universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Assessing type I and type II error probabilities for basic regression methods and an exact monotonicity test, in binary choice environments"

verfasst von / submitted by

Michel Geller Bakk.rer.soc.oec.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science

Wien, 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 913
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Volkswirtschaftslehre UG2002
Betreut von / Supervisor:	Universitäts Professor Karl H. Schlag, Phd

Contents

List of Figures

Chapter 1

Introduction

In experimental economics, small samples are a common trouble when doing statistical inference. Especially for samples below n=100, it is seldom clear how well regression methods perform, just consider Bellemare et al. (2014), to see the large spread in observation sizes for economic experiments. To enlighten this matter a bit, this thesis aims at uncovering small sample properties of 4 basic asymptotic regression methods, joined by a recently developed exact monotonicity test, the binary stratified direction of change test (Working paper Schlag 2016). The underlying models are chosen to be quite similar to an experimental economics setting. All explaining and explained variables are binary. Models are the two most recurrent binary models, a linear probability model accompanied by a logit binary choice model. Both models are simple and easy to combine with binary variables. Additionally, due to their simplicity and threshold behaviour, they find widespread usage in behavioral and experimental economics. When comparing ordinary least squares to White's heteroscedasticity robust least squares, a linear probability model provides a useful and simple test environment.

Analysis focuses on errors of the first kind, i.e. rejecting the null hypothesis when it is in fact true. The probability to commit a type I error is denoted by the size of a test. Type I errors may have devastating consequences, just think of a medicine which is supposed to cure a patient while it has no effect at all. Errors of the second kind, i.e. rejecting the alternative hypothesis, when it is true are also treated in this study, but to a lesser extend. While the probability to commit an error of the second kind is denoted by the Greek letter β, the complementary, i.e., the ability to not reject the alternative when it is true is called the power of a test. While the study is designed to uncover problems in falsely rejecting the nul hypothesis, the power values are corresponding to

this framework.

This paper can be seen as an extension of studies done by Gossner and Schlag in 2013, where they did a similar study using least squares regressors and an exact regression technique. Accordingly, in this study, simulations are run in order to investigate the respective power and size of ordinary least squares, covariance heteroscedasticity robust least squares, logit, probit and an exact monotonicity test. Due to the odds-nature of the results of loigit and probit, it is not possible to compare the estimate from linear regression to those of probit and logit. Hence such results are not included in this study.

In the next chapter, the underlying models are presented, followed by a chapter on the specifics of the different regression methods and tests. The fourth chapter revisits the parameters applied in the simulations, followed by an overview on the most interesting results. The study finishes with a conclusion. Extensive figures of the simulation results may be found in the Appendix.

Chapter 2

Models

The models selected for this study reflect three basic requirements. They are easy to implement and flexible enough to represent various sets of parameters. They are widely used in economics, and they are able to demonstrate the peculiarities of the applied regression techniques. The linear probability model as well as binary response model fulfil these requirements. The corresponding methods to both models find use in experimental and behavioral economics. For example in her famous 1999 paper on the effect of large stakes in the ultimatum game, Lisa Cameron uses both methods, but publishes the results of a linear probability model as it is more convenient to interpret.

2.1 Linear Probability Model

In order to make a fairly general statement on regression methods and their limitations, a basic and simple model is needed. A linear probability model (chapter 13.2 DiNardo et al. 2007) is simple to implement, and offers straightforward interpretation. The basic structure follows a linear model, except, that y_i follows a binary distribution. However, the drawback of the linear probability model, is its unreliability with very small and very high probabilities of a success of y. In effect, the probability for $Y = 1$ might be estimated to be negative or larger than 1 for some x_i. To keep our study simple co-variates were chosen to be binary as well.

This leaves us with a model of the form:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \epsilon$$

Where:

$$x_{i,j} = \begin{cases} 0 & \text{with probability } 1 - p_i \\ 1 & \text{with probability } p_i \end{cases}$$

$$P(y = 1|x) = E(y|x) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2$$

The model uses 2 explanatory variables x_1 and x_2. Both x_1 and x_2 take on binary values, generating one with the probabilities p_1 and p_2. Additionally, both variables are generated with varying correlation-level (Leisch et al. 1998)). The probabilities p_1 and p_2 can be adapted as far as not conflicting with the degree of correlation between x_1 and x_2. The simulation focuses on the ability of various tests to detect if β_1 is zero respective different from zero.

The model is fully specified without imposing an exogenous error term. The error term is not normally distributed and has a heteroscedastic variance:

$$var(\epsilon) = (p_y * (1 - p_y)) = (\beta_0 + \beta_1 * x_i + \beta_2 * x_2) * (1 - \beta_0 - \beta_1 * x_1 - \beta_2 * x_2)$$

This is an interesting ground, to compare heteroscedasticity robust regression to ordinary least squares regression.

2.2 Logit Model

The logit model uses a function to convert $X\beta$ from the linear probability model into a probability ranging between zero and one. Although the logit model is non-linear and has a functional form, its underlying latent variable $y*$ is build upon the same equation as y in the linear probability model above. Remember that $y*$ is an unobserved variable, while the corresponding y is the outcome observed.

$$y* = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \epsilon$$

where :

$$y_i = \begin{cases} 1 & \text{if } y_i* > 0 \\ 0 & \text{else} \end{cases}$$

Again x_1 and x_2 are generated randomly from a bi-variate binary distribution with a given correlation parameter ρ (Leisch et al. 1998). Further, as before the probabilities $p_1 = P(x_1 = 1)$ and $p_2 = P(x_2 = 1)$ might be adapted. Now to transform $x\beta = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2$ into a probability, the inverted logistic distribution is applied onto $x\beta$:

$$Pr(y = 1|x) = G(x\beta) \tag{2.1}$$

where:

$$G(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

The choice of the distribution G distinguishes the logit from the probit. While the logit model uses a logistic distribution, the probit implements a standard normal for G. This gives the logit more weight on the tails of the distribution.

If a logit estimateor is applied to the logit model as specified it is homoscedastic, assuming a constant variance term. However, in applying an linear regression method onto these non-linear data, errors are heteroscedastic. The variance is quite similar to that in the linear probability model, with the difference, that a probability density function is applied to $X\beta$.

Chapter 3

Methods

3.1 Asymptotic regression methods

This section aims to give a short overview on the most basic asymptotic regression techniques and their assumptions. Further, their peculiarities towards this specific simulation study and exact methods will be examined.

3.1.1 Ordinary Least Squares Regression

Ordinary least squares (OLS) is one of the most widely known regression methods in economics and several other sciences. It applies to discrete as well as continuous data of multiple forms. Under the following assumptions on the data, OLS is the best linear unbiased estimator (chapter 3.4 DiNardo et al. 2007).:

- X is non-stochastic or X is stochastic but independent of ϵ
- inference is conditional on X
- $E(x'\epsilon) = 0$
- $E(\epsilon^2 * x'x) = \sigma^2 * E(x'x)$
- $rank(x'x) = K$
  where x is a $1 \times K$ vector

Considering these assumptions, ordinary least squares provides the following unbiased estimator:

$$\beta = (X'X)^{-1}X'y$$

Variance is as follows:

$$var(\beta) = \sigma^2(X'X)^{-1}$$

Due to the fact that our data is not homoscedastic for OLS in neither of our two models, one main assumption on OLS is not met. However heteroscedasticity is a common obstacle in applied work and may be a more realistic situation than homoscedasticity. OLS is still consistent and unbiased, but inefficient. Further, coefficient standard errors are incorrect (chapter 6.1. DiNardo et al. 2007).

As our correlation term is adaptable, for some parameter, the columns are interdependent. This again is a violation of the variance assumption, leading to less efficient OLS outcomes.

3.1.2 Robust regression

The used robust regression technique mirrors the stata command ", robust" which is the most widely used heteroscedasticity robust least squares regression. It makes use of an heteroscedasticity-robust variance covariance matrice by MacKinnon and White (1985). In least squares regression, the variance of $\hat{\beta}$ is:

$$\phi = var(\hat{\beta}) = (X^TX)^{-1}X^T\Omega X(X^TX)^{-1}$$
$$\text{where: } \Omega = \sigma^2 I_n$$

Ordinary least squares now estimates σ^2 by:

$$\hat{\sigma}^2 = (n-k)^{-1} * \sum_{i=1} \hat{u_i}^2$$

Now MacKinnon and White propose to use an alternative estimator for σ^2:

$$\hat{\sigma_i} = \frac{\hat{u_i^2}}{(1-h_i)^2}$$

Where $h_i = P_{ii}$ being the diagonal elements of the projection matrix. This estimator adapts to the different variance structures of the single errors.

3.1.3 Probit and Logit

In contrast to ordinary least squares, probit and logit are non-linear methods. Logit and probit are only applicable on discrete outcome data. They do not have problems from discrete and non-normal error structures. On the other hand, they only provide odds-ratios as results, which are more complicated when interpreting. The model was already presented in some detail in the section on the different models simulated in this study. Probit and logit rely on an unobservable latent variable y_i* with the property $y_i* = X_i\beta + \epsilon_i$. Where y_i* is conditional on X. Now, as we can not observe y_i*, it is modeled by:

$$y_i = \begin{cases} 1 & \text{if } y_i* > 0 \\ 0 & \text{otherwise} \end{cases}$$

As these models are constrained to 0,1 no conventional error term exists. To conclude this refresh, the standard portayal of the Probit is the following $prob(y_i = 1) = \Phi(X_i\frac{\beta}{\sigma})$.

In principle the logit does not differ much from the Probit. Instead of relying on the normal distributed errors it relies on the logistic distribution, which has more weight on its tails (DiNardo, 2007). The logit equation is therefore $prob(y_i = 1) = \frac{exp(X_i\beta)}{1+exp(X_i\beta)}$.

To get a deeper overview on logit and probit, "Econometric Methods" by DiNardo and Johnston provides a compact and well written introduction into the model.

3.2 Finite-Sample exact tests for linear regression

In this section we will introduce a mathematically exact test for finite sample monotonicity testing. This test is under current development by Schlag (2016) and details should be published soon. By identifying the bounds of the dependent variable it is possible for these tests to have a specific type I error probability bound. As our simulation environment is fully binary the bounds on the dependent variable do not represent an obstacle.

3.2.1 Framework for the exact monotonicity test

As the stratified permutation test applies to a wide range of binary, ordinal and cardinal models, let us here specify a framework for the binary case. Consider a nonparametric binary choice model, where Y_i is binary, $Y_i \in 0, 1$ and there is a function $f : \text{Dom} \to [0, 1]$, such that:

$$P(Y_i = 1|X = x) = f(x_i) \text{ for all x.}$$

Where outcomes emerge from the same distribution, i.e. are identically distributed, and *Dom* has a complete order. While f is not specified, it is known to be either strictly monotone increasing, strictly monotone decreasing or independent of x_i. As an interesting side note, f need not necessarily be linear.

In more detail, consider an explained variable Y in $\mathbb{R}^n$ and a matrix of covariates X with dimensions n times k. Now if we have a relation such as f described above, one of 3 cases below is true.

$$\begin{aligned} P(Y_c = 1|X = x) > P(Y_i = 1|X = x) if x_{c,j} > x_{i,j} \\ \text{and } x_{c,k} = x_{i,k} \text{for all} j \neq k. \\ P(Y_c = 1|X = x) = P(Y_i = 1|X = x) if x_{c,j} > x_{i,j} \\ \text{and } x_{c,k} = x_{i,k} \text{for all} j \neq k. \\ P(Y_c = 1|X = x) < P(Y_i = 1|X = x) if x_{c,j} > x_{i,j} \\ \text{and } x_{c,k} = x_{i,k} \text{for all} j \neq k. \end{aligned}$$

Clearly, both of our models are supported by this framework.

3.2.2 Binary stratified direction of change test

As mentioned earlier, the binary stratified direction of change test is able to detect increasing, decreasing or no influence from a certain covariate onto the explained variable. It is exact in the sense, that its probability to reject the null hypothesis when in fact it is true, is bounded above by α. In order to be exact, the test requires discrete and thereby bounded data.
In more detail, the binary stratified direction of change test is based on Fisher's exact test with the corresponding extension by Tocher (1950). This is an exact test on independence of contingency tables, with the Tocher extension it is even the uniformly most powerful unbiased test. However, in the present monotonicity test, it is used to check for different effect intensities.

The binary stratified direction of change test takes 3 steps in order to reject of not reject a null hypothesis.

1. Individual observations are pooled in blocks. Where a block consists of individuals with the same values for all covariates except the covariate for which the hypothesis of monotonicity is checked, i.e. for all $j \neq k$ where j is the covariate in question.
2. Compute the original test-statistic $\sum_{i \text{ s.t. } y_i=1} x_{i,j}$.
3. Values of Y_i are permuted l times within blocks, where $l \to \infty$. For each permutation a test-statistic is computed.
4. Then the proportion of permutations with a corresponding test-statistic above the test-statistic of the original data is computed. If this proportion is above the significance level α, the null hypothesis is rejected.

However, the permutation should lead to an exact randomized test. In order to transform an exact randomized test into an exact non randomized test, the following theorem by Gupta and Hande (1992) is applied:

Theorem 3.1. *Let ϕ be an exact randomized test with level $\theta * \alpha$. Then $\phi\,|_\theta$ is an exact nonrandomized test with level α. The type II error of $\phi\,|_\theta$ is bounded above by the type II error of the underlying randomized test ϕ divided by $(1-\theta)$.*

Needless to say, the presented test is exact non-random, so the above theorem is applied and a θ needs to be chosen. In the working paper on exact monotonicity tests (2008), Karl Schlag recommends a θ of 0.3.

The drawback of this method is that permutation needs more computational power, than estimation by asymtotic regression. Hence, the exact methods were simulated with less rich parameters. However in normal use, this should not present a problem.

In contrast to the asymptotic methods presented above, the binary stratified direction of change test does not rely on assumptions about the asymptotic distribution of the data. Mathematically exact tests offer the advantage, to have a perfectly determined significance level, independent of the underlying distribution of the data. Hence, for particularly sensitive topics, such as medicine or risk analysis, the use of an exact test might prevent potential problems concerning imprecise significance levels. Studies on oversizedness of ordinary least squares and a robust least squares regression in simulated as well as in empirical data can be found in Gossner and Schlag 2013.

Chapter 4

Methodology

4.1 Type I and Type II Errors

In order to describe the power and the size of a test, first errors of the first and of the second kind need to be introduced (cf. Lehman and Romano 2005). A Type I error is occurs if an hypothesis test rejects the null hypothesis, when it is in fact true. As an example, consider a courthouse. Under continental law a suspect is innocent as long as its guilt is not proven. So the null hypothesis is innocence while the alternative hypothesis is guilt. Now if a judge convicts a suspect, he rejects the null. If years later an airtight alibi is found for the suspect. It turns out the judge committed an error of the first type. Similarly, an error of the second kind occurs if an hypothesis test rejects the alternative hypothesis if it is in fact true. Here again the courthouse provides a good example. Consider again a suspect, but now with an alibi at the time of the trial. The judge notes the alibi and pledges for innocence of the suspect. If years later the alibi turns out to be invalid, the judge committed an error of the second kind.

This example shows that the consequences of errors of the first kind respective of the second kind depend on the phrasing of the null hypothesis. For instance, if the null hypothesis in the court would be that the suspect is guilty until its innocence is proven, errors of the first and the second kind would be inverted. However, often the hypothesis test is constructed in a way that an error of the first kind weights heavier. Just like in our original courthouse example, where Napoleonic law assumes innocence until guilt is proven.

4.2 Power and Size

The size of an hypothesis test is the probability of a test to reject the null hypothesis when it is true. In other words, the size is the probability to commit an error of the first kind. Directly linked to the size is the significance level of a test. The significance level of a test is the upper bound on the type I errors which an hypothesis test commits. Often, one wants the size to be below 5 percent, so a test with a significance level of 5 percent, under the given circumstances is chosen.

Conversely, the power of an hypothesis test is the probability to not commit an error of the second kind. In other words, the power is the probability to reject the null hypothesis when it is wrong. Typically it is not possible for a test to have certain lower bound on the power, given a certain significance level. Therefore, the standard procedure is to choose a test with an appropriate significance level while having a power as high as possible. Studies such as the underlying serve as a guidance for this choice.

4.3 Hypothesis Specification

The aim of this paper is to assess the respective size and power of the various methods presented. The linear probability model and the logit model serve as test environment for this benchmarking exercise. The probabilities for type I and type II errors are estimated by simulation.

The null hypothesis of the various tests is that there is no significant impact of β_1 on y, i.e. $\beta_1 = 0$. The alternative hypothesis is defined as y is dependent from β_1. All of the presented methods share the same null hypothesis. However the common alternative hypothesis of the asymptotic regression methods is a specific relationship between y and x_1, while the alternative hypothesis of the exact test is that the y and x_1 have a monotone relation. Still, generalizing the alternative hypothesis to test for dependency of y from β_1 comprises the mentioned alternative hypotheses.

It should be emphasized, that the results provided by these simulations apply only to the specified frameworks. Further only a small fraction of the parameters of this framework are simulated. Still, the simulations should provide a useful assessment for the reliability of the presented methods in a fully binary model.

Chapter 5

Simulation Details

Fortunately, both models can be adapted using the same parameters, so where possible, the same parameters are used. The study runs a simulation count of 1000 repetitions for each parameter setting. The simulation framework is based on a biomedical research paper form Christopher Meany and Rahim Moineddin (2014).

5.1 Simulation parameters for the linear probability model

In order to have a more detailed look on the asymptotic regressors, 2 different parameter settings where chosen. The smaller setting is dedicated to detect the features of the exact regression technique, while enabling a direct comparison to the asymptotic methods. As the binary stratified direction of change test is computationally more demanding, only the small setting could be applied in a reasonable time period. The large parameter setting enables a wider analysis of the behavior of the asymptotic methods.

5.1.1 Reduced simulation parameters

For the linear probability model, $X\beta$ is limited to lie between 0 and 1. This bound does not apply to the logit model, as it is inherently bounded between 0 and 1. β_1 takes the values 0 and 0.3 for the linear probability model and the logit model.

During extensive studies no significant impact of β_2 on the size nor the power could be detected, thereby β_2 is fixed to 0.09. On the other hand, β_0 changes from 0 to .6 in steps of 0.1. Due to the apriori bounding of logit, in this model, β_0 varies from 0 to 1 in steps

of 0.1. The results will show that β_0 has a big impact on errors of robust covariance regression and ordinary least squares. Next the probability of x_1 to be a success will be analysed for 0.1 and 0.5. As the $P(x_2 = 1)$ was found to have no further impact on respective sizes and power, it is fixed to 0.3. This holds for both models. To analyse the impact of correlation, ρ takes the values 0 and 0.5. This value is retained from a paper by Mela and Kopalle (2002). Due to computational limitations, the other interesting case of negative correlation of -0.5 was omitted, but might be analysed in later work. Concerning the binary stratified direction of change test, θ is specified at 0.3, while the number of permutations is fixed at 1000. Finally the observation numbers of the single cohorts variate from 30 over 50 and 100 to 750. These numbers represent a a standard experimental economics framework, as can be found in List et al.(2011), Bellemare et al. (2014) and Meaney et al. (2014). However the present simulation study does not include clustering as it makes no sense in the given randomization framework.

5.1.2 Large simulation parameters

For the large sample, a few changes have been made. Most importantly β_2 is allowed to change in the same manner as β_0. For the linear probability setting, both, β_0 and β_2 change from 0 to 0.6 in steps of 0.05. For the logit model, both change form -1 to 1 in steps of 0.1. Both parameters where chosen to change only if the other parameter is fixed at 0.09. To have a more detailed overview on the $P(x_1 = 1)$ it is extended to (0.1, 0.3, 0.5) for both models.

5.2 Statistical software and simulation matters

Since the software for the binary stratified direction of change test is only available in R, this simulation experiment is based on R version 3.3.0. It uses the inbuilt lm() function for ordinary least squares as well as robust covariance-matrix regression, while the second uses two distinct libraries, "sandwich" and "lmtest". "lmtest" provides the function coeftest() allowing to specify an alternative covariance-matrix for linear regression. This covariance matrix is provided by the library "sandwich" and reproduces the MacKinnon-White heteroscdasticity robust covariance matrix. Further, probit and logit use the inbuild glm() function. The algorithm for the exact monotonicity test runs on a script provided by Karl Schlag. In order to get distinct correlation classes, the library "bindata" was used.

Due to the permutation structure of the binary stratified direction of change test, and its implementation in the relatively slow R-language, the simulation had a massive need

for computational power. This was solved by implementing a parallel structure for the simulation program, running the libraries "foreach" together with "doParallel". Using an eight-core machine, running time for the hole simulation was down to about 15 hours.

To generate a specific correlation structure, an R-library by Leisch et al. is used. It directly converses normally distributed random variables into binary variables with a predefined correlation structure.

In the later analytical phase of the project, the library "ggplot2" was used to create various plots, while "stargazer" enables to create nice tables to export explicitly to Latex.

Chapter 6

Results

The Appendix provides detailed results in figures for each distinct observation count, divided into the large simulation and the small simulation. The figures are augmented by a 5 percent line to easily spot oversizedness. The type of regression method is depicted in different segments of bars. The bars represent different values for β_0. The vertical axis depicts the rejection probability. Further, the figures are segmented in different facets, where β_1 and the correlation divide the figure vertically. The horizontal division is made by different probabilities of a success of x_1.

In the latter the description of the results will be segmented by least squares regression techniques and binomial response models. Further, the text will mostly begin with the low observation sample and finish gradually with the high observation sample, first 30

observations, then 50 observations, then 100 observations.

6.1 Size and Power in the Linear Probability Model

6.1.1 Least squares regression

When looking at the outcomes such as figure 8.1, linear regression techniques attract attention, by showing larger values for both, the probability of a type I error as well as for the probability of a type II error. At 30 observations, ordinary least squares shows increased size for low probability of a success of x_1, i.e. $P(x_1 = 1) = .1$. The lower β_0 the higher the size. White's heteroscedasticity robust regression behaves in the opposite direction, while type I error probabilities are larger for low $P(x_1 = 1)$. The size increases with an increase in β_0. In the worst case, with a low $P(x_1 = 1) = 0.1$, β_0 above .5 and no correlation, variance robust least squares reaches sizes over 20%. Interestingly, while correlation increases the size of OLS at low β_0 it decreases the size of robust at high β_0. Similarly, the power of both regression methods react in the same direction as the respective sizes for β_0. For 50 observations, the pattern described above endures, even though, the rejection probability for the null hypothesis is closer to 5 percent. Ordinary least squares has sizes up to .6 if not controlling for β_0. For small $\beta_0 < 0.1$ and $P(x_1 = 1) = 0.1$ OLS shows still increased sizes, reaching values above 10 percent for β_0. The size for robust has similar behavior than in the small sample case, with sizes above 10 percent for $\beta_0 \geq 0.5$ and $P(x_1 = 1) = .1$. The power again varies with β_0 in opposing directions. In the comparably large sample of 100 observations, both regressors reach normal sizes for most settings tested. Small $P(x_1 = 1) = 0.1$ still drives the size of ordinary least squares up, reaching values above 10 percent for $\beta_0 = 0$. Although less pronounced, correlation still has an opposing effect on sizes. With correlated variables and $P(x_1 = 1) \leq 0.1$ sizes are significantly above 5 percent. The power reacts as described before, while at higher $P(x_1 = 1)$ power is increased for low and high values of β_0 for both regression methods. In the large observation case of 750 observations, ordinary least squares still has problems with the size at $\beta_0 = 0$. Heteroscedasticity robust regression does not show problems with the size. Concerning power, no problems have been detected, even for $P(x_1 = 1) = 0.1$ the null hypothesis is rejected in more than 99 percent of the cases. Concerning correlation, it decreases the power of our linear estimators in all settings by about 10 percent.

6.1.2 Logit and Probit

For all three observation samples, neither probit nor logit show problems with type I error probabilities. The power is quite low compared to least squares regression techniques,

especially low probability of a success of x_1. Considering different β_0, the power is lowest at both low and high values for β_0. At 30 observations, and $P(x_1 = 1) = 0.1$ the power hardly reaches 10 percent. For larger $P(x_1 = 1)$ the power reaches the levels of linear regression techniques, for β_0 around 0.3. At 50 observations and $P(x_1 = 1) = 0.1$, the power is still considerably lower than for least squares regression. At higher observation counts the power catches up. Further, correlation has the same decreasing effect on the power.

6.1.3 Exact monotonicity test

Considering the size, the binary stratified monotonicity test clearly undercuts all of the asymptotic tests. Although, logit and probit perform reasonably well, the exact test has less than a tenth of their corresponding sizes. This holds for all observation numbers and all remaining parameters. The power is a little inferior to that of logit and probit, while it is higher at the ends of the values for β_0. At 30 observations and low $P(x_1 = 1) = 0.1$, the power is below 5 percent. This performance is comparable to that of probit and logit, while robust and OLS reach far higher power. For $P(x_1 = 1) = 0.5$ it is around 10 percent, while it is considerably lower with correlation. When comparing again to logit and probit, this is half their power and less than a third of OLS. At 50 observations the picture of low power prevails, as the power of the binary stratified direction of change test is about half that of ordinary least squares. While the test catches up for 100 observations to about three fourths the power of OLS, at 750 observations 99 percent of the wrong null hypotheses where rejected.

6.1.4 Size and Power in the Logit Model

6.1.5 Least squares regression

The binary response model is not the natural home of linear regression functions, still next to probit and logit and for reasonably high observation counts, they show reasonable size and power. In the 30 observations sample, the probability for type I errors for robust is strongly increased. Ordinary least squares on the other hand does not show problems with its size. Especially with a low probability on the success of x_1 the white-robust regression expresses mean rejection probabilities above 20 percent. Further, in the small sample, rejection probabilities for the null hypothesis are still exaggerated when looking at the results for $P(x_1 = 1) = .3$. Only at large samples of 750 observations, robust reaches a size below 5 percent in all parameter settings. As in the case of a linear probability model, the size of variance robust regression tends to increase for a higher

levels of β_0. Concerning power, OLS, has power levels below 10 percent for the 3 smaller observation samples, even at a fairly large $P(x_1 = 1) = 0.5$. Still this is fairly comparable to logit and probit. At the large sample, OLS has a power of around 20 percent for low $P(x_1 = 1) = 0.1$ while it goes up to about 50 percent for $P(x_1 = 1) = 0.5$. White's robust regression method's power is bad as well, especially regarding the meaningless size at smaller sample sizes. At 100 observations it has a power of 10 percent while at 750 observations its power reaches 20 percent. The power of ordinary least squares is decreasing for larger values of beta0, while robust has increased power.

Correlation has a decreasing effect on power and has an ambivalent effect on size for heteroscedasticity robust regression.

6.1.6 Logit and Probit

Let us consider a detailed analysis of the logit and probit regressors starting with the small sample of 30 observations. Considering the size, logit and probit are not remarkable, their sizes stay below the 5 percent level. Especially in low observation samples, the size seems to be larger for negative $\beta 0$. Power levels are quite comparable to ordinary least squares regression. At the 3 smaller samples, the power lies between 1 and about 10 percent. Only in the large sample of 750 observations, the power of logit an probit reach 50 percent for the $P(x_1 = 1) \leq .5$ case and around 20 percent for $P(x_1 = 1) = .1$. Correlation has the same effect as for ordinary least squares regression, it decreases the power but has no effect on the size

6.1.7 Exact monotonicity test

The binary stratified monotonicity test has a size below 1 percent for all parameter constellations simulated. On the other hand, the power is also the lowest observed, with values below 5 percent for all different success probabilites for x_1 and observation counts of 30, 50 and even 100. At 750 observations, the power lies still between 50 and 60 percent of that of all the other regression techniques. Correlation has again a decreasing effect on the power of the non-parametric monotonicity test.

Chapter 7

Discussion

Some interesting results have been found in this study. Especially concerning White's heteroscedasticity robust least squares estimator on the one hand with large size and power. And on the other hand the binary stratified monotonicity test with low size and low power. Ordinary least squares size was around 5 percent, except in simulations with no constant where OLS showed increased size. Further it provided quite good power results, making it the silent star of this study.

When introducing inexperienced students to econometrics, a large part of the classes is dedicated to heteroscedasticity in data and its "cure". The solution most students retain from these lessons is to put ", robust" behind the regression command. The author of this study himself, lacking experience and knowledge, was prone to just using robust at the slightest sign of heteroscedasticity. Quite shockingly, in the present parameter setting, including inherent heteroscedasticity, OLS fares better than robust. Especially in small samples, White's heteroscedasticity robust least squares estimator reaches unusably high sizes. This was found for both models, where all other estimators, had no, or at least a lot less problems with falsely rejecting the null hypothesis. Ordinary least squares showed its biggest troubles with correlated covariates of low success probability.

This study does not focus on the exact prediction made by the estimators, but rather on their ability to reject or not to reject the null hypothesis, when it is a priori true respectively false. It is possible, that the true outcomes, delivered by robust perform better than those of the other estimators. But when these data where produced as a side product in this study, they did not seem to confirm this.

When comparing to non-robust power levels, the exact test presented in this study did quite well in terms of power. The size is mathematically proven to be below 5 percent, our simulation study just confirmed this. Although considering the very low

size throughout all parameter constellations, and the relatively low power especially in low observation samples, it seem this test could be a bit too conservative. Maybe some improvements are still to be done. In this place it should be mentioned, that an other soon to be published exact monotonicity test (working paper Schlag 2016), fared similarly in preliminary studies leading to this paper.

In the end it needs to be said, that this simulation study did only cover a small fraction of the binary models out there. Even regarding the 2 models considered, many possibilities have been left out. For example higher probabilities for a success in the logit model should be investigated further. The same is true for larger coefficients for x_1 and x_2. A simulation study using only 2 models might fail at underlining the main advantage of an exact test, its independence from distributive assumptions (except for bounds). While OLS showed better results in the underlying setting, a different distribution of the data could show the opposite. No matter the underlying distribution, a mathematically exact test will not be oversized and thereby be at least a valuable reinforcement of an empirical work.

Last but not least, this study constitutes a praise to simple models. In our simulaton stting, ordinary least squares showed the best overall performance. Among the considered asymptotic models, it is by far the simplest and most studied. This holds also when talking about interpretation, where logit and probit have a serious drawback. Especially concerning the publication and spreading of results, logit and probit are prone to misinterpretation. The exact monotonicity test on the other side put up to much more assumption loaded techniques, while it is very simple in both application to data and interpretation.

Bibliography

[1] Bellemare, Charles, Bissonette Luc, Krger, Sabine, "Statistical power of within and between-subjects designs in economic experiments", CESIFO working paper No.5055 cat 12:empirical and theoretical methods 2014.

[2] Cameron, Lisa A., "Raising the stakes in the ultimatum game: Experimental evidence from Indonesia." Economic Inquiry 37.1: 47-59, 1999.

[3] Gupta, S.S. and Hande S.N., "On Some Nonparametric Selection Procedures", Nonparametric Statistics and Related Topics, 1992

[4] Johnston, J., and J. DiNardo. "Econometric methods.", Fourth Edition. McGraw-Hill International Series, 2007.

[5] Erich L. Lehman, Joseph P. Romano, "Testing Statistical Hypotheses", Third Edition. Springer, 2005,

[6] Friedrich Liesch, Andreas Weingessel, Kurt Hornik, "On the Generation of Correlated Artificial Binary Data." Working Paper No. 13, 1998.

[7] Gossner O. and Schlag K., "Finite-sample exact tests for linear regressions with bounded dependent variables" Journal of Econometrics, Volume 177, Issue 1, Pages 75-84, 2013.

[8] J.List, S. Sadoff, M. Wagner, "So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design", Experimental Economics, Springer, vol. 14(4), pages 439-457, 2011.

[9] MacKinnon, James G., and Halbert White. "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." Journal of econometrics 29.3 (1985): 305-325, 1985.

[10] Meaney C. and Rahim M., "A Monte Carlo simulation study comparing linear regression, beta regression, variable-dispersion beta-regression and fractional logit regression at recovering average difference measures in a two sample design", BMC Medical Research Methodology 14:14, 2014.

[11] Mela C. and Praveen K. Kopalle, "The impact of collinearity on regression analysis: the asymmetric effect og negative and positive correlations", Applied Economics, 34, 667-677, 2002.

[12] Schlag, Karl, "Designing non-parametric estimates and tests for means." Economics Working Papers ECO 2006/26, European University Institute, Florence, 2006.

[13] Schlag, Karl, "A New Method for Constructiong Exact Tests without Making any Assumptions", Paper provided by Department of Economics and Business, Universitat Pompeu Fabra in its series Economics Working Papers with number 1109, 2008.

[14] Schlag, Karl, "Exact Nonparametric Binary Choice and Regression Analysis", Working Paper University of Vienna, 2014

[15] Tocher, Keith D., "Extension of the Neyman-Pearson theory of tests to discontinuous variates.", Biometrika 37, 130-144, 1950

Appendices

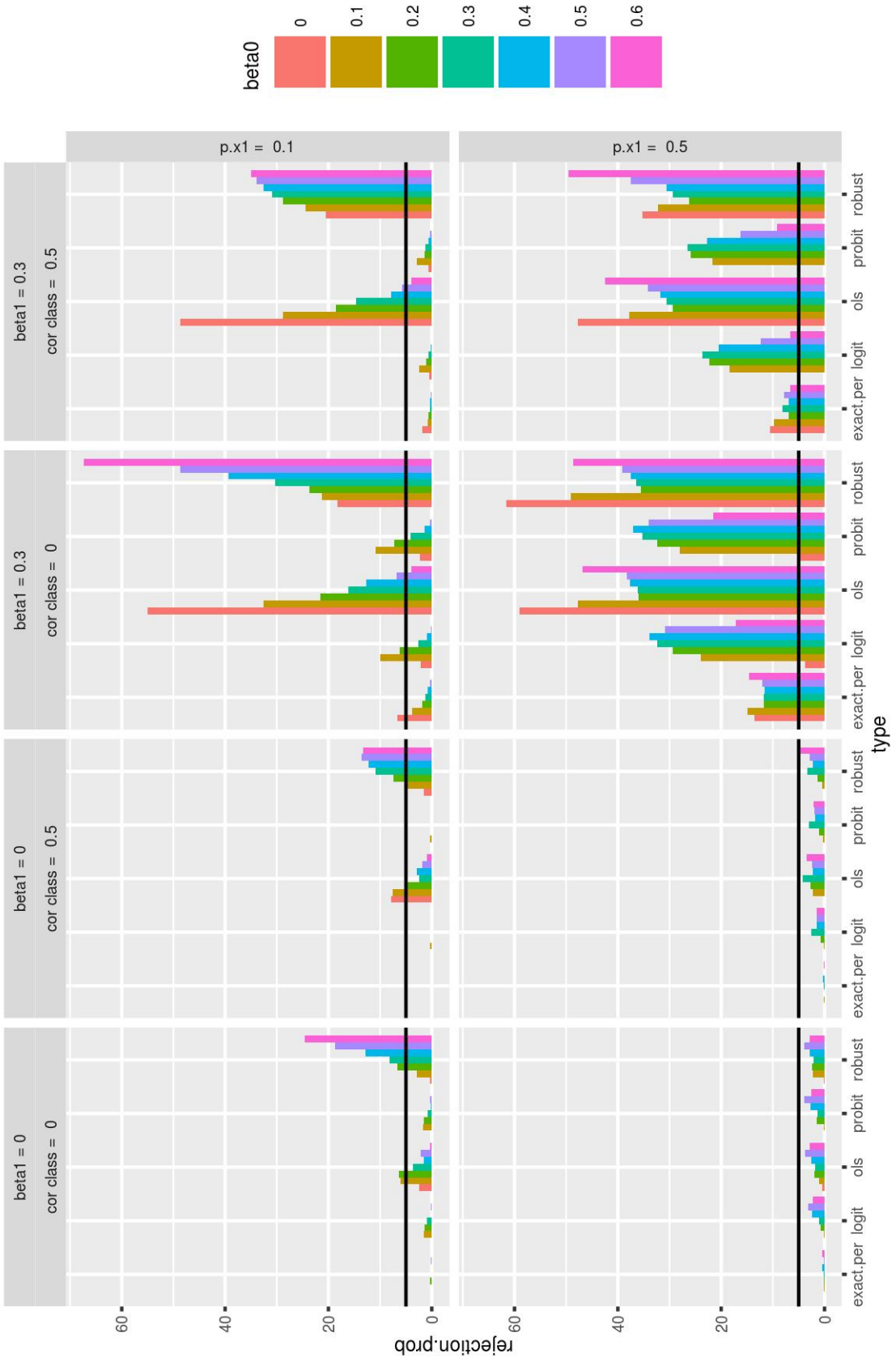

FIGURE 1: small simulation of the linear probability model at 30 observations

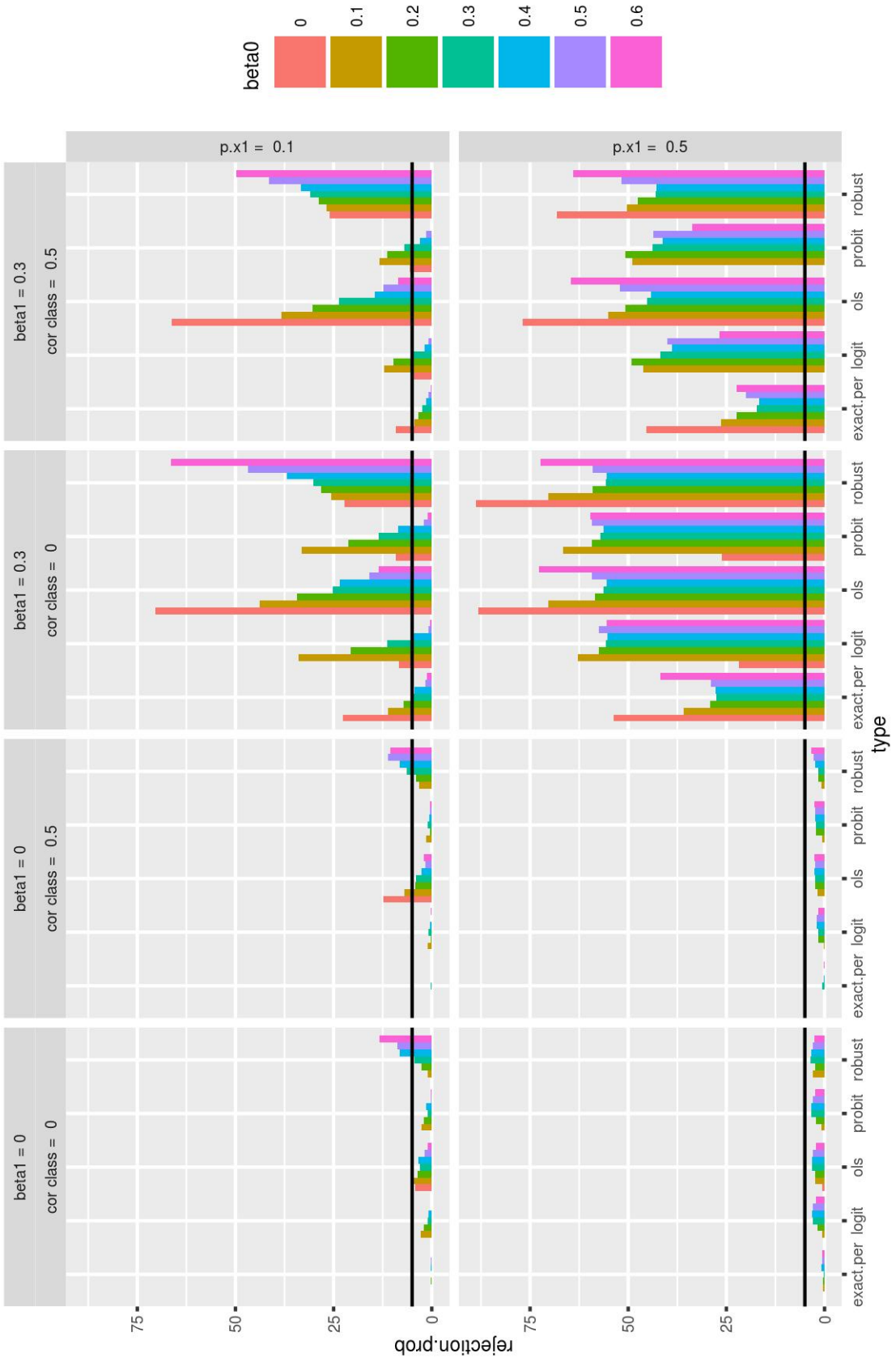

FIGURE 2: small simulation of the linear probability model at 50 observations

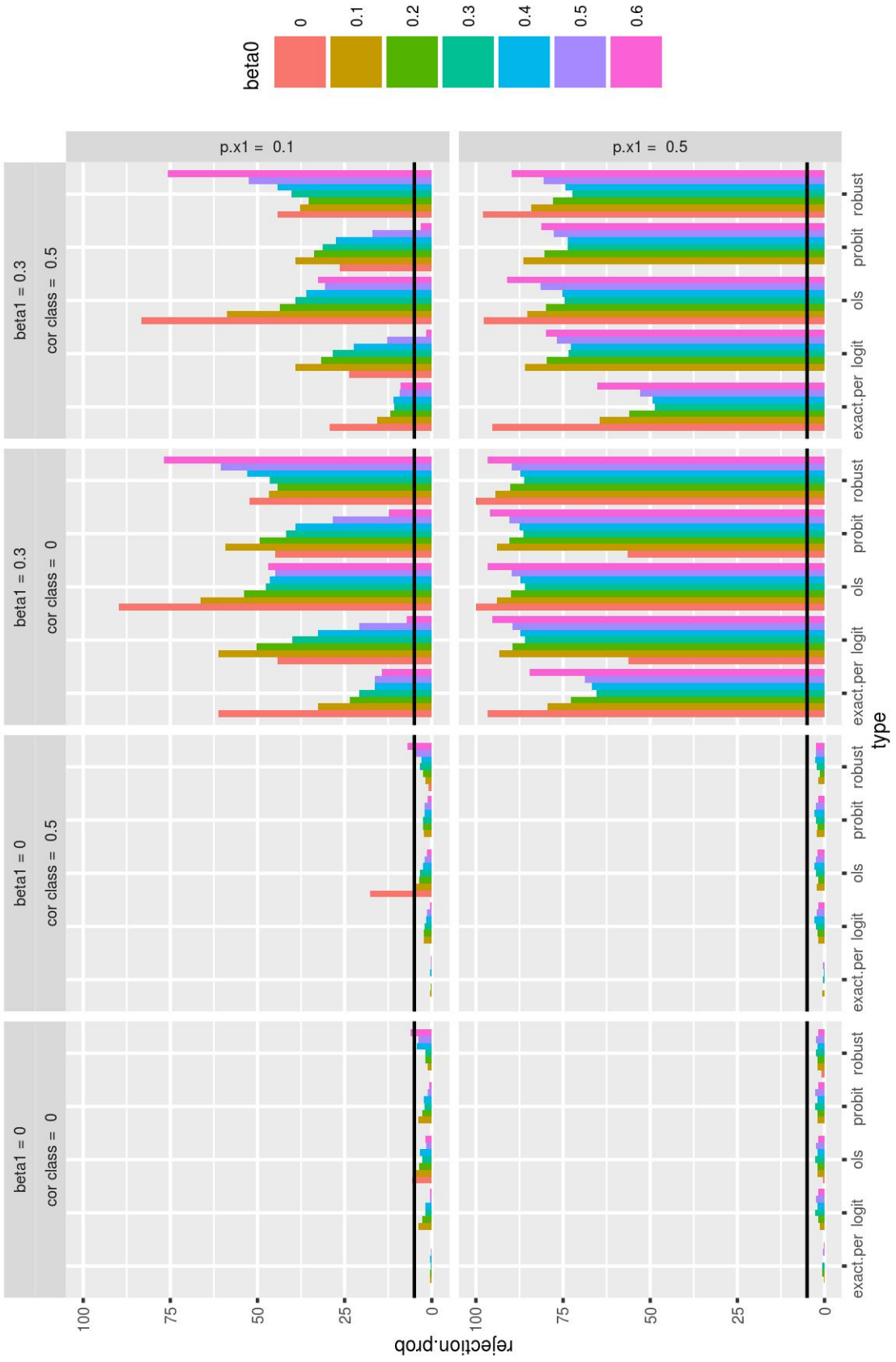

FIGURE 3: small simulation of the linear probability model at 100 observations

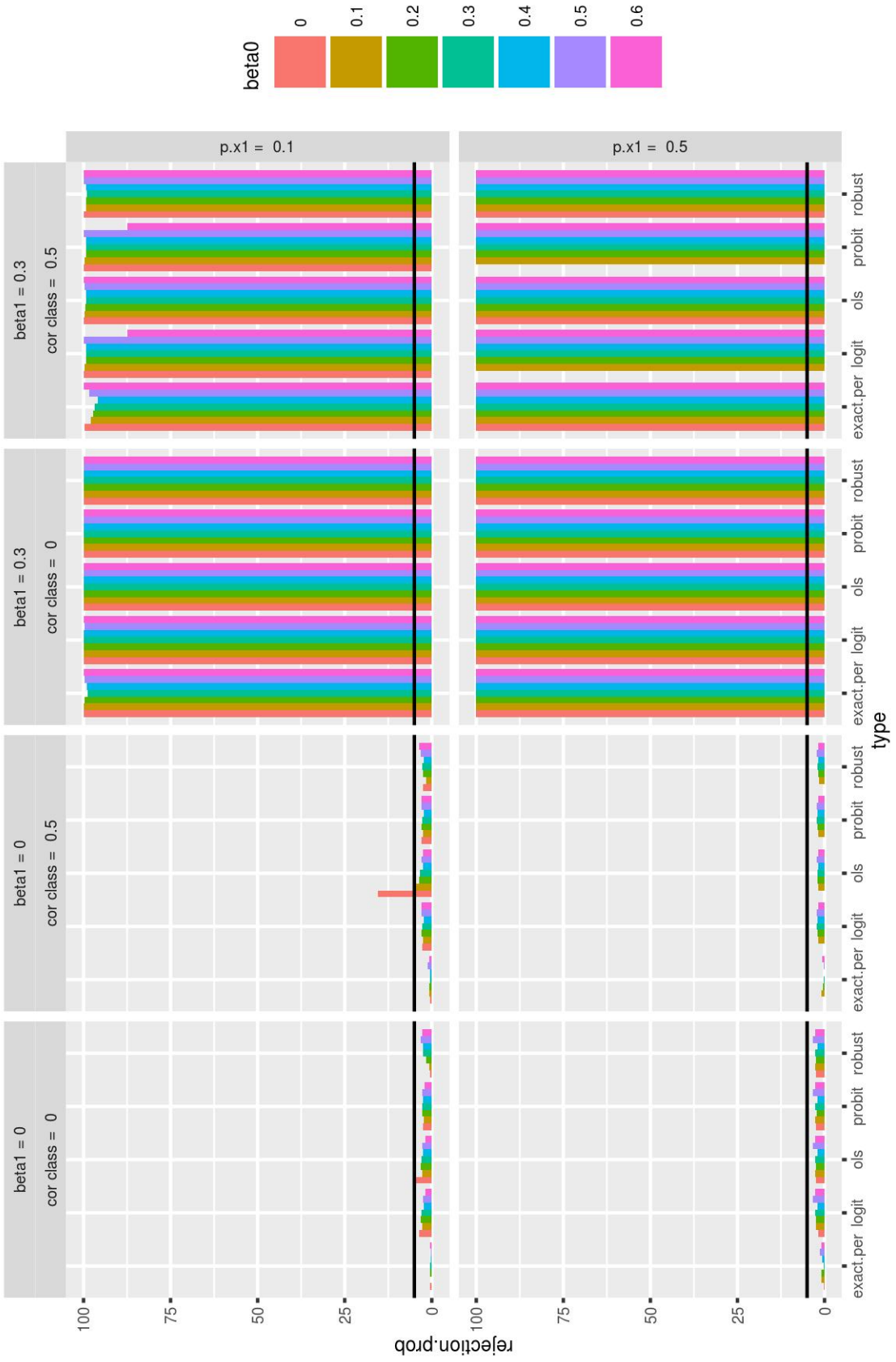

FIGURE 4: small simulation of the linear probability model at 750 observations

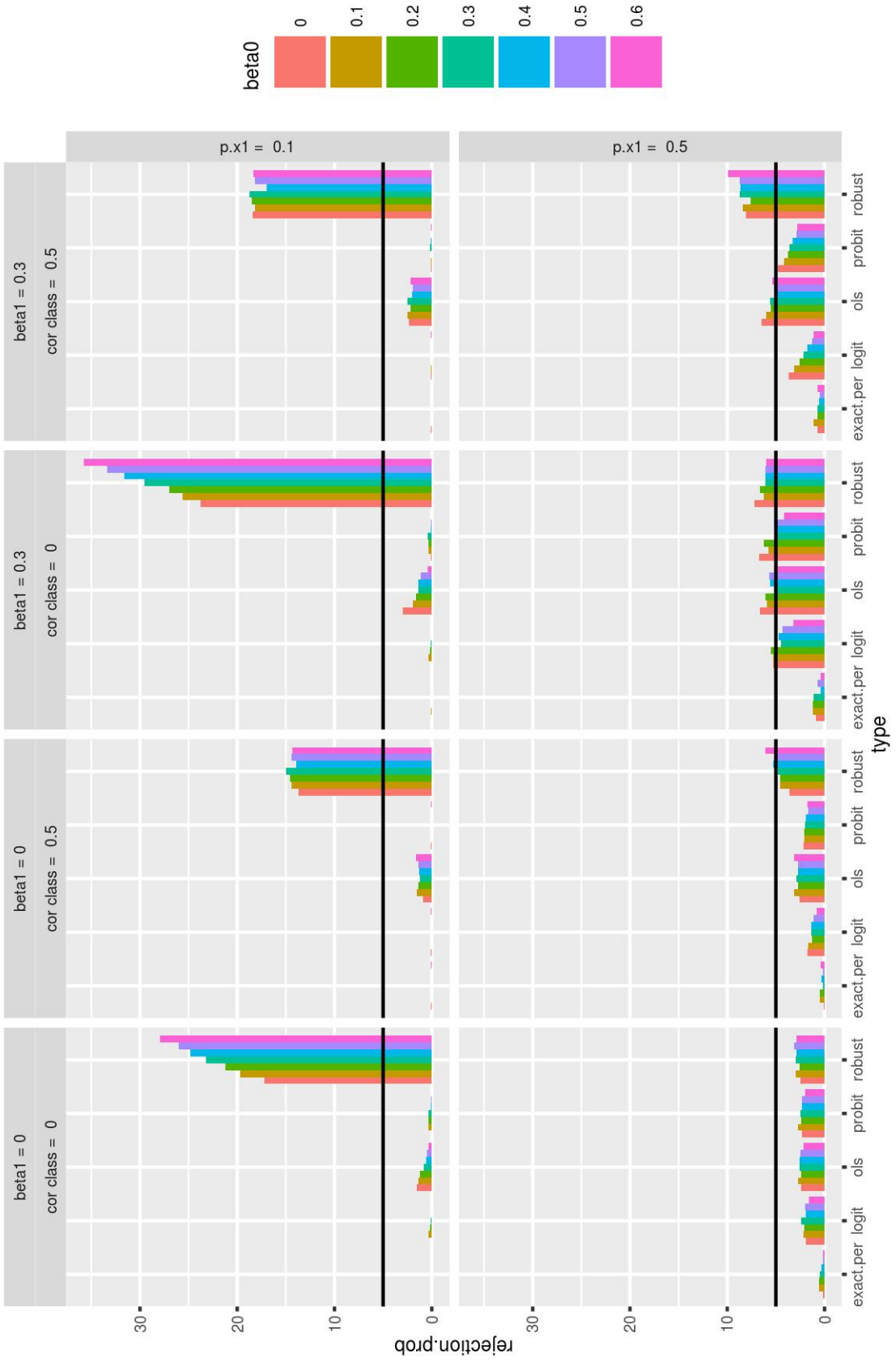

FIGURE 5: small simulation of the logit model at 30 observations

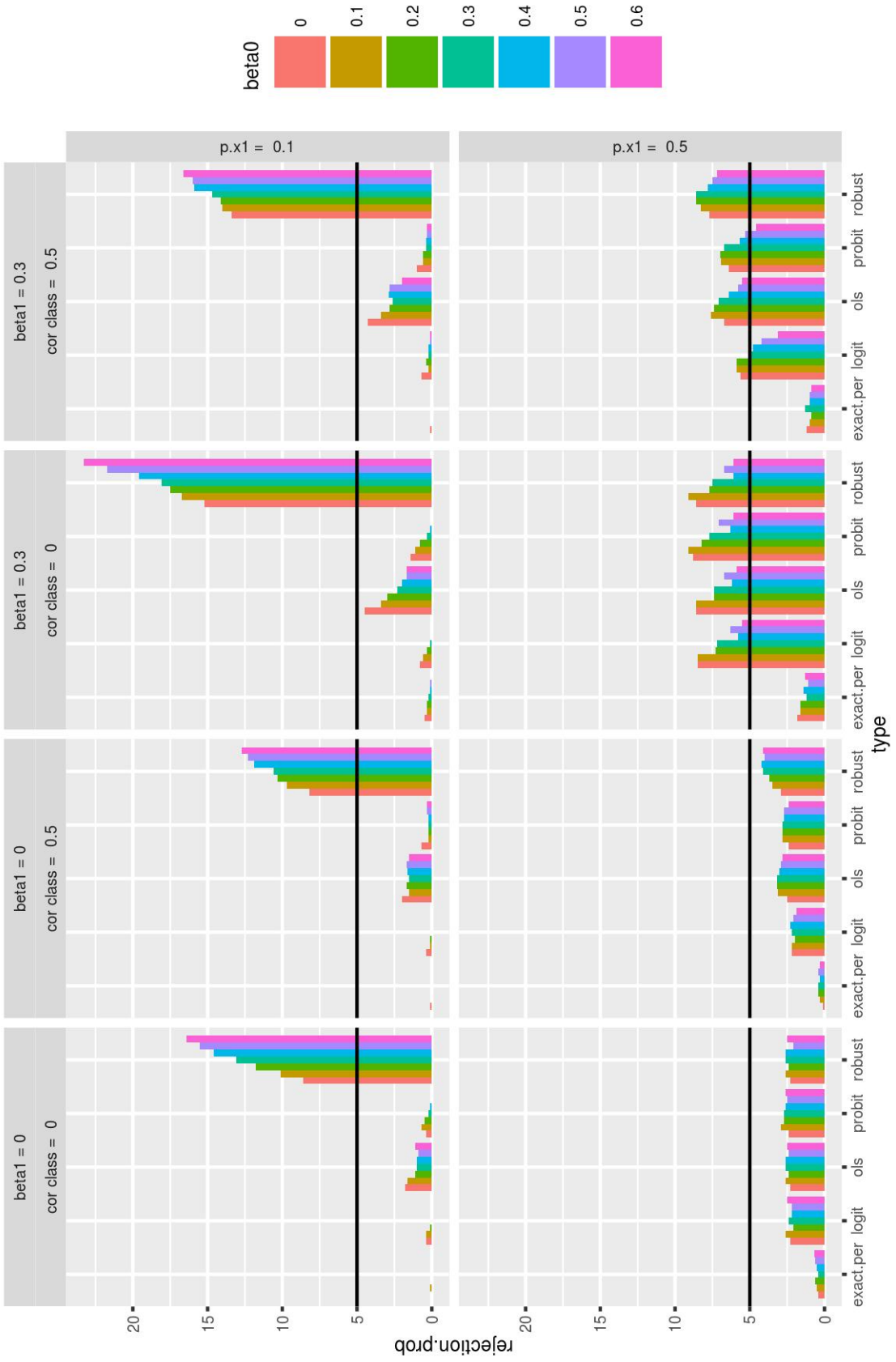

FIGURE 6: small simulation of the logit model at 50 observations

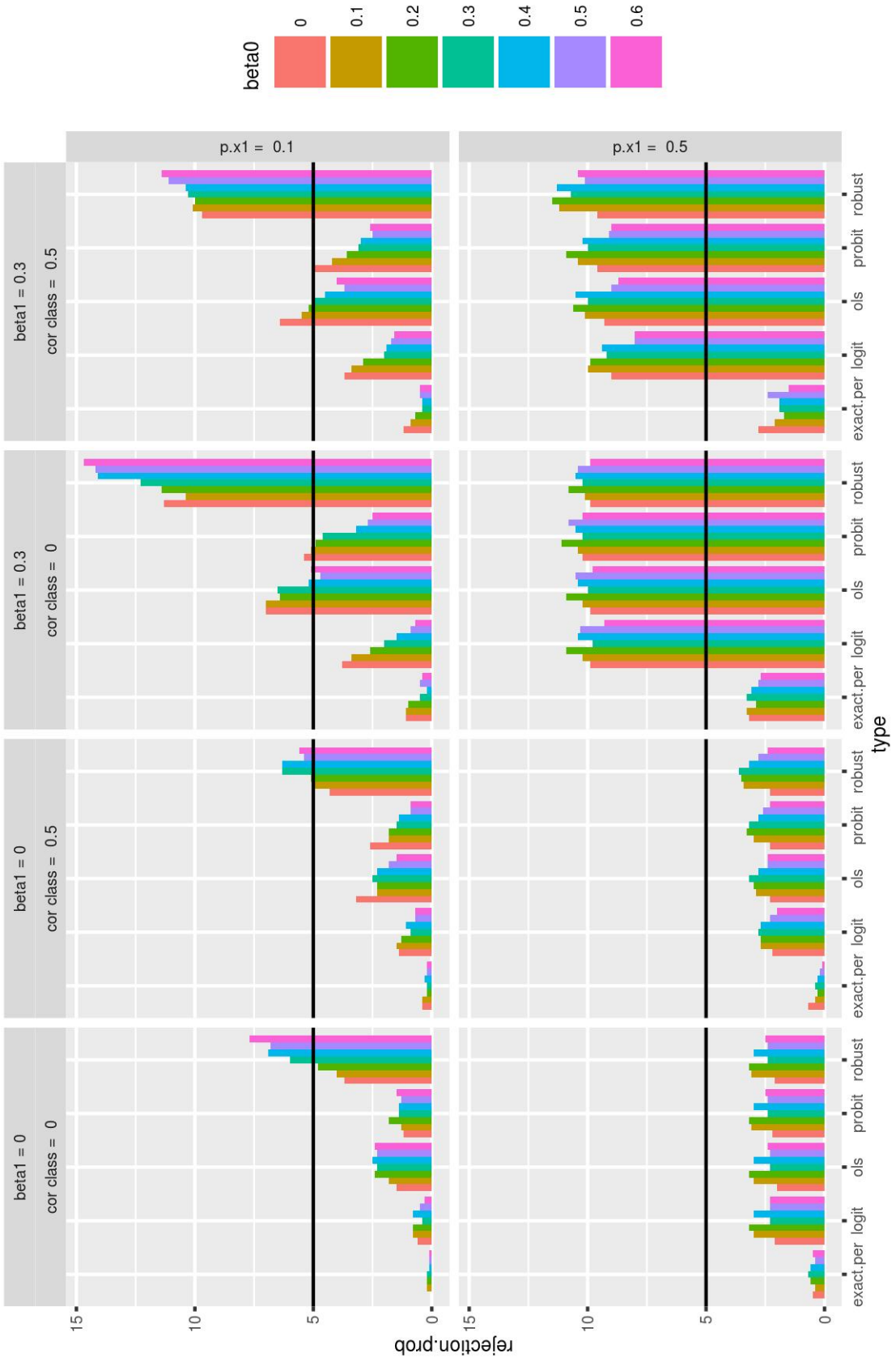

FIGURE 7: small simulation of the logit model at 100 observations

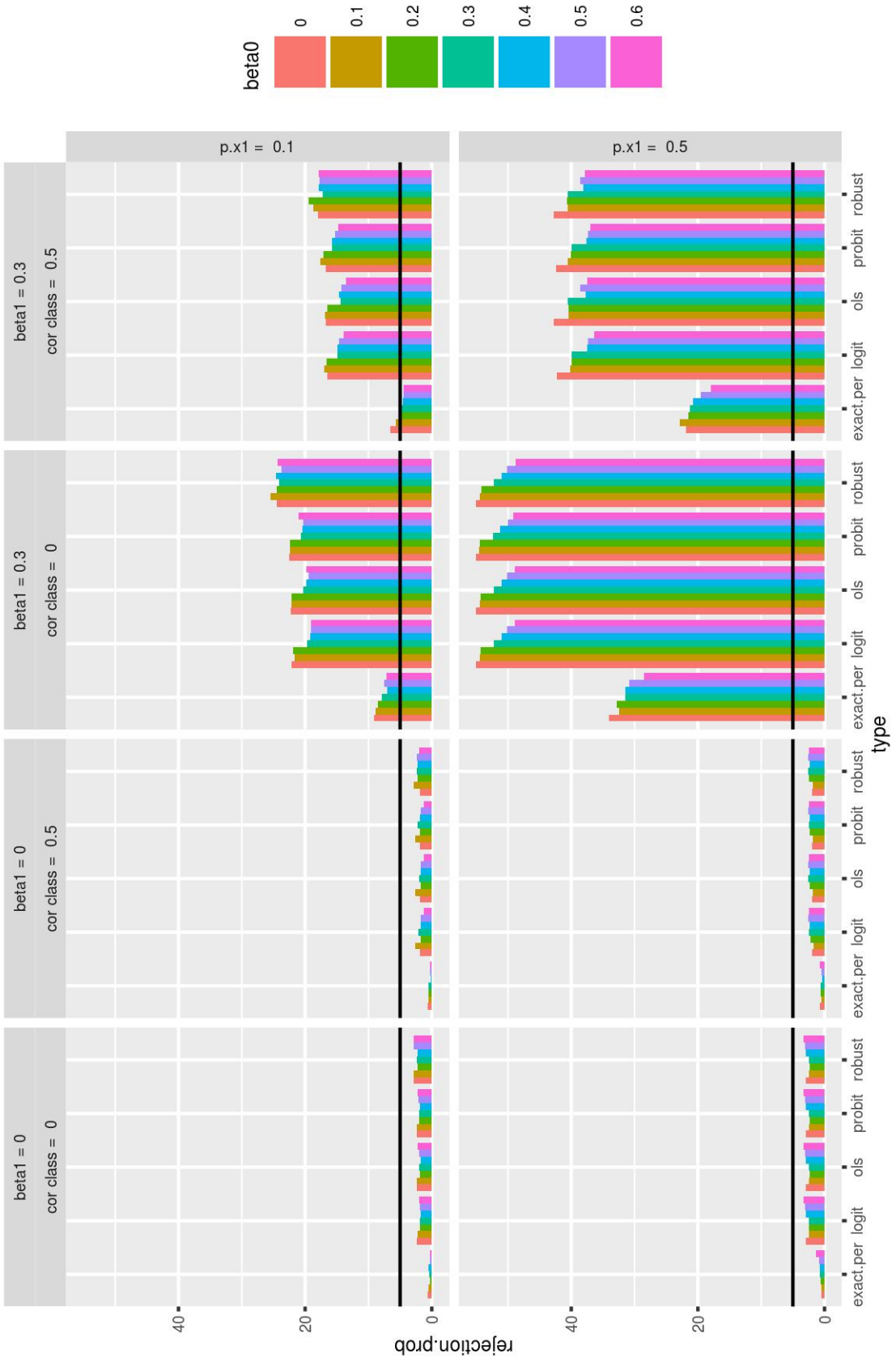

FIGURE 8: small simulation of the logit model at 750 observations

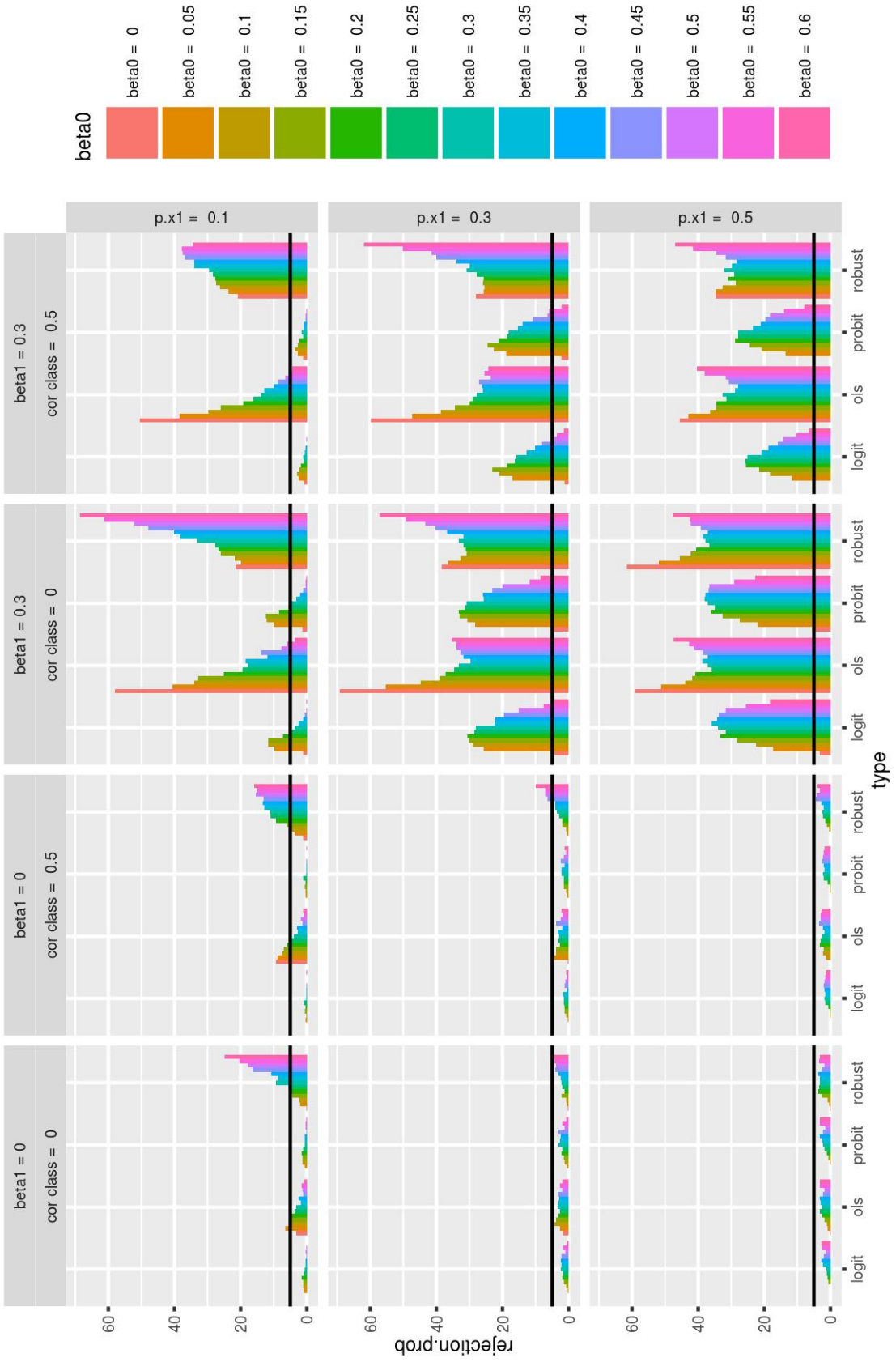

FIGURE 9: large simulation of the LPM model at 30 observations

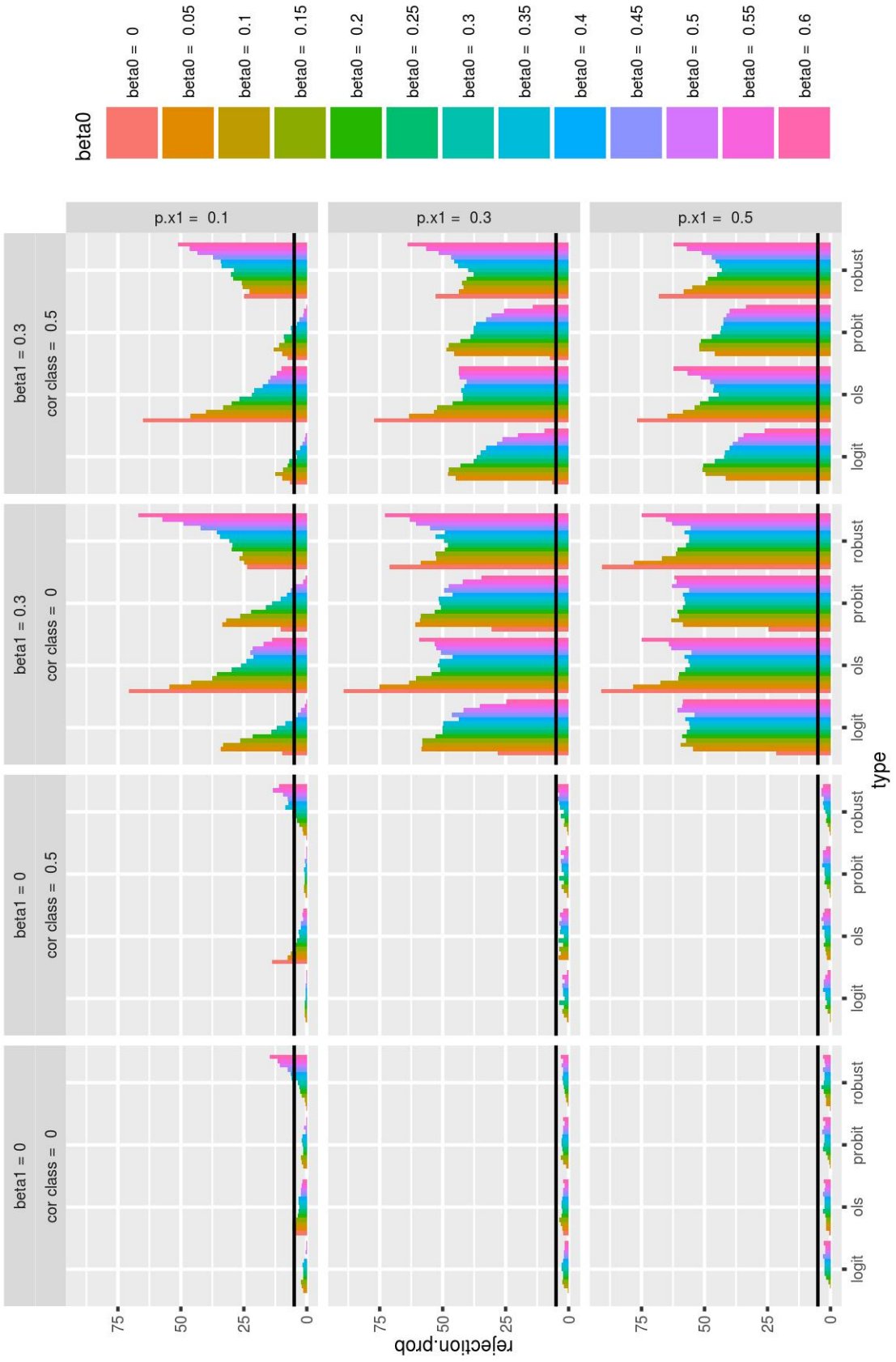

FIGURE 10: large simulation of the LPM model at 50 observations

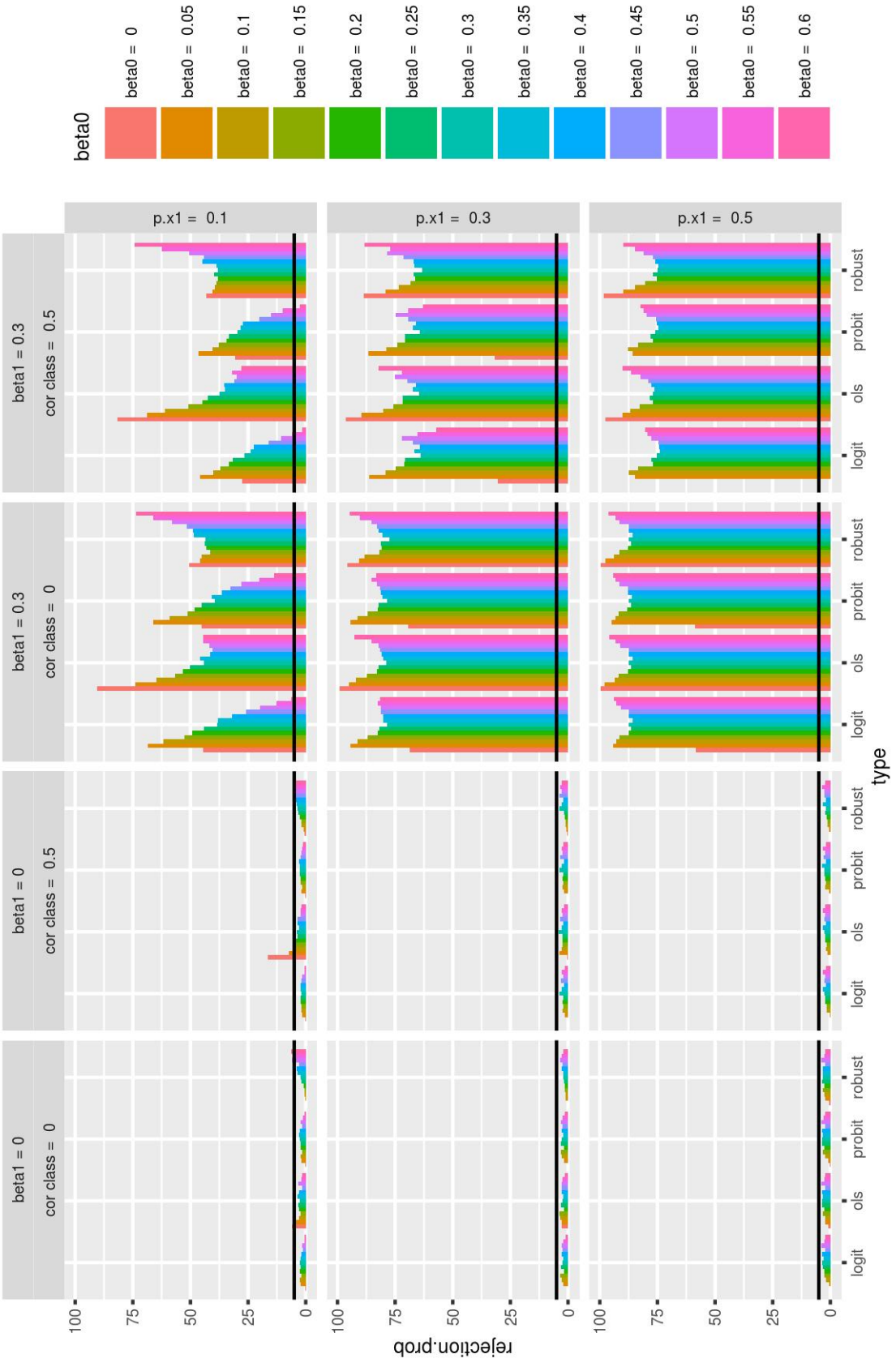

FIGURE 11: large simulation of the LPM model at 100 observations

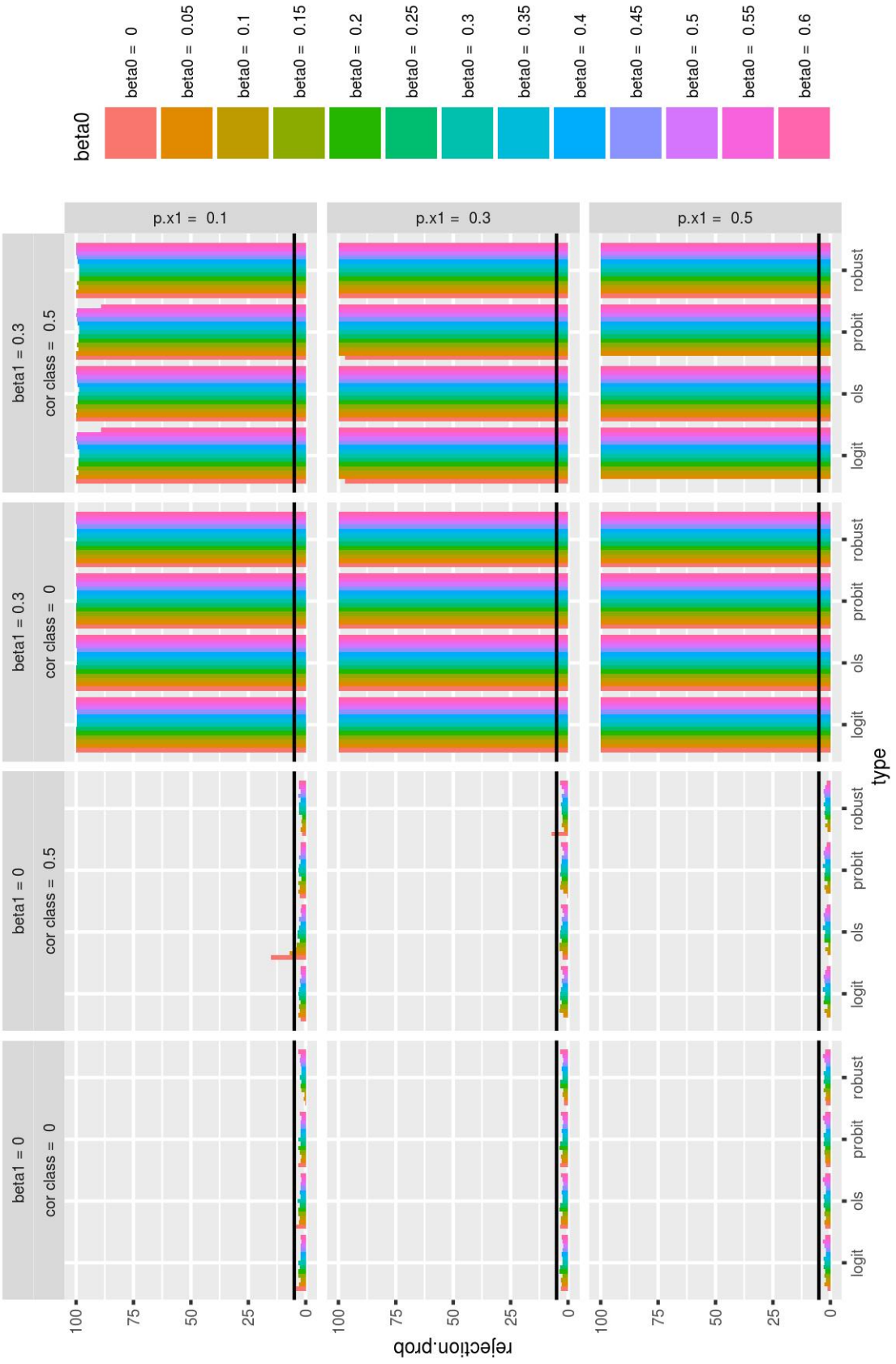

FIGURE 12: large simulation of the LPM model at 750 observations

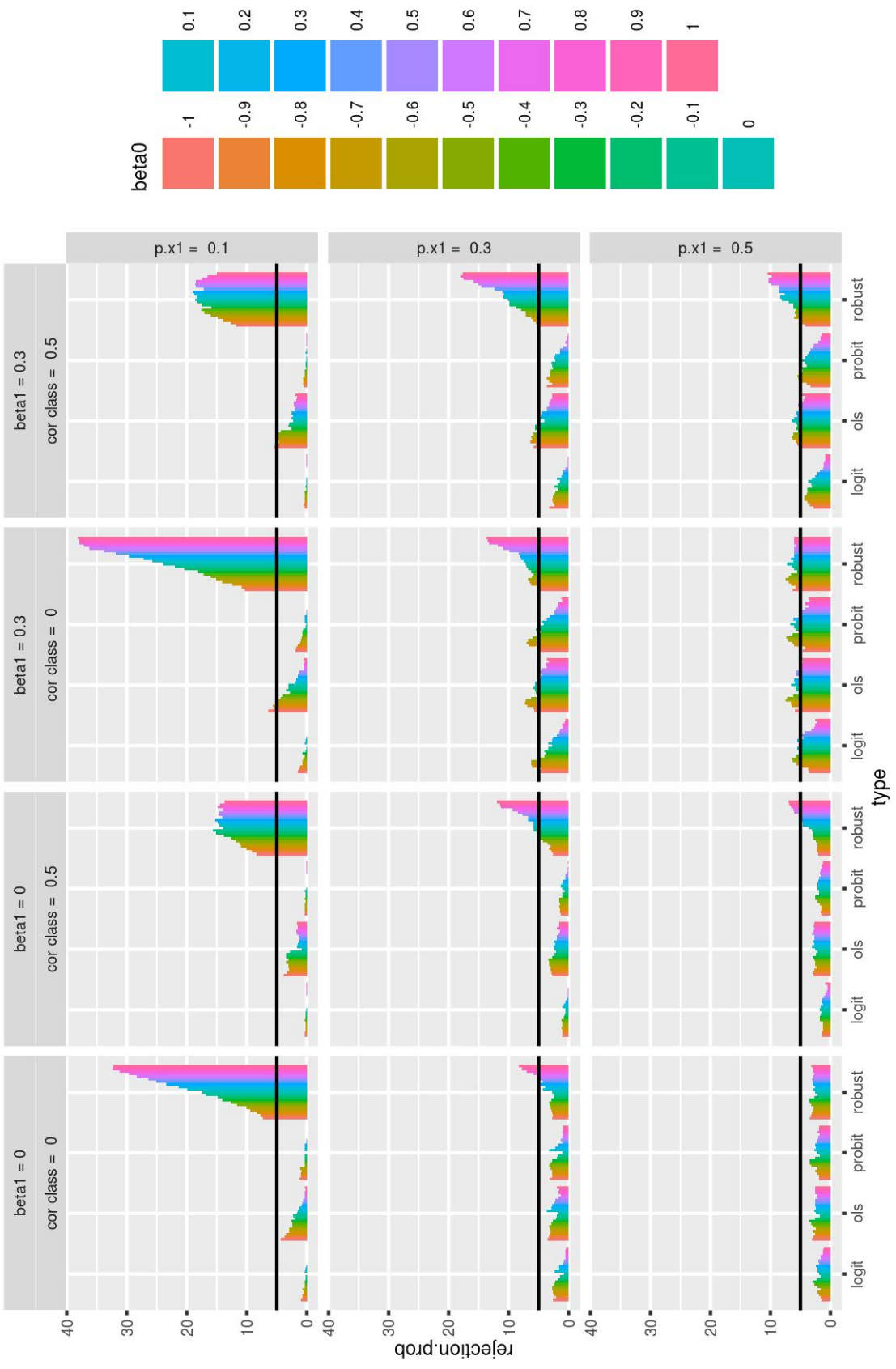

FIGURE 13: large simulation of the logit model at 30 observations

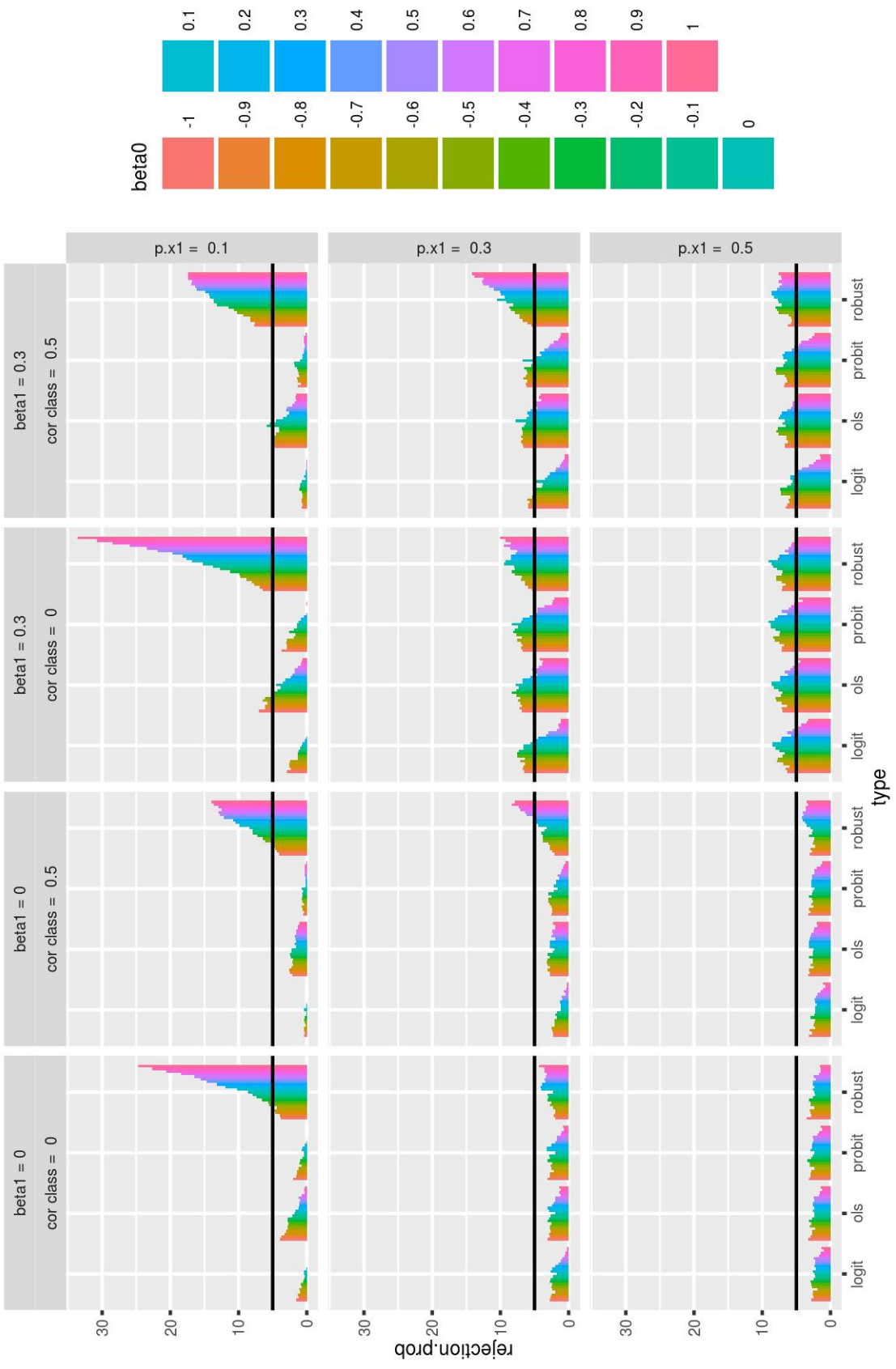

FIGURE 14: large simulation of the logit model at 50 observations

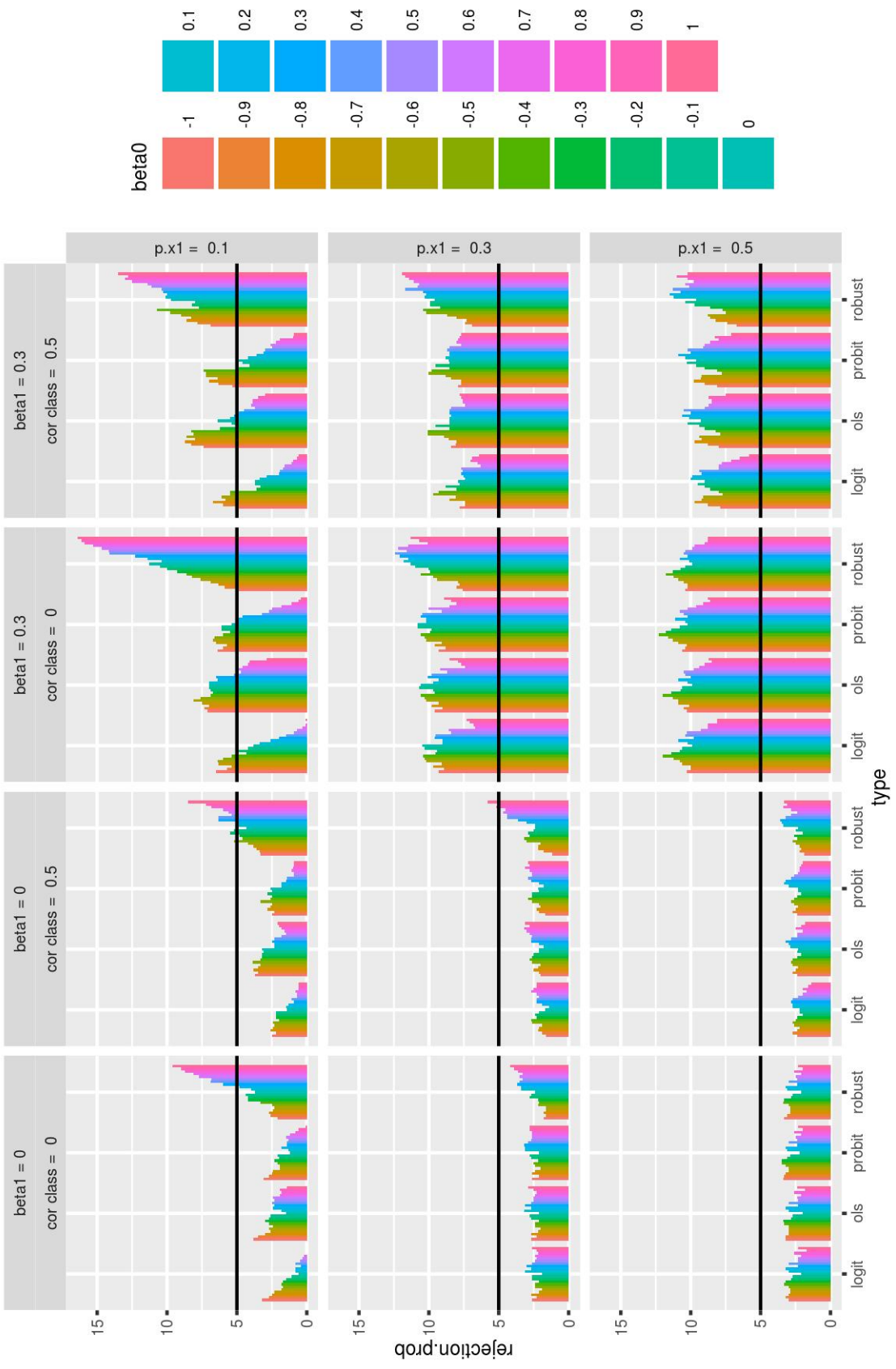

FIGURE 15: large simulation of the logit model at 100 observations

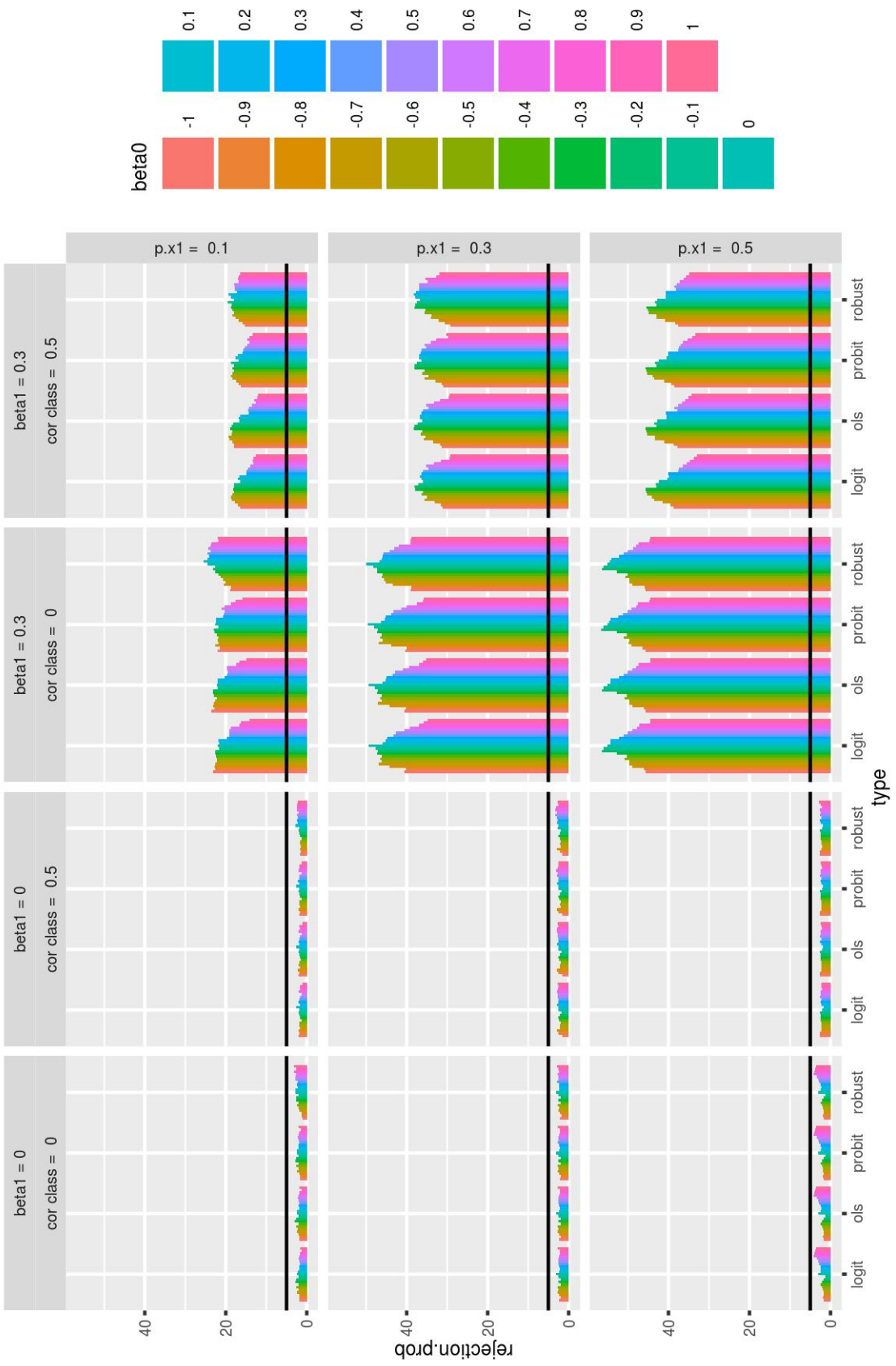

FIGURE 16: large simulation of the logit model at 750 observations

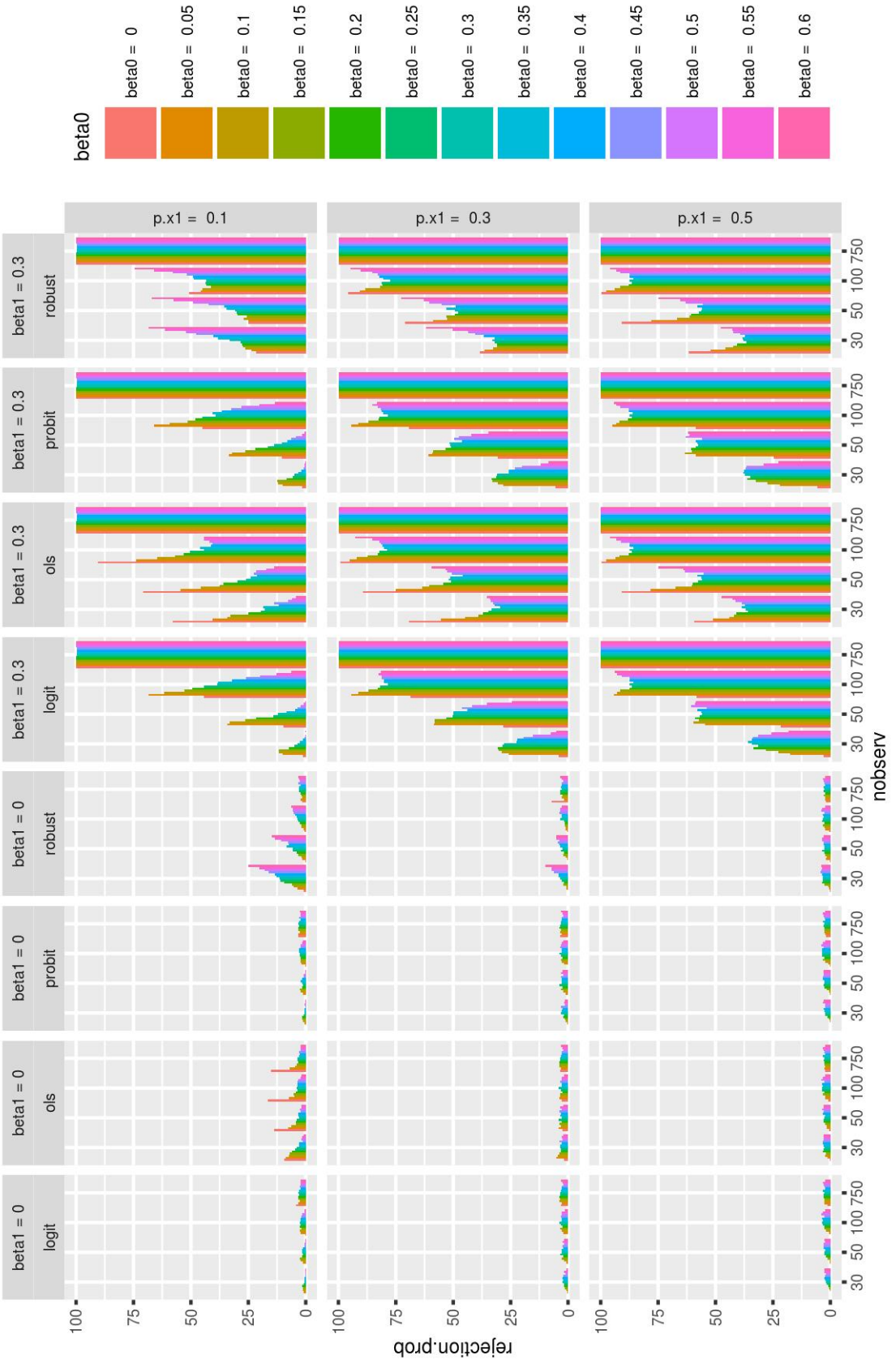

FIGURE 17: Complete overview of the large linear probability model

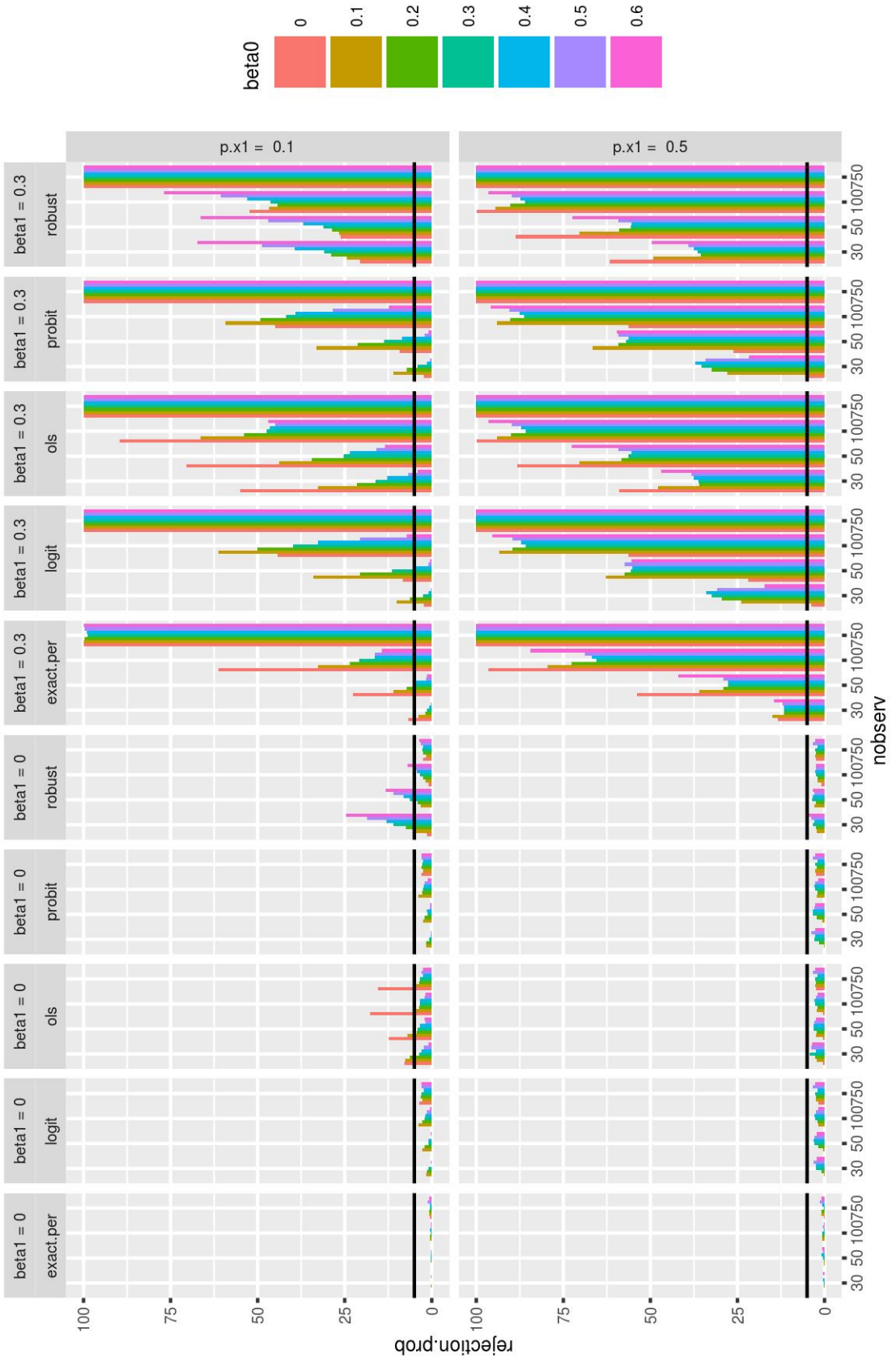

FIGURE 18: Complete overview of the reduced linear probability model

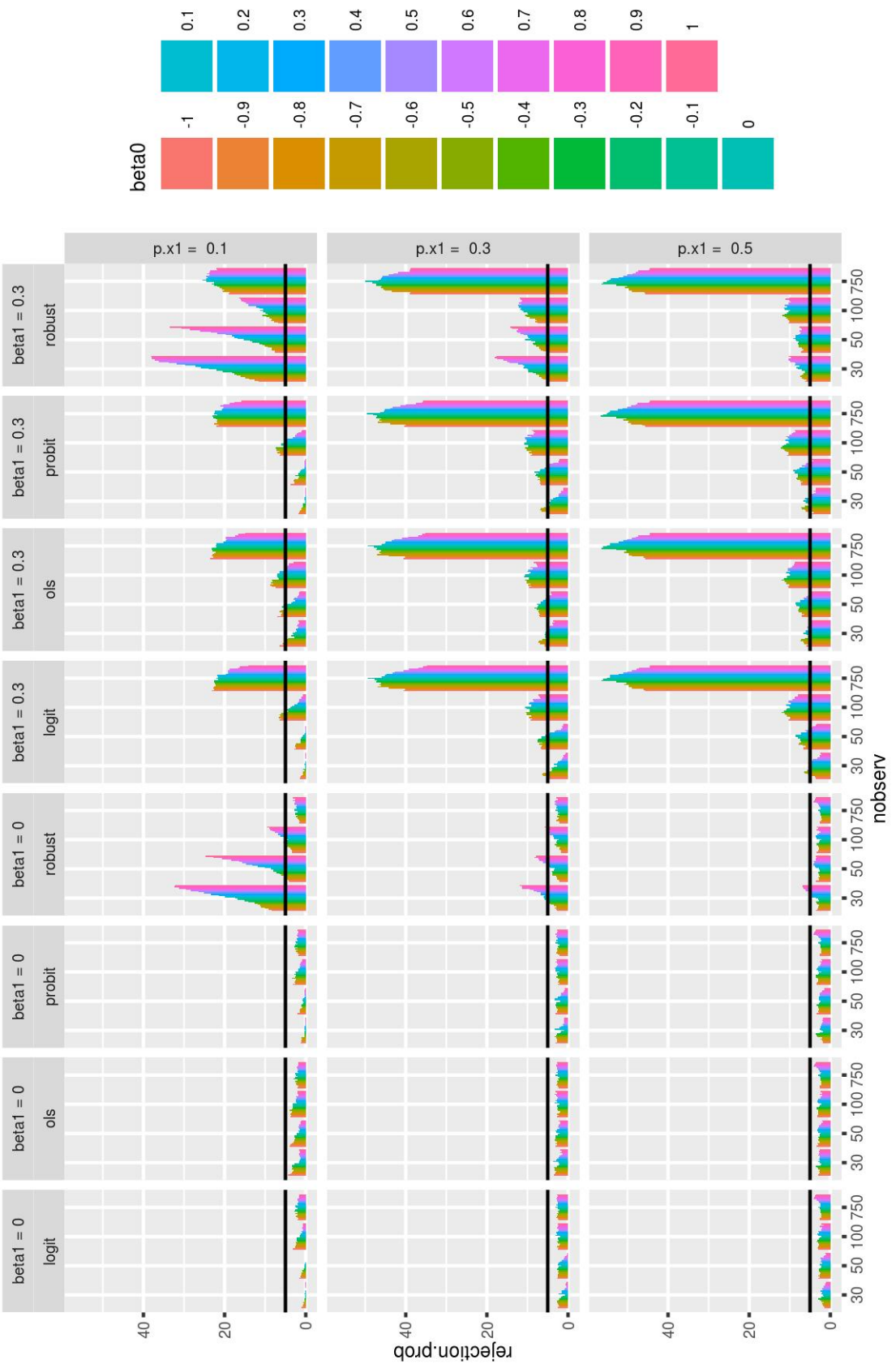

FIGURE 19: Complete overview of the large logit model

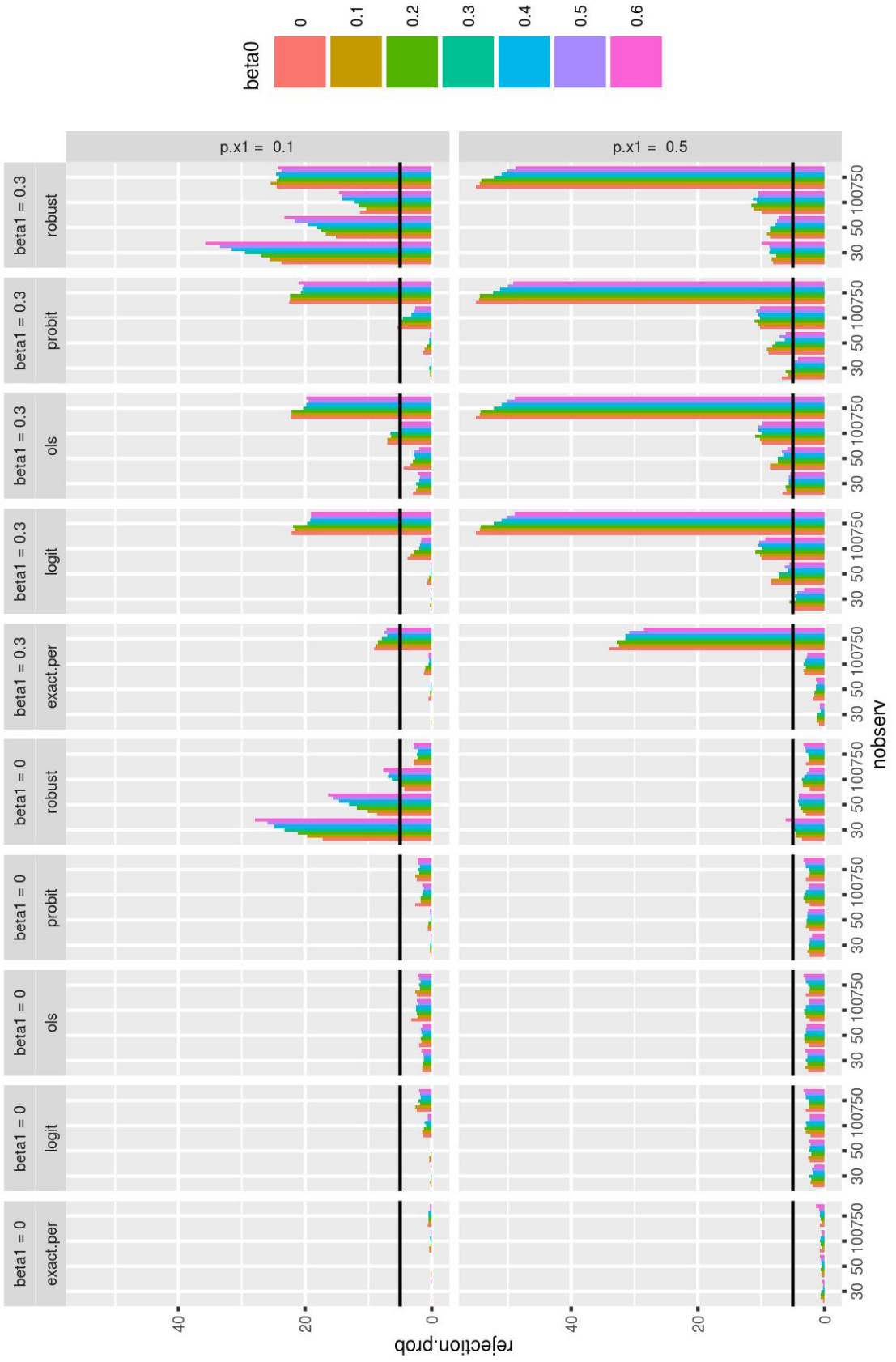

FIGURE 20: Complete overview of the reduced logit model