Successive sample selection and its relevance for management decisions

Web-Appendix

Stephan Wachtel

Thomas Otter

Goethe University, Frankfurt, Germany

Web-Appendix 1: Illustration of MCMC performance using simulated data

To illustrate the performance of our MCMC algorithm we report results from the analyses of two simulated data sets. The first simulated data set features three successive selection stages without unobserved dependence, eight observed covariates in total and 5,000 (multivariate) observations. We analyze this data set with our prior formulated over all possible exclusion restrictions to show that we recover the special case of independence across selection stages correctly. We refer to our proposed prior as the adaptively parsimonious (AP) model. We re-analyze the same data set estimating a saturated model to illustrate the practical problems due to weak empirical identification. The saturated model includes all observed covariates on all stages and employs a standard weakly informative inverse Wishart (IW) prior for the covariance matrix Σ.

Table W1.1 shows that the posterior excludes coefficients that connect stage specific unobservables with following stages most of the time (see column three). Marginally, these coefficients are estimated to be essentially equal to zero (see column four), correctly reflecting the independence that generated the data.

Table W1.1 Data generating λ coefficients (True), posterior inclusion probabilities, and posterior means from the adaptively parsimonious model (AP), posterior standard deviations in parentheses

	True	inclusion prob	posterior means (standard devations)
λ_{21}	0	.211	-0.005 (.069)
λ_{31}	0	.298	0.038 (.103)
λ_{32}	0	.400	-0.010 (.130)

Table W1.2 shows that exclusions (inclusions) of observed covariates on the various stages are almost perfectly recovered.

Table W1.2 True inclusions (1) and exclusions (0) and posterior means of inclusion probabilities from the adaptively parsimonious model (AP)

	Stage 1		Stage 2		Stage 3	
	True	incl. prob	True	incl. prob	True	incl. prob
X_1	1	1.000	0	0.059	0	0.092
X_2	1	1.000	0	0.037	0	0.183
X_3	0	0.024	1	1.000	0	0.059
X_4	0	0.033	1	1.000	0	0.064
X_5	0	0.077	1	0.795	0	0.079
X_6	1	1.000	1	1.000	1	1.000
X_7	0	0.056	0	0.040	1	1.000
X_8	0	0.036	0	0.032	1	1.000

The comparison between the saturated model (S) and the AP-model based on our prior over all possible sets of exclusions reveals that the full model incorrectly infers a highly statistically credible, negative correlation between stage 1 and 3 (see column three of Table W1.3). In contrast, the AP model recovers the independence across selection stages nicely (see column four of Table W1.3) as implied by the results presented in Table W1.1.

Table W1.3 Data generating correlations between selection stages (True) and posterior means from the full model (F) and the adaptively parsimonious model (AP), posterior standard deviations in parentheses

	True	F	AP
ρ21	0	.034 (.214)	-.005 (.069)
ρ31	0	-.962 (.042)	.038 (.103)
ρ32	0	.075 (.201)	-.010 (.130)

Figures W1.1a and W1.1b exhibit trace-plots of implied correlations from the AP model and the saturated (S) model respectively. We ran each sampler for 500,000 iterations and include every hundredth draw in these plots. The posterior mode explored by the saturated model can be traced back to properties of the corresponding multivariate likelihood and the shape of the weakly informative IW prior. The multivariate likelihood increases without bounds as the covariance matrix Σ and the implied correlation matrix approach rank deficiency. Therefore the problem with spuriously extreme correlations in these models is genuine and not just a feature of a specific type of

Bayesian analysis. However, so called weakly informative IW priors do not help here as they are concentrated on covariance matrices exhibiting extreme dependencies by construction (Rossi et al. 2005).

Figure W1.1a Trace-plots of correlations AP model

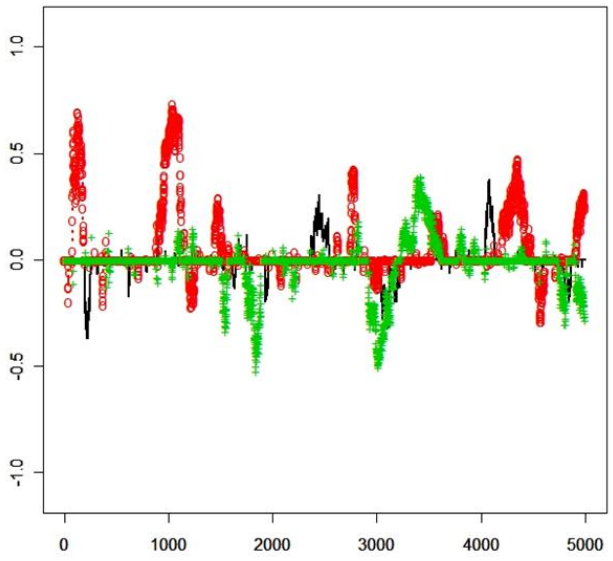
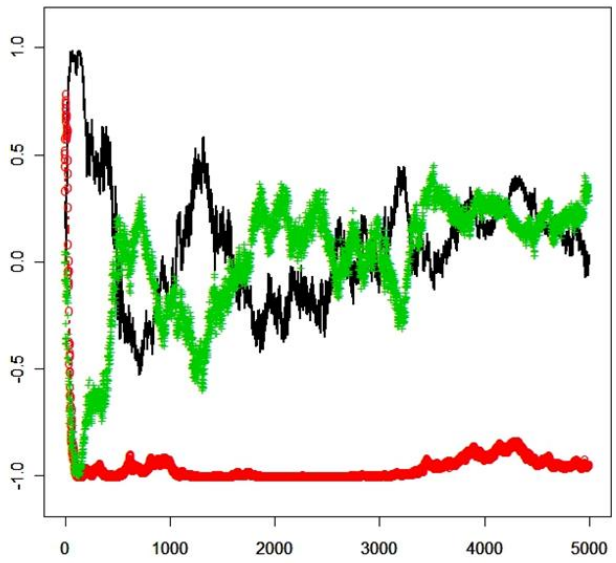


Figure W1.1b Trace-plots of correlations F model



As a consequence, the full model fails to recover the structural effects of observed covariates on stage 3 (Table W1.4). For example, the full model overestimates the constant and spuriously suggests that X_2 and X_5 should be included on stage 3 (see column 9 in Table W1.4). However, there is no causal ordering between the bias in the unobserved dependence structure and the bias in the estimated regression coefficients. In a sense, the bias in the unobserved dependence structure occurs precisely because all observed covariates are entered on all stages by the full model.

Table W1.4 Data generating coefficients (True) and posterior means from the full model (F) and the adaptively parsimonious model (AP), posterior standard deviations in parentheses

	Stage 1			Stage 2			Stage 3		
	True	**F**	**AP**	**True**	**F**	**AP**	**True**	**F**	**AP**
Constant	.032	.052 (.073)	.020 (.047)	.024	.006 (.116)	.018 (.086)	.032	**.614** **(.198)**	.079 (.123)
X_1	-.063	**-.056** **(.015)**	**-.053** **(.014)**	0	-.018 (.022)	-.001 (.005)	0	.020 (.019)	-.002 (.009)
X_2	.126	**.126** **(.015)**	**.130** **(.014)**	0	-.004 (.019)	-.000 (.004)	0	**-.079** **(.020)**	-.006 (.016)
X_3	0	.015 (.014)	.000 (.002)	-.097	**-.066** **(.018)**	**-.069** **(.018)**	0	-.016 (.020)	-.000 (.006)
X_4	0	.016 (.015)	.000 (.003)	.073	**.071** **(.019)**	**.083** **(.020)**	0	-.028 (.021)	-.001 (.006)
X_5	0	-.026 (.015)	-.002 (.007)	.049	**.052** **(.021)**	**.044** **(.027)**	0	**.048** **(.020)**	.002 (.011)
X_6	.063	**.058** **(.014)**	**.060** **(.014)**	.121	**.092** **(.018)**	**.096** **(.019)**	.126	**.059** **(.024)**	**.114** **(.024)**
X_7	0	.005 (.015)	.001 (.005)	0	.019 (.018)	.001 (.004)	-.158	**-.122** **(.019)**	**-.188** **(.023)**
X_8	0	-.009 (.013)	-.000 (.003)	0	-.001 (.017)	.000 (.003)	.095	**.066** **(.018)**	**.101** **(.022)**

Our second simulated data set illustrates how our MCMC algorithm recovers the finer structure in the unobserved dependence among successive selections. The data set features three successive selection stages with structured unobserved dependence, eight observed covariates in total and 10,000 (multivariate) observations. The structure of the unobserved dependence between selection stages is such that unobservables only emerging at stage two do not directly influence the outcome on stage three, i.e. $\lambda_{32} = 0$ (see the column 'True' in Table W1.5). That is, all dependence between stage two and stage three comes through the joint influence of unobservables emerging at stage one. These unobservables have a positive effect on the success on stage two but a negative one on stage three giving rise to a negative correlation between stages two and three, $\rho_{32} = -.210$ (see Table W1.7).

Tables W1.5 through W1.8 demonstrate that the MCMC implementation of our prior over all possible exclusion restrictions nicely recovers both the data generating model structure as well as the data generating parameter values.

Table W1.5 Data generating λ coefficients (True), posterior inclusion probabilities, and posterior means from the adaptively parsimonious model, posterior standard deviations in parentheses

	True	inclusion prob	posterior means (standard devations)
λ21	0.696	1.000	0.612 (.056)
λ31	-0.302	.829	-0.284 (.167)
λ32	0	.228	-0.047 (.112)

Table W1.6 True inclusions (1) and exclusions (0) and posterior means of inclusion probabilities from the adaptively parsimonious model

	Stage 1		Stage 2		Stage 3	
	True	incl. prob	True	incl. prob	True	incl. prob
X_1	1	1.000	0	0.055	0	0.174
X_2	1	1.000	0	0.029	0	0.062
X_3	0	0.016	1	1.000	0	0.088
X_4	0	0.021	1	1.000	0	0.054
X_5	0	0.015	1	1.000	0	0.063
X_6	1	1.000	1	1.000	1	1.000
X_7	0	0.055	0	0.093	1	1.000
X_8	0	0.029	0	0.270	1	1.000

Table W1.7 Data generating correlations between selection stages (True) and posterior means from the adaptively parsimonious model, posterior standard deviations in parentheses

	True	Estimates
ρ21	.661	.611 (.056)
ρ31	-.286	-.284 (.166)
ρ32	-.210	-.137 (.111)

Table W1.8 Data generating coefficients (True) and posterior means from the adaptively parsimonious model, posterior standard deviations in parentheses

	Stage 1		Stage 2		Stage 3	
	True	Estimates	True	Estimates	True	Estimates
Constant	.253	.216 (.028)	.139	.174 (.079)	.241	.229 (.118)
X_1	-.313	-.309 (.011)	0	.001 (.005)	0	-.007 (.018)
X_2	.196	.211 (.011)	0	.000 (.002)	0	-.001 (.005)
X_3	0	.000 (.002)	-.132	-.132 (.014)	0	.002 (.007)
X_4	0	-.000 (.002)	.134	.157 (.013)	0	.001 (.005)
X_5	0	.000 (.001)	.157	.149 (.014)	0	.001 (.005)
X_6	.285	.288 (.011)	.122	.112 (.019)	.217	.225 (.030)
X_7	0	.001 (.004)	0	-.002 (.008)	-.226	-.225 (.015)
X_8	0	-.000 (.003)	0	.008 (.015)	.271	.270 (.015)

Web-Appendix 2: Identification in Selection Models

We discuss identification of regression coefficients and unobserved dependence from the joint distribution of y_1 and y_2 using the example from Figure 1. Here we assume that we know a priori that x_1 affects y_1 and x_2 affects y_2, i.e. we do not have to select variables, and show how exclusion restrictions identify the model.

Table W2.1 Covariate patterns

Pattern	$\mathbf{x_1}$	$\mathbf{x_2}$
A	1	1
B	1	-1
C	-1	1
D	-1	-1

Table W2.1 collects all possible covariate patterns for x_1 and x_2 assuming that both covariates take only two different values. The general case of continuous covariates is messier in terms of notation but in essence identical. The covariates affecting y_1 and y_2 are perfectly orthogonal. Assuming non-zero coefficients, we see that the unobserved dependence between selection stages (λ_{21})[1] is identified through differences in probabilities between response patterns with the same x_2 but different x_1 from the following set of inequalities:

$$
\begin{aligned}
p\left(y_2 \left| \overbrace{x_1 = 1, x_2 = 1}^{A}, y_1 = 1 \right. \right) &\neq p\left(y_2 \left| \overbrace{x_1 = -1, x_2 = 1}^{C}, y_1 = 1 \right. \right) \\
p\left(y_2 \left| \overbrace{x_1 = 1, x_2 = -1}^{B}, y_1 = 1 \right. \right) &\neq p\left(y_2 \left| \overbrace{x_1 = -1, x_2 = -1}^{D}, y_1 = 1 \right. \right)
\end{aligned}
\qquad \text{(W2-1)}
$$

[1] As explained in the paper, parameters {**β**} and {λ} and are not jointly likelihood identified from choice data. We project down to the likelihood identified space by normalizing the variances to 1. Our discussion here implicitly refers to the identified transformations of {**β**} and {λ}.

The regression coefficient on the second stage, β_2, is identified through differences in probabilities among response patterns with the same x_1 but different x_2 values from the following set of inequalities:

$$p\left(y_2 \left| \overbrace{x_1 = 1, x_2 = 1}^{A}, y_1 = 1 \right. \right) \neq p\left(y_2 \left| \overbrace{x_1 = 1, x_2 = -1}^{B}, y_1 = 1 \right. \right)$$
$$p\left(y_2 \left| \overbrace{x_1 = -1, x_2 = 1}^{C}, y_1 = 1 \right. \right) \neq p\left(y_2 \left| \overbrace{x_1 = -1, x_2 = -1}^{D}, y_1 = 1 \right. \right) \qquad \text{(W2-2)}$$

Note that we have two independent inequalities to identify the unobserved dependence and two more independent inequalities to identify β_2. Identification of the regression coefficient on the first stage, β_1, is standard and simply requires sufficient variation in x_1.

Deleting response patterns from Table W2.1 introduces covariance between x_1 and x_2. To jointly identify the regression coefficient β_2 and the unobserved dependence, we need at least two independent inequalities. Deleting pattern A leaves the inequality between B and D, and between C and D to identify dependence and β_2, respectively, similarly for deleting patterns B, C, or D, individually. However, if we delete either B and C, or A and D, we are left with no inequalities to independently identify the two parameters in this example. Note that deleting B and C (A and D) results in perfectly positively (negatively) correlated x_1 and x_2. Finally, jointly deleting A and B or C and D leaves β_1 unidentified. Generalizing this result to multiple variables in $\mathbf{x_1}$ and $\mathbf{x_2}$, we rely on inequalities of the form

$$p\left(y_2 \left| \mathbf{x}'_{1,i}\boldsymbol{\beta}_1 = v_1, \mathbf{x}'_{2,i}\boldsymbol{\beta}_2 = v_2, y_1 = 1 \right. \right) \neq p\left(y_2 \left| \mathbf{x}'_{1,j}\boldsymbol{\beta}_1 = v_1^*, \mathbf{x}'_{2,j}\boldsymbol{\beta}_2 = v_2, y_1 = 1 \right. \right),$$

i.e. observations with the same value $\mathbf{x'}_{2,i}\boldsymbol{\beta}_2 = \mathbf{x'}_{2,j}\boldsymbol{\beta}_2 = v_2$ on the second stage but different histories v_1 and v_1^*, to identify unobserved dependence between selection stages.

Exclusion restrictions help jointly identify regression coefficients and the unobserved dependence structure. Exclusion restrictions result in a set of variables affecting y_1 that is at least partially different from the set of variables affecting y_2, precluding the problematic case of perfect correlation between $\mathbf{x}_1'\boldsymbol{\beta}_1$ and $\mathbf{x}_2'\boldsymbol{\beta}_2$. Specifically, strong identification requires that $\mathbf{x}_1$ contains at least one active variable that is excluded from $\mathbf{x}_2$. Independent variation in this variable, holding $\mathbf{x}_2$ constant, identifies that nature of unobserved dependence between stages.[2]

With multiple explanatory variables, the case where the set of variables in $\mathbf{x}_1$ is identical to the set of variables in $\mathbf{x}_2$ is identified conditional on parametric assumptions, but requires large amounts of data for empirical identification. The identification status with multiple independent variables, identical across different stages, depends on the pattern in the coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The more (less) orthogonal the coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the more (less) independent information about $\boldsymbol{\beta}_2$ and unobserved dependence between selection stages is available. For a similar result in the context of the correlated multinomial probit model, but obtained using a different approach, see Keane (1992) and Zeithammer and Lenk (2009).

When there are more than two successive selections, we need to identify dependence between successive selections as well as those between current stages and stages before the immediately preceding stage. For example, with three successive selections we have λ_{21}, linking the first to the second stage, λ_{32}, linking the second to the third, and finally λ_{31} linking the first and the third stage. The identification of λ_{32} is analogue to that of λ_{21}. Identification of λ_{31} comes through inequalities of the following form:

$$p\left(y_3 \middle| \mathbf{x}_{1,i}'\boldsymbol{\beta}_1 = v_1, \mathbf{x}_{2,i}'\boldsymbol{\beta}_2 = v_2, \mathbf{x}_{3,i}'\boldsymbol{\beta}_3 = v_3, y_1 = 1, y_2 = 1\right) \neq$$

$$p\left(y_3 \middle| \mathbf{x}_{1,j}'\boldsymbol{\beta}_1 = v_1^*, \mathbf{x}_{2,j}'\boldsymbol{\beta}_2 = v_2, \mathbf{x}_{3,j}'\boldsymbol{\beta}_3 = v_3, y_1 = 1, y_2 = 1\right).$$

[2] Independent variation in the 'future', i.e. $\mathbf{x}_2$ holding fixed the 'past' $\mathbf{x}_1$ only identifies functional forms on the second stage.

That is observations with the same value $\mathbf{x'}_{3,i}\boldsymbol{\beta}_3 = \mathbf{x'}_{3,j}\boldsymbol{\beta}_3 = \nu_3$ on the third, the same history $\mathbf{x'}_{2,i}\boldsymbol{\beta}_2 = \mathbf{x'}_{2,j}\boldsymbol{\beta}_2 = \nu_2$ on the second but *different* histories ν_1 and ν_1^* on the first stage identify λ_{31}.

If a particular data set poses structural identification problems such as in a two stage model with only one observed covariate where $x_1 = x_2$, for example, the posterior obtained by updating our prior specified over all possible exclusion restrictions with this data is purely driven by the subjective prior choices for $p(\alpha)$ and $p(\{\boldsymbol{\beta}\}, \{\lambda\})$. However, if the data only weakly identify some models, notably the saturated model entering all covariates on all stages while estimating unstructured dependence, updating our prior with this data extracts all information about the likely model structure, including information from weakly identified models.

Web-Appendix 3: Numerical Illustration of DAG results

We numerically illustrate the theoretical results from Section 3 in the paper comparing dependent analysis to independent analysis in a simple two stage setup under different degrees of unobserved dependence. We also include a comparison to an approach popular in practice that collapses the multi staged process to one stage when measuring probabilities to pass all stages. This 'collapsing-approach' ignores all stages but the final, often managerially most important stage, regressing success on the final stage on covariates describing the population of consumers initially entering the process. Failures on any stage prior to the last stage are treated equivalently to a failure on the last stage, i.e.

$$\left[\left(y_1 = 0\right) \vee \ldots \vee \left(y_{last-1} = 0\right)\right] \Rightarrow y_{last} = 0 \qquad \text{(W3-1)}$$

This is different from the independent analysis of the last stage which only considers consumers that have successfully passed all selections prior to the last.

We generate observations from a model with two stages and three independently generated covariates overall where x_1 affects y_1, x_2 affects y_2, x_{12} affects both y_1 and y_2, and finally y_1 is a selection criterion (Equation W3-2). We investigate four data generating mechanisms characterized by different degrees of (unobserved) dependence between the first stage and the second stage varying the correlation parameter ρ in Equation W3-2 over the grid [0, .1, .5, .9].[3] We generate 10,000 observations from each data generating mechanism and investigate the performance of ignoring dependence and the collapsing-approach relative to the data generating model in terms of scoring, targeting and influencing. The data generating mechanism with $\rho = 0$ corresponds to the independence model.

$$y_{1,i} = \begin{cases} 0 & z_{1,i} < 0 \\ 1 & z_{1,i} \geq 0 \end{cases} \qquad y_{2,i} = \begin{cases} 0 & \left(z_{2,i} < 0\right) \vee \left(y_{1,i} = 0\right) \\ 1 & \left(z_{2,i} \geq 0\right) \wedge \left(y_{1,i} = 1\right) \end{cases}$$

$$\begin{pmatrix} z_{1,i} \\ z_{2,i} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,i} & x_{12,i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_{2,i} & x_{12,i} \end{pmatrix} \begin{pmatrix} -.4 \\ .5 \\ .4 \\ -.1 \\ .6 \\ .4 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix}, \begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \qquad \text{(W3-2)}$$

$$x_{1,i} \sim N(0, 2)$$
$$x_{2,i} \sim N(0, 2)$$
$$x_{12,i} \sim N(0, 2)$$

Table W3.1 summarizes the inference for regression coefficients from the independence model, the collapsing-approach and the dependence model relative to the data generating values. The first column indicates what stage is analyzed. For the second stage, we report two versions of the independence model. One, 'the independence model with causal knowledge' conditions on

[3] We use the marginal representation and not the Wold decomposition here for ease of presentation in the special case of only two successive stages.

knowing which observed variables cause probabilities on stage 1 (x_1 and x_{12}, but not x_2) and on stage 2 (x_2 and x_{12} but not x_1) (see Equation W3-2). The other, 'the independence model with variable selection' does not use this prior knowledge and relies on variable selection instead. On the first stage we do not make this distinction because the difference between using prior causal knowledge and variable selection is only a matter of efficiency before any relevant selection. The second column in Table W3.1 lists the covariates, the third column the data generating coefficients and the following columns the estimates for different levels of unobserved correlation.

As expected, we obtain consistent inferences from the independent analysis of the first stage regardless of the correlation with subsequent stages. On the second stage, i.e. after selection, we observe an increasingly positive bias in the constant as the correlation between the stages increases. This is because of the positive expectation of ε_2 conditional on selection and a positive error correlation. We also see an increasing negative bias in the coefficient for x_{12}, and when we select variables, x_1 appears to have a significant negative effect on the second stage for correlations equal to .5 and .9. The reason is that consumers with large x_1 or x_{12} succeed on the first stage even with, relative to other consumers, larger negative draws of ε_1. Negative draws of ε_1 increase the likelihood of negative ε_2, given the positive error correlation. As a result, we obtain a downward bias in the effect of x_{12} and the spurious negative effect of x_1 on the second stage from the independence model.

Table W3.1 Estimated coefficients on the first and the second stage (standard errors)

	Covariate	True Values	ρ = 0	ρ = .1	ρ = .5	ρ = .9
Independence models						
1st stage	**Constant**	**-0.4**	-0.41 (.02)	-0.39 (.02)	-0.41 (.02)	-0.41 (.02)
	X1	**0.5**	0.51 (.01)	0.50 (.01)	0.51 (.01)	0.49 (.01)
	X2	**0**	0	0	0	0
	X12	**0.4**	0.42 (.01)	0.39 (.01)	0.41 (.01)	0.39 (.01)
… with causal knowledge						
2nd stage	**Constant**	**-0.1**	-0.12 (.03)	-0.02 (.03)	0.18 (.03)	0.53 (.03)
	X1	**0**	0	0	0	0
	X2	**0.6**	0.61 (.02)	0.59 (.02)	0.61 (.02)	0.59 (.02)
	X12	**0.4**	0.39 (.02)	0.37 (.02)	0.37 (.02)	0.28 (.02)
… with variable selection						
2nd stage	**Constant**	**-0.1**	-0.13 (.04)	-0.01 (.03)	0.37 (.04)	0.90 (.04)
	X1	**0**	0.006 (.02)	-0.009 (.02)	-0.13 (.02)	-0.22 (.02)
	X2	**0.6**	0.61 (.02)	0.59 (.02)	0.62 (.02)	0.64 (.02)
	X12	**0.4**	0.39 (.02)	0.37 (.02)	0.34 (.02)	0.26 (.02)
'collapsing-approach'						
stages collapsed	**Constant**	**NA**	-1.21 (.02)	-1.14 (.02)	-1.03 (.02)	-0.84 (.02)
	X1	**NA**	0.32 (.01)	0.31 (.01)	0.29 (.01)	0.27 (.01)
	X2	**NA**	0.31 (.01)	0.29 (.01)	0.26 (.01)	0.22 (.01)
	X12	**NA**	0.45 (.01)	0.42 (.01)	0.41 (.01)	0.37 (.01)
Dependence model						
1st stage	**Constant**	**-0.4**	-0.41 (.02)	-0.39 (.02)	-0.41 (.02)	-0.41 (.02)
	X1	**0.5**	0.51 (.01)	0.50 (.01)	0.51 (.01)	0.49 (.01)
	X2	**0**	0	0	0	0
	X12	**0.4**	0.42 (.01)	0.39 (.01)	0.41 (.01)	0.39 (.01)
2nd stage	**Constant**	**-0.1**	-0.09 (.05)	-0.04 (.05)	-0.18 (.05)	-0.12 (.04)
	X1	**0**	0	0	0	0
	X2	**0.6**	0.61 (.02)	0.60 (.02)	0.61 (.02)	0.57 (.02)
	X12	**0.4**	0.38 (.02)	0.37 (.02)	0.43 (.02)	0.40 (.01)

We obtain consistent inferences for the effect of x_2 on the second stage, independent from the size of the correlation and from whether we select variables or not. This is because covariates are generated independently in this simulation. Biases in the coefficients of causally effective variables on the second stage only occur when these variables are correlated with the selection process. However, the (unobserved) variance of ε_2 conditional on selection is in general smaller than its unconditional counterpart. Rescaling requires knowledge of the data generating ε_2 draws, which we have in this simulation. In practice, without this knowledge, consistency implies that coefficient ratios of variables that are independent from the selection process can be estimated consistently from selected samples.

Next, Table W3.1 reports parameter estimates from the collapsing-approach. Because the collapsing-approach ignores the nonlinearity due to the first stage selection, there exists no general functional relationship between the data generating parameters and the coefficients reported. As the correlation between stages increases, the constant increases and all slope coefficients decrease. Finally, Table W3.1 reports the coefficient estimates from the dependence model that recovers all coefficients for all levels of unobserved dependence up to posterior uncertainty, as it should.

Independent analyses suggest that consumers high on x_1 are more likely to pass the first stage but less likely to pass the second stage when the correlation between stages is larger or equal to .5. Consequently, a decision maker interested in targeting consumers based on covariates will be conflicted regarding which levels of x_1 to target. Numerically solving for the x_1 level that maximizes the probability that the observed population (as characterized by the joint distribution of observed covariates in Equation W3-2) succeeds on stages 1 and 2 results in optimal x_1 targets of 6.85, 3.93 and 3.34 for correlation of .1, .5 and .9 respectively. These target values are by factors of 1.4, 2.2, and 2.5 respectively smaller than the optimal x_1 target implied by dependent analysis. Of course, dependent analysis in this case leads to targeting the maximal x_1 values in the population.

In similarly selected samples of consumers on the second stage, the spurious negative effect of x_1 could be exploited for scoring purposes. However, marketing actions directed at lowering the levels of x_1 on the second stage will result in a null effect because x_1 is not causally effective on this stage.

Table W3.2 summarizes the differences between the data generating probabilities for passing both stages and those approximated by the collapsing-model, independent analyses, and finally those estimated by the dependence model.

Table W3.2 Distribution of (absolute) deviations from data generating probabilities (n=10,000).

		ρ = 0	**ρ = .1**	**ρ = .5**	**ρ = .9**
Mean Absolute Deviation	**Collapsing**	.043	.044	.057	.073
	Independent w. causal knowledge	.004	.005	.016	.032
	Independent w. variable selection	.004	.005	.008	.015
	Dependence Model	.004	.005	.006	.007
[10%, 90%] (Deviation)	**Collapsing**	[-.07, .08]	[-.07, .08]	[-.09, .11]	[-.12, .13]
	Independent w. causal knowledge	[-.004, .008]	[-.009, .009]	[-.02, .03]	[-.05, .05]
	Independent w. variable selection	[-.004, .008]	[-.008, .008]	[-.003, .01]	[-.02, .02]
	Dependence Model	[-.005, .008]	[-.008, .008]	[-.0004, .02]	[-.004 .02]
Stddev(Deviation)	**Collapsing**	.07	.07	.09	.10
	Independent w. causal knowledge	.006	.008	.02	.05
	Independent w. variable selection	.006	.007	.01	.02
	Dependence Model	.006	.007	.007	.009

The discrepancies between the data generating and the approximated probabilities increase as the unobserved dependence between selection stages increases. The collapsing-approach approximates the data generating probabilities much worse than independent analyses of the selection stages across all correlations investigated. The corresponding mean absolute deviation

between estimated and data generating probabilities is large. The 10^{th}-percentile and the 90^{th}-percentile of the deviations between estimated and data generating probabilities are most extreme and the standard deviation of the distribution of approximation errors is largest. This demonstrates that the collapsing-approach is not a viable approximation to the (stochastic) non-linearity introduced by selection, even without unobserved dependence across selection stages.

Comparing independent analyses with causal knowledge to independent analyses relying on variable selection nicely illustrates how the spurious effect of x_1 on the second stage proxies for the unobserved heterogeneity creating dependence between the selection stages. For example, when the correlation is equal to .9 the mean absolute deviation decreases from 3.2% to 1.5% when x_1 is included on the second stage.

Table W3.3 compares the lists of the 1,000, 2,000 and 5,000 top ranked consumers, i.e. rankings of $P(y_{1,i} = 1, y_{2,i} = 1 \mid \mathbf{x}_i, \text{Model})$ from the data generating probabilities to the collapsing-model, independent analyses, and finally to the estimated dependence model in terms of list overlap. 100% overlap indicates that a model includes the very same consumers as the data generating model in the respective list. Zero percent overlap indicates that a model and the data generating model identify completely different lists of consumers.

The collapsing-model performs worst across all list lengths and correlations in terms of overlap with the lists from the data generating model. Moving from left to right in Table W3.3 the correlation between the selection stages increases and this decreases the overlap between estimated lists and the list from the data generating probabilities, except when the data generating dependence model is estimated.

The independence model with variable selection outperforms that using prior causal knowledge that x_1 is ineffective on the second stage when the correlation becomes large, i.e. .5 or .9.

Relative to the expected overlaps of 10%, 20% and 50% from random lists[4], all models add more information in the tails of the consumer distribution. Overall, the close performance of independent analyses when spurious variables are included, i.e. without causal knowledge, compared to estimates from the data generating dependence model is remarkable when the goal is to simply score, with no direct preferences or utility over covariates that describe consumers.

Table W3.3 Consumer scoring, n = 10,000

List	Model	ρ = 0	ρ =.1	ρ =.5	ρ =.9
TOP 1000	**Collapsing selections**	82.4 %	84.4 %	80.7 %	76.6 %
	Independent w. causal knowledge	98.3 %	98.4 %	95.9 %	87.7 %
	Independent w. variable selection	98.8 %	98.5 %	97.9 %	95.9 %
	Dependence Model	98.6 %	98.5 %	98.8 %	98.9 %
TOP 2000	**Collapsing selections**	87.4 %	87.6 %	84.3 %	82.1 %
	Independent w. causal knowledge	99.2 %	98.7 %	97.0 %	91.7 %
	Independent w. variable selection	99.1 %	99.0 %	99.0 %	96.3 %
	Dependence Model	99.2 %	98.9 %	99.2 %	99.0 %
TOP 5000	**Collapsing selections**	93.4 %	93.1 %	91.0 %	88.5 %
	Independent w. causal knowledge	99.5 %	99.2 %	97.6 %	95.2 %
	Independent w. variable selection	99.6 %	99.3 %	99.3 %	97.6 %
	Dependence Model	99.6 %	99.3 %	99.3 %	99.4 %

Customer scoring, i.e. scoring based on $P(y_{2,i} = 1 \mid \mathbf{x}_i, y_{1,i} = 1, \text{Model})$ is only defined for the data generating model and independent analyses. It is not defined in the collapsing-approach because the collapsing-approach glosses over all intermediate selection stages. Table W3.4 compares the lists of top ranked customers conditional on having passed the first selection, i.e. rankings of $P(y_{2,i} = 1 \mid \mathbf{x}_i, y_{1,i} = 1, \text{Model})$ from the data generating probabilities to the independent analysis with prior knowledge of causal effects, that using variable selection instead, and finally the

[4] With a sample size of 10,000 the expected overlap between fixed lists of length 1000, 2000, 5000 and randomly picked lists of the same length is 10%, 20%, 50%, respectively.

estimated dependence model. We use relative list lengths here because the number of customers on the second stage is random.

We find that increasing the correlation decreases the overlap for both independent analyses. Estimates from the dependence model perform stably, as they should. Ignoring causal knowledge and including spurious variables in the independent analysis helps significantly once correlations become larger. The differences between estimates from the data generating dependence model and independent analysis including spurious variables is larger when we score customers that have passed the first stage as in Table W3.4 compared to scoring consumers that are before the first selection as in Table W3.3. This is because the dependence model learns about unobserved heterogeneity through the realized error on stage one when stages are dependent.

Table W3.4 Customer scoring, customer-n (after the first selection) approximately equal to 4,000, consumer n = 10,000 before any selection

	Model	Rho = 0	Rho =.1	Rho =.5	Rho =.9
TOP 10 %	**Ind. w. causal knowledge**	99.0 %	97.8 %	86.8 %	63.5 %
	Ind. w. variable selection	98.5 %	98.3 %	95.5 %	81.0 %
	Dependence Model	98.5 %	98.0 %	98.5 %	97.5 %
TOP 25 %	**Ind. w. causal knowledge**	98.7 %	98.1 %	89.9 %	76.6 %
	Ind. w. variable selection	98.7 %	98.3 %	97.5 %	89.2 %
	Dependence Model	98.5 %	98.3 %	98.8 %	97.7 %
TOP 50 %	**Ind. w. causal knowledge**	99.2 %	98.5 %	95.3 %	88.2 %
	Ind. w. variable selection	99.2 %	98.8 %	98.2 %	95.4 %
	Dependence Model	99.2 %	98.7 %	99.2 %	99.3 %

We conclude from our illustrative simulation that independent analyses of selection stages will result in misleading targeting recommendations and influencing actions when unobserved causes connect the selection stages. Independent analyses with variable selection perform surprisingly well when the goal is a simple measurement of probabilities at the price of including spurious variables.

The collapsing-approach fails at this task because of the non-linearity of the selection process. This contrasts to consistent results obtained when mediators are ignored in linear systems. Thus, a firm may successfully score consumers, ignoring dependence between successive selections in a stable environment, claim only epsilon benefits from more advanced modeling, and be surprised by the failure of targeting and influencing decisions based on the same analysis.