ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/24000235

Fractional Response Models - A Replication Exercise of Papke and Wooldridge (1996)

Article · January 2009

DOI: 10.5709/ce.1897-9254.50 · Source: RePEc

CITATIONS
32

READS
204

2 authors, including:

Harald Oberhofer
Wirtschaftsuniversität Wien

42 PUBLICATIONS **460** CITATIONS

SEE PROFILE

All content following this page was uploaded by Harald Oberhofer on 22 May 2014.

The user has requested enhancement of the downloaded file.

University of Innsbruck

Working Papers in Economics and Statistics

Fractional Response Models - A Replication Exercise of Papke and Wooldridge (1996)

Harald Oberhofer and Michael Pfaffermayr

2009-02

Fractional Response Models - A Replication Exercise of Papke and Wooldridge (1996)

Harald Oberhofer[†] and Michael Pfaffermayr[‡§]

Abstract

This paper replicates the estimates of a fractional response model for share data reported in the seminal paper of Leslie E. Papke and Jeffrey M. Wooldridge published in the Journal of Applied Econometrics 11(6), 1996, pp.619-632. We have been able to replicate all reported estimation results concerning the determinants of employee participation rates in 401(k) pension plans using standard routines provided in Stata. As an alternative, we estimate a two-part model that is able to cope with the excessive number of boundary values of one in the data. The estimated marginal effects are similar to that derived in that paper. A small scale Monte Carlo simulation exercise suggests that the RESET tests proposed by Papke and Wooldridge in their robust form are useful for detecting neglected non-linearities in small samples.

JEL Codes: C15, C21

Keywords: Replication Exercise, Fractional Response Models, Two-Part Models, Monte Carlo Simulation.

[†]Department of Economics, University of Innsbruck, Universitaetsstrasse 15, 6020 Innsbruck, Austria. Tel.:(+43)512 507 7164. E-mail: harald.oberhofer@uibk.ac.at.

[‡]Department of Economics, University of Innsbruck, Universitaetsstrasse 15, 6020 Innsbruck, Austria. Tel.:(+43)512 507 7359. E-mail: michael.pfaffermayr@uibk.ac.at.

[§]Austrian Institute of Economic Research, PO-Box 91, A-1103-Vienna, Austria.

Financial support from the 'Oesterreichische Nationalbank' (OeNB, grant no. 12831) is gratefully acknowledged.

1 Introduction[1]

In many applications one has to deal with share data confined to the $[0, 1]$ interval and, in addition, with a significant amount of observations of the dependent variable taking on values at the boundaries, 0 or 1. While share data can be handled using log-odds transformed variables, the combination of these two issues is tricky. In their seminal paper Leslie E. Papke and Jeffrey M. Wooldridge (1996) propose a fractional response model that extends the generalized linear model (GLM) literature from statistics.[2] They introduce a quasi-maximum likelihood estimator (QLME) to obtain a robust method to estimate fractional response models without an *ad hoc* transformation of boundary values. The paper shows that the proposed QLME is consistent as long as the conditional mean function is correctly specified (see their equation 4). In addition, the authors introduce robust Ramsey RESET tests for the correct specification of the mean function. Lastly, the paper provides an application of this estimation procedure, estimating a model of employee participation rates in 401(k) pensions plans.

Papke and Wooldridge (1996) consider the following model for the conditional expectation of the fractional response variable:

$$E(y_i|\mathbf{x}_i) = G(\mathbf{x}_i\beta), \qquad i = 1, ..., N, \tag{1}$$

[1]We are grateful to Jeffrey M. Wooldridge for his detailed comments and suggestions on a previous draft.

[2]In a recent paper, Papke and Wooldridge (2008) introduce fractional response models for panel data.

where $0 \leq y_i \leq 1$ denotes the dependent variable and (the $1 \times k$ vector) $\mathbf{x}_i$ refers to the explanatory variables of observation i. Typically, $G(.)$ is a distribution function like the logistic function $G(z) = \exp(z)/(1 + \exp(z))$ which maps z to the $(0, 1)$ interval. The authors follow McCullagh and Nelder (1991) and suggest to maximize the Bernoulli log likelihood with the individual contribution given by:[3]

$$l_i(\beta) = y_i \log[G(\mathbf{x}_i\beta)] + (1 - y_i)\log[1 - G(\mathbf{x}_i\beta)]. \tag{2}$$

In this formulation of the likelihood function, the number of draws (here the number of eligible employees of each firm) drops out, since it does not depend on the parameters. Rather, the share of successes, i.e. the participation rate, enters the likelihood directly (see McCullagh and Nelder, 1991, p. 114).

The consistency of the QLME follows from Gourieroux et al. (1984), since the density upon which the likelihood function is based on is a member of the linear exponential family and because of the assumption that the conditional expectation of y_i is correctly specified. In fact, the QLME is $\sqrt{N}$-asymptotically normal regardless of the distribution of y_i conditional on $\mathbf{x}_i$. Papke and Wooldridge (1996) provide valid (robust) estimators of the asymptotic variance of β based on the well known sandwich formula (see Cameron and Trivedi, 2005) and the non-linear conditional mean $G(.)$.

Papke and Wooldridge (1996) introduce and apply extended Ramsey RESET tests for $H_0 : \gamma_1 = 0, \gamma_2 = 0$ in the augmented model $G(\mathbf{x}_i\beta +$

[3] Papke and Wooldridge (1993) also consider the case where group size is known and given by n_i. They show that in this case the conditional likelihood for observation i is the same as in (2), but it is weighted by n_i.

$\gamma_1(\mathbf{x}_i\beta)^2 + \gamma_2(\mathbf{x}_i\beta)^3)$. Their first RESET test is non-robust as it maintains the GLM variance assumption: $Var(y_i|\mathbf{x}_i) = \sigma^2 G(\mathbf{x}_i\beta)[1 - G(\mathbf{x}_i\beta)]$. The robust RESET test only requires the correct specification of the conditional mean. Details on calculating the RESET test are given on pages 623-625 in their paper.

In many applications, including the present one, there is a significant share of boundary values. Taking the data generating process in the paper of Papke and Wooldridge (1996) literally, one would use the number of eligible employees as the number of Bernoulli draws. However, in the full sample the mean firm size is 4621 and the median firm size is 628. Basing the Bernoulli draws on these numbers makes a boundary value of 1 in PRATE a very rare event. Thus, in the presence of 42.7 percent boundary values of 1 in the data, it seems plausible to assume that firms that exhibit 100 percent participation rates in their pension plans behave differently and are not well described by the Bernoulli model.

Following Wooldridge (2002, Problem 19.8) and Ramalho and Vidigal da Silva (2008), we alternatively consider a two-part model that accounts for an excessive number of boundary values of ones.[4] We define:

$$y_i^* = \begin{cases} 0 & \text{if } y_i \,\epsilon\, [0,1) \\ 1 & \text{if } y_i = 1 \end{cases} \tag{3}$$

and assume for the first part of the model that $P(y_i^* = 1|\mathbf{x}_i) = P(y_i = 1|\mathbf{x}_i) = G(\mathbf{x}_i\gamma)$, where $G(\mathbf{x}_i\gamma)$ denotes the cumulative logistic distribution

[4] See also Pohlmeier and Ulrich (1995) for an early application of a two-part model for count data.

function. The second part is the fractional response model that refers to observations $y_i \epsilon [0, 1)$. Then, the conditional mean of the two-part model is specified as:

$$\begin{aligned} E[y_i|\mathbf{x}_i] &= P(y_i^* = 0|\mathbf{x}_i)E[y_i|\mathbf{x}_i, y_i^* = 0] + P(y_i^* = 1|\mathbf{x}_i) \\ &= (1 - G(\mathbf{x}_i\gamma))G(\mathbf{x}_i\beta) + G(\mathbf{x}_i\gamma). \end{aligned} \quad (4)$$

The marginal effects of the explanatory variables can be derived as:

$$\begin{aligned} \frac{\partial E[y_i|\mathbf{x}_i]}{\partial x_{ij}} &= \frac{\partial P(y_i^* = 1|\mathbf{x}_i)}{\partial x_{ij}}(1 - E[y_i|\mathbf{x}_i, y_i^* = 0]) \\ &+ (1 - P(y_i^* = 1|\mathbf{x}_i))\frac{\partial E[y_i|\mathbf{x}_i, y_i^* = 0]}{\partial x_{ij}}. \end{aligned} \quad (5)$$

This model allows the explanatory variables to affect the outcome $y_i = 1$ and the size of y_i at $y_i \epsilon [0, 1)$ in a different way. More importantly, the explanatory variables in the first and second part of the model need not be the same. Under this specification (quasi) maximum likelihood estimation is straight forward, since it separates into the estimation of the logit model explaining $P(y_i^* = 1|\mathbf{x}_i)$ using all observations and the estimation of parameters of the conditional density $f(y_i|\mathbf{x}_i, y_i^* = 0)$ based only on the observation with $y_i < 1$.[5] In fact, the second part is defined as the fractional response model introduced above. Again the critical assumption to obtain consistent

[5]Actually, the conditional distribution of $y_i|\mathbf{x}_i, y_i^* = 0$ is derived from the unconditional binomial distribution through division by $1 - G(\mathbf{x}_i\beta)^{n_i}$ so that $f(y_i|\mathbf{x}_i, y_i^* = 0) = \begin{pmatrix} n_i \\ n_i y_i \end{pmatrix} G(\mathbf{x}_i\beta)^{n_i y_i}(1 - G(\mathbf{x}_i\beta))^{n_i(1-y_i)}(1 - G(\mathbf{x}_i\beta)^{n_i})^{-1}$, where n_i is the number of eligible employees (see also Papke and Wooldridge, 1993). In case n_i is large, the last term will be approximately 1. In the following we neglect this term.

parameters is the correct specification of the conditional mean, which now requires the correct specification of $P(y_i^* = 1|\mathbf{x}_i)$ and $E[y_i|\mathbf{x}_i, y_i^* = 0]$.

2 The replication exercise

In their application, Papke and Wooldridge (1996) are interested in an econometric model of participation rates in 401(k) pension plans. These are employer sponsored pension plans, where employees are permitted to make pre-tax contributions and the employer may match part of the contribution. The dependent variable ($PRATE$) is defined as the number of active pension accounts divided by the number of employees eligible to participate for a sample of US manufacturing firms. The explanatory variables of their model include the plan match rate of the employer ($MRATE$), log size of the firm measured in terms of employment and the square of it, the plan's age and its square and a dummy called $SOLE$ that indicates whether the 401(k) pension plan is the only one offered by the firm. To sum up, the following specification is estimated in Tables II and III of the paper:

$$\begin{aligned} E(PRATE|\mathbf{x}) &= G(\beta_1 + \beta_2 MRATE + \beta_3 \log(EMP) + \beta_4 \log(EMP)^2 \\ &+ \beta_5 AGE + \beta_6 AGE^2 + \beta_7 SOLE). \end{aligned} \tag{6}$$

The linear specification assumes $G(z) = z$, while in the non-linear fractional response regression G(.) is specified as logistic function, i.e. $G(z) = \exp(z)/(1 + \exp(z))$. In a second specification the authors additionally include $MRATE^2$ as explanatory variable.

Tables II and III in the paper report simple OLS estimates and the QMLE of the fractional response model. The estimates in Table II use only observations with $MRATE < 1$, while the estimation results in Table III are based on all observations. There are no zeros in the dependent variable, but 42.7 percent of the sample refer to firms, where all employees participate in 401(k) pension plans so that $PRATE = 1$.

In their Table II, the authors report a significant positive impact of the firm's matching rate. Log firm size and the age of the plan enter non-linearly. The impact of log firm size is significantly negative, but increases for large firms. AGE turns out significantly positive, but also with a decreasing effect. Lastly, the variable $SOLE$ is insignificant.

In Table II of the paper the OLS estimates are rejected by both the non-robust and the robust RESET test, suggesting that the linear model misses important non-linearities. However, the signs of the estimated parameters are the same for the OLS and the QLME estimates for all variables. There is an important difference between the OLS and QMLE estimates, since the RESET tests (both in their robust and non-robust version) do not reject the fractional response model. Further, the R^2 of the fractional response model is by 6 percentage points higher as compared to the linear model.

From an economic point of view the difference between the two models is important, since the fractional response model implies a decreasing marginal effect of $MRATE$. The authors also conclude that simply adding ($MRATE^2$) in the linear model is not sufficient to capture this non-linearity. The results in their Table III show that the basic story does not change if the models are estimated over the entire sample. The only noticeable difference is that the quadratic term in $MRATE$ is now significant and that the RESET test does not reject the fractional response model that includes $MRATE^2$, while it rejects the baseline specification.

The authors estimated and tested the fractional response model using GAUSS-code. We could replicate and verify their estimated results easily using now available standard Stata code and specifically the Stata procedure glm with options fam(bin), link(logit) and scale(x2) for non-robust standard errors and the options fam(bin), link(logit) and rob for robust standard errors. In this way, we have been able to replicate each and any entry in Tables II and III. So the fractional response model is attractive as it can be estimated easily using standard econometric software.[6]

We also estimated the two-part model using the basic specification of Papke and Wooldridge (1996) reported in the first two columns of their Table II. As noted above, these estimates exclude observations with $MRATE > 1$. For comparison we reproduce the corresponding estimates in Table 1. In the logit model of the the two-part model the same variables that enter the fractional response model determine whether all employees participate in the 401(k) pension plans or not. Almost all explanatory variables are significant and for $MRATE, log(EMP)$ and $log(EMP)^2$ we obtain the same signs as in the fractional response model. In contrast to the results of the fractional response model, AGE turns out insignificant, while AGE^2 is positive at a p-value slightly higher than 0.05. The variable $SOLE$ is significantly positive, which is also in contrast to the estimate in the fractional response model.

[6]The Stata code is available upon request from the authors.

The second part fractional response model uses the observations with $PRATE < 1$. With exception of the significant negative impact of $SOLE$, we obtain qualitatively similar results as Model 2 in Table II of Papke and Wooldridge (1996). However, in quantitative terms the parameter estimates are quite different. The fit of the two-part model is comparable to the original estimates with R^2 amounting to 0.153.[7] However, both the robust and the non-robust RESET tests are rejected indicating possible misspecification of the second part fractional response model.

*** Table 1 ***

The main advantage of the of the fractional response model and the two-part model is their ability to capture non-linearities, especially the decreasing effect of the matching rate. Table 2 reproduces the marginal effects of the matching rate of the estimated model of columns 1 and 2 in Table II in the paper. For this $SOLE$ is set to 0, $AGE = 13$ and $EMP = 200; 4,620; 100,000$. The partial effects are computed at the values 0, 0.5 an 1 of the matching rate. While the marginal effect under the linear model amounts to 0.156, it is diminishing for the fractional response model and the two-part model. The fractional response model implies an increase in $PRATE$ by 2.9 percentage point as a response to an increase in MRATE from 0 to 0.1. Under the two-part-model the effect is smaller and amounts to 2.1 percentage points. Conversely, at $MRATE = 1$ the marginal effect of the two-part model is 1.3 percent as compared to 1.2 percent implied by the

[7]Similar to the R^2 of the non-linear fractional response model in Papke and Wooldridge (1996) the R^2 of the tow-part model is based on the predicted values of all observations including the boundary values.

two-part model. Generally, the two-part model leads to somewhat smaller marginal effects at low values of $MRATE$, but to a less pronounced decrease of the marginal effects as $MRATE$ becomes larger.

*** Table 2 ***

When looking at the in-sample predictions of the estimated models, we find two puzzling results. First, it can easily be seen from the specification of the conditional mean under the logistic link assumption, i.e. $G(z) = \exp(z)/(1 + \exp(z))$, that both considered models rule out values of 1 in the dependent variable. Put differently, the models by definition always predict a value lower than one for those observations of $PRATE$ that fall on the boundary 1.

Table 3 below calculates the mean of the residuals resulting from the estimates in Table II in the paper as well as for the two-part model within each quintile of $PRATE$ and, separately, for the values on the boundary cases with $PRATE = 1$. As expected the residuals are positive for the values of $PRATE = 1$ for both, the OLS estimation and the QMLE. Also, there is virtually no difference between the considered models.

*** Table 3 ***

Secondly, we find systematic effects in the residuals of both the linear and the non-linear models. For the observations with $PRATE < 1$ all considered models overpredict in the lower three quintiles of $PRATE$ and underpredict in the two upper ones. The same pattern is found for the residuals of the two-part model. In fact, the residuals of the four estimated models in Table

II of Papke and Wooldridge (1996) and that of the two-part model are highly correlated with correlations as high as 0.99. As in many applications, there is only a minor difference between the linear and non-linear models in terms of root mean squared prediction error and using a logistic link function leads to only small improvements.

3 A small scale Monte Carlo exercise on the performance of the proposed RESET tests

To investigate the performance of the proposed RESET test, we set up a small Monte-Carlo simulation exercise. We generate Bernoulli random variables using the predicted participation rates of column 4 of Table II in the paper assuming that the reported parameters are the true ones (see Equations 2 and 3). Since the Bernoulli random variable measures the number of successes in n trials, we set $n = 10$ in the first experiment to generate a large share of ones (approximately 20 percent). To obtain share variables we divided the resulting Bernoulli random number by n (and similarly in the other experiments). The drawback of this design is that we obtained only 9 different realizations of the generated random variable. Experiment 2 sets $n = 1000$, while the Experiment 3 allows n to vary and assumes $n = EMP$. The latter experiment introduces additional heterogeneity and violates the nominal variance assumption, since the log of the number of employees and its square are used as regressors (see equation 6 in the paper and the discussion below). Experiments 4 and 5 are the same as Experiments 2 and 3, but assume that the estimated logit model is the true data generating precess for the boundary values. We generate a uniformly distributed random variable and set the simulated value of $PRATE$ to 1 if this random variable is lower than the predicted probability as implied by the logit model. Then, we apply the two-part model and estimate a fractional response model using only the non-boundary values.

We calculated the bias and the root mean squared error (RMSE) of the estimated parameters resulting from 10000 replications of the Monte Carlo experiments. Following Kelejian and Prucha (1999), we define the bias as $med(\widehat{\theta} - \theta)$ and RMSE by $(Bias^2 + (IQ/1.35)^2)^{0.5}$, where IQ is the interquantile range. In all experiments the estimated parameters are virtually unbiased. With exception of experiment 1 the RMSEs are quite small. In particular, they are considerable smaller than the standard errors reported in the paper, which come from estimated models with a significant share of boundary values.

To obtain the power curves of the RESET tests, we assume that γ_1 takes values in $\{-0.025, -0.015, -0.005, 0, 0.005, 0.015, 0.025\}$ and γ_2 is 1/5-th of γ_1. Since, the power turned out very low in experiment 1, we scaled the γ- values for this experiment by a factor 10. In each experiment, we added $\gamma_1(\mathbf{x}_i\beta)^2 + \gamma_2(\mathbf{x}_i\beta)^3$ to the linear predictor. Therefore, at $\gamma_1 = \gamma_2 = 0$ the share of rejections in the respective experiment is an estimate of the size of the RESET tests and at $\gamma_1 \neq 0$ or $\gamma_2 \neq 0$ on obtains the power of the test.

*** Table 4 ***

*** Table 5 ***

In Tables 4 and 5 the simulated size (in bold figures) and power of the RESET tests are displayed for a nominal size of 0.05. For each value of $g1$ the first line in the tables referes to the non-robust RESET test and the second line to the robust one. While the RESET tests are properly sized under Experiment 1 and 2, we find the correct size only for the robust RESET test under Experiment 3 as one would expect. Although the construction of the share variable often remains unobserved empirically, its calculation matters for estimating and testing of fractional response models as argued by Papke and Wooldridge (1996). In this respect, our findings confirm the discussion of the RESET tests in the paper. The results of Experiments 4 and 5 referring to the two-part confirm the findings of experiments 2 and 3.[8]

Generally, the reset RESET tests exhibit enough power to detect neglected non-linearities. Only at small n as in experiment 1 the power is not satisfactory. For this experiment we get power figures comparable to the other experiments, when scaling γ_1 and γ_2 by a factor 10. The highest power of the RESET test is observed when either γ_1 or γ_2 is zero and the corresponding non-zero value is high in absolute value. However, at large absolute values of γ_1 and γ_2 but different signs the power of the the RESET test turns out very low. This holds for the robust and non-robust version of the RESET-test.

[8] We also investigated the case where a fractional response model using all observations is estimated in case of a large share of boundary values. The results, which are available upon request from the authors, indicate that in this case RESET tests are oversized and their power tends to be considerably lower, even when taking into account that the tests are oversized. However, this has to be expected since this set-up violates the conditional mean assumption.

4 Conclusions

This paper has replicated the results of the seminal paper of Leslie E. Papke and Jeffrey M. Wooldridge (1996) concerning a fractional response model for employee participation rates in 401(k) pension plans in US-manufacturing firms. Using the now available standard Stata code, we have been able to replicate each and any estimation result of the paper.

An important feature of their dependent variable is that more than 40 percent of these data are ones, indicating full employee participation. To cope with the excessive number of boundary values, we additionally estimated a two- part model. The first part models the probability of a boundary observation by a simple logit model. The second part refers to non-boundary values and is estimated by the same fractional response model. The estimation of the second part model yields somewhat different results. However, the marginal effects of the matching rate that take both parts into account are of comparable size. They are slightly smaller and the diminishing impact of the matching rate is less pronounced. Therefore, in the presence of a high share of boundary values, the two-part model is a useful alternative to the fractional response model. Moreover, it is as easy to calculate with available standard software.

Looking at the in-sample predictions of the estimated model reveals some puzzles. First, for all observations with a boundary value of one in the dependent variable, the corresponding predictions by definition are smaller than one. Second, in all estimated models there are systematic differences in the residuals left, depending on the size of the participation rate. A small scale Monte Carlo simulation exercise confirms that the proposed RESET tests are useful for detecting neglected non-linearities in small samples. In their robust form the RESET tests are always properly sized and equipped with power in almost all considered cases.

References

Cameron C.A, Trivedi P. K. 2005. *Microeconometrics : Methods and Applications*. Cambridge University Press: Cambridge, UK.

Gourieroux C., Monfort A, Trognon A. 1984. Pseudo-maximum likelihood methods: theory. *Econometrica* **52** : 681-700.

McCullagh P, Nelder J. A. 1991. *Generalized Linear Models*. 2nd ed. Chapman and Hall: London, UK.

Kelejian H.H, Prucha I. R. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **40** : 509-533. DOI: 10.1111/1468-2354.00027

Papke L. E, Wooldridge J. M 1993. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. National Bureau of Economic Research Technical Working Paper No. 147.

Papke L. E, Wooldridge J. M 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* **11** : 619-632. DOI: 10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1.

Papke L. E, Wooldridge J. M 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* **145** : 121-133. DOI: 10.1016/j.jeconom.2008.05.009.

Pohlmeier, W.and Ulrich V. 1995. An econometric model of the two-part decisionmaking process in the demand for health care. *Journal of Human Resources* **30** : 339-361.

Ramalho, J.J.S., J.V. Silva 2008. A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. *Quantitative Finance* forthcoming.

Wooldridge J. M 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT: Cambridge, MA.

Table 1: Results for the Restricted Sample

Variable	(1) OLS	(2) QMLE	(3) Two-Part Model Logit	(4) Two-Part Model QMLE
$MRATE$	0.156	1.390	1.504	0.895
	(0.012)	(0.100)	(0.160)	(0.089)
	[0.011]	[0.107]	[0.166]	[0.097]
$\log(EMP)$	−0.112	−1.002	−0.852	−0.690
	(0.014)	(0.111)	(0.200)	(0.092)
	[0.013]	[0.110]	[0.197]	[0.094]
$\log(EMP)^2$	0.052	0.054	0.039	0.037
	(0.001)	(0.007)	(0.013)	(0.006)
	[0.001]	[0.007]	[0.013]	[0.006]
AGE	0.006	0.050	−0.006	0.054
	(0.001)	(0.009)	(0.014)	(0.007)
	[0.001]	[0.009]	[0.016]	[0.006]
AGE^2	−0.000	−0.001	0.001	−0.001
	(0.000)	(0.000)	(0.000)	(0.000)
	[0.000]	[0.000]	[0.000]	[0.000]
$SOLE$	−0.000	0.008	0.585	−0.215
	(0.006)	(0.047)	(0.078)	(0.039)
	[0.006]	[0.050]	[0.078]	[0.040]
ONE	1.213	5.058	2.316	3.420
	(0.051)	(0.427)	(0.740)	(0.354)
	[0.048]	[0.421]	[0.741]	[0.352]
Observations	3,784	3,784	3,784	2,489
SSR	93.666	92.695	-	92.506
SER	0.157	0.438	-	0.390
R^2	0.142	0.152	-	0.153
RESET	39.55	0.606	-	29.55
	(0.000)	(0.738)	-	(0.000)
Robust RESET	45.36	0.782	-	23.85
	(0.000)	(0.676)	-	(0.000)

Notes: See Table II in Papke and Wooldridge (1996). In the logit model the dependent variable is one if all employees participate in the 401(k) pension plan and zero otherwise. The QMLE of the two-part model is estimated only for $PRATE < 1$.

Table 2: Marginal Effects from QMLE and Two-Part Model

	$EMP = 200$		$EMP = 4,620$		$EMP = 100,000$	
MRATE	QMLE	Two-Part	QMLE	Two-Part	QMLE	Two-Part
0	0.172	0.164	0.288	0.214	0.273	0.195
0.5	0.100	0.113	0.197	0.176	0.182	0.157
1	0.054	0.063	0.118	0.127	0.106	0.115

Table 3: Residuals from OLS, QMLE and Two-Part Model

	Prate	OLS	QMLE	Two-Part
1^{st} Quintile	0.552	-0.284	-0.280	-0.280
2^{nd} Quintile	0.720	-0.111	-0.111	-0.111
3^{rd} Quintile	0.802	-0.035	-0.038	-0.037
4^{th} Quintile	0.876	0.032	0.030	0.031
5^{th} Quintile	0.949	0.087	0.084	0.085
$PRATE = 1$	1.000	0.127	0.127	0.127
Total	0.937	0.000	0.000	-0.000

Note: The figures are based on the means within the respective quintile.

Table 4: Power and Size of the RESET tests under the Fractional Response Model, Experiments 1,2 and 3

Experiment		$g2$	-0.050	-0.030	-0.010	0.000	0.010	0.030	0.050
	$g1$								
1	-0.250		1.000	1.000	0.939	0.250	0.993	0.549	0.292
1	-0.250		1.000	1.000	0.919	0.227	0.992	0.536	0.342
1	-0.150		1.000	0.363	1.000	0.989	0.668	0.103	0.791
1	-0.150		1.000	0.331	1.000	0.986	0.644	0.133	0.835
1	-0.050		0.999	1.000	0.754	0.192	0.049	0.565	0.965
1	-0.050		0.999	1.000	0.716	0.164	0.057	0.626	0.975
1	0.000		1.000	0.995	0.229	**0.046**	0.152	0.780	0.984
1	0.000		1.000	0.992	0.191	**0.049**	0.191	0.822	0.990
1	0.050		1.000	0.801	0.053	0.121	0.378	0.890	0.993
1	0.050		1.000	0.747	0.046	0.151	0.434	0.912	0.995
1	0.150		0.813	0.082	0.277	0.545	0.779	0.969	0.997
1	0.150		0.745	0.055	0.317	0.602	0.819	0.977	0.998
1	0.250		0.120	0.225	0.676	0.836	0.924	0.987	0.998
1	0.250		0.070	0.256	0.720	0.868	0.942	0.991	0.999
		$g2$	-0.005	-0.003	-0.001	0.000	0.001	0.003	0.005
	$g1$								
2	-0.025		1.000	1.000	1.000	1.000	0.986	0.409	0.184
2	-0.025		1.000	1.000	1.000	1.000	0.986	0.406	0.183
2	-0.015		1.000	1.000	0.993	0.887	0.473	0.097	0.808
2	-0.015		1.000	1.000	0.993	0.884	0.467	0.098	0.807
2	-0.005		1.000	0.998	0.577	0.164	0.056	0.680	0.998
2	-0.005		1.000	0.998	0.567	0.158	0.054	0.679	0.998
2	0.000		1.000	0.951	0.200	**0.051**	0.190	0.929	1.000
2	0.000		1.000	0.947	0.195	**0.052**	0.193	0.928	1.000
2	0.005		0.999	0.699	0.050	0.153	0.538	0.994	1.000
2	0.005		0.999	0.689	0.050	0.154	0.541	0.993	1.000
2	0.015		0.801	0.093	0.432	0.842	0.984	1.000	1.000
2	0.015		0.791	0.085	0.435	0.844	0.984	1.000	1.000
2	0.025		0.169	0.363	0.966	0.998	1.000	1.000	1.000
2	0.025		0.160	0.362	0.966	0.998	1.000	1.000	1.000
3	-0.025		1.000	1.000	0.985	0.925	0.753	0.264	0.200
3	-0.025		1.000	1.000	0.983	0.916	0.729	0.209	0.083
3	-0.015		1.000	0.988	0.790	0.534	0.291	0.160	0.454
3	-0.015		1.000	0.985	0.753	0.472	0.209	0.059	0.312
3	-0.005		0.991	0.814	0.345	0.181	0.146	0.380	0.814
3	-0.005		0.987	0.769	0.234	0.087	0.051	0.263	0.762
3	0.000		0.953	0.597	0.200	**0.138**	0.197	0.586	0.927
3	0.000		0.935	0.499	0.093	**0.050**	0.098	0.501	0.910
3	0.005		0.851	0.390	0.140	0.179	0.333	0.779	0.977
3	0.005		0.792	0.250	0.049	0.092	0.242	0.740	0.975
3	0.015		0.449	0.158	0.286	0.494	0.721	0.973	0.999
3	0.015		0.283	0.053	0.215	0.440	0.701	0.970	0.999
3	0.025		0.190	0.254	0.687	0.862	0.962	0.997	1.000
3	0.025		0.064	0.199	0.671	0.856	0.962	0.998	1.000

Notes: The DGP is assumed to be Model 4 reported in Table II of Papke and Wooldridge(1996). Bold figures refer to the size of the test, the other ones to the power. For each value of $g1$ the first line in the table referes to the non-robust version of the RESET test and the second line to the robust one.

Experiment 1: Bernoulli random variable scaled by 10.

Experiment 2: Bernoulli random variable scaled by 1000.

Experiment 3: Bernoulli random variable scaled by employment.

Table 5: Power and Size of the RESET tests under the Two-Part Model, Experiments 4 and 5

Experiment		$g2$	-0.005	-0.003	-0.001	0.000	0.001	0.003	0.005
	$g1$								
4	-0.025		1.000	1.000	0.997	0.977	0.862	0.269	0.093
4	-0.025		1.000	1.000	0.997	0.977	0.857	0.264	0.091
4	-0.015		1.000	0.999	0.879	0.609	0.290	0.065	0.434
4	-0.015		1.000	0.999	0.871	0.600	0.280	0.062	0.431
4	-0.005		0.999	0.904	0.328	0.104	0.053	0.354	0.895
4	-0.005		0.999	0.897	0.315	0.097	0.054	0.354	0.893
4	0.000		0.988	0.676	0.119	**0.051**	0.115	0.631	0.974
4	0.000		0.987	0.660	0.112	**0.049**	0.117	0.628	0.973
4	0.005		0.922	0.375	0.051	0.106	0.300	0.859	0.996
4	0.005		0.914	0.358	0.050	0.105	0.298	0.859	0.996
4	0.015		0.430	0.069	0.265	0.553	0.826	0.994	1.000
4	0.015		0.410	0.064	0.263	0.555	0.823	0.994	1.000
4	0.025		0.091	0.245	0.787	0.943	0.991	1.000	1.000
4	0.025		0.086	0.246	0.788	0.943	0.991	1.000	1.000
5	-0.025		1.000	0.993	0.892	0.771	0.588	0.259	0.162
5	-0.025		0.999	0.986	0.854	0.706	0.492	0.152	0.060
5	-0.015		0.990	0.902	0.601	0.411	0.270	0.153	0.274
5	-0.015		0.982	0.854	0.481	0.290	0.149	0.051	0.150
5	-0.005		0.907	0.624	0.284	0.188	0.153	0.243	0.546
5	-0.005		0.851	0.485	0.137	0.070	0.047	0.143	0.453
5	0.000		0.791	0.456	0.205	**0.154**	0.164	0.361	0.689
5	0.000		0.683	0.285	0.073	**0.053**	0.067	0.268	0.623
5	0.005		0.647	0.316	0.160	0.164	0.239	0.517	0.822
5	0.005		0.481	0.147	0.047	0.068	0.141	0.438	0.788
5	0.015		0.344	0.169	0.222	0.334	0.486	0.804	0.961
5	0.015		0.151	0.052	0.146	0.269	0.432	0.780	0.954
5	0.025		0.201	0.225	0.473	0.636	0.785	0.956	0.995
5	0.025		0.061	0.150	0.436	0.614	0.767	0.956	0.995

Notes: The DGP is assumed to be Model 4 reported in Table II of Papke and Wooldridge(1996). Bold figures refer to the size of the test, the other ones to the power. For each value of $g1$ the first line in the table referes to the non-robust version of the RESET test and the second line to the robust one.

Experiment 4: Bernoulli random variable scaled by 1000. Logit model of Table 1 is assumed to be the DGP.

Experiment 5: Bernoulli random variable scaled by employment. Logit model of Table is assumed to be the DGP.

University of Innsbruck – Working Papers in Economics and Statistics
Recent papers

2009-02	**Harald Oberhofer and Michael Pfaffermayr:** Fractional Response Models - A Replication Exercise of Papke and Wooldridge (1996)
2009-01	**Loukas Balafoutas:** How do third parties matter? Theory and evidence in a dynamic psychological game.

2008-27	**Matthias Sutter, Ronald Bosman, Martin Kocher and Frans van Winden:** Gender pairing and bargaining – Beware the same sex!
2008-26	**Jesus Crespo Cuaresma, Gernot Doppelhofer and Martin Feldkircher:** The Determinants of Economic Growth in European Regions.
2008-25	**Maria Fernanda Rivas and Matthias Sutter:** The dos and don'ts of leadership in sequential public goods experiments.
2008-24	**Jesus Crespo Cuaresma, Harald Oberhofer and Paul Raschky:** Oil and the duration of dictatorships.
2008-23	**Matthias Sutter:** Individual behavior and group membership: Comment. *Revised Version forthcoming in American Economic Review.*
2008-22	**Francesco Feri, Bernd Irlenbusch and Matthias Sutter:** Efficiency Gains from Team-Based Coordination – Large-Scale Experimental Evidence.
2008-21	**Francesco Feri, Miguel A. Meléndez-Jiménez, Giovanni Ponti and Fernando Vega Redondo:** Error Cascades in Observational Learning: An Experiment on the Chinos Game.
2008-20	**Matthias Sutter, Jürgen Huber and Michael Kirchler:** Bubbles and information: An experiment.
2008-19	**Michael Kirchler:** Curse of Mediocrity - On the Value of Asymmetric Fundamental Information in Asset Markets.
2008-18	**Jürgen Huber and Michael Kirchler:** Corporate Campaign Contributions as a Predictor for Abnormal Stock Returns after Presidential Elections.
2008-17	**Wolfgang Brunauer, Stefan Lang, Peter Wechselberger and Sven Bienert:** Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna.
2008-16	**Harald Oberhofer, Tassilo Philippovich:** Distance Matters! Evidence from Professional Team Sports.
2008-15	**Maria Fernanda Rivas and Matthias Sutter:** Wage dispersion and workers' effort.
2008-14	**Stefan Borsky and Paul A. Raschky:** Estimating the Option Value of Exercising Risk-taking Behavior with the Hedonic Market Approach. *Revised version forthcoming in Kyklos.*
2008-13	**Sergio Currarini and Francesco Feri:** Information Sharing Networks in Oligopoly.
2008-12	**Andrea M. Leiter:** Age effects in monetary valuation of mortality risks - The relevance of individual risk exposure.
2008-11	**Andrea M. Leiter and Gerald J. Pruckner:** Dying in an Avalanche: Current Risks and their Valuation.
2008-10	**Harald Oberhofer and Michael Pfaffermayr:** Firm Growth in Multinational Corporate Groups.
2008-09	**Michael Pfaffermayr, Matthias Stöckl and Hannes Winner:** Capital Structure, Corporate Taxation and Firm Age.
2008-08	**Jesus Crespo Cuaresma and Andreas Breitenfellner:** Crude Oil Prices and the Euro-Dollar Exchange Rate: A Forecasting Exercise.
2008-07	**Matthias Sutter, Stefan Haigner and Martin Kocher:** Choosing the carrot or the stick? – Endogenous institutional choice in social dilemma situations.
2008-06	**Paul A. Raschky and Manijeh Schwindt:** Aid, Catastrophes and the Samaritan's Dilemma.

2008-05 **Marcela Ibanez, Simon Czermak and Matthias Sutter:** Searching for a better deal – On the influence of group decision making, time pressure and gender in a search experiment. *Revised version published in Journal of Economic Psychology, Vol. 30 (2009): 1-10.*

2008-04 **Martin G. Kocher, Ganna Pogrebna and Matthias Sutter:** The Determinants of Managerial Decisions Under Risk.

2008-03 **Jesus Crespo Cuaresma and Tomas Slacik:** On the determinants of currency crises: The role of model uncertainty. *Revised version accepted for publication in Journal of Macroeconomics)*

2008-02 **Francesco Feri:** Information, Social Mobility and the Demand for Redistribution.

2008-01 **Gerlinde Fellner and Matthias Sutter:** Causes, consequences, and cures of myopic loss aversion - An experimental investigation. *Revised version forthcoming in Economic Journal.*

2007-31 **Andreas Exenberger and Simon Hartmann:** The Dark Side of Globalization. The Vicious Cycle of Exploitation from World Market Integration: Lesson from the Congo.

2007-30 **Andrea M. Leiter and Gerald J. Pruckner:** Proportionality of willingness to pay to small changes in risk - The impact of attitudinal factors in scope tests. *Revised version forthcoming in Environmental and Resource Economics.*

2007-29 **Paul Raschky and Hannelore Weck-Hannemann:** Who is going to save us now? Bureaucrats, Politicians and Risky Tasks.

2007-28 **Harald Oberhofer and Michael Pfaffermayr:** FDI versus Exports. Substitutes or Complements? A Three Nation Model and Empirical Evidence.

2007-27 **Peter Wechselberger, Stefan Lang and Winfried J. Steiner:** Additive models with random scaling factors: applications to modeling price response functions.

2007-26 **Matthias Sutter:** Deception through telling the truth?! Experimental evidence from individuals and teams. *Revised version published in The Economic Journal, Vol. 119 (2009), 47-60.*

2007-25 **Andrea M. Leiter, Harald Oberhofer and Paul A. Raschky:** Productive disasters? Evidence from European firm level data. *Revised version forthcoming in Environmental and Resource Economics.*

2007-24 **Jesus Crespo Cuaresma:** Forecasting euro exchange rates: How much does model averaging help?

2007-23 **Matthias Sutter, Martin Kocher and Sabine Strauß:** Individuals and teams in UMTS-license auctions. *Revised version with new title "Individuals and teams in auctions" published in Oxford Economic Papers.*

2007-22 **Jesus Crespo Cuaresma, Adusei Jumah and Sohbet Karbuz:** Modelling and Forecasting Oil Prices: The Role of Asymmetric Cycles. *Revised version accepted for publication in The Energy Journal.*

2007-21 **Uwe Dulleck and Rudolf Kerschbamer:** Experts vs. discounters: Consumer free riding and experts withholding advice in markets for credence goods. *Revised version published in International Journal of Industrial Organization, Vol. 27, Issue 1 (2009): 15-23.*

2007-20 **Christiane Schwieren and Matthias Sutter:** Trust in cooperation or ability? An experimental study on gender differences. *Revised version published in Economics Letters, Vol. 99 (2008): 494-497.*

2007-19 **Matthias Sutter and Christina Strassmair:** Communication, cooperation and collusion in team tournaments – An experimental study. *Revised version forthcoming in: Games and Economic Behavior.*

2007-18 **Michael Hanke, Jürgen Huber, Michael Kirchler and Matthias Sutter:** The economic consequences of a Tobin-tax – An experimental analysis.

2007-17 **Michael Pfaffermayr:** Conditional beta- and sigma-convergence in space: A maximum likelihood approach. *Revised version forthcoming in* *Regional Science and Urban Economics.*

2007-16 **Anita Gantner:** Bargaining, search, and outside options. *Published in:* *Games and Economic Behavior, Vol. 62 (2008), pp. 417-435.*

2007-15 **Sergio Currarini and Francesco Feri:** Bilateral information sharing in oligopoly.

2007-14 **Francesco Feri:** Network formation with endogenous decay.

2007-13 **James B. Davies, Martin Kocher and Matthias Sutter:** Economics research in Canada: A long-run assessment of journal publications. *Revised version published in:* *Canadian Journal of Economics, Vol. 41 (2008), 22-45.*

2007-12 **Wolfgang Luhan, Martin Kocher and Matthias Sutter:** Group polarization in the team dictator game reconsidered. *Revised version forthcoming in:* *Experimental Economics.*

2007-11 **Onno Hoffmeister and Reimund Schwarze:** The winding road to industrial safety. Evidence on the effects of environmental liability on accident prevention in Germany.

2007-10 **Jesus Crespo Cuaresma and Tomas Slacik:** An “almost-too-late” warning mechanism for currency crises.

2007-09 **Jesus Crespo Cuaresma, Neil Foster and Johann Scharler:** Barriers to technology adoption, international R&D spillovers and growth.

2007-08 **Andreas Brezger and Stefan Lang:** Simultaneous probability statements for Bayesian P-splines.

2007-07 **Georg Meran and Reimund Schwarze:** Can minimum prices assure the quality of professional services?.

2007-06 **Michal Brzoza-Brzezina and Jesus Crespo Cuaresma:** Mr. Wicksell and the global economy: What drives real interest rates?.

2007-05 **Paul Raschky:** Estimating the effects of risk transfer mechanisms against floods in Europe and U.S.A.: A dynamic panel approach.

2007-04 **Paul Raschky and Hannelore Weck-Hannemann:** Charity hazard - A real hazard to natural disaster insurance. *Revised version forthcoming in:* *Environmental Hazards.*

2007-03 **Paul Raschky:** The overprotective parent - Bureaucratic agencies and natural hazard management.

2007-02 **Martin Kocher, Todd Cherry, Stephan Kroll, Robert J. Netzer and Matthias Sutter:** Conditional cooperation on three continents. *Revised version published in:* *Economics Letters Vol. 101 (2008): 175-178.*

2007-01 **Martin Kocher, Matthias Sutter and Florian Wakolbinger:** The impact of naïve advice and observational learning in beauty-contest games.

University of Innsbruck

Working Papers in Economics and Statistics

2009-02

Harald Oberhofer and Michael Pfaffermayr

Fractional Response Models - A Replication Exercise of Papke and Wooldridge (1996)

Abstract
This paper replicates the estimates of a fractional response model for share data reported in the seminal paper of Leslie E. Papke and Jeffrey M. Wooldridge published in the Journal of Applied Econometrics 11(6), 1996, pp.619-632. We have been able to replicate all reported estimation results concerning the determinants of employee participation rates in 401(k) pension plans using standard routines provided in Stata. As an alternative, we estimate a two-part model that is able to cope with the excessive number of boundary values of one in the data. The estimated marginal effects are similar to that derived in that paper. A small scale Monte Carlo simulation exercise suggests that the RESET tests proposed by Papke and Wooldridge in their robust form are useful for detecting neglected non-linearities in small samples.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)

View publication stats