MIS Quarterly

RESEARCH ARTICLE

WHEN DOES REPOSITORY KMS USE LIFT PERFORMANCE? THE ROLE OF ALTERNATIVE KNOWLEDGE SOURCES AND TASK ENVIRONMENTS

Seung Hyun Kim
School of Business, Yonsei University, Seoul 120-749 KOREA {seungkim@yonseiac.kr}

Tridas Mukhopadhyay
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213 U.S.A., and
Carnegie Mellon University in Qatar, Education City, Doha QATAR {tridas@cmu.edu}

Robert E. Kraut
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213 U.S.A. {robert.kraut@cmu.edu}

Appendix

This appendix provides additional details of the robustness checks that are not fully explained in the paper due to limited space. The first section of this appendix discusses alternative estimations of the baseline model, Equation (1). The second section provides the details about the propensity score matching (PSM) Analysis. The third section presents more details about our two-stage least squares (2SLS) estimation results.

Alternate Specifications

We present the results of additional specifications of Equation (1) in Columns (4) through (6) of Table A1, and compare them with those presented in the main body of the paper, Column (1) through (3). Model (4) employs the FGLS estimator allowing cross-sectional correlation across departments and panel-specific AR(1). Although the coefficient size is smaller than that in Model (3), our selected model, the effect of the KMS usage is still positive and significant ($\beta_{KMSU} = 1.339$, $p < 0.001$). Model (5) is a dynamic panel model using the Arellano-Bond estimator by including the lagged dependent variable as an explanatory variable. The estimated coefficient of the cumulative KMS usage is slightly larger than that in two other fixed effects models ($\beta_{KMSU} = 4.528$, p < 0.001). Model (6) assumes that knowledge accumulated from the KMS usage depreciates over time at the rate of five percent per week.[1] The effect of repository KMS usage is still positive and significant ($\beta_{KMSU} = 2.499$, p < 0.01). Thus the effect of the KMS usage is not only positive and highly significant, but also robust to various alternative specifications.

[1]We have tested with alternative rates such as 1, 2, and 10 percent, and obtained qualitatively similar results.

Table A1. Estimation of the Effect of KMS Usage	(1) FE with Robust SE	(2) 2SLS	(3) FE with AR(1)	(4) FGLS	(5) Dynamic Panel	(6) FE with Depreciation
Intercept	35.534* (19.445)	24.949*** (8.108)	47.286*** (1.245)	-15.849*** (0.403)	-19.802*** (3.557)	40.115** (19.040)
Log of Department Employees	6.019 (7.796)	7.890*** (1.516)	3.095*** (1.083)	45.872*** (0.161)	6.156*** (0.954)	5.314 (7.702)
Log of Cumulative Repository Use	4.318*** (1.559)	10.886** (4.873)	2.877*** (0.456)	1.339*** (0.034)	4.528*** (0.274)	2.499*** (0.845)
Log of Data Warehouse Use	1.353 (2.704)	-0.541 (1.416)	1.728*** (0.425)	-2.685*** (0.041)	0.734** (0.305)	1.888 (2.551)
Department Manager Training	-2.261*** (0.533)	-2.237*** (0.166)	-1.792*** (0.382)	-2.870*** (0.032)	-3.482*** (0.332)	-2.285*** (0.540)
Department Employee Turnover Rate	0.067 (2.191)	-2.096 (2.920)	-0.429 (2.760)	-10.452*** (0.150)	1.117 (3.590)	0.816 (1.992)
Trade Area Income	0.260 (2.468)	-0.294 (0.678)	1.076 (1.173)	-7.708*** (0.087)	0.476 (0.785)	0.271 (2.471)
Trade Area Competition	2.925 (2.943)	2.573*** (0.650)	1.753 (1.277)	7.537*** (0.121)	5.418*** (0.875)	2.938 (2.942)
Lagged Sales					0.650*** (0.004)	
Number of Observations	39,858	39,858	39,585	39,858	39,312	39,858
Number of Department Managers	273	273	273	273	273	273
R-Squared (Within)	15.97%	N/A	23.20%	N/A	N/A	15.54%
R-Squared (Between)	13.22%	N/A	8.84%	N/A	N/A	17.54%
R-Squared (Overall)	10.38%	N/A	6.19%	N/A	N/A	11.40%

***Significant at 1%;** significant at 5 %; *significant at 10%. The numbers in parentheses are standard errors.

There are three things to note from the additional robustness checks. The coefficient of KMS usage in Model (3) using the first order autoregressive error terms is still the smallest among the estimated fixed effects models, indicating that the estimate in the selected model is conservative. Furthermore, Model (5) is not preferred since we believe that the lagged dependent variable should be included as an independent variable only when there is a strong theoretical justification. Despite the flexibility in the error structure, Model (4) using FGLS estimator with the random effects model is not selected due to the possible bias of the random effects model owing to correlations between department-specific effects and other covariates. Thus, we prefer Model (3) to other alternative specifications. Recall that we selected Model (3) since we found a strong evidence of the first order autoregressive error while Model (1) using robust standard errors clustered within each department manager is less suitable in our case. Moreover, our Hausman test confirmed that the estimates from 2SLS are not statistically different from the estimates from the fixed effects model.

Details about the Propensity Score Matching (PSM) Analysis

As described in the paper, our interest in the PSM analysis is to estimate the change in performance due to a change in KMS usage while addressing a selection bias. The outcome measure is the change in yearly sales in thousand dollars, $\Delta SALE_i = [SALE_{iY} - SALE_{iY-1}]$, where i indexes department manager ($i = 1, 2, \ldots, 273$), Y indexes year ($Y = 1$ or 2), and Δ denotes the difference between the two years. We redefine the treatment variable. *Treatment* = 1 if $[Log(KMSU_{iY}) - Log(KMSU_{iY-1})]$ is greater than the median of 273 managers. Otherwise, we code *Treatment* = 0.

We first estimate the treatment effect using the nearest neighbor (NN) matching method based on the differenced time-variant covariates used in Equation (1) as matching variables. That is, $\Delta Log(LEMP)_i$, $\Delta Log(DWHU)_i$, $\Delta TRNG_i$, $\Delta TINC_i$, $\Delta TCOM_i$, and $\Delta TURN_i$ used as matching variables. We add two more variables which may explain the treatment state. $SURV_i$ indicates whether a manager has participated in the KMS survey as we explained in the "Model Specification" section about our 2SLS estimation. $\Delta Log(OKMS)_i$ captures the possible influence by peers in adopting KMS and measures the difference in annual KMS usage by other department managers in the manager i's store. In the second model, we add sales in year 1 ($SALE_{i1}$). That is, it is possible that department managers may increase their KMS usage level based on their current sales. Finally, we estimate the treatment effect using Kernel matching method as an alternative matching algorithm. Table A2 summarizes the measurements of the outcome ($\Delta SALE_i$) and matching variables used in our PSM analysis.

Table A2. Description of Variables

Variable	Description
$\Delta SALE_i$	Difference in aggregate sales in manager *i*'s department between year Y and year Y-1 (in thousand dollars)
$Treatment_i$	1 if the difference in log of cumulative KMS usage for manager *i* is greater than that of the median of 273 managers, and 0, otherwise.
$\Delta Log(LEMP)_i$	Difference in log of the average number of employees in manager *i*'s department between year Y and year Y-1
$\Delta TRNG_i$	Difference in the cumulative number of days of computer-related training by department manager *i* between year Y and year Y-1
$\Delta TINC_i$	Difference in income level in department manager *i*'s trade area between year Y and year Y-1
$\Delta TCOM_i$	Difference in competition level in department manager *i*'s trade area between year Y and year Y-1
$\Delta Log(DWHU)_i$	Difference in log of the cumulative number of data warehouse reports viewed by department manager *i* between year Y and year Y-1
$\Delta TURN_i$	Difference in the average employee turnover rate in the manager *i*'s department between year Y and year Y-1
$SURV_i$	1 if manager *i* has participated in the internal user satisfaction survey, and 0, otherwise.
$\Delta LOG(OKMS)_i$	Difference in log of the cumulative KMS usage by department manager *i*'s peer managers in the same store between year Y and year Y-1
$SALE_{iY\text{-}1}$	Manager *i*'s departmental sales in year Y-1

Next, Tables A3 and A4 present the descriptive statistics and correlations of variables used in our PSM analysis.

Table A3. Descriptive Statistics

Variable	N	Mean	Std Dev
Δ(Yearly Department Sales)	273	355.62	1093.07
Treatment (Increase in KMS Usage)	273	0.50	0.50
Δ(Department Employees)	273	0.00	0.15
Δ(Log of Data Warehouse Use)	273	0.44	0.61
Δ(Trade Area Income)	273	-0.01	0.23
Δ(Trade Area Competition)	273	0.03	0.23
Δ(Department Manager Training)	273	0.31	0.90
Δ(Department Employee Turnover Rate)	273	0.00	0.02
Participation in Internal Survey	273	0.30	0.46
Δ(Log of Repository Use by Other Managers)	273	0.41	0.10
Department Yearly Sales in Year 1	273	3.96	6.09

Table A4. Correlation Matrix											
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	Δ(Yearly Department Sales)										
(2)	Treatment (Increase in KMS Usage)	0.20									
(3)	Δ(Department Employees)	0.28	0.10								
(4)	Δ(Log of Data Warehouse Use)	0.03	0.19	0.16							
(5)	Δ(Trade Area Income)	0.03	0.06	-0.09	0.02						
(6)	Δ(Trade Area Competition)	-0.02	-0.02	-0.01	0.00	0.01					
(7)	Δ(Department Manager Training)	-0.10	-0.02	-0.09	-0.11	0.07	-0.10				
(8)	Δ(Department Employee Turnover Rate)	-0.01	-0.05	-0.14	-0.03	-0.05	0.00	-0.01			
(9)	Participation in Internal Survey	-0.01	-0.06	-0.02	0.12	-0.07	0.09	0.01	0.00		
(10)	Δ(Log of Repository Use by Other Managers)	-0.23	-0.10	0.09	-0.13	-0.29	-0.01	-0.08	0.04	0.00	
(11)	Department Yearly Sales in Year 1	0.48	0.22	0.07	1.00	0.00	0.00	0.00	0.00	0.00	0.00

Logit Regression Results

Table A5 presents the logit regression results of the PSM analysis. Model (1) is the result from the logit regression that does not include the current sales ($SALE_i$) as a matching variable and is used to calculate the probability of treatment state for Models (1) and (3) in Table 8 of the paper.

Table A5. Logit Regression Results of PSM Analysis		
	Model (1)	**Model (2)**
Intercept	-0.189 (0.441)	-0.641 (0.466)
Δ(Department Employees)	0.909 (0.560)	0.774 (0.569)
Δ(Log of Data Warehouse Use)	1.316*** (0.427)	1.342*** (0.424)
Δ(Trade Area Income)	0.341 (0.360)	0.346 (0.362)
Δ(Trade Area Competition)	-0.166 (0.350)	-0.179 (0.356)
Δ(Department Manager Training)	-0.015 (0.090)	0.020 (0.091)
Δ(Department Employee Turnover Rate)	-0.481 (4.658)	0.105 (4.687)
Participation in Internal Survey	-0.153 (0.171)	-0.235 (0.177)
Δ(Log of Repository Usage by Other Managers)	-0.613 (0.939)	-0.013 (0.969)
Department Yearly Sales in Year 1		0.056*** (0.016)
Number of Observations	273	273
Pseudo R^2	0.06	0.1

Only the difference in data warehouse usage in Model (1) is positively associated with the treatment state (i.e., the difference in KMS usage). This may reflect that a selection bias is not a major concern in our model. It is interesting that participation in the chain's internal user satisfaction survey is not significant although the same variable is an effective instrumental variable in our 2SLS analysis in Table 6. We believe that this is because the internal survey was conducted rather early, and its effect was already reflected in the first year of the PSM analysis. Thus it does not explain the difference in KMS between the two years. During the two-year period taken for the PSM analysis, the internal survey was conducted at the 19th week of the first year. In Model (2) of Table A5, department sales in year 1 are significant. That is, a manager tends to increase its KMS usage more when her department sales are greater. The variable may account for the capability of managers.

Matching Quality and Sensitivity Analysis

In this section, we examine the indicators to assess the matching quality of our PSM analysis. The details about our PSM results and matching quality are summarized in Table A6. To examine whether our matching balances the distribution of covariates between the treated and non-treated group, standardized biases (SB) before and after matching can be compared (Rosenbaum and Rubin 1983). While the SB after matching has substantially decreased compared to the SB before matching in all three models, the Kernel matching method balances the covariate distribution most satisfactorily with its lowest score among the three specifications. The pseudo-R^2 should be lowered after matching since any systematic difference in covariates between groups should be reduced (Caliendo and Kopeinig 2008; Sianesi 2004). The pseudo-R^2 in all three models decreases after matching and is fairly low in all three models. Referring to the likelihood ratio test, all three models do not reject the null hypothesis of satisfactory balancing. Furthermore, the null hypothesis that the mean of the treatment and control groups do not differ after matching cannot be rejected for all the variables at the 10% significance level except the data warehouse use. In Models (1) and (2) with the NN (nearest neighbor) Matching, the two group's data warehouse use did not differ at the 5% significance level. In Model (3) with the Kernel matching, all the variables do not differ even at the 10% level, which shows the most satisfactory matching as discussed above. In general, we conclude that all three models satisfy the balancing property reasonably well.

Table A6. Propensity Score Matching Method Results

	(1)	(2)	(3)
	NN Matching	**NN Matching with Current Sales**	**Kernel Matching**
Average treatment effect on treated	378.71	417.32	405.52
Standard errors	142.23	165.2	137.5
t-value	2.66	2.53	2.95
Number of observations	273	273	265
Outside the common Support	0	0	8
MSB before Matching*	12.3	12.9	12.3
MSB after Matching	6.4	8.9	3.5
Pseudo-R^2 before matching	0.063	0.102	0.063
Pseudo-R^2 after matching	0.02	0.029	0.005
Likelihood Ratio (Pr > χ^2 after Matching)	7.7 (0.47)	10.9 (0.28)	2.0 (0.98)

*Median standardized bias before matching.

The sensitivity analysis was performed using the Rosenbaum bounds for average treatment effects on treated when unobserved heterogeneity is present (Rosenbaum 2002). This analysis addresses the typical assumption of the PSM method that treatment and control groups differ only on the observed variables by varying the level of gamma (Γ), log odds of differential assignment due to unobserved factors. A high level of gamma means that unobserved variables are more likely to determine assignment to a treatment group. In all three models shown in Table 8, under the scenario that we might have underestimated the true treatment effect, the effect is significant at $\Gamma = 1$, and becomes even more significant with an increase in Γ. Under the scenario that we might have overestimated the true treatment effect, the effect at $\Gamma = 1.3$ remains significant at the 10 percent level in Model (1). In Model (2), the effect at $\Gamma = 1.6$ remains significant at the 10 percent level. However, in Model (3), the effect is significant near $\Gamma = 1$. We believe that this increased sensitivity is partly due to the reduced sample size in Model (3). Thus, the effect size in (3) should be interpreted with more caution.

Comparison with Regression Approach

In this section, we estimate a regression model comparable to the PSM analysis as a benchmark to compare with the treatment effect estimated by the PSM method. We first estimate the following:

$$\Delta SALE_i = \alpha + \beta_1 \cdot Treatment_i + \beta_2 \cdot \Delta Log(LEMP_i) + \beta_3 \cdot \Delta Log(DWHU)_i \\ + \beta_4 \cdot \Delta TRNG_i + \beta_5 \cdot \Delta TINC_i + \beta_6 \cdot \Delta TCOM_i + \beta_7 \cdot \Delta TURN_i + \varepsilon_{4i} \tag{A1}$$

Note that this model includes all of the time variant variables from Equation (1) but does not include three other variables, $SURV_i$, $\Delta Log(OKMS_i)$, and $SALE_i$, used as matching covariates to explain the treatment state. We estimated the model using the OLS (ordinary least squares).

Table A7. Description of Variables

	Estimate
Intercept	213.45** (98.92)
Δ(Department Employees)	1,977.45*** (433.49)
Treatment (Increase in KMS Usage)	381.18*** (129.09)
Δ(Log of Data Warehouse Use)	-76.32 (106.89)
Δ(Trade Area Income)	203.24 (281.93)
Δ(Trade Area Competition)	-21.62 (272.11)
Δ(Department Manager Training)	-77.23 (71.01)
Δ(Department Employee Turnover Rate)	1,728.69 (3,750.31)
Number of Observations	273
R-Squared	11.7%

***Significant at 1%; **significant at 5%; and *significant at 10%. Numbers in parentheses are standard errors.

From the results in Table A7, the treatment effect is 381.2. That is, if a manager increases her usage of repository KMS more than 50 percent of other users, she can improve her annual department sales by 381,180 dollars. This effect size is very similar to the treatment effect from the PSM analysis shown in Table 8. Thus, we conclude that the selection bias is not a serious concern in our model.

Detailed Information about the 2SLS Estimation Results

In this section, we provide more details about our 2SLS estimation, Model (2) in Table 6. The first stage results in Table A8 show that department managers' participation in the internal survey had a significant impact on their future usage. Department managers' participation in the survey in the immediate previous week (*SRVP*) has a positive effect on their repository KMS usage (β_{SRVP} = 0.140, $p < 0.01$). Thus, participation in the internal survey seemed to have provided the managers with an opportunity to enhance their awareness of the KMS and explore it more in the short run. However, their participation in the survey (*SRVA*) has a negative effect in the longer term (β_{SRVA} = -0.081, $p < 0.001$).

Table A8. First Stage Estimation for 2SLS	
	Estimate
Intercept	1.605*** (0.049)
Log of Department Employees	-0.281*** (0.015)
Log of Data Warehouse Use	0.288*** (0.004)
Trade Area Income	0.083*** (0.013)
Trade Area Competition	0.059*** (0.015)
Department Manager Training	-0.004 (0.004)
Department Employee Turnover Rate	0.320*** (0.060)
Survey Answered by Department Manager Any Time Before	-0.081*** (0.010)
Survey Answered by Department Manager in the Immediate Previous Week	0.140*** (0.054)
Number of Observations	39,858
Number of Departments	273
R-Squared (Within)	72.58%
R-Squared (Between)	22.66%
R-Squared (Overall)	36.47%

***Significant at 1%;** significant at 5 %; *significant at 10%. The numbers in parentheses are standard errors.

The significance of the two instrumental variables indicates that our instrumental variables are not weak and adequately correlated with the potentially endogenous regressor, cumulative repository KMS usage. We believe that our instrumentation is very strong. The Cragg-Donald Wald *F*-statistic is 34.73, which is far above the suggested threshold of 10 (Staiger and Stock 1997) and larger than 10% critical value of 19.93 when there is one endogenous regressor and two instrumental variables (Stock and Yogo 2005). The Sargan's test statistic, which tests the instrument exclusion restriction, is 1.453 ($p = 0.228$), showing that the null hypothesis that the instruments are valid cannot be rejected. From the endogeneity test (Baum et al. 2007), we cannot reject the null hypothesis that the potentially endogenous regressor can be treated as exogenous ($\chi^2(1) = 1.875$, p = 0.171). Similarly, the Hausman test shows that the estimates in 2SLS are not statistically different from those in the fixed effects model ($\chi^2(150) = 1.82$, $p \approx 1.0$). We conclude that not only our instrumental variables satisfy the conditions of instrument relevance and exogeneity, but also the endogeneity is not a serious concern in our estimates.

References

Baum, C. F., Schaffer, M. E., and Stillman, S. 2007. "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing," *The Stata Journal* (7:4), pp. 465-506.

Caliendo, M., and Kopeinig, S. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys* (22:1), pp. 31-72.

Rosenbaum, P. R. 2002. *Observational Studies* (2nd ed.), New York: Springer.

Rosenbaum, P. R., and Rubin, D. B. 1983. "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," *Journal of the Royal Statistical Society Series B-Methodological* (45:2), pp. 212-218.

Sianesi, B. 2004. "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *Review of Economics and Statistics* (86:1), pp. 133-155.

Staiger, D., and Stock, J. H. 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica* (65:3), pp. 557-586.

Stock, J. H., and Yogo, M. 2005. "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* D. W. K. Andrews and J. H. Stock (eds.), Cambridge, UK: Cambridge University Press.