Replication summary of ***Peukert (2019)***

"Determinants and Heterogeneity of Switching Costs in IT Outsourcing: Estimates from Firm-Level Data"

Daniel Peredo Siles Theo Herold

December 11, 2023

Preface

We have opted to move away from the initial paper we presented (MacKay, 2022) in class as the replication package is 'too complete', i.e., the provided code allows us to hit enter to get the desired results and leaves little room for analysis and investigation. We instead choose to replicate (Peukert, 2019), *"Determinants and heterogeneity of switching costs in IT outsourcing: Estimates from firm-level data"*. This replication package produces the main results from the paper, but requires some tweaking by the replicator in order to work. The replication package and data were made available to us after correspondence with the author and are provided on the Aalto Mycourses-page.

In brief, Peukert (2019) investigates how firms decide whether to continue an existing firm-vendor relationship or switch to an alternative vendor. In focus are switching costs, i.e., the costs to a firm associated with changing from one vendor to another. Factors that characterize switching costs are search and information costs (Dahlman, 1979; MacKay, 2022), contract-specific characteristics (Abraham & Taylor, 1996; Peukert, 2019), compatibility and learning costs (Xie & Sirbu, 1995; Jeon, Menicucci, & Nasr, 2023), and information asymmetries (Baccara, 2007; Mookherjee, Motta, & Tsumagari, 2020), to name a few. The author develops a model to identify the determinants of firms' switching costs and their relative importance, as well as share of each component on total switching costs. The largest portion of switching costs comes from relationship- and firm-specific factors, as opposed to market-specific variables. Switching costs of those that decide to switch vendors are 50 percent lower than those that decide to stay with their current vendor.

1 Replication

Objective of the paper and code

The objective of the paper is to estimate the determinants of switching costs.[1] The challenge is that whether a firm stays with their current vendor is deliberate or non-deliberate is unobserved. I.e., we can only observe whether a firm switches from a vendor to another but not whether they have renegotiated contractual terms in their current agreement. In such a scenario, estimated switching costs of non-switchers will most likely be biased.

We wish to estimate latent value of switching $v_{i,t+1}$ in a standard probit model, characterized by equation 3 in Peukert (2019), according to

$$v_{i,t+1} = \alpha p_{i,\tau} - \mathbf{x'}_{i,j,t}\beta + \mathbf{z'}_{i,j,t}\zeta + \epsilon_{i,j,t}$$

where $p_{i,\tau}$ is IT outsourcing expenditure in an arbitrary reference year τ, $\mathbf{x'}_{i,j,t}$ are observable characteristics that define switching costs, $\mathbf{z'}_{i,j,t}$ are observable characteristics that explain short-run changes in the long-run trajectory for IT outsourcing expenditure, and $\epsilon_{i,j,t}$ is a normally distributed, zero-mean and constant variance error term. Parameters are only identified up to a scale, i.e., β/σ. This is because the ability to make inference about a latent continuous variable from an observed discrete variable is hampered. In essence, we wish to infer β from $P(v_{i,t+1}|p_{i,\tau};\mathbf{x}_{i,j,t}) = \Phi((\alpha p_{i,\tau} - \mathbf{x'}_{i,j,t}\beta)/\sigma$, where Φ is the standard normal cdf. To estimate the size of switching costs, we need to first estimate σ and then use this to back out β.[2]

The challenge is that we cannot infer about σ from the probit model. But, similarly to β/σ, α is also identifiable only up to a scale, i.e., α/σ. α is the coefficient for IT outsourcing expenditure $p_{i,\tau}$, so given a two-stage method where we first estimate a model pinning down α, we can infer about σ and thus infer about β. To obtain an estimate for α, akin to a standard IV-procedure, we utilize a variable that is correlated with IT outsourcing expenditure $p_{i,\tau}$, in this case expenditure for professional services $c_{i,t}$, from which we can back out α. We therefore fit the following OLS with the dependent variable being the change in expenditure for professional services $\Delta c_{i,t}$ for firm

[1]Descriptive statistics of these variables are provided in table 4 in section 2.

[2]See appendix B.1.2. Economic Identification for more details.

i of the form

$$\Delta c_{i,t} = (\alpha - \gamma) p_{i,\tau} + \gamma c_{i,\tau} + \mathbf{z'}_{i,j,t}\zeta + \tilde{\epsilon}_{i,j,t}$$

Here, $\alpha - \gamma$ follows from the fact that $p_{i,t} \subseteq c_{i,t}$[3]. Together with the estimate from the probit model, $\hat{\alpha} = \alpha/\sigma$, we can derive $\hat{\sigma} = (\widehat{\alpha - \gamma} + \hat{\gamma})/\hat{\alpha}_v$, which allows us to back out $\hat{\beta}$ from the probit model, i.e., $\hat{\beta}/\hat{\sigma} = \frac{\hat{\beta}}{(\widehat{\alpha - \gamma} + \hat{\gamma})/\hat{\alpha}_v}$. With this, we can successfully infer about the size of switching costs directly as a prediction from the model, $\hat{\beta}$. We start by replicating the two-step procedure (table 2 and table 3 in the paper) and then the estimated switching costs between firms that stay in their current contracts with firms that do not.

To the code: Replication of the main results in table 2 and table 3

To start our replication, we run the do-file "main_results.do". Lines 10-14 define the global macro "textpath", in which outregs are stored. "data.dta" is the provided data set, with "data_modified.dta" being the dataset we save for our modifications in the code when replicating.

```
1  *********************
2  * Determinants and heterogeneity of switching costs in IT outsourcing: estimates from firm-level data *
3  * Christian Peukert *
4  * European Journal of Information Systems, 2019 *
5  * https://doi.org/10.1080/0960085X.2018.1529374 *
6  *********************
7
8  * This code reproduces the results reported in Table 2, 3 and 5
9
10 set more off
11 version 11
12 global texpath="results"
13 use "data.dta", clear
14 use "data_modified.dta", clear //Modified data when replicating
15
16
17
18
19
20
```

We start by defining global macros on lines 31-35. The macros we define follow the main variables of interest in the paper. v in the code is $v_{i,t+1}$ in the paper, z in the code is $z_{i,j,t}$ in the paper,

[3]See the technical appendix B.1.3. Parameter Identification, p.48-49.

ind captures all other covariates and *other* are year and vendor specific fixed effects.[4] *c* in the code is $\Delta c_{i,t}$ in the paper and captures the change in expenditure for professional services and is used as dependent variable in the OLS model. The contents of these macros are not clearly specified in the code nor in the paper, and are left to the replicators to figure out. We ran the code with different specifications to achieve as accurate results as possible.

```
25
26  **Some good to haves
27  //macro drop _all
28  //sort join_number year
29
30  **The main macros used, as done in the paper
31  global v vswm
32  global z llogbarg hhi_cont lhhi dnostd dvrepd
33  global ind lspec lvex lvex2 logtravelemp swex swex2 logemp
34  global other yd1-yd11 vmatch_1-vmatch_35 lvmatch_1-lvmatch_33
35  global c deltac
36
37
38  **We need to install some packages:
39  ssc install utest
40  ssc install erepost
41
```

On lines 58 to 82, we perform the probit and OLS models as specified in the objective of the paper. Note that the probit model is performed on line 58 to assure that the shock $\tilde{\epsilon}_{i,j,t}$ follows in the OLS model.[5] On line 59, we perform the OLS model to obtain parameters $(\widehat{\alpha - \gamma})$ for *px* and $\hat{\gamma}$ for a closely correlated variable *cx*. On lines 63-71 we run the probit model with these parameter values stored and obtain estimates for the latent value of switching *v*, as well as $\hat{\alpha} = \alpha/\sigma$ for *px*. With this, we have identified our desired parameters. The "utest" command is a custom package testing the null of a u-shaped relationship, which is done for *lvex* (relationship duration) and *swex* (switching experience). The global macro $k is the count of parameters.

We run the models with White robust standard errors for *c* and *v* separately on lines 74 and

[4] We take data sources and quality as given in this replication, but detailed explanations are provided on pages 16-20 in Peukert (2019).

[5] See the technical appendix, B.1.2 Economic Identification and B.1.3. Parameter Identification for more. In short, $\tilde{\epsilon}_{i,j,t} = (\alpha - \gamma)\nu_{i,\tau} + \vartheta_{i,j,t}$ and $p_{i,\tau} = \bar{p}_\tau + \nu_{i,\tau}$.

76, and with combined standard errors that take unobserved correlation between the two equations in to account on line 78.[6] Further in the code we also estimate a model using errors that deal with potential correlation within firm-vendor clusters, thus estimating three models in total. Row vector b on line 80 contains the parameter estimates and symmetric matrix V on line 79 contains the variance and covariance estimates. These result corresponds to column 1 and column 2 in table 2.

```
54   ***************************************************
55   ************Replication of table 2 and 3************
56   ***************************************************
57
58   probit $v px $z $ind $other
59   reg $c px cx $z $other if e(sample), nocons
60   estimates store c
61
62
63   version 10
64   probit $v px $z $ind $other if e(sample)
65   utest lvex lvex2, fieller
66   utest swex swex2, fieller
67   version 11
68
69   probit $v px $z $ind $other if e(sample)
70   global k=e(k)
71   estimates store v
72
73   gen smpl=e(sample)
74   suest c
75   estimates store crobust
76   suest v
77   estimates store vrobust
78   suest c v
79   matrix V=e(V)
80   matrix b=e(b)
81   erepost b=b V=V
82   estimates store sandwich
```

These results can only be interpreted in terms of the coefficient estimates' respective signs, informing us only of whether some covariate positively or negatively affects the latent value of switching $v_{i,t+1}$ and the change in expenditure for professional services $\Delta c_{i,t}$. To obtain inter-

[6]"suest" is a postestimation command that combines the results—parameter estimates and associated variance-covariance matrices into one parameter vector and simultaneous variance-covariance matrix. In essence, it is us running the models with our desired parameters and their designated results.

pretability of our estimates in dollar units, we need to 1) adjust the coefficients, and 2) standardize some of them.

Lines 88-99 adjusts the coefficient estimates. Lines 88-89 define local macros of α and σ according to the economic model and the estimated parameter values from earlier. Lines 91-96 (quietly) defines a global macro *nlcom* of non-linear combinations of coefficients, with the results stored in *nlcom* corresponding to results in table 3. All variables in '*ind*' are scaled with α and σ on line 94 according to

$$\begin{aligned} `ind' &: \quad `sigma' * [v\_\$v]`ind' \\ `ind' &: \quad ((`alpha')/[v\_\$v]px) * [v\_\$v]`ind' \\ `ind' &: \quad ([c\_mean]px + [c\_mean]cx)/[v\_\$v]px * [v\_\$v]`ind' \end{aligned}$$

which for an arbitrarily chosen variable g in '*ind*' estimates

$$g : \quad ([c\_mean]px + [c\_mean]cx)/[v\_vswm]px * [v\_vswm]g$$

Line 97 does the same for the constant, with line 98 then conducting the estimation. The estimates are stored in *nlcom_sandwich*.

```
86  //------------------------------------------
87  //Adjusting coefficients
88  local alpha="[c_mean]px+[c_mean]cx"
89  local sigma="(`alpha')/[v_${v}]px"
90
91      qui {
92          global nlcom "(alpha: `alpha') (sigma: `sigma')"
93          foreach ind in $z $ind {
94              global nlcom "${nlcom} (`ind': `sigma'*[v_${v}]`ind')"
95          }
96      }
97  global nlcom "$nlcom (_cons: `sigma'*[v_${v}]_cons)"
98  nlcom ${nlcom}, post
99  estimates store nlcom_sandwich
```

We still need to cluster our standard errors on the firm-vendor level. This is done for column 3 in table 2 (the previously mentioned third estimation of the probit and OLS model with combined and clustered standard errors) on lines 104-115, where we estimate combined and clustered standard errors errors. Clustering happens by storing the joint variance estimates with firm and vendor clusters in V on line 112. This is similarly done for column 1 in table 3 on line 121.

```
101 //---------------------------------------------
102 //Standard error clustering for column 3 in table 2 and column 1 in table 3.
103
104 qui {
105     suest c v,vce(cluster join_number)
106     matrix V_1=e(V)
107     suest c v,vce(cluster vidt)
108     matrix V_2=e(V)
109     suest c v,vce(cluster join_number_vidt)
110     matrix V_12=e(V)
111     matrix b=e(b)
112     matrix V=V_1+V_2-V_12
113     erepost b=b V=V
114     estimates store cluster
115 }
116
117 local alpha="[c_mean]px+[c_mean]cx"
118 local sigma="(`alpha')/[v_${v}]px"
119
120
121 nlcom ${nlcom}, post
122 estimates store nlcom_cluster
123     matrix B=e(b)
```

While the coefficient estimates are now intepretable in dollar units, they are not directly comparable as we have both binary and continuous regressors. To maintain coherence, we multiply each coefficient by 2 standard deviations of the corresponding variable. Standard deviations are acquired on line 149 from the local macro '*var*' by defining _sd of each variable in the sample and returning standard deviations through r(sd). These are then estimated on lines 155-161. Similarly, the effect of non-linear coefficients are further standardized by computing the average effect according to $\beta_x 2\sigma_x + \beta_x(2\sigma_x)^2$. Dummy variables are acquired on line 150 from the same local macro by defining sample _dummy for each variable that shows observations taking on values of 0 or 1. These are then scaled on lines 162-164 and the model is then estimated with adjusted and standardized coefficients (with combined and clustered standard errors) on line 167. This result corresponds to column 2 in table 3.

```
//--------------------------------------------
//Standardizing the coefficients with their standard deviations and formula for non-linear coefficients.

qui {
    suest c v,vce(cluster join_number)
    matrix V_1=e(V)
    suest c v,vce(cluster vidt)
    matrix V_2=e(V)
    suest c v,vce(cluster join_number_vidt)
    matrix V_12=e(V)
    matrix b=e(b)
    matrix V=V_1+V_2-V_12
    erepost b=b V=V
    estimates store cluster
}

local alpha="[c_mean]px+[c_mean]cx"
local sigma="(`alpha')/[v_${v}]px"

qui{
foreach var in $z $ind {
    qui su `var' if smpl
    local `var'_sd=r(sd)
    local `var'_dummy=r(min)==0 & r(max)==1
}
}

global nlcom_std "(_cons: (`sigma')*[v_${v}]_cons)"
global nlcom_std "${nlcom_std} (llogbarg: 2*`llogbarg_sd'*`sigma'*[v_${v}]llogbarg)"
global nlcom_std "${nlcom_std} (hhi_cont: 2*`hhi_cont_sd'*`sigma'*[v_${v}]hhi_cont)"
global nlcom_std "${nlcom_std} (dvrepd: `dvrepd_sd'*`sigma'*[v_${v}]dvrepd)"
global nlcom_std "${nlcom_std} (dnostd: 2*`dnostd_sd'*`sigma'*[v_${v}]dnostd)"
global nlcom_std "${nlcom_std} (logemp: 2*`logemp_sd'*`sigma'*[v_${v}]logemp)"
global nlcom_std "${nlcom_std} (lhhi: 2*`lhhi_sd'*`sigma'*[v_${v}]lhhi)"
global nlcom_std "${nlcom_std} (lspec: 2*`lspec_sd'*`sigma'*[v_${v}]lspec)"
global nlcom_std "${nlcom_std} (lvex_s: 2*`lvex_sd'*`sigma'*[v_${v}]lvex+2*`lvex_sd'^2*`sigma'*[v_${v}]lvex2)"
global nlcom_std "${nlcom_std} (logtravelemp: 2*`logtravelemp_sd'*`sigma'*[v_${v}]logtravelemp)"
global nlcom_std "${nlcom_std} (swex_s: 2*`swex_sd'*`sigma'*[v_${v}]swex+2*`swex_sd'^2*`sigma'*[v_${v}]swex2)"


nlcom ${nlcom_std}, post
estimates store nlcom_cluster_std
```

With this, we have the output of the main results in table 2 and table 3 in (Peukert, 2019), which we replicate in table 1 and table 2 below. These are exported from Stata using esttab commands as provided in the code[7]. We manage to almost perfectly replicate the results, with the caveat that the constant in the second stage result differs by little over 1000 dollars.

[7] The estttab commands are however omitted here.

Table 1: First stage results of table 2 on p.24

	(1) White Robust Standard Errors		(2) Combined Standard Errors		(3) Combined & Clustered Standard Errors	
Expenditure Model (dependent variable: Δc)						
$\alpha - \gamma$	-0.0434***	(0.0111)	-0.0434**	(0.0220)	-0.0434***	(0.0167)
γ	-0.0451***	(0.00277)	-0.0451***	(0.00779)	-0.0451***	(0.00543)
Bargaining Power	4514.3***	(1235.6)	4514.3**	(1902.0)	4514.3***	(1651.6)
Vendor Market HHI at Signing	4851.4	(37743.7)	4851.4	(37339.9)	4851.4	(33289.8)
Vendor Market HHI	-7884.3	(30925.1)	-7884.3	(28427.0)	-7884.3	(14671.9)
Higher Visibility	-5424.0	(3848.6)	-5424.0	(3372.2)	-5424.0**	(2634.9)
More Features	68.31	(109.1)	68.31	(111.0)	68.31	(81.10)
Switching Model (dependent variable: v)						
α/σ	-0.00000557***	(0.00000162)	-0.00000557***	(0.00000151)	-0.00000557***	(0.00000137)
Bargaining Power	0.303***	(0.0831)	0.303***	(0.0950)	0.303**	(0.131)
Vendor Market HHI at Signing	6.278**	(2.627)	6.278*	(3.431)	6.278	(4.133)
Vendor Market HHI	-6.650**	(2.588)	-6.650**	(2.871)	-6.650**	(2.605)
Higher Visibility	1.716***	(0.280)	1.716***	(0.283)	1.716***	(0.453)
More Features	0.0233***	(0.00675)	0.0233***	(0.00626)	0.0233***	(0.00734)
Asset Specificity	-0.346***	(0.0864)	-0.346***	(0.0759)	-0.346***	(0.0632)
Relationship Duration	-0.917***	(0.123)	-0.917***	(0.121)	-0.917***	(0.109)
Relationship Duration sq.	0.0360***	(0.00940)	0.0360***	(0.00869)	0.0360***	(0.00779)
Search Effort	0.315***	(0.0907)	0.315***	(0.0811)	0.315***	(0.0523)
Switching Experience	-0.546***	(0.176)	-0.546***	(0.152)	-0.546***	(0.129)
Switching Experience sq.	0.0844***	(0.0328)	0.0844***	(0.0252)	0.0844***	(0.0210)
Firm Size	0.470***	(0.160)	0.470***	(0.141)	0.470***	(0.154)
Constant	2.356	(1.470)	2.356	(1.526)	2.356	(1.725)
Year fixed effect	Yes		Yes		Yes	
Vendor fixed effect t	Yes		Yes		Yes	
Vendor fixed effect t-1	Yes		Yes		Yes	
adj. R2/pseudo R2	0.0799/0.658					
Observations	6879		6879		6879	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Second stage results of table 3 on p.26

	(1) Adjusted Coefficients		(2) Adjusted and Standardized	
Parameters				
α	-0.0885***	(0.0153)		
σ	15902.4***	(5777.7)		
Fee Shifters				
Bargaining Power	4811.2**	(2057.6)	15895.7**	(6798.2)
Vendor Market HHI at Signing	99838.2	(80441.3)	4634.5	(3734.1)
Switchng Costs				
Vendor Market HHI	-105758.0**	(50853.2)	-6686.2**	(3215.0)
Higher Visibility	27291.9**	(10786.0)	6987.1**	(2761.4)
More Features	370.8**	(165.2)	7013.9**	(3124.6)
Asset Specificity	-5507.6**	(2196.5)	-9224.7**	(3679.0)
Relationship Duration	-14584.0***	(4997.4)		
Relationship Duration sq.	572.6***	(219.1)		
Search Effort	5006.6**	(2019.6)	9790.4**	(3949.3)
Switching Experience	-8675.0**	(4034.1)		
Switching Experience sq.	1342.4**	(623.3)		
Firm Size	7473.9**	(3139.3)	13722.6**	(5764.0)
Constant	37461.8*	(19548.4)	37461.8*	(19548.4)
Year fixed effect	Yes			
Vendor fixed effect t	Yes			
Vendor fixed effect t-1	Yes			
Observations	6879			

Combined and clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back to the code: replicating table 5

The other table we have received a replication package for concerns table 5 from the paper (p.29), which summarizes the average size of switching costs, both unconditional and conditional on switching. In other words, after estimating the coefficients for the first column of the Second Stage Results, focusing on the panel for Switching Costs, we can directly estimate the individual switching costs since:

$$\hat{s}_{i,j,t} = \mathbf{x'}_{i,j,t}\hat{\beta}$$

```
global ind lspec lvex lvex2 logtravelemp swex swex2 logemp
global s lhhi dnostd dvrepd $ind
gen s=0
local i=0



foreach var in 0 0 0 0 $s 0{
    local i=`i'+1
    di B[1,`i'] "`var'"
    replace s=s-B[1,`i']*`var' if smpl
}

drop srel
gen srel=s/mean_acct_290d
drop srelt
gen srelt=s/mean_acct_290d/vex
drop srelpx
gen srelpx=s/px
```

With the results from the Second Stage stored in Matrix B (dimension 1,15), the loop looks for the coefficient $B(1, i)$, extracting coefficient values for each variable and multiplying it with the observation value for that variable, then it subtracts from the new variable s (since the coefficients affect the switching cost with a negative sign) before passing to the next position/coefficient, the zero at the end relates to the constant being the last variable stored in B. Doing so, the s variable represents the switching costs for each observation, allowing us to look into the heterogeneity of results.

It's evident that the switching costs for the ones who switched were smaller compared to the ones who don't switch (54% smaller in absolute terms). The minimum being negative for all categories indicates a possibility that the vendors pay firms to switch. The distribution of switching costs is fairly asymmetric since the standard deviation is around 25% of the mean.

Table 3: Estimated Switching Costs, Switchers vs Stayers

	Mean	SD	Min	Max	Median	IQR
Total						
Absolute	81684	17111	-29680	128561	84316	17849
Relative	.37	.107	-.1342	.8653	.3604	.09349
Relative per Contract Year	.04467	.02279	-.04502	.5549	.04032	.01354
Non-Switchers						
Absolute	82531	15707	-19920	128561	84640	17320
Relative	.3737	.103	-.09004	.8653	.3617	.09176
Relative per Contract Year	.04399	.01583	-.04502	.336	.04032	.01324
Switchers						
Absolute	38087	27093	-29680	99519	37181	34639
Relative	.176	.128	-.1342	.5549	.1676	.1514
Relative per Contract Year	.07964	.115	-.03794	.5549	.03568	.06686

Relative switching costs are expressed in relation to annual expenditure for professional and outside services

2 Challenges/problems in replicating the paper

As mentioned in the Preface, we contacted the author to ask for the replication package, adding a layer of difficulty to the task at hand. Fortunately enough, the author was kind enough to provide both data and code.

The main challenges we found when replicating the paper where that the code didn't include important steps such as defining what variables are stored in the global macros that the author uses thought the code ($v, z, ind, other, z$ and s, as seen in the code snippets). At first sight this didn't seem difficult to fill in since we had read the paper and went through the equations there, but some variables weren't labeled and the equations syntax in the code was slightly different than the paper (for implementation purposes). Nevertheless, we inferred what variables to use following the tables for First and Second Stage Results and a table in the Appendix containing the descriptive statistics of some of the variables used in the econometric model, and we are confident about our choices since we can replicate the First and Second Stage Results almost entirely. But when looking at the descriptive statistics, we noticed in table 4 that there is a typo in the paper, as the descriptive statistics for "More Features" differ from the ones reported in the paper, in the table below we compare the means for the stayers, switchers and total sample.

	Replication	Paper
Stayers	4.519	4.877
Switchers	8.557	11.54
Total	4.596	5.004

On a similar note, we mentioned in Section 1 that the constant we estimated for the Second Stage results (Table 2) differs from the one reported in page 26 of the paper, being the only result that doesn't match (Peukert, 2019), differing by 4%.

	Replication	Paper	Diff.
Constant	37461.8	36116.5	1345.3
standard error	19548.4	18781.0	767.4

Regarding the results shown in Table 3, we tried defining the global ”s” on line 194 as:

1. Only the variables defined in $\mathbf{x}'_{i,j,t}$: Asset specificity, relationship duration (level and squared), search effort, switching experience (level and squared) and firm size.
2. **Switching Costs** panel of the Second Stage Results: This includes the variables in $X'_{i,j,t}$ plus some variables from $\mathbf{z}'_{i,j,t}$ (Vendor market HHI, more features and higher visibility). We tried this specification because as discussed in section 5.2.2 of the paper, these variables are the determinants of Switching Costs.

Where the first attempt resonates with the code provided by the author, as it runs, albeit not finding the results we look for and instead reporting average switching costs of around 40.000 $, considerably lower than the 81.684 $ that are correct. To try the second approach, the one that allowed us to replicate the results as in the paper, we had to do a slight modification of the code provided by the author. It looks trivial once its done, but it took us time to figure out how to make things work in an efficient way, as opposed to extracting coefficient values for each variable manually (not in a loop) and generating the variable s as the sum of those, which we had to try before figuring out how to tweak the code so it would work.

```
****trial to get table 5 withouth using the code provided******
matrix list B

gen coef_llogbarg = B[1,3] * -llogbarg
gen coef_hhi_cont = B[1,4] * -hhi_cont
gen coef_lhhi = B[1,5] * -lhhi
gen coef_dnostd = B[1,6] * -dnostd
gen coef_dvrepd = B[1,7] * -dvrepd
gen coef_lspec = B[1,8] * -lspec
gen coef_lvex = B[1,9] * -lvex
gen coef_lvex2 = B[1,10] * -lvex2
gen coef_logtravelemp = B[1,11] * -logtravelemp
gen coef_swex = B[1,12] * -swex
gen coef_swex2 = B[1,13] * -swex2
gen coef_logemp = B[1,14] * -logemp
gen coef_cons = B[1,15] *-1

gen s = coef_lhhi + coef_dnostd + coef_dvrepd +coef_lspec + coef_lvex + ///
                coef_lvex2 + coef_logtravelemp + coef_swex + coef_swex2 + ///
                coef_logemp
```

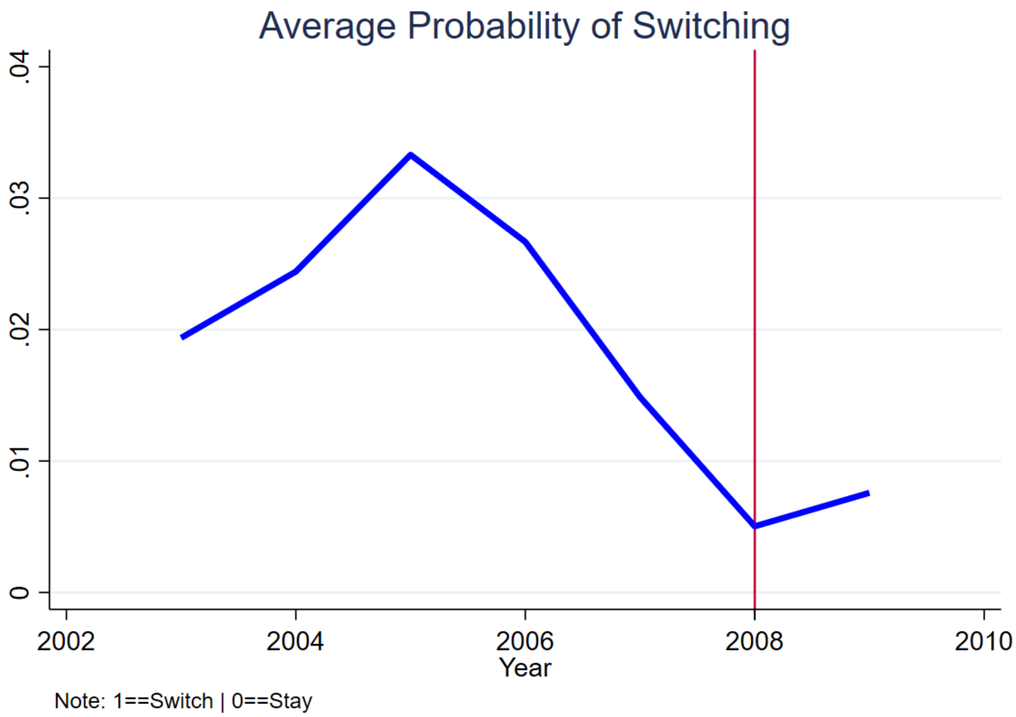
Table 4: Descriptive statistics

	Mean	SD	Min	Max
Stayers (n=6748)				
$\bar{p}_\tau$ (in US $)	135712	121734	5094	1680263
Δc (in US $)	-8295	71691	-1641123	1588515
c_τ	305594	410217	0	5422685
Bargaining Power at Signing*	-6.664	1.658	-11.92	-.1534
Bargaining Power*	-6.669	1.639	-11.92	0
Vendor Market HHI at Signing	.4062	.02277	.331	.508
Vendor Market HHI	.414	.03151	.331	.508
More Features	4.519	9.389	0	44
Higher Visibility	.06669	.2495	0	1
Asset Specificity	2.364	.835	0	4
Relationship Duration (years)	8.893	2.05	2	11
Search Effort* (in US $)	6.976	.9812	1.609	9.914
Switching Experience	.3048	1.034	0	7
Firm Size* (FTE employees)	2.534	.9189	0	5.707
Switchers (n=131)				
$\bar{p}_\tau$ (in US $)	130621	87766	39825	364799
Δc (in US $)	-19504	78703	-465937	383121
c_τ	277210	346611	9448	2750119
Bargaining Power at Signing*	-5.196	2.186	-9.359	-.6819
Bargaining Power*	-5.458	1.89	-8.973	-.8317
Vendor Market HHI at Signing	.4113	.03965	.3183	.5083
Vendor Market HHI	.402	.03456	.331	.508
More Features	8.557	11.91	0	41
Higher Visibility	.2672	.4442	0	1
Asset Specificity	2.293	.9548	0	4
Relationship Duration (years)	4.076	2.615	1	10
Search Effort* (in US $)	7.173	.7582	4.045	8.789
Switching Experience	1.008	1.778	0	6
Firm Size* (FTE employees)	2.661	.8641	1.253	4.58
Total (n=6879)				
$\bar{p}_\tau$ (in US $)	135615	121173	5094	1680263
Δc (in US $)	-8508	71841	-1641123	1588515
c_τ	305054	409095	0	5422685
Bargaining Power at Signing*	-6.636	1.681	-11.92	-.1534
Bargaining Power*	-6.646	1.652	-11.92	0
Vendor Market HHI at Signing	.4063	.02321	.3183	.5083
Vendor Market HHI	.4138	.03161	.331	.508
More Features	4.596	9.458	0	44
Higher Visibility	.0705	.256	0	1
Asset Specificity	2.362	.8375	0	4
Relationship Duration (years)	8.801	2.165	1	11
Search Effort* (in US $)	6.979	.9778	1.609	9.914
Switching Experience	.3182	1.058	0	7
Firm Size* (FTE employees)	2.537	.918	0	5.707

*Logarithm used in econometric model but level data shown in Table A.1 in the paper

3 Suggested extensions and modifications

When plotting the average probability of switching, we find a downwards-sloping trend since 2005. The author makes no note of this nor the potential effects of the great 2008 financial crisis in the paper, which is why we choose to investigate this further. In particular, we focus the post-2007 period (i.e., 2008-2009). The propensity to switch has been oscillating between 1.5 percent and 3 percent up until 2007, which is arguably different enough from the subsequent period.



We re-estimate the entire model presented by interacting our time-variant variables on a post-2007 dummy to gauge if there are any significant effects due to the financial crisis. The post-2007 dummy captures approximately 28 percent of all observations. We provide the modifications to the code in screenshots in the appendix and focus on the output here. We estimate the time-variant variables that are fee shifters (i.e., not switching costs), but in the interest of time, we choose to only focus on switching costs here. Note that the econometric specification is such that a positive (negative) sign implies decreased (increased) switching costs; see the equation for $v_{i,t+1}$ in section 1.

The sign of vendor market concentration flips, indicating decreasing switching costs post-2007 and a higher propensity to switch. Given the turmoil of the financial crisis, it is reasonable to assume that vendors became more willing to accept renegotiations of cost-sensitive buyers. Higher quality of a vendor gets a similar sign-flip, indicating increasing switching costs post-2007 and thus a lower propensity to switch. In the paper, the author hypothesizes that higher quality is a trade-off between the oppotunity cost of working with the higher quality vendor vis-a-vis the migration-, search- and opportunity costs of switching. The negative coefficient in the main results in table 3 indicates that buyers deem the opportunity cost of higher quality to be greater than migration-, search- and opportunity costs over the entire sample period. The interaction term would indicate the opposite following the financial crisis. When market concentration of vendors increases (i.e., fewer vendors), the number of vendors offering higher quality presumably decreases. Similarly, the burden of tangible costs (such as a the cost of migration) presumably weighs heavier than intangible costs (such as opportunity costs) for cost-constrained buyers. Taken together, the economic context does give credence to the interaction terms' coefficient estimates. Higher brand recognition post-2007 is omitted due to multicollinearity.

Table 5: Modification with interaction term for time-variant regressors and post-2007.

Dep. var: $v_{i,t+1}$ (Probability to switch)	(1) Modified model	
Vendor Market HHI	-7305.4***	(2565.1)
Vendor Market HHI post-2008	6141.2	(56545.1)
Higher quality	7457.9**	(3188.2)
Higher quality post-2008	-4401.3	(2911.9)
Higher Brand Recognition	7584.2**	(3071.8)
Higher Brand Recognition post-2008	0	(.)
Observations	6879	

Standard errors in parentheses* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This modification is rather minor, but it does highlight the importance of model specification and economic context. Given the sample period of the paper, controlling for the financial crisis could be a valid form of robustness check.

A further avenue of modification would be to challenge **Assumption 1** for economic identification (p.45) by changing the functional form for the prices to a quadratic term. The entire economic identification builds on the assumption that outsourcing expenditure follows a stable trend in the long-run, with short-run fluctuations being caused by renegotiation and vendor switching. The author assumes that these expenditures on outsourcing affects the value of switching only linearly. An argument could be made that large previous changes in expenditure impacts the value of switching more than smaller previous changes in expenditure do. If so, the relationship would not be linear. Allowing for higher-order relationships in IT-expenditure would require us to use $p_{i,t}$ as further assumptions are placed on the persistence of (unobserved) shocks, but the replication package only gives us $\Delta p_{i,t}$. We cannot therefore properly model such higher-order relationships in accordance with the economic model.[8]

Internally, we discussed a range of modifications and inquiries, such as changing the distribution for the maximum likelihood estimation, performing power calculations to assess validity of results. In the interest of time, these were not further pursued.

[8]Simply including higher-order extensions of px has only marginal changes to the results in no particularly meaningful way.

4 Appendix - Screenshots of code for post-2007 model modification

For our modification with time-variant switching costs interacted with a post-2008 dummy, we start by generating the post-2008 dummy on lines 249-251. We then create separate interaction terms for the time-variant variables on lines 252-256. On line 263, we define the global macro *crisis*, which enters in the probit and OLS on lines 266-278.

```
247 //----------------------------------------------------
248 //Interact the post-2008 period with variables that are time variant (i.e., z)
249 gen crisisy = 0
250 replace crisisy=1 if year==2008
251 replace crisisy=1 if year==2009
252 gen crisishhi = crisisy*lhhi //relevant market in general
253 gen crisishhicont = crisisy*hhi_cont //relevant market at signing
254 gen crisisbarg = crisisy*logbarg //bargaining power of the firm
255 gen crisisfeatures = crisisy*dvrepd //more features
256 gen crisisvisibility = crisisy*dvrepd //more visibility
257
258 global v vswm
259 global z llogbarg hhi_cont lhhi dnostd dvrepd
260 global ind lspec lvex lvex2 logtravelemp swex swex2 logemp
261 global other yd1-yd11 vmatch_1-vmatch_35 lvmatch_1-lvmatch_33
262 global c deltac
263 global crisis crisishhi crisishhicont crisisbarg crisisfeatures crisisvisibility
```

```
266 probit $v px $z $ind $other $crisis
267 reg $c px cx $z $other $crisis if e(sample), nocons
268 estimates store c
269
270 version 10
271 probit $v px $z $ind $other $crisis if e(sample)
272 utest lvex lvex2, fieller
273 utest swex swex2, fieller
274 version 11
275
276 probit $v px $z $ind $other $crisis if e(sample)
277 global k=e(k)
278 estimates store v
```

When creating the modified results from table 3 in paper, table 2 in this replication summary, we again need to adjust and standardize the interaction terms defined in the *crisis*-macro. The adjustment happens on lines 305-306, while the standardization with two standard deviations happen on lines 377-383. Other than that, the code is identical to the initial code in the replication package.

```
299 //We modify on lines 312-314 to include these new interaction terms.
300     qui {
301         global nlcom "(alpha: `alpha') (sigma: `sigma')"
302         foreach ind in $z $ind {
303             global nlcom "${nlcom} (`ind': `sigma'*[v_${v}]`ind')"
304         }
305         foreach crisis in $crisis {
306             global nlcom "${nlcom} (`crisis': `sigma'*[v_${v}]`crisis')"
307         }
308     }
309 
310 global nlcom "$nlcom (_cons: `sigma'*[v_${v}]_cons)"
311 nlcom ${nlcom}, post
312 estimates store nlcom_sandwich
```

```
360 qui{
361 foreach var in $z $ind $crisis {
362     qui su `var' if smpl
363     local `var'_sd=r(sd)
364     local `var'_dummy=r(min)==0 & r(max)==1
365 }
366 }
367 
368 //The interaction terms in $crisis are adjusted and standardized
369 global nlcom_std "(_cons: (`sigma')*[v_${v}]_cons)"
370 global nlcom_std "${nlcom_std} (llogbarg: 2*`llogbarg_sd'*`sigma'*[v_${v}]llogbarg)"
371 global nlcom_std "${nlcom_std} (hhi_cont: 2*`hhi_cont_sd'*`sigma'*[v_${v}]hhi_cont)"
372 global nlcom_std "${nlcom_std} (dvrepd: `dvrepd_sd'*`sigma'*[v_${v}]dvrepd)"
373 global nlcom_std "${nlcom_std} (dnostd: 2*`dnostd_sd'*`sigma'*[v_${v}]dnostd)"
374 global nlcom_std "${nlcom_std} (logemp: 2*`logemp_sd'*`sigma'*[v_${v}]logemp)"
375 global nlcom_std "${nlcom_std} (lhhi: 2*`lhhi_sd'*`sigma'*[v_${v}]lhhi)"
376 global nlcom_std "${nlcom_std} (lspec: 2*`lspec_sd'*`sigma'*[v_${v}]lspec)"
377 global nlcom_std "${nlcom_std} (crisishhi: 2*`crisishhi_sd'*`sigma'*[v_${v}]crisishhi)" //for lhhi post 2008
378 global nlcom_std "${nlcom_std} (crisishhicont: 2*`crisishhicont_sd'*`sigma'*[v_${v}]crisishhicont)" //for hhi_cont post 2008
379 global nlcom_std "${nlcom_std} (crisisbarg: 2*`crisisbarg_sd'*`sigma'*[v_${v}]crisisbarg)" //for logbarg post 2008
380 global nlcom_std "${nlcom_std} (crisisfeatures: 2*`crisisfeatures_sd'*`sigma'*[v_${v}]crisisfeatures)" //for dvrepd post 2008
381 global nlcom_std "${nlcom_std} (crisisvisibility: 2*`crisisvisibility_sd'*`sigma'*[v_${v}]crisisvisibility)" //for dnostd post 2008
382 global nlcom_std "${nlcom_std} (lvex_s: 2*`lvex_sd'*`sigma'*[v_${v}]lvex+2*`lvex_sd'^2*`sigma'*[v_${v}]lvex2)"
383 global nlcom_std "${nlcom_std} (logtravelemp: 2*`logtravelemp_sd'*`sigma'*[v_${v}]logtravelemp)"
384 global nlcom_std "${nlcom_std} (swex_s: 2*`swex_sd'*`sigma'*[v_${v}]swex+2*`swex_sd'^2*`sigma'*[v_${v}]swex2)"
```

References

Abraham, K. G., & Taylor, S. K. (1996). Firms' use of outside contractors: Theory and evidence. Journal of Labor Economics, 14(3), 394–424. Retrieved 2023-09-25, from http://www.jstor.org/stable/2535360

Baccara, M. (2007). Outsourcing, information leakage, and consulting firms. The RAND Journal of Economics, 38(1), 269–289. Retrieved 2023-09-25, from http://www.jstor.org/stable/25046303

Dahlman, C. J. (1979). The problem of externality. Journal of Law and Economics, 22(1), 141-62. Retrieved from https://EconPapers.repec.org/RePEc:ucp:jlawec:v:22:y:1979:i:1:p:141-62

Jeon, D.-S., Menicucci, D., & Nasr, N. (2023, February). Compatibility choices, switching costs, and data portability. American Economic Journal: Microeconomics, 15(1), 30-73. Retrieved from https://www.aeaweb.org/articles?id=10.1257/mic.20200309 doi: 10.1257/mic.20200309

MacKay, A. (2022, August). Contract duration and the costs of market transactions. American Economic Journal: Microeconomics, 14(3), 164-212. Retrieved from https://www.aeaweb.org/articles?id=10.1257/mic.20200128 doi: 10.1257/mic.20200128

Mookherjee, D., Motta, A., & Tsumagari, M. (2020). Consulting collusive experts. Games and Economic Behavior, 122, 290-317. Retrieved from https://www.sciencedirect.com/science/article/pii/S0899825620300646 doi: https://doi.org/10.1016/j.geb.2020.04.011

Peukert, C. (2019). Determinants and heterogeneity of switching costs in it outsourcing: estimates from firm-level data. European Journal of Information Systems, 28(3), 291-317. Retrieved from https://doi.org/10.1080/0960085X.2018.1529374 doi: 10.1080/0960085X.2018.1529374

Xie, J., & Sirbu, M. (1995). Price competition and compatibility in the presence of positive demand externalities. Management Science, 41(5), 909-926. Retrieved from https://doi.org/10.1287/mnsc.41.5.909 doi: 10.1287/mnsc.41.5.909