MIS Quarterly

RESEARCH ARTICLE

WHEN DO IT SECURITY INVESTMENTS MATTER? ACCOUNTING FOR THE INFLUENCE OF INSTITUTIONAL FACTORS IN THE CONTEXT OF HEALTHCARE DATA BREACHES

Corey M. Angst
IT, Analytics, and Operations Department, University of Notre Dame, 348 Mendoza College of Business,
Notre Dame, IN 46556 U.S.A. {cangst@nd.edu}

Emily S. Block
Department of Strategic Management and Organization, University of Alberta, 4-21 F Alberta School of Business,
Edmonton, AB T6G 2R6 CANADA {eblock@ualberta.ca}

John D'Arcy
Department of Accounting and MIS, University of Delaware, 356 Purnell Hall,
Newark, DE 19716 U.S.A. {jdarcy@udel.edu}

Ken Kelley
IT, Analytics, and Operations Department, University of Notre Dame, 363 Mendoza College of Business,
Notre Dame, IN 46556 U.S.A. {kkelley@nd.edu}

Appendix A

IT Security Included in this Study

Technology	Description
Biometric systems	Authentication mechanisms that determine whether a user is authorized to access a particular IT system based on his/her physical characteristics.
ID management	Used to electronically identify users and control their access to IT resources based on certain access privileges.
Intrusion detection	Monitoring systems designed to detect an attack on a network or computer system.
Anti-virus software	Software programs used to detect and remove computer viruses.
Single sign-on technology	Software authentication that enables a user to authenticate once and gain access to the resources of multiple systems, reducing the need to track and manage multiple passwords.
Non-biometric user authentication systems	Used to verify the identity of a user through non-physical means (e.g., user ID and password, electronic tokens or smart cards, responses to short questions, or some combination).
Data encryption	Technologies that encode electronic data in such a way that non-authorized users cannot read it but authorized parties can.
Internet firewalls	Hardware and/or software technologies that control incoming and outgoing network traffic by analyzing data packets.
Spyware filters	Software programs used to detect and deter unwanted spyware programs that monitor internal systems.

Appendix B

Correlation Table

Variable	1	2	3	4	5	6	7	8	9
(1) SystemSize	1.00								
(2) HospitalSize	-0.02	1.00							
(3) Age	**-0.34**	0.01	1.00						
(4) BusinessModel	**0.62**	**-0.11**	**-0.29**	1.00					
(5) Teaching	**-0.13**	**0.51**	**0.04**	**-0.12**	1.00				
(6) Mission	**0.03**	**0.17**	**-0.16**	**-0.18**	**-0.04**	1.00			
(7) EntrepMindset	**0.03**	**0.39**	**-0.07**	**-0.27**	**0.23**	**0.15**	1.00		
(8) ITSec	**-0.04**	**0.19**	**0.01**	**-0.10**	**0.07**	**0.07**	**0.46**	1.00	
(9) Breach	**0.21**	**0.28**	**-0.07**	**0.10**	**0.22**	0.02	**0.25**	**0.15**	1.00

Note: Bold represents statistically significant coefficients at $p < 0.05$.

Appendix C

Statistical Specification of Our GMM Model

In the path diagram below (Figure C1), squares are measured variables and circles are latent variables. Arrows represent a presumed causal relationship. The triangle on the left of the model represents the intercept and the one on the right, the fixed effects of the predictors. The 1's along the paths for the intercept are the constant effect the intercept has on each time point, and the numbers along the paths for the slope denote the particular value of time. Zero is used for the first time point so that the intercept can be more easily interpreted as the "baseline," in which the intercept term represents the logit (or probability if rescaled) of breach in 2005.

In terms of mapping this diagram to our conceptual model (Figure 1 in the paper), the squares on the left side represent the firm-specific institutional factors (covariates) that predict latent class (specifically the symbolic latent class, as per H1a through H1g). Moving to the right, the arrows from IT Security Investment (*ITSec*) to Intercept of Breach and Slope of Breach represent the influence of *ITSec* on these growth factors (which are derived from the repeated measures of *Breach* from 2005–2013). Note that *ITSec* is held constant for testing H2 (as described in footnote 19) to assess its influence on the combined classes. The arrows from Latent Class to the Intercept of Breach and Slope of Breach indicate that the influence of *ITSec* on these growth factors varies by Latent Class, as tested in H3a and H3b (also described in footnote 19). This corresponds to the regressions of the Intercept of Breach and Slope of Breach on a dummy variable representing the latent class categories (symbolic and substantive adoption, in our case).

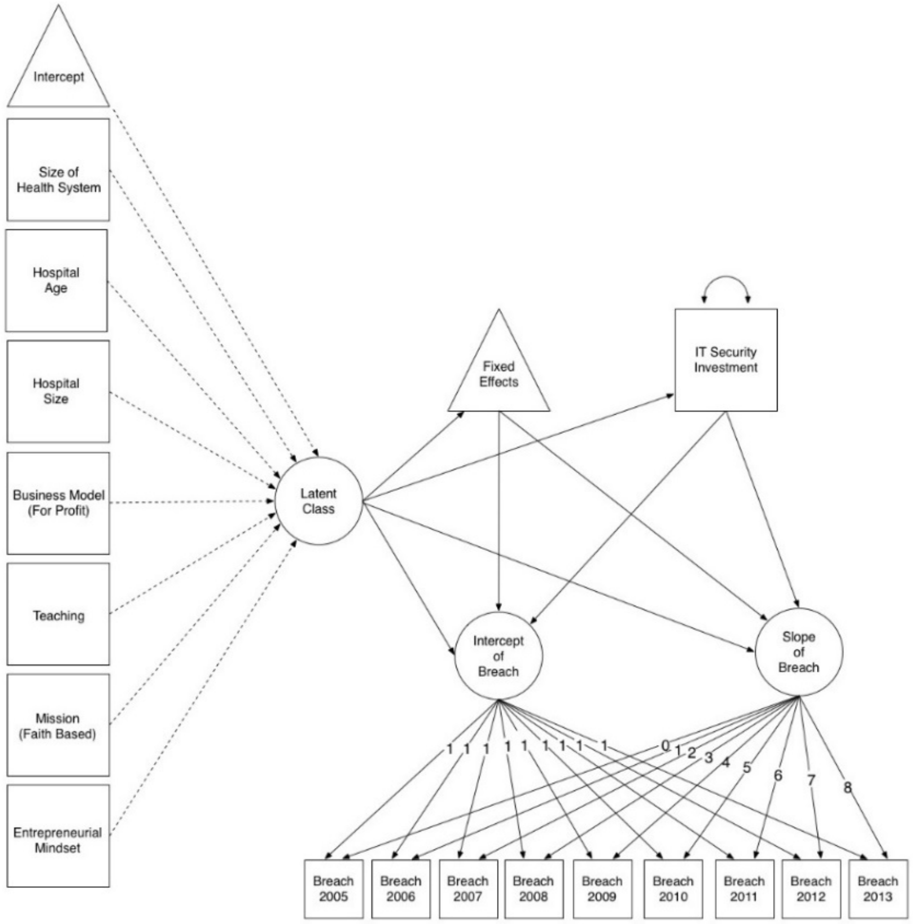

Figure C1. Path Model Showing Statistical Specifications of Our GMM Model

Appendix D

Detailed Description of Analysis Including Mplus Syntax

The Mplus syntax below fits the measurement (intercept and slope) and structural model (effect of *ITSec*). Note that an exclamation point denotes a commented line. As noted earlier, we scaled time such that 0 represents 2005. The value of time can be seen in the Mplus model statement below (Part 2; note that this model statement is for the model in which *ITSec* is held constant across classes) where, for example, *Breach* at 2005 is denoted as BrchY_05@0, *Breach* at 2006 is denoted as BrchY_06@1, *Breach* at 2007 is denoted as BrchY_07@2 and so forth up until 2013. Also, M_ITS is the mean of our *ITSec* variable; other variable names are self-explanatory. The R3STEP labeling on the variables listed under the AUXILIARY heading (Part 1) indicates that these variables will be treated as latent class predictors per the three-step method described earlier (in the Mplus program, AUXILIARY is an option of the VARIABLE command and for the three-step method, a variable is specified as R3STEP if it is to be included in this procedure).

As recommended (e.g., Jung and Wickrama 2008), we use multiple random starts (1,000) and multiple optimization attempts (250). This can be seen in the ANALYSIS section of the Mplus syntax below (Part 3). We took these steps to help ensure that our solutions yield the global minimum log likelihood discrepancy function rather than (as can happen in nontrivial models) the optimization procedure converging at a local minimum log likelihood discrepancy function. The global log likelihood minimum discrepancy function is what produces the estimates that maximize the likelihood (i.e., that yield the maximum likelihood solution).

Part 1: Specifying the Variables

```
VARIABLE:  NAMES ARE
   SeqID
   ln_HosAg
   FaithBsd
   P_Acad
   ln_StBed
   P_Prof
   M_SysSz
   M_SaidT
   M_ITS
   ;
   AUXILIARY ARE
   M_SaidT(R3STEP)
   M_SysSz(R3STEP)
   ln_StBed(R3STEP)
   ln_HosAg(R3STEP)
   FaithBsd(R3STEP)
   P_Prof(R3STEP)
   P_Acad(R3STEP)
   ;
   USEVARIABLES ARE
   BrchY_05 BrchY_06 BrchY_07 BrchY_08 BrchY_09 BrchY_10 BrchY_11 BrchY_12  BrchY_13
   M_ITS
   ;
   CATEGORICAL ARE
   BrchY_05 BrchY_06 BrchY_07 BrchY_08 BrchY_09 BrchY_10 BrchY_11 BrchY_12 BrchY_13;

   IDVARIABLE is SeqID;
   MISSING ARE all (9999);
   CLASSES = C(2);
```

Part 2: Mplus Model Statement

```
Model:   %OVERALL%
   i s | BrchY_05@0 BrchY_06@1 BrchY_07@2
      BrchY_08@3 BrchY_09@4 BrchY_10@5
      BrchY_11@6 BrchY_12@7 BrchY_13@8;

   i on M_ITS(I_MITS);
   s on M_ITS(S_MITS);

   [
   i@0
   s(Slope1)
   BrchY_05$1(Thres)
   BrchY_06$1(Thres)
   BrchY_07$1(Thres)
   BrchY_08$1(Thres)
   BrchY_09$1(Thres)
   BrchY_10$1(Thres)
   BrchY_11$1(Thres)
   BrchY_12$1(Thres)
   BrchY_13$1(Thres)
   ];
```

```
! Model differences for Class 2, in which differences are between the overall model (here Class 1 because there are only two classes),

%C#2%

   [
   i(int2)
   s(Slope2)
   BrchY_05$1(Thres)
   BrchY_06$1(Thres)
   BrchY_07$1(Thres)
   BrchY_08$1(Thres)
   BrchY_09$1(Thres)
   BrchY_10$1(Thres)
   BrchY_11$1(Thres)
   BrchY_12$1(Thres)
   BrchY_13$1(Thres)
   ];
```

Part 3: Model Options

```
ANALYSIS:  TYPE=MIXTURE;
STARTS =1000 250; ! = (#Random Starts; #Optimizations)
Estimator=ML;
```

Part 4: Abbreviated Model Output

Tests of Categorical Latent Variable Multinomial Logistic Regressions Using the Three-Step Procedure

```
                                         Two-Tailed
                      Estimate        S.E.       Est./S.E.      P-Value
  C#1     ON
     M_SAIDT             0.068        0.030          2.228        0.026
     M-SYSSZ             0.017        0.008          2.207        0.027
     LN_STBED            0.475        0.167          2.835        0.005
     LN_HOSAG           -0.535        0.13          -3.840        0.000
     FAITHBSD           -0.677        0.353         -1.920        0.055
     P_PROF             -4.415        1.309         -3.374        0.001
     P_ACAD              1.878        0.382          4.921        0.000
  Intercepts
     C#2                 5.035        0.933         -5.399        0.000
```

Parameterization Using Reference Class 1 *(the results under this heading correspond with Table 4 in the paper)*

```
  C#1     ON
     M_SAIDT            -0.068        0.030         -2.228        0.026
     M-SYSSZ            -0.017        0.008         -2.207        0.027
     LN_STBED           -0.047        0.167         -2.835        0.005
     LN_HOSAG            0.535        0.139          3.840        0.000
     FAITHBSD            0.677        0.353          1.920        0.055
     P_PROF              4.415        1.309          3.374        0.001
     P_ACAD             -1.878        0.382         -4.921        0.000
  Intercepts
     C#2                 5.035        0.933         -5.399        0.000
```

Model Results *(these are for the model in which ITSec is held constant across classes; results displayed in Panel 1 of Table 5 in the paper)*

				Two-Tailed
	Estimate	S.E.	Est./S.E.	P-Value
Latent Class 1				
1 ON				
M_ITS	**0.344**	**0.047**	**7.351**	**0.000**
S ON				
M_ITS	**0.017**	**0.010**	**1.633**	**0.102**
Intercepts				
I	0.000	0.000	999.000	999.000
S	0.371	0.061	6.099	0.000
Thresholds				
BRCHY_05$1	4.914	0.318	15.444	0.000
BRCHY_06$1	4.914	0.318	15.444	0.000
BRCHY_07$1	4.914	0.318	15.444	0.000
BRCHY_08$1	4.914	0.318	15.444	0.000
BRCHY_09$1	4.914	0.318	15.444	0.000
BRCHY_10$1	4.914	0.318	15.444	0.000
BRCHY_11$1	4.914	0.318	15.444	0.000
BRCHY_12$1	4.914	0.318	15.444	0.000
BRCHY_13$1	4.914	0.318	15.444	0.000
Latent Class 2				
1 ON				
M_ITS	**0.344**	**0.047**	**7.351**	**0.000**
S ON				
M_ITS	**0.017**	**0.010**	**1.633**	**0.102**
Intercepts				
I	-0.173	0.310	-0.558	0.577
S	-0.145	0.049	-2.936	0.003
Thresholds				
BRCHY_05$1	4.914	0.318	15.444	0.000
BRCHY_06$1	4.914	0.318	15.444	0.000
BRCHY_07$1	4.914	0.318	15.444	0.000
BRCHY_08$1	4.914	0.318	15.444	0.000
BRCHY_09$1	4.914	0.318	15.444	0.000
BRCHY_10$1	4.914	0.318	15.444	0.000
BRCHY_11$1	4.914	0.318	15.444	0.000
BRCHY_12$1	4.914	0.318	15.444	0.000
BRCHY_13$1	4.914	0.318	15.444	0.000

Model Results *(these are for the model in which ITSec is allowed to vary across classes; results displayed in Panels 2 and 3 of Table 5 in the paper)*

			Two-Tailed	
	Estimate	S.E.	Est./S.E.	P-Value
Latent Class 1				
1 ON				
M_ITS	**-0.118**	**0.185**	**-0.637**	**0.524**
S ON				
M_ITS	**0.061**	**0.031**	**1.944**	**0.052**
Intercepts				
I	0.000	0.000	999.000	999.000
S	0.195	0.102	1.915	0.056
Thresholds				
BRCHY_05$1	3.288	0.641	5.129	0.000
BRCHY_06$1	3.288	0.641	5.129	0.000
BRCHY_07$1	3.288	0.641	5.129	0.000
BRCHY_08$1	3.288	0.641	5.129	0.000
BRCHY_09$1	3.288	0.641	5.129	0.000
BRCHY_10$1	3.288	0.641	5.129	0.000
BRCHY_11$1	3.288	0.641	5.129	0.000
BRCHY_12$1	3.288	0.641	5.129	0.000
BRCHY_13$1	3.288	0.641	5.129	0.000
Latent Class 2				
1 ON				
M_ITS	**0.379**	**0.052**	**7.244**	**0.000**
S ON				
M_ITS	**0.020**	**0.013**	**1.453**	**0.146**
Intercepts				
I	-1.923	0.731	-2.632	0.008
S	-0.173	0.071	-2.436	0.015
Thresholds				
BRCHY_05$1	3.288	0.641	5.129	0.000
BRCHY_06$1	3.288	0.641	5.129	0.000
BRCHY_07$1	3.288	0.641	5.129	0.000
BRCHY_08$1	3.288	0.641	5.129	0.000
BRCHY_09$1	3.288	0.641	5.129	0.000
BRCHY_10$1	3.288	0.641	5.129	0.000
BRCHY_11$1	3.288	0.641	5.129	0.000
BRCHY_12$1	3.288	0.641	5.129	0.000
BRCHY_13$1	3.288	0.641	5.129	0.000

Appendix E

Model Comparisons, Two-Factor Versus One- and Three-Factor Solutions

Recent research has offered insight into the determination of the number of classes to use in a latent class analysis (Diallo et al. 2017; Nylund et al. 2007). While the central argument still holds that theory should guide the choice (Diallo et al. 2017; Tofighi and Enders 2008), these new approaches offer an empirical test and guidance for comparing models with different numbers of classes. Importantly, Diallo et al. (2017) evaluate the effect of including covariates in this type of model comparison and find that models should be compared in the absence of covariates. Following their guidance, we use the GMM for binary outcomes in the absence of covariates and compare one- and three-class models to our baseline two-class model. We find that our theoretically derived two-class solution performs better than the other two models. Specifically, the three-class model does not converge due to singularity (i.e., the matrix that is being optimized has a determinant of zero), meaning that it is ill-conditioned, likely as a result of being over-fitted. Thus, we cannot compare it to the two-class model. The fit criteria for the other classes are shown below.

Fit Criteria[1]	**One-Class Model**	**Two-Class Model (i.e., baseline model)**
Akaike (AIC)	8867.8	8857.7
Bayesian (BIC)	8901.2	8891.1
Sample-size adjusted BIC	8885.3	8875.2

[1] Lower numbers represent a better fitting model

References

Diallo, T. M. O., Morin, A. J. S., and Lu, H. 2017. "The Impact of Total and Partial Inclusion or Exclusion of Active and Inactive Time Covariants on the Class Enumeration Process of Growth Mixture Models," *Psychological Methods* (22), pp. 166-190.

Jung, T., and Wickrama, K. A. S. 2008. "An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling," *Social and Personality Psychology Compass* (2:1), pp. 302-317.

Nylund, K. L., Asparouhov, T., and Muthén, B. 2007. "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Analysis," *Structural Equation Modeling* (14:4), pp. 535-569.

Tofighi, D., and Enders, C. K. 2008. "Identifying the Correct Number of Classes in Growth Mixture Models," in *Advances in Latent Variable Mixture Models,* G. R. Hancock and K. M. Samuelson (eds.), Charlotte, NC: Information Age Publishing, pp. 317-341.