Online Appendix
Stuck in the Adoption Funnel: The Effect of Interruptions in the Adoption Process on Usage

Anja Lambrecht
London Business School
alambrecht@london.edu

Katja Seim
University of Pennsylvania
kseim@wharton.upenn.edu

Catherine Tucker
MIT Sloan School of Management
cetucker@mit.edu

September 28, 2010

Abstract

This appendix explores in detail issues dealt with only briefly in the paper 'Stuck in the Adoption Funnel: The Effect of Interruptions in the Adoption Process on Usage'. It provides additional information and background on the German vacation system. It then derives the empirical estimation equation more fully. Last, it presents several alternative specifications to the main results in the paper as robustness checks.

Keywords: Online Banking, Technology Adoption, Adoption Process, Online Security, Self-Service Technology.

1 German School Vacations and Holidays

Our instrumenting strategy exploits significant variation in the timing and length of state-specified school vacations across German states over the course of the year. This variation stems from official federal government policy to stagger school vacations across states. By ensuring that not everyone goes on vacation at the same time (see Bundesministerium für Verkehr, Bau und Stadtentwicklung, "LLW-Fahrverbot in der Ferienreisezeit", `http://www.bmvbs.de/-,302.2221/Lkw-Fahrverbot-in-der-Ferienre.htm` (accessed 08/16/07)), the government thereby hopes to reduce traffic congestion.

Germans exploit these vacation times extensively. They take on average 26 vacation days a year, compared to the 11 days taken by Americans,[1] and go on roughly 1.6 vacation trips per year.[2] The prevalence of traffic congestion at the onset and end of the various states' school vacation periods[3] suggests that such trips frequently coinicide with state-specified school vacations and that these potentially serve as a source of disengagement from online activity, as we assume.

Public holidays are primarily Christian holidays, their observance varying between states with a predominantly Catholic population and those with a predominantly Protestant population. The number of public holidays increases with the share of a state's Catholic population ($p<0.001$). In addition, the number of public holidays is higher in former East German states ($p<0.01$). In our regressions we control for differences between states by using state fixed effects. Travel is less frequent on such dates, but there is no mail delivery.

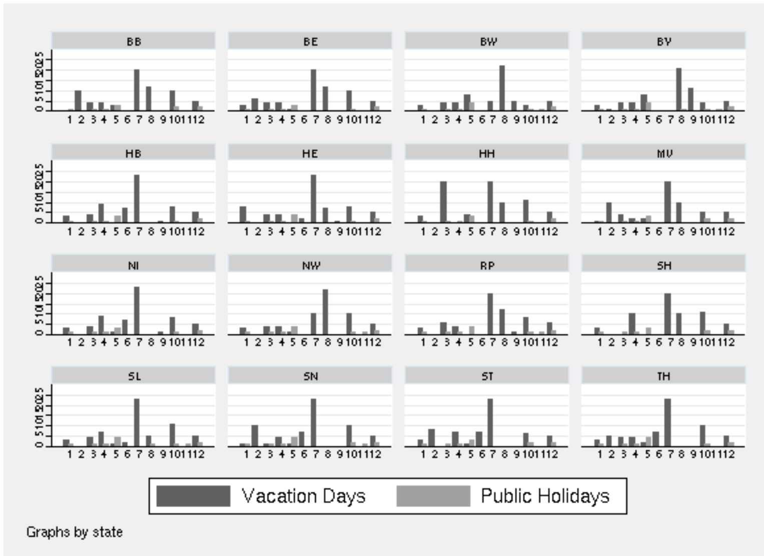
Figure A-1 displays the extensive variation in vacations and public holidays both across states and within states over time for 2002.

[1]Expedia.com, "2007 International Vacation Deprivation Survey Results", `http://www.expedia.com/vacationdeprivation` (accessed 08/16/07).

[2]Axel Springer Marketing Anzeigen, "Tourismus 2002", `http://www.mediapilot.de/cda` (accessed 08/16/07).

[3]See `www.adac.de/Verkehr/Staukalender/default.asp` (accessed 08/16/07).

Figure A-1: Variation in Vacation and Public Holidays by State, 2002



2 Empirical Model of the Multistage Adoption Process: Details

Here, we provide a detailed derivation of the empirical likelihood function in Equation (2) in the main paper. We assume a simple utility model for the comparison of benefits and costs to using online banking. This allows us to derive the customer's decision to make at least one online transaction in a given month t, $t = 1, ..., T_i$, following the initial trial. We observe only whether the customer uses the online banking service, U_{it}, but not the underlying latent utility from doing so, U_{it}^*:

$$U_{it} = \begin{cases} 1 & \text{if } U_{it}^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{A-1}$$

where U_{it}^* reflects the benefits and costs of using the service,

$$U_{it}^* = f(\mathit{Benefits}_{it}, \mathit{Costs}_{it}). \tag{A-2}$$

As described in the main paper, we capture the $\mathit{Benefits}_{it}$ of using the service by a vector of customer attributes, $X_{it}\beta$. These customer attributes include the number of bank branches near the customer and whether they hold a brokerage account, to proxy for the likely usefulness of not having to go to a branch; the customer's age and gender, to reflect possible differences in cost of time across these demographic groups; and their overall demand for banking services, which we approximate through the number of offline transactions the customer makes in that month. We further include seasonal and state controls, and the number of public holidays and vacation days in a month, to allow for systematic differences in the attractiveness of online banking in different months or in different locations within Germany. The Costs_{it} of using the service reflect the customer's perceived cost of returning to the website that month and reusing the service.

$$\mathit{Costs}_{it} = f(\zeta_{it}, U_{i,t-1}, \mathit{Int\_Login}_i, \mathit{Int\_Trans}_i) \tag{A-3}$$

We include in Equation (A-3) a vector of indicator variables ζ_{it} that reflect how long the customer has been making transactions using online banking, to capture the likely decline in learning costs over time. We include an indicator for whether a customer used the service in the last month, $U_{i,t-1}$, as a separate cost shifter, reflecting the potential for state dependence in cognitive costs. $\mathit{Int\_Login}_i$ captures whether that customer experienced an interruption between signup and initial login. Int_Trans_i similarly captures an interruption between initial login and the first online transaction. We allow the effect of interruptions to vary flexibly with the time spent in the usage stage, by including non-parametric interactions between ζ_{it} and the interruption indicators, Int_Login_i and Int_Trans_i.

We further include in the utility specification an unobserved individual effect, ϵ_i^3, and random customer- and month-specific shock, ν_{it}^3, which we assume to be distributed i.i.d. according to a standard normal distribution with mean zero and standard deviation one.

We add to the T_i equations describing the monthly usage decisions subsequent to the month of the customer's first online transaction two equations that specify the likelihood of being interrupted in adoption as a function of customer attributes and, importantly, exogenous determinants of interruptions. This yields the system of equations (1) in the main paper:

$$\begin{aligned} Int\_Login_i &= I(\beta_{10} + X_i^1\beta_{11} + Z_i^1\gamma_1 + \epsilon_i^1 > 0) = I(\overline{u}_i^1 + \epsilon_i^1 > 0) \\ Int\_Trans_i &= I(\beta_{20} + X_i^2\beta_{21} + \alpha_{21}Int\_Login_i + Z_i^2\gamma_2 + \epsilon_i^2 > 0) = I(\overline{u}_i^2 + \epsilon_i^2 > 0) \\ U_{it} &= I(\beta_{30} + X_{it}^3\beta_{31} + \zeta_{it}\alpha_{30} + Int\_Login_i(\alpha_{31}^1 + \zeta_{it}\alpha_{32}^1) + Int\_Trans_i(\alpha_{31}^2 + \zeta_{it}\alpha_{32}^2) \\ &\quad + \alpha_{33}U_{i,t-1} + \epsilon_i^3 + \nu_{it}^3 > 0) = I(\overline{u}_{it}^3 + \epsilon_i^3 + \nu_{it}^3 > 0) \end{aligned} \tag{A-4}$$

As described in the paper, we assume that the unobserved customer attributes ϵ_i are correlated across stages and follow a trivariate normal distribution:

$$\begin{pmatrix} \epsilon_i^1 \\ \epsilon_i^2 \\ \epsilon_i^3 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & \sigma_3 \end{bmatrix} \right). \tag{A-5}$$

Under these assumptions, the likelihood of observing customer i's stream of interruption and usage outcomes is given by:

$$\begin{aligned} L_i &= \Pr\left(Int\_Login_i = int_i^L, Int\_Trans_i = int_i^T, U_{i1} = u_{i1}, ..., U_{iT_i} = u_{iT_i}\right) \\ &= \Pr\left(\epsilon_i^1 < \left(2int_i^L - 1\right)\overline{u}_i^1, \epsilon_i^2 < \left(2int_i^T - 1\right)\overline{u}_i^2,\right. \\ &\qquad \left.\epsilon_i^3 + \nu_{it}^3 < (2u_{it} - 1)\,\overline{u}_{it}^3 \quad \forall t = 1, ..., T_i\right) \end{aligned} \tag{A-6}$$

where int_i^L and int_i^T denote customer i's observed outcomes for $\{Int\_Login_i, Int\_Trans_i\}$ and u_{it} their observed transaction decisions. We simplify customer i's likelihood by applying

Bayes' rule and exploiting the *i.i.d.* assumption for ν_{it}^3:

$$L_i = \int_{-\infty}^{\infty} \Big[\Pr\left(\epsilon_i^1 < \left(2int_i^L - 1\right)\overline{u}_i^1, \epsilon_i^2 < \left(2int_i^T - 1\right)\overline{u}_i^2 \left|\epsilon_i^3\right.\right) \\ \times \prod_{t=1,\ldots T_i} \Pr\left(\nu_{it}^3 < \left(2u_{it} - 1\right)\left(\overline{u}_{it}^3 + \epsilon_i^3\right)\left|\epsilon_i^3\right.\right) \Big] f(\epsilon_i^3) d\epsilon_i^3, \tag{A-7}$$

where $f(\epsilon_i^3)$ denotes the marginal normal pdf of ϵ_i^3. Conditional on ϵ_i^3, the first part of the likelihood, corresponding to the propensities of a customer experiencing interruptions, is a bivariate probit probability for each of the four possible interruption outcomes, $\{Int\_Login_i, Int\_Trans_i\}$=$(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. Similarly, the likelihood of observing each transaction decision, conditional on ϵ_i^3, is a univariate probit probability. This yields customer i's likelihood in Equation (2) in the paper:

$$L_i = \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\left(2int_i^T - 1\right)\overline{u}_i^2} \int_{-\infty}^{\left(2int_i^L - 1\right)\overline{u}_i^1} f(\epsilon_i^1, \epsilon_i^2 | \epsilon_i^3) d\epsilon_i^1 d\epsilon_i^2 \\ \times \prod_{t=1,\ldots T_i} \left(1 - \Phi(\overline{u}_{it}^3 + \epsilon_i^3)\right)^{1-u_{it}} \Phi(\overline{u}_{it}^3 + \epsilon_i^3)^{u_{it}} \Big] f(\epsilon_i^3) d\epsilon_i^3. \tag{A-8}$$

We use Monte Carlo simulation techniques to calculate the joint likelihood of observing each customer's usage stream and interruption outcomes. We then find parameters that maximize the aggregate log-likelihood across customers,

$$L = \sum_{i=1,\ldots,N} \ln L_i. \tag{A-9}$$

3 Robustness Checks

In this section, we discuss an alternative approach to modeling the interaction between different stages of the adoption process on the ultimate adoption outcome and present the results from two robustness checks for our main empirical specification. First, we consider an alternative specification for the decay in delays over time. Second, we examine the implications of restricting attention to customers who make at least one online transaction and the possibility of sample selection effects that results.

3.1 Alternative Modeling Approach

Our model considers the effect of a discrete interruption in the adoption process on usage. An alternative would be to consider the effect of the length of an interruption on usage. Given the data at hand, it is difficult to specify a model that treats delays as continuous endogenous variables. In the adoption process we consider, each customer goes through each stage of the adoption process only once. As a result, a simultaneous-equations hazard specification for the time a customer spends in each stage of the adoption funnel is not viable: multiple observations for each customer's adoption decision in a given stage are necessary to identify the individual-level heterogeneity typically employed in simultaneous hazard models such as Lillard (1993). In addition, our instruments of vacation and public holidays are strongly correlated with whether someone is interrupted in the adoption process, but they do not predict the length of this interruption well.

3.2 Linear Decay Specification

Table A-1 specifies a linear decay of the effect of interruptions on regular usage. Here, we interact the incidence of delays with the number of days a customer has spent in the regular usage stage. Similar to our main specification, the effect of interruptions is significantly negative, but declines in magnitude in the time since the customer's first arrival in the regular usage stage. A likelihood ratio test indicates, however, that our main specification explains the decision to make transactions online significantly better than this alternative.

Table A-1: Linear Specification for Decay of Interruptions

	System-of-Equations FIML Model with Linear Decay Interaction	
	Coef.	Std. Err.
Regular Usage		
Interruption bef. Login	-0.394	0.053***
Interruption bef. Login × Days since 1st Transaction	0.002	1.5E-4***
Interruption bef. Transaction	-1.088	0.059***
Interruption bef. Transaction × Days since 1st Transaction	0.002	1.6E-4***
Days since 1st Transaction	3.7E-4	1.4E-4***
Age	0.038	0.007***
Age Squared	-0.412	0.078***
Male	0.070	0.028**
Brokerage	-0.002	0.031
Bank Branches	0.187	0.056***
Public Holidays (t)	-0.003	0.054
Vacation Days (t)	-0.003	0.002*
No. Offline Transactions	0.012	3.2E-4***
Transaction Decision ($t-1$)	0.924	0.017***
Interruption bef. Login		
Age	-0.008	0.013
Age Squared	0.078	0.151
Male	-0.140	0.056**
Brokerage	0.054	0.064
Bank Branches	0.176	0.124
Public Holidays Signup Month	0.321	0.067***
Vacation Days Signup Month	0.019	0.007***
Interruption bef. Transaction		
Age	-0.011	0.013
Age Squared	0.149	0.152
Male	0.243	0.058***
Brokerage	-0.051	0.066
Bank Branches	0.210	0.128
Public Holidays Login Month	-0.236	0.057***
σ_3	0.625	0.018***
ρ_{21}	-0.044	0.065
ρ_{31}	0.232	0.050***
ρ_{32}	-0.406	0.047***
Log-Likelihood	-16,333.30	
Observations	32,206	

* p <0.10, ** p <0.05, *** p <0.01.
Estimation: Trivariate normal specification for individual unobserved effects across equations, with a random effect in the equation for the decision to use the technology.
Sample: 2,130 customers who made at least one online transaction during the 23-month sample period.

3.3 Customers 'Stuck in the Funnel'

Table 4 only captures the impact of interruptions on subsequent transactions for customers who actually did make a transaction. It does not include the 392 customers who never made an online transaction, but experimented with online banking. In parallel to Table 1, we display summary statistics for these customers in Table A-2. At 37.5 years old, these customers are slightly older than the customers who made at least one online transaction. At 85%, they are also far more likely to have delayed their first login.

Table A-2: Summary Statistics for Customers 'Stuck in the Funnel'

	Mean	Std Dev	Min	Max
Interruption bef. Login	0.85	0.36	0	1
Months between signup and login	0.58	2.10	0	24
Age	37.50	13.00	16	99
Age squared / 1000	1.58	1.14	0.26	9.80
Male	0.51	0.50	0	1
Brokerage account	0.30	0.46	0	1
Branches in Zip	0.94	0.24	0	1
Avg. no. offline transactions / month	9.38	12.10	0	98

Selected sample of 392 customers who logged in but did not make an online transaction during the 23-month sample period. Average number of offline transactions calculated over the full sample period.

We investigate whether such selection effects affect our results by extending our analysis to those customers who never made an online transaction. Table A-3 tracks how their propensity to log in to the bank's website was affected by an initial interruption before the first login. We use a similar strategy of instrumenting for this initial interruption with public holidays and vacations as in Table 4. However, our specification simplifies since we now consider only two stages. Model (1) reports results for a random effects panel model that does not adjust for the potential endogeneity of interruptions. Model (2) reports results of a bivariate probit where Int_Login_i is the interruption before a first login and U_{it} indicates whether or not the customer logged in during the months following the initial login. The results across Models (1)-(2) in Table A-3 echo those in Tables 4 and A-1. We find that the interruption before first login has a negative effect on the propensity to log in again in future months. Again, this effect diminishes over time. This suggests that for those customers who were 'stuck in the funnel', the effects of an interruption were negative, as in our main sample. Hence, our results are not an artifact of not observing their actual outcomes.

Table A-3: Investigating Selection: The Effect of Interruptions on Logging into the Website

	Rnd Eff Probit Model		System-of-Equations FIML Model	
	Coef.	Std. Err.	Coef.	Std. Err.
	(1)		(2)	
Regular Login Decision				
Interruption bef. Login	-1.086	0.173***	-1.705	0.192***
Interruption bef. Login × Months since 1st Login				
Months 4-6	0.513	0.160***	0.479	0.164***
Months 7-9	0.531	0.177***	0.461	0.185**
Months 10+	0.731	0.164***	0.599	0.146***
Months 4-6	-0.684	0.675	-0.510	0.522
Months 7-9	-0.953	0.678	-0.762	0.527
Months 10+	-1.232	0.677*	-1.011	0.513**
Age	0.008	0.029	0.008	0.022
Age Squared	0.095	0.326	0.032	0.246
Male	0.284	0.148*	0.298	0.103***
Brokerage	-0.258	0.156*	-0.338	0.107***
Bank Branches	-0.360	0.265	-0.255	0.188
Public Holidays (t)	0.033	0.089	0.073	0.074
Login Decision ($t-1$)	0.484	0.067***	0.537	0.061***
Interruption bef. Login				
Age			-0.010	0.025
Age Squared			0.052	0.274
Male			-0.035	0.133
Brokerage			0.103	0.142
Bank Branches			-0.590	0.312*
Public Holidays Signup Month			0.421	0.181**
Vacation Days Signup Month			0.061	0.016***
σ_2	1.195	0.083***	1.185	0.076***
ρ_{21}			-0.404	0.093***
Log-Likelihood	-1,649.0		-1,893.9	
Observations	4,247		4,620	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Estimation: Model (1) is a random effects specification of the customer's decision to log into the website in each month in the months following initial login. Model (2) assumes a bivariate normal distribution for individual unobserved effects across the two equations, allowing for a random effect in the panel equation for logins.
Sample: 373 customers who signed up and logged in but did not conduct an online transaction.

References

Lillard, L. A. (1993). Simultaenous equations for hazards: Marriage duration and fertility timing. *Journal of Econometrics 56*, 187–217.