.0001	<.0001
GENDER	1	14E3	557.16	557.16	<.0001	<.0001
Days*GENDER	1	14E3	31.91	31.91	<.0001	<.0001
Days3	1	14E3	283.06	283.06	<.0001	<.0001
Days3*GENDER	1	14E3	23.54	23.54	<.0001	<.0001

Table 16: Type 3 Fixed Effects Result for Unstructured Model 2

Exponential Structure

All main effects were significant in the model with exponential covariance structure, while all interaction terms were not.

Fit Statistics	
-2 Res Log Likelihood	540076.7
AIC (Smaller is Better)	540080.7
AICC (Smaller is Better)	540080.7
BIC (Smaller is Better)	540095.8

Table 17: Fit Statistics for Exponential Model 2

Solution for Fixed Effects						
Effect	GENDER	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		3.0543	0.03906	14E3	78.19	<.0001
Days		0.004592	0.001021	97E3	4.50	<.0001
GENDER	0	-1.2684	0.04974	14E3	-25.50	<.0001
GENDER	1	0	.	.	.	.
Days*GENDER	0	-0.00082	0.001299	97E3	-0.63	0.5302
Days*GENDER	1	0	.	.	.	.
Days3		-0.00439	0.001035	97E3	-4.24	<.0001
Days3*GENDER	0	0.000016	0.001318	97E3	0.01	0.9904
Days3*GENDER	1	0	.	.	.	.

Table 18: Fixed Effects for Exponential Model 2

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Days	1	97E3	41.48	41.48	<.0001	<.0001
GENDER	1	14E3	650.41	650.41	<.0001	<.0001
Days*GENDER	1	97E3	0.39	0.39	0.5302	0.5302
Days3	1	97E3	44.16	44.16	<.0001	<.0001
Days3*GENDER	1	97E3	0.00	0.00	0.9904	0.9904

Table 19: Type 3 Fixed Effects for Exponential Model 2

	-2 (REML)log-likelihood	AIC	df	log-likelihood diff	df diff	P
Unstructured	495400.6	495472.6	36			
Exponential	540076.7	540080.7	2	44676.1	34	00E+00

Table 20: Log Likelihood Ratio Test

Again, we compared two model by looking at their log-likelihood ratio test. This time, the p-value was even smaller, suggesting the unstructured model was still better than the exponential one.

Result

Final Model

The procedure described above illustrates (a) our initial search for an adequate model of the relationship between between gender, days into the recession, and confidence in the economy, then (b) the resulting coefficients produced by that model. We performed each of these tasks twice: once using complete case analysis and once using multiple imputation to adjust for data that may not have been missing completely at random (MCAR). Regarding model selection, results show that the relationship between confidence ratings at different points in time is best accounted for by an unstructured covariance model. This is true both in the complete cases model and in the model augmented by multiple imputation - both times, the unstructured model demonstrated significantly better fit to the data than the exponential covariance model. Given that multiple imputation involves more plausible assumptions about the pattern of missingness observed in our data, the remainder of this section only utilizes parameter estimates from the multiply imputed model using the unstructured covariance matrix. The final model is depicted below:

$$Log((Confidence +1) / (100-Confidence+1)) =$$
$$\beta_0 + \beta_1 * G_i + \beta_2 * Days_{ij} + \beta_3 * Days_{3ij+} + \beta_4 * (G_i*Days_{ij}) + \beta_5 * (G_i*Days_{3ij+}) + e_{ij}$$

In other words,

$$Confidence =$$
$$102 / (e^{-(\beta 0 + \beta 1 * Gi + \beta 2 * Daysij + \beta 3 * Days3ij+ + \beta 4 * (Gi*Daysij) + \beta 5 * (Gi*Days3ij+) + eij)} + 1) - 1$$

Model interpretation

Results of the spline regression demonstrated several significant effects. First, results showed a significant effect of gender, this suggests that men and women were reporting different levels of confidence in the economy at baseline, with women reporting significantly lower confidence than men. Second, results indicated a significant interaction between gender and days from baseline, as well as the interaction between gender and days since the third measurement occasion (i.e., the “knot” in the regression). This suggests the rate of change from measurement occasion to measurement occasion was different for men and women. Based on the way gender was coded, the significant effect of days and days3 suggest that men’s scores initially increased over time prior to the third measurement occasion then slowed after that point. In fact, given the estimates for days and days3 are almost perfectly counterbalanced, our results suggest men’s mean confidence in the economy remained roughly unchanged after time-point 3. Similarly, women’s confidence grew over the course of the initial phase of the recession, but grew more slowly than men's, based on the coefficient for the gender*days interaction term. Again, after the knot at time-point 3, the growth in women's confidence slowed like men's.

Detailed interpretation by different scenarios are listed as follows in Table 21.

Gender	Days	Transformed Response Variable =	Interpretation
G = 0, Female		$\beta_0 + \beta_2 * Days_{ij} + \beta_3 * Days_{3ij+} + e_{ij}$	
	1	$\beta_0 + \beta_2 + e_{ij}$	β_2: rate of change in transformed response between measurement time points before days3 among females
	2	$\beta_0 + 2*\beta_2 + e_{ij}$	
	3	$\beta_0 + 3*\beta_2 + e_{ij}$	
	4	$\beta_0 + 4*\beta_2 + 2*\beta_3 + e_{ij}$	*$\beta_2 + \beta_3$: rate of change in transformed response between measurement time points after days3 among females*
	5	$\beta_0 + 5*\beta_2 + 3*\beta_3 + e_{ij}$	
	6	$\beta_0 + 6*\beta_2 + 4*\beta_3 + e_{ij}$	
	7	$\beta_0 + 7*\beta_2 + 5*\beta_3 + e_{ij}$	
	8	$\beta_0 + 8*\beta_2 + 6*\beta_3 + e_{ij}$	
G = 1, Male			
	1	$\beta_0 + \beta_1 + \beta_2 + \beta_4 + e_{ij}$	*$\beta_2 + \beta_4$: rate of change in transformed response between measurement time points before days3 among males*
	2	$\beta_0 + \beta_1 + 2*\beta_2 + 2*\beta_4 + e_{ij}$	
	3	$\beta_0 + \beta_1 + 3*\beta_2 + 3*\beta_4 + e_{ij}$	
	4	$\beta_0 + \beta_1 + 4*\beta_2 + \beta_3 + 4*\beta_4 + \beta_5 + e_{ij}$	$\beta_2 + \beta_4 + \beta_5$: rate of change in transformed response between measurement time points after days3 among males
	5	$\beta_0 + \beta_1 + 5*\beta_2 + 2*\beta_3 + 5*\beta_4 + 2*\beta_5 + e_{ij}$	
	6	$\beta_0 + \beta_1 + 6*\beta_2 + 3*\beta_3 + 6*\beta_4 + 3*\beta_5 + e_{ij}$	
	7	$\beta_0 + \beta_1 + 7*\beta_2 + 4*\beta_3 + 7*\beta_4 + 4*\beta_5 + e_{ij}$	
	8	$\beta_0 + \beta_1 + 8*\beta_2 + 5*\beta_3 + 8*\beta_4 + 5*\beta_5 + e_{ij}$	

Table 21: Model Interpretation by Scenarios

Conclusion

This investigation was focused on the potential for different rates of change in economic confidence across men and women in the United States during the most recent recession (i.e., ‘The Great Recession’). Results show that the confidence men and women reported in the economy at 8 measurement occasions across the recession improved significantly over the first 3 occasions, but then slowed to a nearly flat line for every measurement occasion after that. The rate of improvement was slower for women than for men, women also had lower confidence at baseline than men. Taken together, these results suggest that people’s confidence in the economy improves over time, but only early on in a recession. As a recession takes longer to remit, confidence will not likely decrease much from previous levels, but it will also nearly cease to increase as people become adjusted to what might be called a “new normal.”

Reference

Burns, William. "Financial Crisis: A Longitudinal Study of Public Response (ICPSR 36341)."*Financial Crisis: A Longitudinal Study of Public Response*. ICPSR, 25 Jan. 2016. Web. 25 Apr. 2017.

Fitzmaurice, Garrett M., Nan M. Laird, and James H. Ware. *Applied Longitudinal Analysis*. Hoboken: Wiley, 2012. Print.

Appendix

SAS Code

```
/*missing MI*/
proc mi data=d2 seed=346895 nimpute=25 out=d3;
 var gender logC1 logC2 logC3 logC4 logC5 logC6 logC7 logC8;
 mcmc nbiter=5000 niter=500;
run;

proc export data=d3
dbms=csv outfile="I:\project\7670\missed.csv" replace;
run;

proc import out=d1
 datafile='I:\project\7670\data.csv'
 dbms=csv replace;
run;

/*creat Tcat*/
data d2;
 set d1;
if T=1 then Days=0;
if T=2 then Days=11;
if T=3 then Days=39;
if T=4 then Days=80;
if T=5 then Days=178;
if T=6 then Days=281;
if T=7 then Days=375;
if T=8 then Days=1050;
Tcat=T;
t2=t*t;
t3=t*t*t;
t4=t*t*t*t;
time3=max(t-3,0);
Days3=max(Days-39,0);
t3_2=time3*time3;
proc sort data=d2;
 by T;
run;
/*get the mean of the group*/
```

```
proc sort data=d2;
 by g descending Days;
run;

proc means noprint data=d2 n mean std stderr;
 var logc;
 by g descending Days;
 output out=meand1 mean=mean std=sd;
run;

/*plot*/
ods listing gpath = 'I:\project\7670';
symbol1 value = circle interpol = join;
symbol2 value = star interpol = join;
proc gplot data = meand1;
   plot mean*Days=g;
run; quit;


/*random linear regression*/
proc reg data=d2;
 model logc=t;
 by g;
run; quit;


proc mixed data=d2;
 class id Tcat g;
 model logc= Days g g*Days  /s chisq;
 repeated Tcat/type=un subject=id r rcorr;
run; quit;

/*profile (time conti)*/
proc mixed data=d2;
 class id Tcat g;
 model logc= Days g g*Days Days3 g*Days3 /s chisq;
 repeated Tcat/type=un subject=id r rcorr;
run; quit;

ods trace on;
ods trace off;
/*Hetero top*/
proc mixed data=d2;
```

```
 class id Tcat g;
 model logc= Days g g*Days Days3 g*Days3 /s chisq;
 repeated Tcat/type=toeph subject=id r rcorr;
 ods select SolutionF Tests3 FitStatistics;
run; quit;

/*Hetero Ex*/
proc mixed data=d2;
 class id Tcat g;
 model logc= Days g g*Days Days3 g*Days3 /s chisq;
 repeated Tcat/type=sp(exp)(Days) subject=id r rcorr;
 ods select SolutionF Tests3 FitStatistics;
run; quit;

/*hetero AR(1)*/
proc mixed data=d2;
 class id Tcat g;
 model logc= Days g g*Days Days3 g*Days3 /s chisq;
 repeated Tcat/type=ARH(1) subject=id r rcorr;
 ods select SolutionF Tests3 FitStatistics;
run; quit;


/*p-value*/
data P;
 p_t=1-probchi(71.1,21);
 p_e=1-probchi(174.5,27);
 p_a=1-probchi(266.1,34);
 run;
proc print data=p;
run;

/*random int&t&time3*/
proc mixed data=d2;
 class id Tcat g;
 model logc= days g g*days days3 g*days3 /s chisq;
 random intercept days days3/type=un subject=id g gcorr v vcorr;
run; quit;
```