University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 0704

Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach

Stefan Boes

March 2007

Socioeconomic Institute
University of Zurich

Working Paper No. 0704

Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach

March 2007

Authors' addresses: Stefan Boes
E-mail: boes@sts.unizh.ch

Publisher Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-44-634 21 37
Fax: +41-44-634 49 82
URL: www.soi.unizh.ch
E-mail: soilib@soi.unizh.ch

Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach

Stefan Boes*
University of Zurich

March 2007

Abstract

As previously argued, the correlation between included and omitted regressors generally causes inconsistency of standard estimators for count data models. Using a specific residual function and suitable instruments, a consistent generalized method of moments estimator can be obtained under conditional moment restrictions. This approach is extended here by fully exploiting the model assumptions and thereby improving efficiency of the resulting estimator. Empirical likelihood estimation in particular has favorable properties in this setting compared to the two-step GMM procedure, which is demonstrated in a Monte Carlo experiment. The proposed method is applied to the estimation of a cigarette demand function.

JEL Classification: C14, C25, D12

Keywords: Nonparametric likelihood, Poisson model, nonlinear instrumental variables, optimal instruments, approximating functions, semiparametric efficiency.

*Address for correspondence: University of Zurich, Socioeconomic Institute, Zuerichbergstrasse 14, CH-8032 Zurich, Switzerland, phone: +41 44 634 2301, email: boes@sts.uzh.ch. I thank Joao Santos Silva, Rainer Winkelmann, and participants of meetings in Zurich, Dresden and Lisbon for valuable comments. This is a substantially revised version of SOI Discussion Paper No. 0404 "Empirical Likelihood in Count Data Models: The Case of Endogenous Regressors".

1 Introduction

Regression models for count data have become a standard tool in empirical analyses with applications in all fields of economics. Examples include the number of patents applied for by a firm (Hausman *et al.* 1984), the number of doctor visits (Pohlmeier and Ulrich 1995), the number of children borne to a woman (Winkelmann and Zimmermann 1995), and the number of days a worker is absent from his job (Delgado and Kniesner 1997).

Count data models should, in some way, incorporate the special feature of the dependent variable y being a nonnegative integer. One possibility is to specify a conditional probability model of y given a vector of observed explanatory variables x, such as in the Poisson regression model. The Poisson model, although very popular in applied work, presumes that the researcher is able to account for the full amount of individual heterogeneity just by including x. Additional unobserved heterogeneity is not allowed for, unlike for example in the linear regression model, where an additive error term captures such unobservable factors.

Various generalizations of the Poisson model have been proposed that account for unobserved heterogeneity. Standard approaches employ mixture distributions, either parametrically by introducing for example Gamma distributed unobservables (the negative binomial models), or semiparametrically by leaving the mixing distribution unspecified (e.g., Gurmu *et al.* 1998). Winkelmann (2003: Ch. 4.2) gives an overview. Mullahy (1997) extends the discussion to the important case when statistical independence between observed and unobserved heterogeneity fails. He focuses on the conditional expectation function, formally $E(y|x, v)$, specified as the exponential of a linear predictor $x'\beta$, with multiplicative unobserved heterogeneity v. Mullahy (1997) points out that, given nonzero correlation between x and v, standard estimators like Poisson pseudo maximum likelihood or non-linear least squares will generally be inconsistent for β because the usual residual function will not be orthogonal to x. Also, a non-linear instrumental variables (IV) strategy based on this residual function will be inconsistent due to the non-separability of the observable and the unobservable factors.

Fortunately, a simple transformation of the model yields a residual function, say $\rho(y, x; \beta)$, that is additively separable in the parametric structural part and the problematic unobservables, and the assumption of mean independence between the latter and instruments z can be used to construct conditional moment restrictions of the form $E[\rho(y,x;\beta)|z]=0$. As proposed by Mullahy (1997), estimation can be based on the generalized method of moments (GMM) using moment functions $g(y,x,z;\beta)=a(z)\rho(y,x;\beta)$ for some function $a(z)$, and the GMM estimator will be consistent for β and asymptotically normally distributed. The resulting estimator is not necessarily efficient, though, because the asymptotic variance depends on $a(z)$.

The aim of this paper is to extend Mullahy's (1997) approach using optimal instruments $a^*(z)$ that fully utilize the information given by the conditional moment restrictions. In this, I follow Donald *et al.* (2003) who approximate conditional moment restrictions by a series of unconditional moments using a general vector of approximating functions. From a theoretical point of view, semiparametric efficiency is achieved as linear combinations of these functions may well approximate the optimal instrument matrix of Chamberlain (1987) and as the dimension of the vector is increased with the sample size. As a practical matter, I select the number of unconditional moments according to the mean squared error criteria in Donald *et al.* (2005).

Clearly, the idea of using functions of the conditioning variables as additional instruments is not new; for a non-technical discussion see Wooldridge (2001). In fact, one motivation of GMM is that all possible information — as it is given by the conditional moment restrictions — can be used in an efficient manner by choosing the "right" weighting matrix. A general vector of approximating functions like the one employed here has the advantage of systematically using the information at hand. If cautiously implemented, this will in general improve the efficiency of the resulting estimator compared to a baseline where $a(z)=z$, or compared to any other vague choice of $a(z)$. On the downside, many approximating functions, and thus unconditional moment conditions, may be needed to obtain the optimal estimator in practice.

Recent work on the finite-sample properties of GMM, however, emphasizes the poor performance of the two-step procedure with increasing number of moment conditions, and several alternatives have been proposed, for example the empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994) and Imbens (1997). Other moment estimators exist as well (e.g., Hansen *et al.* 1996, Kitamura and Stutzer 1997, Imbens *et al.* 1998). Smith (1997) introduces the class of generalized empirical likelihood (GEL) estimators that include the forementioned estimators as special cases, and asymptotic equality of GEL and GMM was shown. Further studies by Newey and Smith (2004) and Imbens and Spady (2006) examine the higher order properties of GEL and GMM estimators and evidence the relative advantage of EL compared to two-step GMM in terms of higher order asymptotic bias and higher order efficiency (after bias correction) in the case of increasing degree of overidentification.

The novelty of this paper is the application of the approximating functions to an inherently non-linear IV model, first in a generated data experiment and then with real data in a model for cigarette demand. The model and moment conditions will be laid out in the next section. Section 3 briefly discusses EL and GMM estimation, and the moment selection criteria. Section 4 compares the properties of the estimators in a simulated data environment. The results indicate that the EL estimator has indeed favorable properties in terms of bias and efficiency, as it was to be expected from earlier theoretical results. Section 5 applies the method to estimate a cigarette demand function similar to Mullahy (1997). Fully exploiting the model assumptions considerably improves the efficiency of the estimators. For example, just by including the optimal vector of approximating functions for one instrument, the t-statistic for the parameter of interest is more than doubled compared to the baseline IV estimator. Section 6 concludes.

2 Count Data Models with Unobserved Heterogeneity

Let y denote a random variable with support being the non-negative integers, let x denote a $k \times 1$ vector of explanatory variables (including a constant), and let z denote a $q \times 1$ vector of instruments ($q \geq k$) with properties to be defined below. Assume that n observations of (y, x, z) form a random sample of the population, and suppose that the main objective is to estimate the effect of elements of x on y.

The paper focuses on the relationship between y and x as summarized in the conditional expectation function (CEF). Specifically, assume that the data-generating process is consistent with the CEF

$$E(y|x, v; \beta) = \exp(x'\beta)v \tag{1}$$

where β is the $k \times 1$ vector of unknown parameters, and $v = \exp(u) > 0$ is unobservable to the researcher. Without loss of generality the normalization $E(v) = 1$ can be invoked as a constant term is included in x. Note that observable and unobservable characteristics are treated symmetrically in (1) because the CEF is log-linear in both x and u. The specific functional form of the CEF might appear restrictive at first, but there is no *a priori* reason for x and u to enter the CEF asymmetrically. Moreover, the linear index $x'\beta$ is sufficiently flexible to approximate any non-linear function in the regressors arbitrarily close, and the exponential function ensures (1) to be positive, as required for a count dependent variable. Strictly speaking, it is not necessary for (1) to be fulfilled that y is a count. What follows is equally relevant to any other data-generating process consistent with such an exponential CEF.

The specification of the CEF in (1) implies the nonlinear regression model

$$y = \exp(x'\beta)v + \varepsilon \tag{2}$$

where the regression error ε has property $E(\varepsilon|x, v) = 0$, by construction. Windmeijer and Santos Silva (1997) consider estimation of models like (2) in situations where some of the regressors may be simultaneously determined with the dependent count. In this case, there is a crucial distinction between additive and multiplicative (for that matter structural) errors, the two otherwise being observationally equivalent (Wooldridge 1992). Grogger (1990) discusses the additive approach and testing for exogeneity of the regressors using a Hausman-type test.

In the given context, it is natural to maintain the notation in (2) to distinguish between regression error and unobservable characteristics, the latter not being accounted for in the regression and potentially correlated with x. Mullahy (1997) gives conditions for consistent estimation of β in such a model. In a nutshell, if v and x are mean independent, then pseudo maximum likelihood (PML) estimation of the Poisson model is consistent for β (see Gourieroux *et al.* 1984, Wooldridge 1997). Contrary to that, if mean independence fails, then PML will generally be inconsistent, and estimation with instrumental variables based on appropriately defined residuals is suggested alternatively. Mullahy (1997) imposes two key assumptions on the instrument vector z. The first assumption is an independence condition that v and z must be mean independent, formally $E(v|z) = E(v)$. The second assumption imposes the restriction

$E(y|x,v,z) = E(y|x,v)$ which implies for the regression error that $E(\varepsilon|x,z,v) = 0$.

Let $w = (y,x)$ to simplify notation. With the assumptions on z, conditional moment restrictions can be constructed via the residual function $\rho(w;\beta) = y\exp(-x'\beta) - 1$ since

$$E[\rho(w;\beta)|z] = E[y\exp(-x'\beta) - 1|z] = 0 \qquad (3)$$

by iterated expectations. As noted by Mullahy (1997), the crucial step in deriving such a residual function is that v needs to be additively separable from x which can be achieved by dividing both sides of equation (2) by $\exp(x'\beta)$. The conditional moment restriction is assumed to uniquely identify the true parameter value β. Now let $a(z)$ denote a matrix-valued function of z. It is common practice to derive unconditional (population) moment restrictions from (3) as

$$E\left[a(z)\rho(w;\beta)\right] = 0$$

and the estimator of β is obtained as the solution to sample counterparts $\sum_i a(z_i)\rho(w_i;\hat{\beta}) = 0$, as it is applied for example in GMM or nonlinear IV estimation. Such a procedure, however, is suboptimal for at least two reasons. First, the conditional moment restriction is stronger than the unconditional one implying that an estimator based on the latter does not fully exploit the available information. Second, the procedure is only valid under the presumption that $a(z)$ identifies β, which must not necessarily be so; see Dominguez and Lobato (2004).

A recent paper by Donald, Imbens, and Newey (2003) overcomes both problems considering an approach directly based on the conditional moment restriction. Given the information in (3), Chamberlain (1987) shows that an estimator with optimal instruments

$$a^*(z) = E[\partial\rho(w;\beta)/\partial\beta|z]\{E[\rho(w;\beta)^2|z]\}^{-1}$$

would achieve the semiparametric efficiency bound. In general, the estimator using optimal instruments is not feasible as both expectations forming $a^*(z)$ are unknown. Furthermore, even if the functional form of the expectations were known, identification of β via $a^*(z)$ may fail, see Dominguez and Lobato (2004) for an example. Donald *et al.* (2003) use a series of functions of z to form unconditional moment restrictions, and let the dimension K of the vector of approximating functions grow with the sample size. Let $q^K(z)$ denote such a vector. Under certain regularity conditions, the sequence of unconditional moment restrictions

$$E[q^K(z)\rho(w;\beta)] = 0 \tag{4}$$

is equivalent to the conditional moment restriction in (3). Efficiency is established if linear combinations of $q^K(z)$ can approximate $a^*(z)$, with approximation error diminishing as K grows, since the asymptotic variance of the optimal GMM estimator with instruments $a^*(z)$ reaches the semiparametric efficiency bound (Newey 1993).

Donald *et al.* (2003) suggest using splines as approximating functions. If z is univariate, the s-th order spline with knots $t_1, \ldots, t_{K-s-1}$ is given by

$$q^K(z) = (1, z, \ldots, z^s, [1(z > t_1)z]^s, \ldots, [1(z > t_{K-s-1})z]^s)' \tag{5}$$

with indicator function $1(\cdot)$. Common choice is $s = 3$ for cubic splines. For z multivariate, the approximating functions may be generated by products of univariate splines for each element of z. Under the assumption that z is continuously distributed with compact support and density bounded away from zero, Donald *et al.* (2003) derive limits on the growth rate of K to obtain asymptotic efficiency. The method can be easily implemented in existing procedures that utilize unconditional moment restrictions, a potential advantage over alternative approaches such as Kitamura *et al.* (2004) and Dominguez and Lobato (2004).

3 Estimation Methods and Moment Selection

3.1 Generalized Method of Moments

The GMM principle has become a well-established estimation technique for moment conditions such as (4) since Hansen (1982); see also Hall (2005). To describe it, let $g_i(\beta) = q^K(z_i)\rho(w_i;\beta)$ and $\hat{g}_n(\beta) = \sum_{i=1}^n g_i(\beta)/n$. The GMM estimator $\hat{\beta}_{gmm}$ minimizes the weighted squared distance of sample and population moments, algebraically

$$\hat{\beta}_{gmm} = \arg\min_{\beta} \hat{g}_n(\beta)' W \hat{g}_n(\beta) \tag{6}$$

where W is a $K \times K$ weighting matrix. For optimal GMM, the weighting matrix is chosen such that $W = \hat{\Omega}_n(\tilde{\beta})^{-1}$ with $\hat{\Omega}_n(\beta) = \sum_{i=1}^n g_i(\beta)g_i(\beta)'/n$ and preliminary consistent estimator $\tilde{\beta}$. Under mild regularity conditions the resulting estimator $\hat{\beta}_{gmm}$ is consistent and the stabilizing transformation $\sqrt{n}(\hat{\beta}_{gmm} - \beta)$ is asymptotically normal with zero expectation and estimated covariance matrix

$$\hat{\Sigma}_{gmm} = \left[\hat{G}_n(\hat{\beta}_{gmm})'\hat{\Omega}_n(\hat{\beta}_{gmm})^{-1}\hat{G}_n(\hat{\beta}_{gmm})\right]^{-1}$$

where $G_i(\beta) = \partial g_i(\beta)/\partial\beta'$ and $\hat{G}_n(\beta) = \sum_{i=1}^n G_i(\beta)/n$.

Accumulating empirical evidence and recent theoretical work on the properties of two-step GMM, however, reveals that point estimates and inference based on the asymptotic normal distribution may be highly unreliable in finite samples (Hansen *et al.* 1996 and Hall 2005, among others). Newey and Smith (2004) discuss higher order asymptotic properties of GMM as possible explanation for the finite sample behavior. In particular, note that the optimization problem for two-step GMM implies first order conditions

$$\hat{G}_n(\hat{\beta}_{gmm})'\hat{\Omega}_n(\tilde{\beta})^{-1}\hat{g}_n(\hat{\beta}_{gmm}) = 0$$

and thus, in the optimum, a linear combination of sample equivalents to (4) must equal zero. It is shown, *inter alia*, that asymptotic (higher order) bias of the two-step GMM estimator arises from estimating the Jacobian matrix (left term) and the matrix of second moments (middle term) by sample averages, and the weighting matrix depending on a first step (inefficient) estimator.

As the asymptotic bias formulae are known, an analytical bias correction of $\hat{\beta}_{gmm}$ becomes available. The bias arising from estimation of the Jacobian matrix is particularly important, and a bias corrected GMM estimator can be obtained as

$$\hat{\beta}_{bcgmm} = \hat{\beta}_{gmm} + \hat{\Sigma}_{gmm}\sum_{i=1}^n \hat{G}_i\hat{P}\hat{g}_i/n \tag{7}$$

where $\hat{g}_i = g_i(\hat{\beta}_{gmm})$, $\hat{G}_i = G_i(\hat{\beta}_{gmm})$, and $\hat{P} = \hat{\Omega}^{-1} - \hat{\Omega}^{-1}\hat{G}\hat{\Sigma}_{gmm}\hat{G}'\hat{\Omega}^{-1}$ with $\hat{G} = \hat{G}_n(\hat{\beta}_{gmm})$, $\hat{\Omega} = \hat{\Omega}_n(\hat{\beta}_{gmm})$; see Newey and Smith (2004) and Donald *et al.* (2005) for details.

In comparison to two-step GMM, other moment estimators imply first order conditions in which the Jacobian and second moment matrix are estimated more efficiently. Among the alternatives, the empirical likelihood estimator received considerable attention and was found to possess some desirable higher order properties. In particular, it was shown that the asymptotic bias of GMM grows with the number of overidentifying restrictions, whereas the bias of EL is bounded. I will therefore discuss EL estimation of β next.

3.2 Empirical Likelihood

Empirical likelihood estimation was first introduced in the biostatistics literature, see Owen (1988, 1991) and Qin and Lawless (1994, 1995) for details on EL and its application to moment condition models; see also Owen (2001) for a monograph on empirical likelihood. More recent surveys by Imbens (2002) and Kitamura (2006) point out the richness of the EL approach, in particular as an alternative to the two-step GMM procedure.

Let p_i denote an unknown probability weight assigned to the sample outcome (y_i, x_i, z_i) of one observation i with $0 < p_i < 1\ \forall i$, impose the normalization $\sum_i p_i = 1$, and let $p = (p_1, \ldots, p_n)'$. A nonparametric likelihood estimator of p is obtained by maximizing the nonparametric log-likelihood function, algebraically

$$\hat{p} = \arg\max_{p} \sum_{i=1}^{n} \ln p_i \quad \text{s.t.} \quad \sum_{i=1}^{n} p_i = 1 \tag{8}$$

Without further restrictions, optimal probability weights are given by $\hat{p}_i = 1/n$. In order to incorporate special features of the data-generating process, one may impose empirical moments as additional restrictions, which can be specified from (4) as $\sum_i p_i g_i(\beta) = 0$. Following Kitamura (2006), the optimization problem yields the Lagrangian function

$$\mathcal{L} = \sum_{i=1}^{n} \ln p_i + \eta \left(1 - \sum_{i=1}^{n} p_i\right) - n\lambda' \sum_{i=1}^{n} p_i g_i(\beta) \tag{9}$$

where λ and η denote Lagrangian multipliers. It can be shown that the first order conditions are solved by $\hat{\eta} = n$,

$$\begin{aligned} \hat{p}_i(\beta) &= \frac{1}{n\left[1 + \hat{\lambda}(\beta)' g_i(\beta)\right]} \\ \hat{\lambda}(\beta) &= \arg\min_{\lambda} \ -\sum_{i=1}^{n} \ln\left[1 + \lambda' g_i(\beta)\right] \end{aligned}$$

Optimal probability weights $\hat{p}_i$ and optimal Langrangian multipliers $\hat{\lambda}$ both depend on the unknown parameter vector β. Plugging the optimality conditions into the objective function in (8) yields the empirical log-likelihood function for β

$$\ln L_{el}(\beta) = \min_{\lambda} \; -\sum_{i=1}^{n} \ln \left[1 + \lambda' g_i(\beta)\right] - n \ln n$$

and the EL estimator is defined as

$$\hat{\beta}_{el} = \arg\max_{\beta} \; \ln L_{el}(\beta) = \arg\max_{\beta} \; \min_{\lambda} \; -\sum_{i=1}^{n} \ln \left[1 + \lambda' g_i(\beta)\right] \tag{10}$$

Since maximization of (10) does not have a simple closed form solution, numerical methods have to be applied to obtain the value of $\hat{\beta}_{el}$. Owen (2001) and Kitamura (2006) provide details on computational algorithms that have stable convergence properties in the above problem.

Under similar regularity conditions as in the GMM framework, Qin and Lawless (1994) show consistency of the empirical likelihood estimator and prove asymptotic normality of the stabilizing transformation $\sqrt{n}(\hat{\beta}_{el} - \beta)$ with zero expectation and estimated covariance matrix

$$\hat{\Sigma}_{el} = [\hat{G}_p(\hat{\beta}_{el})' \hat{\Omega}_p(\hat{\beta}_{el})^{-1} \hat{G}_p(\hat{\beta}_{el})]^{-1}$$

where $\hat{G}_p(\beta) = \sum_{i=1}^{n} \hat{p}_i(\beta) \partial g_i(\beta)/\partial\beta'$ and $\hat{\Omega}_p(\beta) = \sum_{i=1}^{n} \hat{p}_i(\beta) g_i(\beta) g_i(\beta)'$. Note that the terms in the EL covariance matrix are estimated using probability weights $\hat{p}_i(\hat{\beta}_{el})$ obtained from an empirical likelihood optimization, whereas the terms in the GMM variance are estimated using sample weights $1/n$.

It can be shown that optimal probability weights $\hat{p}_i$ and Langrangian multipliers $\hat{\lambda}$, both evaluated at the EL estimator, imply first order conditions

$$\hat{G}_p(\hat{\beta}_{el})' \hat{\Omega}_p(\hat{\beta}_{el})^{-1} \hat{g}_n(\hat{\beta}_{el}) = 0$$

As with two-step GMM, a linear combination of sample moments must equal zero. EL uses empirical moments for the Jacobian term and the matrix of second moments, and probability weights p_i are chosen efficiently. Moreover, the EL estimator does not depend on a preliminary, possibly inefficient estimator $\tilde{\beta}$. Based on these properties, Newey and Smith (2004) show that the EL estimator is preferable to the GMM estimator in terms of higher order asymptotic bias, and higher order efficiency after bias correction.

3.3 Moment Selection Criteria

To describe the moment selection criteria of Donald *et al.* (2005), some further notation needs to be introduced. Let $\hat{\beta}_K$ denote any of the three estimators — GMM, bias corrected GMM, or EL — given that the vector of approximating functions has dimension K. Let $t'\hat{\beta}_K$ denote a linear combination of $\hat{\beta}_K$ for some linear combination coefficients t. Let

$$\hat{\rho} = \rho(w_i; \hat{\beta}_K), \qquad \hat{G} = \hat{G}_n(\hat{\beta}_K), \qquad \hat{\Omega} = \hat{\Omega}_n(\hat{\beta}_K), \qquad \hat{\Sigma} = [\hat{G}'\hat{\Omega}^{-1}\hat{G}]^{-1}, \qquad \hat{\tau} = \hat{\Sigma}t$$

$$\hat{d}_i = \hat{G}' \left[\sum_{j=1}^{n} q^K(z_j)q^K(z_j)'/n\right]^{-1} q^K(z_i), \qquad \hat{\eta}_i = \partial\hat{\rho}/\partial\beta - \hat{d}_i$$

$$\hat{\xi}_i = q^K(z_i)'\hat{\Omega}q^K(z_i)/n, \qquad \hat{\Lambda}(K) = \sum_{i=1}^{n}(\hat{\tau}'\hat{\eta}_i)^2\hat{\xi}_i, \qquad \hat{\Pi}(K) = \sum_{i=1}^{n}(\hat{\tau}'\hat{\eta}_i)\hat{\xi}_i\hat{\rho}$$

$$\hat{\Phi}(K) = \hat{\Lambda}(K) - \hat{\tau}'\hat{\Sigma}^{-1}\hat{\tau}, \qquad \hat{Q} = \sum_{i=1}^{n} q^K(z_i)\hat{\rho}(\hat{\tau}'\hat{\eta}_i)q^K(z_i)'$$

$$\hat{\Pi}_b(K) = tr(\hat{\Omega}^{-1/2}\hat{Q}\hat{\Omega}^{-1}\hat{Q}\hat{\Omega}^{-1/2}), \qquad \hat{D}_i = \hat{G}'\hat{\Omega}^{-1}q^K(z_i)$$

$$\hat{\Xi}(K) = \sum_{i=1}^{n}\{5(\hat{\tau}'\hat{d}_i)^2 - \hat{\rho}^4(\hat{\tau}'\hat{D}_i)^2\}\hat{\xi}_i$$

$$\hat{\Xi}_{el}(K) = \sum_{i=1}^{n}\{3(\hat{\tau}'\hat{d}_i)^2 - \hat{\rho}^4(\hat{\tau}'\hat{D}_i)^2\}\hat{\xi}_i$$

The selection criteria are

$$\begin{aligned} S_{gmm}(K) &= \hat{\Pi}(K)^2/n + \hat{\Phi}(K) \\ S_{bcgmm}(K) &= [\hat{\Lambda}(K) + \hat{\Pi}_b(K) + \hat{\Xi}(K)]/n + \hat{\Phi}(K) \\ S_{el}(K) &= [\hat{\Lambda}(K) - \hat{\Pi}_b(K) + \hat{\Xi}(K) - 2\,\hat{\Xi}_{el}(K)]/n + \hat{\Phi}(K) \end{aligned} \tag{11}$$

The optimal dimension K^* of the vector of approximating functions is chosen such that $S(K)$ is minimal, i.e., $K^* = \arg\min_K S(K)$, which is shown to minimize the higher-order mean squared error (MSE) of each estimator. The terms in each criterion contain second and higher order moments, for details on the interpretation see Newey and Smith (2004) and Donald *et al.* (2005).

4 Monte Carlo Evidence

In this section, I compare the finite sample behavior of EL and GMM in a generated count data experiment with correlated unobserved heterogeneity. The model imposes a conditional moment restriction as the one introduced in the discussion above, and I investigate the performance of the proposed estimators with increasing dimension of the vector of approximating functions.

The sampling process is based on the Poisson model with Gamma distributed heterogeneity. The model is non-standard compared to the well-known negative binomial models in that the heterogeneity term is correlated with the single observed regressor x. Specifically, consider the following data-generating process

$$(r, s) \sim BVN(0, 0, 1, 1, 0), \qquad w = r + \gamma s - (1 + \gamma^2)/2$$

$$z \sim N(0, 1) \quad \text{or} \quad z \sim LN(0, 1)$$

$$x = (1,\, \alpha z + s)', \qquad \mu = \exp(x'\beta), \qquad v|w \sim Gamma[1, \exp(w)]$$

$$y|x, v \sim Poisson(\mu v)$$

where $BVN(\cdot)$ stands for the bivariate normal distribution with zero means, unit variances, and zero correlation, $N(0, 1)$ stands for the standard normal, and $LN(0, 1)$ for the standard log-normal distribution. It is assumed that only (y, x, z) are observed. The conditional distribution of $v|w$ is normalized such that $E(v|w) = \exp(w)$ and $Var(v|w) = \exp(2w)$. The location normalization of w implies that $E(v) = E[E(v|w)] = E[\exp(w)] = 1$. For α fixed, the parameter γ determines the correlation between x and w. If γ equals zero, the unobserved heterogeneity is independent of the regressor and PML consistently estimates β. For nonzero γ, the conditional expectation $E(v|x)$ is non-constant in x, and PML estimation will generally be inconsistent. Since v and z are statistically independent, an assumption somewhat stronger than required, and $\alpha \neq 0$, moment estimation as outlined above using the instrument z can be applied.

The parameter vector β is fixed at $(0,\, 1)'$, and γ is set to 0.5. In order to vary the correlation between instrument and regressor, two different values of α are chosen — 0.3 and 0.7. Two different sample sizes are considered — $n = 500$ and $n = 2000$ — and samples are drawn for all variables in each of 1000 Monte Carlo replications. Since $\gamma \neq 0$, PML estimation will be inconsistent for β in each of the settings. The experiment shows that, depending on the variation in x, the median bias in the estimated slope $\hat{\beta}_{1,pml}$ varies between 0.264 and 0.381 in the normal case, and between 0.377 and 0.446 in the log-normal case. These numbers need to be compared with the results for the other estimators, that are displayed in Tables 1–4.

Consider Tables 1 and 2 with $n = 500$ observations first. The columns in Table 1 correspond to the median of the estimated standard error of $\hat{\beta}_1$ (Med.SE) and the rejection rate for an overidentifying test (in the case of $K > 2$) with 5% significance level. Table 2 shows the median bias (Med.Bias) and the median absolute deviation (MAD) from the true value, and the probabilities of $\hat{\beta}_1$ deviating from 1 by more than 0.1 and 0.2, respectively. Robust measures of central tendency and dispersion are presented as the existence of (finite-sample) moments might be an issue (e.g., Kunitomo and Matsushita 2003, Guggenberger 2005, Guggenberger and Hahn 2005, Davidson and MacKinnon 2006). Five different specifications of $q^K(z)$ are presented. The first, as a benchmark, is basic IV with instrument z, i.e., the vector of approximating functions is simply $q^2(z) = (1, z)'$. The next three rows give the results with augmented instrument vector having dimensions $K = 4, 8, 16$, and optimal K^*. The approximating functions are chosen such that they form a basis for the set of cubic splines, i.e., $s = 3$, and the knots $t_1, \ldots, t_{K-4}$ are set equal to the quantiles of the empirical z-distribution. For the selection criteria, the linear combination coefficients pick the slope as parameter of interest.

The results in Table 1 indicate that there are considerable efficiency gains by increasing the dimension of the vector of approximating functions. These gains are higher with a low value of α and for the EL estimator more than for the GMM estimators. If z is normally distributed, EL seems to perform better than GMM, if z follows a log-normal distribution the differences between the three estimators are less clear. In all cases, the optimal K^* yields the lowest median standard error. Due to the variation in K^*, it is suggestive to choose the dimension of $q^K(z)$ according to the MSE criteria, as opposed to a rule-of-thumb fixed choice of K. The rejection rate for the overidentifying restrictions test is always close to the nominal level.

Despite the efficiency gains, it is important to note that the estimators behave quite differently when looking at the summary statistics of $\hat{\beta}_1$ in Table 2. In all cases, the basic IV estimator produces consistent results, which is reflected in almost zero median bias. As it was expected from previous theoretical results, the GMM estimator exhibits significant bias if K and thus the number of overidentifying restrictions grows, and even under the optimal choice K^* the bias remains. Bias correction helps to improve upon the standard two-step GMM procedure, but in all settings the EL estimator has lowest bias. With respect to the median absolute deviation and the deviation probabilities, there are only minor differences between the three estimators.

Tables 3 and 4 report the simulation results for $n = 2000$ observations. In this case, GMM and EL perform similarly, which was to be expected as they are all first order asymptotically equivalent. It is noteworthy that even with 2000 observations, the two-step GMM estimator with large degree of overidentification exhibits bias that does not occur with bias corrected GMM and EL. The efficiency gains from augmenting the vector of approximating functions, however, are much smaller in the large sample than they are in the small sample experiment.

5 Cigarette Demand and Smoking Habits

As a final exercise, I apply the proposed methods to the estimation of a cigarette demand function. Cigarette demand is measured as the number of cigarettes smoked per day, and thus y has the character of a count dependent variable. Mullahy (1985) studies the dynamic link between today's demand for cigarettes and an individual's smoking habits amassed over lifetime. If included in a regression model, such habits can be interpreted as a lagged dependent variable, and there is good reason to believe that unobserved smoking determinants are also dynamically linked. One would thus suspect that, given a positive correlation between unobservables over time, the smoking habit dynamics may be overestimated in a simple Poisson regression model, and IV estimation as outlined above may help to avoid such problems.

The analysis is based on a subsample of $n = 1140$ male observations of the data used in Mullahy (1997); see also Mullahy (1985) for a description. The data stem from the Smoking Supplement of the 1979 US National Health Interview Survey and contain information on the respondent's socioeconomic characteristics as well as information on various health topics and smoking behavior. For the regressions, the dependent variable has been scaled to the number of cigarette packs smoked per day (number of cigarettes divided by 20). Mullahy (1985) constructs the smoking habit measure from the total time smoked and the number of cigarettes consumed. This measure is zero for non-smokers, and positive for smokers, the exact value depending on the discount rate (here 10 percent) and not having direct unit interpretation. Apart from the smoking habit measure as the key variable of interest, the estimated models control for age (in years), the years of schooling, a dummy variable indicating race, family income (in thousand US Dollars), household size, average state-level cigarette price (in US Dollars per pack in 1979), and an indicator whether smoking in restaurants had been restricted (in 1979).

The excluded instruments are the cigarette price in 1978 and the total number of years smoking in restaurants had been restricted (before and with 1979). The rationale for the instruments is that both should affect smoking habits, i.e., smoking behavior in 1978 and before, but they should not have a direct effect on current cigarette demand. The latter exclusion restriction is plausible, since cigarette prices and indicators of smoking restrictions in 1979, i.e., at the time current cigarette demand is recorded, are explicitly controlled for, and thus there is no reason to believe why the instruments should have an effect on y other than the habits channel. Compared to the data in Mullahy (1997), I restrict the sample to individuals aged younger than 25, as those are the most responsive to changes in the instruments.

Table 5 displays the results for the smoking habit coefficient. The columns correspond to the Poisson pseudo maximum likelihood (PML) estimator, the two-step and bias-corrected GMM estimators, and the EL estimator. For the ease of exposition, the estimated parameters and standard errors have been multiplied by 1000. The PML estimate shows a value of 12.53 with estimated standard error 0.81. This value indicates that the expected number of packages smoked per day increases by $100[\exp(12.53/1000) - 1] = 1.26$ percent for an unit increase in the smoking habit measure. Multiplied by the average value of the smoking habits (35.65), this gives an elasticity of 0.45, i.e., if the smoking habit measure increases by 1 percent, then the expected number of cigarettes smoked per day (measured in packs) increases by 0.45 percent. The elasticity may of course be evaluated at other values than the average smoking habits.

Using the basic IV setting with instruments all regressors except the smoking habits plus the cigarette price in 1978 and the number of years the smoking restrictions had been in place, the estimated parameters drop by around 5 to 10 percent with much larger standard error. The IV point estimates confirm the expectation that PML might overestimate the true smoking habit effect. On the downside, from a statistical point of view, smoking habits do not significantly affect current smoking behavior, which contradicts the perspective of smoking habits entering cigarette demand as a psychological and/or physiological addiction. Note that the overidentifying test statistic is sufficiently small as to not reject the null hypothesis of valid instruments. Note too that the basic setting does not fully exploit the model assumptions and, given that the instruments fulfill mean independence, an improvement over these results might be possible.

The remaining of Table 5 shows the estimation results for various specifications of the vector of approximating functions. Among the many options to specify this vector, a reasonable working guess is to first find the optimal dimension, say K_l^* for the l-th element of the instrument vector, given basic specification for all other instruments, and then gradually combine the optimal K_l^* including interactions if suitable. The table first reports the results for the optimal specification of the excluded instruments, i.e., the number of years smoking restrictions had been in place and the cigarette price in 1978, respectively. In curly brackets is the number of additional approximating functions, e.g., for the cigarette price in 1978 its square has been additionally included. This number plus one are the degrees of freedom for the overidentifying restrictions test with test statistic reported in square brackets.

The point estimates of the smoking habit coefficient drop compared to PML and basic IV. Using the square of cigarette prices in 1978 as additional instrument even turns the sign of the coefficient negative for bias-corrected GMM and EL. Although the overidentifying restrictions are not rejected, there is only a minor gain in the value of the moment selection criteria in this case. For the restaurant smoking restrictions, the overidentifying restrictions are not rejected either, but there is a considerable drop in the value of the selection criteria, indicating higher potential efficiency gains by adding the approximating functions. Note that in both cases the null hypothesis of a zero coefficient cannot be rejected. Clearly, the element-wise optimization may be done for the included instruments as well.

Next, I combine the optimal approximating functions for each excluded instrument to further explore the model assumptions. It turns out that the optimal number of approximating functions K_l^* for each instrument can be simply combined to obtain the optimal number of approximating functions when both instruments are considered simultaneously. Presumably, this result is specific to the data and does not hold in general, but in any case, such a strategy might be a good starting point to explore the validity of mean independence. Using the additional approximating functions and including interactions does not change the point estimates by much but the standard errors become smaller due the additional information that is used.

Finally, combining the optimal dimension K_l^* for excluded and included instruments and adding interactions if indicated the optimal vector of approximating functions for GMM has five additional terms for restaurant smoking restrictions, household size and the cigarette price in 1979, two additional terms for family income, and the square of cigarette price in 1978. For the EL estimator, the interaction between smoking restrictions and cigarette prices in 1978 and an additional term for family income has been included. The results show point estimates of 8.86 for two-step GMM, 7.61 for bias corrected GMM, and 7.08 for EL. In terms of elasticities, a one percent increase in the smoking habit measure leads to an increase in the expected number of cigarette packs consumed per day by about 0.32 percent for the GMM estimator, 0.27 percent for the bias corrected GMM estimator, and 0.25 percent for the EL estimator, respectively. In all cases, the estimated coefficients are statistically different from zero at the 5 percent level.

6 Concluding Remarks

This paper extends Mullahy's (1997) IV approach for the estimation of count data models with correlated unobserved heterogeneity. Based on transformed residuals and a mean independence assumption, the model implies conditional moment restrictions that can be estimated by common moment estimators. As the asymptotic variance typically depends on the choice of instruments, the paper proposes the use of a general vector of approximating functions, opting ideas of Donald *et al.* (2003), to improve efficiency of the resulting estimator.

A small Monte Carlo experiment points out the benefits of the method and outlines the relative advantage of EL compared to two-step GMM. Finally, the approach is applied to estimate the effect of smoking habits on cigarette demand. Compared to the standard Poisson PML estimator, the estimated elasticities of cigarette demand with respect to smoking habits change from 0.45 to 0.32 (GMM) and 0.25 (EL), respectively, a drop that is conformable to previous findings. Importantly, since the methods applied here fully exploit the model assumptions, the parameters have been estimated with much higher precision than before.

References

Cameron, A.C. and P.K. Trivedi (1998): *Regression Analysis of Count Data*, Econometric Society Monograph No. 30, Cambridge University Press.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

Davidson, R. and J.G. MacKinnon (2006): "Moments of IV and JIVE Estimators," *unpublished manuscript*.

Delgado, M.A. and T.J. Kniesner (1997): "Count Data Models with Variance of Unknown Form: An Application to a Hedonic Model of Worker Absenteeism," *Review of Economics and Statistics*, 79, 41-49.

Dominguez, M. and I. Lobato (2004): "Consistent estimation of models defined by conditional moment restrictions," *Econometrica*, 72, 1601-1615.

Donald, S.G., G.W. Imbens and W.K. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55-93.

Donald, S.G., G.W. Imbens and W.K. Newey (2005): "Choosing the Number of Moments in Conditional Moment Restriction Models," *unpublished manuscript*.

Gourieroux, C., A. Monfort and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701-720.

Grogger, J.T. (1990): "A Simple Test for Exogeneity in Probit, Logit, and Poisson Regression Models," *Economics Letters*, 33, 329-332.

Guggenberger, P. (2005): "Monte-carlo evidence suggesting a no moment problem of the continuous updating estimator," *Economics Bulletin*, 3, 1-6.

Guggenberger, P. J. Hahn (2005): "Finite Sample Properties of the 2step Empirical Likelihood Estimator" *Econometric Reviews*, 24, 247-263.

Gurmu, S., P. Rilstone and S. Stern (1998): “Semiparametric Estimation of Count Regression Models,” *Journal of Econometrics*, 88, 123-150.

Hall, A.R. (2005): *Generalized Method of Moments*, Oxford University Press.

Hansen L.P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029-1054.

Hansen, L.P., J. Heaton and A. Yaron (1996): “Finite-Sample Properties of Some Alternative GMM Estimators,” *Journal of Business & Economic Statistics*, 14, 262-280.

Hausman, J., B.H. Hall and Z. Griliches (1984): “Econometric Models for Count Data with an Application to the Patents - R&D Relationship,” *Econometrica*, 52, 909-938.

Imbens, G.W. (1997): “One-Step Estimators for Over-Identified Generalized Method of Moment Models,” *Review of Economic Studies*, 64, 359-383.

Imbens, G.W. (2002): “ Generalized Method of Moments and Empirical Likelihood,” *Journal of Business & Economic Statistics*, 20, 493-506.

Imbens, G.W. and R.H. Spady (2006): “The Performance of Empirical Likelihood and its Generalizations,” in: D.W.K. Andrews (ed.) *Identification and Inference for Econometric Models: Essays in Honour of Thomas Rothenberg.*

Imbens, G.W., R.H. Spady and P. Johnson (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica*, 66, 333-357.

Kitamura, Y. (2006): “Empirical Likelihood Methods in Econometrics: Theory and Practice”, *Cowles Foundation Discussion Paper No. 1569.*

Kitamura, Y. and M. Stutzer (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica*, 65, 861-874.

Kunitomo, N. and Y. Matsushita (2003): “Finite Sample Distributions of the Empirical Likelihood Estimator and the GMM Estimator,” *CIRJE Discussion paper F-200.*

Mullahy, J. (1985): "Cigarette Smoking: Habits, Health Concern, and Heterogeneous Unobservables in a Microeconometric Analysis of Consumer Demand," Ph.D. Dissertation, University of Virginia.

Mullahy, J. (1997): "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics*, 79, 586-593.

Newey, W. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in: G. Maddala, C. Rao and H. Vinod (eds.) *Handbook of Statistics Vol. 11*, Elsevier Science, North Holland.

Newey, W.K. and R.J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

Owen, A.B. (1988): "Empirical Likelihood Ratio Confidence Regions for a Single Functional," *Biometrika*, 75, 237-249.

Owen, A.B. (1991): "Empirical Likelihood for Linear Models," *Annals of Statistics*, 19, 1725-1747.

Owen, A.B. (2001): *Empirical Likelihood*, Chapman & Hall/CRC, Boca Raton.

Pohlmeier, W. and V. Ulrich (1995): "An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care," *Journal of Human Resources*, 30, 339-361.

Qin, J. and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300-325.

Qin, J. and J. Lawless (1995): "Estimating Equations, Empirical Likelihood, and Constraints on Parameters," *Canadian Journal of Statistics*, 23, 145-159.

Smith, R.J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation," *Economic Journal*, 107, 503-519.

Windmeijer, F.A.G. and J.M.C. Santos Silva (1997): "Endogeneity in Count Data Models: An Application to Demand for Health Care," *Journal of Applied Econometrics*, 12, 281-294.

Winkelmann, R. (2003): *Econometric Analysis of Count Data*, Springer Verlag, Berlin.

Winkelmann, R. and K.F. Zimmermann (1994): "Count Data Models for Demographic Data," *Mathematical Population Studies*, 4, 205-221.

Wooldridge, J.M. (1992): "Some Alternatives to the Box-Cox Regression," *International Economic Review*, 33, 935-955.

Wooldridge, J.M. (1997): "Quasi-Likelihood Methods for Count Data," in: M.H. Pesaran and P. Schmidt (eds.) *Handbook of Applied Econometrics Vol. 2 - Microeconomics*, Blackwell.

Wooldridge, J.M. (2001): "Applications of Generalized Method of Moments Estimation," *Journal of Economic Perspectives*, 15, 87-100.

Tables

Table 1: Simulation Results for $se(\hat{\beta}_1)$ and χ^2-test; $n = 500$

	GMM		BCGMM		EL	
	Med.SE	Overid.	Med.SE	Overid.	Med.SE	Overid.
$z \sim N(0,1)$, $\alpha = .3$						
$K = 2$	.359	—	.359	—	.359	—
$K = 4$	.287	.058	.285	.051	.289	.063
$K = 8$	.246	.051	.246	.048	.243	.046
$K = 16$	.206	.063	.201	.052	.198	.058
$K = K^*$	.187	.053	.186	.061	.179	.059
$z \sim N(0,1)$, $\alpha = .7$						
$K = 2$	.158	—	.158	—	.158	—
$K = 4$	.141	.062	.141	.051	.140	.064
$K = 8$	.146	.065	.145	.049	.143	.060
$K = 16$	.146	.048	.139	.058	.137	.052
$K = K^*$	.125	.052	.125	.052	.125	.053
$z \sim LN(0,1)$, $\alpha = .3$						
$K = 2$	.141	—	.141	—	.141	—
$K = 4$	.107	.056	.106	.049	.104	.058
$K = 8$	.117	.054	.118	.051	.114	.062
$K = 16$	.116	.043	.114	.044	.108	.047
$K = K^*$	.069	.052	.070	.051	.073	.051
$z \sim LN(0,1)$, $\alpha = .7$						
$K = 2$	.061	—	.061	—	.061	—
$K = 4$	.045	.052	.045	.052	.043	.055
$K = 8$	.054	.053	.054	.046	.049	.054
$K = 16$	.053	.049	.052	.053	.048	.046
$K = K^*$	.032	.052	.032	.049	.031	.053

Notes: Med.SE is the median of the estimated standard error of $\hat{\beta}_1$, Overid. is the rejection rate for an overidentifying restrictions test with 5% nominal level. $K = 2$ is the basic IV setting, i.e., only the instrument z is included. Values $K > 2$ specify the fixed number of elements in the $q^K(z)$ vector, K^* is the optimal number of elements (which may vary draw by draw).

Table 2: Simulation Results for $\hat{\beta}_1$; $n = 500$

	GMM				BCGMM				EL			
	Med.Bias	MAD	$P(\text{AD} > d)$		Med.Bias	MAD	$P(\text{AD} > d)$		Med.Bias	MAD	$P(\text{AD} > d)$	
			d=.1	d=.2			d=.1	d=.2			d=.1	d=.2
$z \sim N(0,1)$, $\alpha = .3$												
$K = 2$	.005	.247	.819	.609	.005	.246	.819	.609	.005	.246	.819	.609
$K = 4$	-.021	.207	.717	.510	-.005	.216	.741	.537	-.006	.220	.740	.545
$K = 8$	-.024	.180	.692	.452	.006	.225	.756	.543	.001	.241	.778	.574
$K = 16$	-.043	.135	.617	.321	.006	.180	.683	.465	.002	.177	.707	.442
$K = K^*$	-.020	.127	.589	.303	-.001	.143	.621	.335	.001	.145	.655	.339
$z \sim N(0,1)$, $\alpha = .7$												
$K = 2$	.005	.111	.538	.244	.005	.111	.538	.244	.005	.111	.538	.244
$K = 4$	-.012	.110	.540	.253	-.001	.113	.547	.250	.003	.121	.593	.249
$K = 8$	-.033	.120	.573	.269	.001	.130	.609	.308	.005	.119	.561	.276
$K = 16$	-.088	.120	.569	.268	-.008	.132	.602	.314	-.007	.116	.564	.279
$K = K^*$	-.028	.092	.474	.164	-.010	.094	.470	.168	-.004	.096	.477	.171
$z \sim LN(0,1)$, $\alpha = .3$												
$K = 2$	-.002	.113	.542	.236	-.002	.113	.542	.236	-.002	.112	.542	.236
$K = 4$	-.026	.116	.534	.283	.001	.115	.557	.294	.001	.105	.509	.234
$K = 8$	-.032	.107	.514	.204	-.009	.116	.552	.232	-.009	.116	.558	.256
$K = 16$	-.050	.095	.460	.153	-.013	.099	.496	.200	-.007	.110	.535	.232
$K = K^*$	-.018	.091	.458	.195	-.009	.091	.464	.189	-.006	.088	.441	.154
$z \sim LN(0,1)$, $\alpha = .7$												
$K = 2$	-.005	.049	.192	.030	-.005	.049	.192	.030	-.005	.049	.192	.030
$K = 4$	.004	.048	.196	.031	-.001	.048	.192	.029	.002	.043	.157	.018
$K = 8$	-.010	.052	.208	.026	-.007	.051	.209	.026	-.004	.050	.211	.022
$K = 16$	-.040	.044	.164	.018	-.015	.041	.140	.010	-.009	.048	.176	.024
$K = K^*$	-.019	.048	.188	.028	-.009	.048	.170	.028	-.003	.043	.177	.028

Notes: See the notes of Table 1. Med.Bias is the median bias of the estimated $\hat{\beta}_1$ from the true value $\beta_1 = 1$, MAD is the median absolute deviation from the true value, and $P(AD > d)$ is the probability of $\hat{\beta}_1$ deviating from 1 by more than d.

Table 3: Simulation Results for $se(\hat{\beta}_1)$ and χ^2-test; $n = 2000$

	GMM		BCGMM		EL	
	Med.SE	Overid.	Med.SE	Overid.	Med.SE	Overid.
$z \sim N(0,1)$, $\alpha = .3$						
$K = 2$	.172	—	.172	—	.171	—
$K = 4$	.167	.051	.168	.052	.169	.051
$K = 8$	.166	.050	.166	.055	.163	.047
$K = 16$	.157	.049	.157	.048	.155	.052
$K = K^*$	.143	.051	.143	.049	.145	.048
$z \sim N(0,1)$, $\alpha = .7$						
$K = 2$	.092	—	.091	—	.091	—
$K = 4$	.079	.050	.079	.049	.078	.052
$K = 8$	.084	.053	.083	.048	.083	.051
$K = 16$	.091	.051	.091	.051	.089	.051
$K = K^*$	.074	.048	.074	.047	.074	.053

Notes: See the notes of Table 1.

Table 4: Simulation Results for $\hat{\beta}_1$; $n = 2000$

	GMM				BCGMM				EL			
	Med.Bias	MAD	$P(\text{AD} > d)$		Med.Bias	MAD	$P(\text{AD} > d)$		Med.Bias	MAD	$P(\text{AD} > d)$	
			d=.1	d=.2			d=.1	d=.2			d=.1	d=.2
$z \sim N(0,1)$, $\alpha = .3$												
$K = 2$	.005	.131	.572	.326	.005	.131	.572	.326	.005	.131	.572	.326
$K = 4$	.010	.116	.547	.293	.001	.122	.558	.312	-.002	.123	.575	.307
$K = 8$	.014	.118	.584	.268	.008	.140	.620	.325	.001	.144	.594	.355
$K = 16$	.027	.108	.536	.231	.002	.143	.658	.351	.002	.156	.684	.386
$K = K^*$	.023	.091	.470	.144	.003	.094	.473	.150	.001	.092	.485	.155
$z \sim N(0,1)$, $\alpha = .7$												
$K = 2$	.003	.092	.431	.087	.003	.092	.431	.087	.003	.092	.431	.087
$K = 4$	.004	.060	.237	.027	.007	.062	.247	.028	.007	.059	.250	.022
$K = 8$	-.016	.067	.284	.042	-.007	.067	.308	.044	-.002	.070	.332	.055
$K = 16$	-.034	.072	.375	.068	-.011	.082	.411	.093	.005	.077	.357	.070
$K = K^*$	-.008	.052	.205	.009	-.007	.053	.203	.011	-.005	.052	.209	.011

Notes: See the notes of Tables 1 and 2.

Table 5: The Effect of Smoking Habits on Cigarette Demand

	Poisson ML	GMM	BCGMM	EL
	12.53			
	(0.81)			
Basic instruments		11.33	12.04	11.63
		(14.35)	(14.98)	(14.97)
		[0.59]	[0.58]	[0.58]
Optimized over				
(a) rest. smoking restrictions {5}		8.05	7.23	7.17
		(6.13)	(5.89)	(5.90)
		[2.04]	[2.05]	[1.95]
(b) cigarette price in 1978 {1}		1.42	-0.57	-0.98
		(7.27)	(6.87)	(6.90)
		[2.28]	[2.04]	[2.04]
(a) and (b) {6}		7.04	5.82	5.58
		(5.80)	(5.49)	(5.35)
		[7.70]	[7.52]	[7.92]
(a) and (b) plus interaction {7}		6.34	4.47	7.09
		(5.50)	(5.11)	(5.40)
		[7.69]	[7.73]	[8.18]
all variables		8.86	7.61	7.08
{GMM, BCGMM: 19; EL: 21}		(4.04)	(3.85)	(3.51)
		[26.91]	[24.25]	[24.88]

Notes: All models control for age, years of schooling, two dummy variables indicating race and whether smoking restrictions had been in place in 1979, cigarette price in 1979, household income, and household size. The first value is the estimated coefficient; the second value (in round brackets) is the estimated asymptotic standard error; the third value (in square brackets) is the overidentifying test statistic with degrees of freedom the number in curly brackets plus one.

Excluded instruments: Cigarette price in 1978; number of years restaurant restrictions had been in place. In curly brackets is the number of additional elements, compared to the basic set of instruments, according to the specification of the $q^K(z)$ vector. Optimization over all variables adds functions of the included instruments and interactions.

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.unizh.ch/research/wp/index2.html

0704	Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach, Stefan Boes, March 2007, 26p.
0703	Risk and Rationality: The Effect of Incidental Mood on Probability Weighting, Helga Fehr, Thomas Epper, Adrian Bruhin, and Renate Schubert, February 2007, 27p.
0702	Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Family, Adrian Bruhin and Rainer Winkelmann, February 2007, 20p.
0701	On the Geographic and Cultural Determinants of Bankruptcy, Stefan Buehler, Christian Kaiser, and Franz Jaeger, February 2007, 35 p.
0610	A Product-Market Theory of Industry-Specific Training, Hans Gersbach and Armin Schmutzler , November 2006, 28 p.
0609	How to induce entry in railway markets: The German experience, Rafael Lalive and Armin Schmutzler, November 2006, 17 p.
0608	Does Greater Competition Increase R&D Investments? Evidence from a Laboratory Experiment, Dario Sacco and Armin Schmutzler, November 2006, 15 p.
0607	Merger Negotiations and Ex-Post Regret, Dennis Gärtner and Armin Schmutzler, September 2006, 28 p.
0606	Foreign Direct Investment and R&D offshoring, Hans Gersbach and Armin Schmutzler, June 2006, 34 p.
0605	The Effect of Income on Positive and Negative Subjective Well-Being, Stefan Boes and Rainer Winkelmann, May 2006, 23p.
0604	Correlated Risks: A Conflict of Interest Between Insurers and Consumers and Its Resolution, Patrick Eugster and Peter Zweifel, April 2006, 23p.
0603	The Apple Falls Increasingly Far: Parent-Child Correlation in Schooling and the Growth of Post-Secondary Education in Switzerland, Sandra Hanslin and Rainer Winkelmann, March 2006, 24p.
0602	Efficient Electricity Portfolios for Switzerland and the United States, Boris Krey and Peter Zweifel, February 2006, 25p.
0601	Ain't no puzzle anymore: Comparative statics and experimental economics, Armin Schmutzler, February 2006, 45 p.
0514	Money Illusion Under Test, Stefan Boes, Markus Lipp and Rainer Winkelmann, November 2005, 7p.
0513	Cost Sharing in Health Insurance: An Instrument for Risk Selection? Karolin Becker and Peter Zweifel, November 2005, 45p.
0512	Single Motherhood and (Un)Equal EducationalOpportunities: Evidence for Germany, Philippe Mahler and Rainer Winkelmann, September 2005, 23p.
0511	Competition for Railway Markets: The Case of Baden-Württemberg, Rafael Lalive and Armin Schmutzler, September 2005, 30p.
0510	The Impact of Aging on Future Healthcare Expenditure; Lukas Steinmann, Harry Telser, and Peter Zweifel, September 2005, 23p.
0509	The Purpose and Limits of Social Health Insurance; Peter Zweifel, September 2005, 28p.
0508	Switching Costs, Firm Size, and Market Structure; Simon Loertscher and Yves Schneider, August 2005, 29p.
0507	Ordered Response Models; Stefan Boes and Rainer Winkelmann, March 2005, 21p.

0506 Merge or Fail? The Determinants of Mergers and Bankruptcies in Switzerland, 1995-2000; Stefan Buehler, Christian Kaiser, Franz Jaeger, March 2005, 18p.
0505 Consumer Resistance Against Regulation: The Case of Health Care
Peter Zweifel, Harry Telser, and Stephan Vaterlaus, February 2005, 23p.
0504 A Structural Model of Demand for Apprentices
Samuel Mühlemann, Jürg Schweri, Rainer Winkelmann and Stefan C. Wolter, February 2005, 25p.
0503 What can happiness research tell us about altruism? Evidence from the German Socio-Economic Panel
Johannes Schwarze and Rainer Winkelmann, February 2005, 26p.
0502 Spatial Effects in Willingness-to-Pay: The Case of Nuclear Risks
Peter Zweifel, Yves Schneider and Christian Wyss, January 2005, 37p.
0501 On the Role of Access Charges Under Network Competition
Stefan Buehler and Armin Schmutzler, January 2005, 30p.
0416 Social Sanctions in Interethnic Relations: The Benefit of Punishing your Friends
Christian Stoff, Dezember 2004, 51p.
0415 Single Motherhood and (Un)equal Educational Opportunities: Evidence from Germany
Philippe Mahler and Rainer Winkelmann, November 2004, 23p.
0414 Are There Waves in Merger Activity After All?
Dennis Gärtner and Daniel Halbheer, September 2004, 39p.
0413 Endogenizing Private Information: Incentive Contracts under Learning By Doing
Dennis Gärtner, September 2004, 32p.
0412 Validity and Reliability of Willingness-to-pay Estimates: Evidence from Two Overlapping Discrete-Choice Experiments
Harry Telser, Karolin Becker and Peter Zweifel. September 2004, 25p.
0411 Willingness-to-pay Against Dementia: Effects of Altruism Between Patients and Their Spouse Caregivers
Markus König und Peter Zweifel, September 2004, 22p.
0410 Age and Choice in Health Insurance: Evidence from Switzerland
Karolin Becker and Peter Zweifel, August 2004, 30p.
0409 Vertical Integration and Downstream Investment in Oligopoly
Stefan Buehler and Armin Schmutzler, July 2004, 30p.
0408 Mergers under Asymmetric Information – Is there a Lemons Problem?
Thomas Borek, Stefan Buehler and Armin Schmutzler, July 2004, 38p.
0407 Income and Happiness: New Results from Generalized Threshold and Sequential Models
Stefan Boes and Rainer Winkelmann, June 2004, 30p.
0406 Optimal Insurance Contracts without the Non-Negativity Constraint on Indemnities Revisited
Michael Breuer, April 2004, 17p.
0405 Competition and Exit: Evidence from Switzerland
Stefan Buehler, Christian Kaiser and Franz Jaeger, March 2004, 28p.
0404 Empirical Likelihood in Count Data Models: The Case of Endogenous Regressors
Stefan Boes, March 2004, 22p.
0403 Globalization and General Worker Training
Hans Gersbach and Armin Schmutzler, February 2004, 37p.
0402 Restructuring Network Industries: Dealing with Price-Quality Tradeoffs
Stefan Bühler, Dennis Gärtner and Daniel Halbheer, January 2004, 18p.
0401 Deductible or Co-Insurance: Which is the Better Insurance Contract under Adverse Selection? Michael Breuer, January 2004, 18p.