Journal of Econometrics 165 (2011) 128–136

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Linear programming-based estimators in simple linear regression

Daniel Preve[a,*], Marcelo C. Medeiros[b]

[a] Department of Statistics, Uppsala University and School of Economics, Singapore Management University, Box 513, 751 20 Uppsala, Sweden
[b] Pontifical Catholic University of Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Available online 12 May 2011

Keywords:
Linear regression
Endogeneity
Linear programming estimator
Quasi-maximum likelihood estimator
Exact distribution

ABSTRACT

In this paper we introduce a linear programming estimator (LPE) for the slope parameter in a constrained linear regression model with a single regressor. The LPE is interesting because it can be superconsistent in the presence of an endogenous regressor and, hence, preferable to the ordinary least squares estimator (LSE). Two different cases are considered as we investigate the statistical properties of the LPE. In the first case, the regressor is assumed to be fixed in repeated samples. In the second, the regressor is stochastic and potentially endogenous. For both cases the strong consistency and exact finite-sample distribution of the LPE is established. Conditions under which the LPE is consistent in the presence of serially correlated, heteroskedastic errors are also given. Finally, we describe how the LPE can be extended to the case with multiple regressors and conjecture that the extended estimator is consistent under conditions analogous to the ones given herein. Finite-sample properties of the LPE and extended LPE in comparison to the LSE and instrumental variable estimator (IVE) are investigated in a simulation study. One advantage of the LPE is that it does not require an instrument.

1. Introduction

The use of certain linear programming estimators in time series analysis is well documented. See, for instance, Davis and McCormick (1989), Feigin and Resnick (1994) and Feigin et al. (1996). LPEs can yield much more precise estimates than traditional methods such as conditional least squares (e.g. Datta et al., 1998; Nielsen and Shephard, 2003). The limited success of these estimators in applied work can be partially explained by the fact that their point process limit theory complicates the use of their asymptotics for inference (e.g. Datta and McCormick, 1995).

In regression analysis, it is well known that the ordinary least squares estimator is inconsistent for the regression parameters when the error term is correlated with the explanatory variables of the model. In this case an instrumental variables estimator or the generalized method of moments may be used instead. In economics, such endogenous explanatory variables could be caused by measurement error, simultaneity or omitted variables. To the authors' knowledge, however, there has so far been no attempt to investigate the statistical properties of LP-based estimators in a cross-sectional setting. In this paper we show that LPEs can, under certain circumstances, be a preferable alternative to LS and IV estimators for the slope parameter in a simple linear regression model. We look at two types of regressors which are likely to be of practical importance. First, we introduce LPEs to the simple case of a non-stochastic regressor. Second, we consider the general case of a stochastic, and potentially endogenous, regressor. For both cases we establish the strong consistency and exact finite-sample distribution of a LPE for the slope parameter.

The LPE can be used in situations where the regressor is strictly positive. For example, in empirical finance, we can consider regressions involving volatility and volume. In labor economics a possible application is the regression between income and schooling, for example.

The remainder of the paper is organized as follows. In Section 2, we establish the strong consistency and exact finite-sample distribution of the LPE when (1) the explanatory variable is non-stochastic, and (2) the explanatory variable is stochastic and potentially endogenous. In Section 3, we discuss how our results can be extended to other endogenous specifications and give conditions under which the LPE is consistent in the presence of serially correlated, heteroskedastic errors. We also describe how the LPE can be extended to the case with multiple regressors. Section 4 reports the simulation results of a Monte Carlo study comparing the LPE and extended LPE to the LSE and IVE. Section 5 concludes. Mathematical proofs are collected in the Appendix. An extended Appendix available on request from the authors contains some results mentioned in the text but omitted from the paper to save space.

* Corresponding author.
E-mail addresses: daniel.preve@statistics.uu.se (D. Preve), mcm@econ.puc-rio.br (M.C. Medeiros).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.05.011

2. Assumptions and results

2.1. Non-stochastic explanatory variable

The first regression model we consider is

$$\begin{cases} y_i = \beta x_i + u_i \\ u_i = \alpha + \varepsilon_i, \quad i = 1, \ldots, n \end{cases}$$

where the response variable y_i and the explanatory variable x_i are observed, and u_i is the unobserved non-zero mean random error. β is the unknown regression parameter of interest. We assume that $\{x_i\}$ is a nonrandom sequence of strictly positive reals, whereas $\{u_i\}$ is a sequence of independent identically distributed (i.i.d.) nonnegative random variables (RVs). For ease of exposition we assume that $E(u_i) = \alpha$. The potentially unknown distribution function F_u of u_i is allowed to roam freely subject only to the restriction that it is supported on the nonnegative reals. A well known continuous probability distribution with nonnegative support is the Weibull distribution, which can approximate the shape of a Gaussian distribution quite well.

A 'quick and dirty' estimator of the slope parameter, based on the nonnegativity of the random errors, is given by

$$\hat{\beta} = \min\left\{\frac{y_1}{x_1}, \ldots, \frac{y_n}{x_n}\right\}. \tag{1}$$

This estimator has been used to estimate β in certain constrained first-order autoregressive time series models, $y_i = \beta x_i + u_i$, with $x_i = y_{i-1}$ (e.g. Datta and McCormick, 1995; Nielsen and Shephard, 2003). As it happens, (1) may be viewed as the solution to the linear programming problem of maximizing the objective function $f(\beta) = \beta$ subject to the n constraints $y_i - \beta x_i \geq 0$. Because of this we will sometimes refer to $\hat{\beta}$ as a LPE. Regardless if the regressor is stochastic or non-stochastic, (1) is also the maximum likelihood estimator (MLE) of β when the errors are exponentially distributed. What is interesting, however, is that $\hat{\beta}$ consistently estimates β for a wide range of error distributions, thus the LPE is also a quasi-MLE.

Assumption 1 holds throughout the section.

Assumption 1. Let $y_i = \beta x_i + u_i$ $(i = 1, \ldots, n)$ where $u_i = \alpha + \varepsilon_i$ and

(i) $\{x_i\}$ is a nonrandom sequence of strictly positive reals,
(ii) 0 is not a limit point of $S \equiv \{x_1, x_2, \ldots\}$,
(iii) $\{u_i\}$ is an i.i.d. sequence of nonnegative RVs,
(iv) $\inf\{u : F_u(u) > 0\} = 0$,
(v) $E(\varepsilon_i) = 0$.

Note that β can be any real number and that conditions (iii) and (v) combined imply that the mean of u_i is $\alpha \geq 0$. Since $\hat{\beta}_n - \beta = R_n$, where $R_n = \min\{u_i/x_i\}$, it is clear that $P(\hat{\beta}_n - \beta \leq z) = 0$ for all $z < 0$ and, hence, the LPE is positively biased. Moreover, as (1) is nonincreasing in the sample size its accuracy either remains the same or improves as n increases. Proposition 1 gives the exact distribution of the LPE in the case of a non-stochastic regressor.

Proposition 1. *Under Assumption 1,*

$$P(\hat{\beta}_n - \beta \leq z) = 1 - \prod_{i=1}^{n}[1 - F_u(x_i z)].$$

The proof of the proposition follows from the observation that

$$P(\hat{\beta}_n - \beta \leq z) = P(R_n \leq z) \overset{\text{(i)}}{=} 1 - P(u_1 > x_1 z, \ldots, u_n > x_n z),$$

and condition (iii) of Assumption 1. By condition (iv), $F_u(u) > 0$ for every $u > 0$ implying that $\hat{\beta}$ consistently estimates β.[1]

[1] If x_i instead is assumed to be strictly negative then the estimator $\max\{y_i/x_i\}$ is strongly consistent for β.

Table 1
Ratio distributions with accompanying moments of $\hat{\beta}_n$. $F_z(z)$ is the cdf of the ratio $z = u_1/x_1$, with parameter $\theta = \theta_u/\theta_x$, on which the moments are based. Results hold for $\gamma = 0$ and $n > 2$. $\Gamma(\cdot)$ is the gamma function.

Ratio	$F_z(z), z > 0$	$E(\hat{\beta}_n - \beta)$	$\mathrm{Var}(\hat{\beta}_n - \beta)$
$\frac{\mathrm{Exp}(\theta_u)}{\mathrm{Exp}(\theta_x)}$	$1 - \frac{1}{1+\theta^{-1}z}$	$\frac{\theta}{n-1}$	$\frac{\theta^2 n}{(n-2)(n-1)^2}$
$\frac{U(0,\theta_u)}{U(0,\theta_x)}$	$\frac{1}{2\theta}z, z \leq \theta$; $1 - \frac{\theta}{2z}, z > \theta$	$\frac{2\theta}{n+1}\left[1 + \frac{1}{(n-1)2^n}\right]$	$O\left(\frac{1}{n^2}\right)$
$\frac{\mathrm{Ra}(\theta_u)}{\mathrm{Ra}(\theta_x)}$	$1 - \frac{1}{1+\theta^{-2}z^2}$	$\frac{\theta\sqrt{\pi}}{2}\frac{\Gamma(n-1/2)}{\Gamma(n)}$	$\theta^2\left[\frac{1}{n-1} - \frac{\pi}{4}\frac{\Gamma^2(n-1/2)}{\Gamma^2(n)}\right]$

Intuitively, this is because the left-tail condition on u_i implies that the probability of obtaining an error arbitrarily close to 0 is non-zero and, hence, that (1) is likely to be precise in large samples.

Corollary 1. *Under Assumption 1, $\hat{\beta}_n \xrightarrow{a.s.} \beta$ as $n \to \infty$.*

From Corollary 1 it follows that α (the unknown mean of the error term) can be consistently estimated by

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}x_i), \tag{2}$$

the sample mean of the residuals, under fairly weak conditions.[2]

It is worth noting that the MLE of β satisfies the stochastic inequality $\hat{\beta}_{\mathrm{ML}} \leq \hat{\beta}$. Regardless if x_i is stochastic or non-stochastic, in some cases the LPE will be equal to $\hat{\beta}_{\mathrm{ML}}$. For instance, it is readily verified that if the random errors are (1) exponentially distributed with non-zero density function $(1/a)\exp\{-u/a\}$ for $u \geq 0$

$$\hat{\beta}_{\mathrm{ML}} = \hat{\beta}, \qquad \hat{a}_{\mathrm{ML}} = \hat{\alpha}, \tag{3}$$

and (2) uniformly distributed on the interval $[0, b]$

$$\hat{\beta}_{\mathrm{ML}} = \hat{\beta}, \qquad \hat{b}_{\mathrm{ML}} = \max\{y_i - \hat{\beta}x_i\}. \tag{4}$$

As an illustration of Proposition 1 in action, Corollary 2 shows that the exact distribution of $\hat{\beta} - \beta$ when the errors are Weibull distributed, and the regressor is non-stochastic, is also Weibull. The Weibull distribution, with distribution function $1 - \exp\{-(u/a)^b\}$ for $u \geq 0$, nests the well known exponential ($b = 1$) and Rayleigh ($b = 2$) distributions.

Corollary 2. *Let the regression errors be Weibull distributed. Then, under Assumption 1,*

$$P(\hat{\beta}_n - \beta \leq z) = 1 - \exp\left\{-\left[\frac{z}{a\left(\sum_{i=1}^{n} x_i^b\right)^{-1/b}}\right]^b\right\},$$

if $z \geq 0$ and 0 otherwise. Hence, $\hat{\beta}_n - \beta$ is Weibull with scale parameter $a\left(\sum_{i=1}^{n} x_i^b\right)^{-1/b}$ and shape parameter b.

For example, in view of Corollary 2 with $b = 1$ it is clear that

$$\sum_{i=1}^{n} x_i(\hat{\beta}_n - \beta),$$

is exponentially distributed with scale parameter a. Moreover, by (3) and basic results of large sample theory, the statistic

$$\frac{1}{\hat{\alpha}_n}\sum_{i=1}^{n} x_i(\hat{\beta}_n - \beta),$$

is asymptotically standard exponential.

[2] If, under Assumption 1, $\alpha < \infty$ and if $n^{-1}\sum_{i=1}^{n} x_i$ is $O(1)$ as $n \to \infty$.

Table 2
Simulation results: univariate regression with i.i.d. uniformly distributed errors—specification (i). Each table entry, based on 1000 Monte Carlo replications, reports the empirical bias/mean squared error (MSE) of different estimators for the slope parameter $\beta_1 = 2.5$ in the univariate regression $y_i = 2.5x_{1i} + u_i$, where $x_{1i} = v_{1i} + \gamma u_i$, $v_{1i} \sim \chi^2(3)$ and $u_i \sim U(0, 10)$. The following estimators are considered: the ordinary least squares estimator (LSE), the instrumental variable estimator (IVE) and the linear programming estimator (LPE). For the IVE, the variable v_{1i} is used as an instrument. Finally, for both the LSE and IVE an intercept is included in the regression equation. Different sample sizes (n) and levels of endogeneity (γ) are considered.

n	$\gamma = 0$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	−0.001	0.031	–	–	−0.001	0.031	–	–	0.067	0.009	–	–
100	0.008	0.015	–	–	0.008	0.015	–	–	0.032	0.002	–	–
200	−0.002	0.007	–	–	−0.002	0.007	–	–	0.016	0.001	–	–
500	0.002	0.003	–	–	0.002	0.003	–	–	0.007	0.000	–	–
1000	−0.000	0.002	–	–	−0.000	0.002	–	–	0.003	0.000	–	–
2000	0.000	0.001	–	–	0.000	0.001	–	–	0.002	0.000	–	–
n	$\gamma = 0.25$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.335	0.144	–	–	−0.019	0.033	–	–	0.065	0.008	–	–
100	0.332	0.124	–	–	−0.003	0.014	–	–	0.033	0.002	–	–
200	0.323	0.112	–	–	−0.003	0.008	–	–	0.017	0.001	–	–
500	0.322	0.107	–	–	0.001	0.003	–	–	0.007	0.000	–	–
1000	0.320	0.104	–	–	−0.001	0.001	–	–	0.003	0.000	–	–
2000	0.325	0.103	–	–	−0.001	0.001	–	–	0.002	0.000	–	–
n	$\gamma = 0.5$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.542	0.316	–	–	−0.023	0.035	–	–	0.065	0.008	–	–
100	0.538	0.291	–	–	−0.006	0.015	–	–	0.033	0.002	–	–
200	0.519	0.275	–	–	−0.006	0.007	–	–	0.017	0.001	–	–
500	0.519	0.272	–	–	−0.002	0.003	–	–	0.007	0.000	–	–
1000	0.519	0.270	–	–	0.001	0.001	–	–	0.003	0.000	–	–
2000	0.516	0.267	–	–	−0.001	0.001	–	–	0.002	0.000	–	–
n	$\gamma = 1$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.590	0.356	–	–	−0.025	0.043	–	–	0.058	0.006	–	–
100	0.586	0.348	–	–	−0.016	0.019	–	–	0.032	0.002	–	–
200	0.583	0.342	–	–	−0.003	0.008	–	–	0.016	0.001	–	–
500	0.584	0.339	–	–	−0.002	0.003	–	–	0.007	0.000	–	–
1000	0.583	0.340	–	–	−0.002	0.001	–	–	0.003	0.000	–	–
2000	0.581	0.338	–	–	−0.002	0.001	–	–	0.002	0.000	–	–

2.2. Stochastic explanatory variable

The second regression model we consider is

$$\begin{cases} y_i = \beta x_i + u_i \\ x_i = v_i + \gamma u_i \\ u_i = \alpha + \varepsilon_i \end{cases}$$

where $\{v_i\}$ is an i.i.d. sequence of nonnegative RVs, $\{u_i\}$ and $\{v_i\}$ are mutually independent, and $\gamma \geq 0$ such that $\mathrm{Cov}(x_i, u_i) = \gamma\,\mathrm{Var}(u_i)$. The parameter γ is potentially unknown. For this model the explanatory variable and error are uncorrelated if and only if $\gamma = 0$. In this case $E(y_i|x_i) = \beta x_i + \alpha$. Here $\gamma > 0$ is a typical setting in which the LSE of β is inconsistent.[3]

Assumption 2 holds throughout the section.

Assumption 2. Let $y_i = \beta x_i + u_i$ ($i = 1, \ldots, n$) where $u_i = \alpha + \varepsilon_i$ and

(i) $x_i = v_i + \gamma u_i$ for some $\gamma \geq 0$,
(ii) $\{u_i\}$ and $\{v_i\}$ are mutually independent i.i.d. sequences of nonnegative RVs,
(iii) $\inf\{u : F_u(u) > 0\} = 0$,
(iv) $P(v_i = 0) = 0$,
(v) $E(\varepsilon_i) = 0$.

Conditions (i) through (iv) ensures that x_i is strictly positive and, hence, that (1) is well-defined. Also for this case the exact distribution of the LPE can be obtained. For ease of exposition, we only give the result for the important special case when the related distributions are continuous.

Proposition 2. *Let u_i and v_i be (absolutely) continuous RVs with pdfs f_u and f_v, respectively, and let $\mathbf{1}_{\{\cdot\}}$ denote the indicator function. Then, under Assumption 2,*

$$P(\hat{\beta}_n - \beta \leq z) = 1 - [1 - F_z(z)]^n, \quad (5)$$

where

$$F_z(z) = \mathbf{1}_{\{z>0\}} \int_0^z \int_0^\infty x f_v(x) f_u(tx)\,\mathrm{d}x\,\mathrm{d}t,$$

[3] More specifically, $\hat{\beta}_{\mathrm{LS}} \xrightarrow{p} \beta + \gamma\,\mathrm{Var}(u_i)/[\mathrm{Var}(v_i) + \gamma^2\,\mathrm{Var}(u_i)]$ as $n \to \infty$ provided the variances exist.

Table 3
Simulation results: bivariate regression with i.i.d. uniformly distributed errors—specification (ii). Each table entry, based on 1000 Monte Carlo replications, reports the empirical bias/mean squared error (MSE) of different estimators for the slope parameters $\beta_1 = 2.5$ and $\beta_2 = -1.5$ in the bivariate regression $y_i = 2.5x_{1i} - 1.5x_{2i} + u_i$, where $x_{1i} = v_{1i} + \gamma u_i$, $x_{2i} = v_{2i} + \gamma u_i$, with $v_{1i} \sim \chi^2(3)$ and $v_{2i} \sim \chi^2(4)$, and $u_i \sim U(0, 10)$. The following estimators are considered: the ordinary least squares estimator (LSE), the instrumental variable estimator (IVE) and the extended linear programming estimator (LPE). For the IVE, the variables v_{1i} and v_{2i} are used as instruments. Finally, for both the LSE and IVE an intercept is included in the regression equation. Different sample sizes (n) and levels of endogeneity (γ) are considered.

n	LSE β_1 Bias	LSE β_1 MSE	LSE β_2 Bias	LSE β_2 MSE	IVE β_1 Bias	IVE β_1 MSE	IVE β_2 Bias	IVE β_2 MSE	LPE β_1 Bias	LPE β_1 MSE	LPE β_2 Bias	LPE β_2 MSE
$\gamma = 0$												
50	0.000	0.031	−0.006	0.024	0.000	0.031	−0.006	0.024	0.098	0.035	0.012	0.019
100	−0.008	0.015	−0.004	0.012	−0.008	0.015	−0.004	0.012	0.043	0.007	0.008	0.004
200	−0.001	0.007	0.001	0.006	−0.001	0.007	0.001	0.006	0.026	0.002	0.002	0.001
500	0.002	0.003	0.001	0.002	0.002	0.003	0.001	0.002	0.011	0.000	0.001	0.000
1000	0.000	0.001	−0.001	0.001	0.000	0.001	−0.001	0.001	0.005	0.000	0.000	0.000
2000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.003	0.000	0.000	0.000
$\gamma = 0.25$												
50	0.313	0.128	0.237	0.076	−0.008	0.034	−0.010	0.024	0.092	0.028	0.015	0.014
100	0.311	0.110	0.228	0.062	−0.006	0.015	−0.009	0.012	0.049	0.009	0.003	0.005
200	0.308	0.102	0.230	0.058	0.000	0.007	−0.000	0.005	0.025	0.002	0.003	0.001
500	0.301	0.094	0.228	0.054	−0.002	0.003	0.001	0.002	0.010	0.000	0.001	0.000
1000	0.301	0.092	0.226	0.052	−0.002	0.001	−0.001	0.001	0.005	0.000	0.000	0.000
2000	0.303	0.092	0.227	0.052	0.001	0.001	0.000	0.001	0.003	0.000	0.000	0.000
$\gamma = 0.5$												
50	0.446	0.220	0.333	0.126	−0.014	0.033	−0.020	0.026	0.088	0.027	0.012	0.014
100	0.448	0.213	0.329	0.115	−0.009	0.017	−0.002	0.011	0.049	0.008	0.006	0.004
200	0.438	0.197	0.325	0.109	−0.005	0.008	−0.005	0.006	0.024	0.002	0.003	0.001
500	0.436	0.192	0.325	0.107	−0.000	0.003	−0.002	0.002	0.010	0.000	0.001	0.000
1000	0.435	0.190	0.323	0.105	0.001	0.002	−0.003	0.001	0.006	0.000	0.000	0.000
2000	0.433	0.188	0.324	0.106	0.000	0.001	−0.001	0.001	0.003	0.000	0.000	0.000
$\gamma = 1$												
50	0.408	0.176	0.310	0.104	−0.039	0.050	−0.021	0.036	0.078	0.022	0.018	0.013
100	0.411	0.174	0.303	0.096	−0.008	0.017	−0.012	0.013	0.049	0.009	0.003	0.005
200	0.406	0.167	0.305	0.095	−0.004	0.008	−0.005	0.006	0.025	0.002	0.002	0.001
500	0.404	0.164	0.306	0.094	−0.006	0.003	−0.002	0.002	0.010	0.000	0.002	0.000
1000	0.407	0.164	0.304	0.093	−0.000	0.001	−0.000	0.001	0.004	0.000	0.001	0.000
2000	0.406	0.165	0.304	0.092	−0.001	0.001	−0.001	0.001	0.003	0.000	0.000	0.000

if $\gamma = 0$ and

$$F_z(z) = \mathbf{1}_{\{0<z<1/\gamma\}} \int_0^z \int_0^\infty x f_v(x - \gamma t x) f_u(tx) \mathrm{d}x\mathrm{d}t + \mathbf{1}_{\{z \geq 1/\gamma\}},$$

otherwise.

For a simple example, consider the case when u_i and v_i are standard exponentially distributed RVs and γ is non-zero. Then, in view of Proposition 2,

$$F_z(z) = \mathbf{1}_{\{0<z<1/\gamma\}} \int_0^z \int_0^\infty x \exp\{-x[1 + (1-\gamma)t]\} \mathrm{d}x\mathrm{d}t + \mathbf{1}_{\{z \geq 1/\gamma\}}.$$

Hence, if $\gamma \neq 1$

$$F_z(z) = \mathbf{1}_{\{0<z<1/\gamma\}} \int_0^z \frac{1}{[1 + (1-\gamma)t]^2} \mathrm{d}t + \mathbf{1}_{\{z \geq 1/\gamma\}}$$
$$= \mathbf{1}_{\{0<z<1/\gamma\}} \left(\frac{1}{1-\gamma} - \frac{1}{(1-\gamma)[1 + (1-\gamma)z]} \right) + \mathbf{1}_{\{z \geq 1/\gamma\}}.$$

Similarly, if $\gamma = 1$ then $F_z(z) = \mathbf{1}_{\{0<z<1\}} z + \mathbf{1}_{\{z \geq 1\}}$ and z_i is uniformly distributed on $(0, 1)$.

Once $F_z(z)$ is obtained the mean and variance of $\hat{\beta}_n$ may be calculated from Eq. (5). To illustrate that the LPE can be superconsistent (and hence superior to the LSE), Table 1 reports the exact mean and variance of $\hat{\beta}_n$ under various distributional specifications for u_i and x_i. More specifically, the table gives three examples of the ratio distribution $F_z(z)$ of $z_i = u_i/x_i$ where u_i and x_i are independent ($\gamma = 0$) and follow the same family of distributions. The first case is the exponential distribution. The second and third cases are the uniform and Rayleigh distributions, respectively. The results for the mean can be used to bias-correct $\hat{\beta}_n$. The results for the variance imply that $\hat{\beta}_n$ is n-consistent in the first two cases, and $\sqrt{n}$-consistent in the last. It is easy to see that the LPE can be superconsistent also in the presence of an endogenous regressor. For instance, if $\gamma = 1$ in the example following Proposition 2 then $\mathrm{Var}(\hat{\beta}_n) = n(n+1)^{-2}(n+2)^{-1}$. Next we establish the strong consistency of $\hat{\beta}$.

Proposition 3. *Under Assumption 2, $\hat{\beta}_n \xrightarrow{\text{a.s.}} \beta$ as $n \to \infty$.*

Table 4
Simulation results: univariate regression with heteroskedastic errors—specification (iii). Each table entry, based on 1000 Monte Carlo replications, reports the empirical bias/mean squared error (MSE) of different estimators for the slope parameter $\beta_1 = 2.5$ in the univariate regression $y_i = 2.5x_{1i} + \sigma_i u_i$, where $x_{1i} = v_{1i} + \gamma u_i$, $v_{1i} \sim \chi^2(3)$, $\sigma_i^2 = 0.25 + 0.75\frac{i}{n}$ and $u_i \sim U(0, \sqrt{12})$. The following estimators are considered: the ordinary least squares estimator (LSE), the instrumental variable estimator (IVE) and the linear programming estimator (LPE). For the IVE, the variable v_{1i} is used as an instrument. Finally, for both the LSE and IVE an intercept is included in the regression equation. Different sample sizes (n) and levels of endogeneity (γ) are considered.

n	$\gamma = 0$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	−0.000	0.003	–	–	−0.000	0.003	–	–	0.018	0.008	–	–
100	−0.000	0.001	–	–	−0.000	0.001	–	–	0.008	0.000	–	–
200	0.002	0.001	–	–	0.002	0.001	–	–	0.004	0.000	–	–
500	0.000	0.000	–	–	0.000	0.000	–	–	0.002	0.000	–	–
1000	0.000	0.000	–	–	0.000	0.000	–	–	0.001	0.000	–	–
2000	−0.000	0.000	–	–	−0.000	0.000	–	–	0.000	0.000	–	–

n	$\gamma = 0.25$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.031	0.004	–	–	−0.000	0.003	–	–	0.017	0.001	–	–
100	0.029	0.002	–	–	−0.001	0.001	–	–	0.009	0.000	–	–
200	0.029	0.001	–	–	−0.001	0.001	–	–	0.004	0.000	–	–
500	0.028	0.001	–	–	−0.001	0.000	–	–	0.002	0.000	–	–
1000	0.028	0.001	–	–	−0.000	0.000	–	–	0.001	0.000	–	–
2000	0.029	0.001	–	–	0.000	0.000	–	–	0.000	0.000	–	–

n	$\gamma = 0.5$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.060	0.006	–	–	−0.001	0.003	–	–	0.017	0.001	–	–
100	0.056	0.004	–	–	−0.003	0.001	–	–	0.009	0.000	–	–
200	0.057	0.004	–	–	0.001	0.001	–	–	0.004	0.000	–	–
500	0.056	0.003	–	–	−0.000	0.000	–	–	0.002	0.000	–	–
1000	0.056	0.003	–	–	0.000	0.000	–	–	0.001	0.000	–	–
2000	0.056	0.003	–	–	0.000	0.000	–	–	0.000	0.000	–	–

n	$\gamma = 1$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.108	0.014	–	–	−0.005	0.003	–	–	0.016	0.001	–	–
100	0.107	0.013	–	–	−0.001	0.001	–	–	0.009	0.000	–	–
200	0.104	0.011	–	–	−0.002	0.001	–	–	0.004	0.000	–	–
500	0.104	0.011	–	–	0.000	0.000	–	–	0.002	0.000	–	–
1000	0.104	0.011	–	–	0.000	0.000	–	–	0.001	0.000	–	–
2000	0.103	0.011	–	–	−0.000	0.000	–	–	0.000	0.000	–	–

Hence, under Assumption 2, the LPE is strongly consistent in the presence of an endogenous regressor. Once more, it follows that $\hat{\alpha}$ in (2) is consistent for α under fairly weak conditions.[4]

3. Extensions

In the previous section we aimed for clarity at the expense of generality. For example, in the case with a stochastic regressor, we assumed that $x_i = v_i + \gamma u_i$ even though other endogenous specifications, such as $x_i = v_i u_i^{\gamma}$, also are possible. In this section we discuss how the results of Section 2 can be extended.

3.1. Serially correlated, heteroskedastic errors

A proof similar to that of Theorem 3.1 in Preve (2011) shows that the LPE remains consistent for certain serially correlated error specifications such as

$$u_i = \alpha + \varepsilon_i + \sum_{k=1}^{q} \psi_k \varepsilon_{i-k},$$

or $u_i = \alpha + \varepsilon_i + \psi \varepsilon_{i-1}\varepsilon_{i-2}$. Consistency also holds for certain heteroskedastic specifications. Because of this, $\hat{\beta}$ can be used to seek sources of misspecification in the errors.

Proposition 4. *Let $y_i = \beta x_i + \sigma_i u_i$ $(i = 1, \ldots, n)$ where*

(i) *$x_i = v_i + \gamma h(\sigma_i) u_i$ for some $\gamma \geq 0$ and $h : (0, \infty) \to (0, \infty)$,*
(ii) *$\{v_i\}$ is an i.i.d. sequence of nonnegative RVs, mutually independent of $\{u_i\}$, with $P(v_i = 0) = 0$,*
(iii) *$\{\sigma_i\}$ is a deterministic sequence of strictly positive reals with $\sup\{\sigma_i\} < \infty$,*
(iv) *$\{u_i\}$ is a sequence of m-dependent identically distributed nonnegative RVs with $\inf\{u : F_u(u) > 0\} = 0$.*

Then, $\hat{\beta}_n \xrightarrow{p} \beta$ as $n \to \infty$. The endogeneity parameter γ, the map $h(\cdot)$ and the specification of $\{\sigma_i\}$ are potentially unknown. $m \in \mathbb{N}$ is finite and also potentially unknown.

4 If, under Assumption 2, both $E(v_i)$ and α are finite.

Table 5
Simulation results: bivariate regression with heteroskedastic errors—specification (iv). Each table entry, based on 1000 Monte Carlo replications, reports the empirical bias/mean squared error (MSE) of different estimators for the slope parameters $\beta_1 = 2.5$ and $\beta_2 = -1.5$ in the bivariate regression $y_i = 2.5x_{1i} - 1.5x_{2i} + \sigma_i u_i$, where $x_{1i} = v_{1i} + \gamma u_i$, $x_{2i} = v_{2i} + \gamma u_i$, with $v_{1i} \sim \chi^2(3)$ and $v_{2i} \sim \chi^2(4)$, $\sigma_i^2 = 0.25 + 0.75\frac{i}{n}$ and $u_i \sim U(0, \sqrt{12})$. The following estimators are considered: the ordinary least squares estimator (LSE), the instrumental variable estimator (IVE) and the extended linear programming estimator (LPE). For the IVE, the variables v_{1i} and v_{2i} are used as instruments. Finally, for both the LSE and IVE an intercept is included in the regression equation. Different sample sizes (n) and levels of endogeneity (γ) are considered.

n	$\gamma = 0$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	−0.002	0.003	−0.001	0.002	−0.002	0.003	−0.001	0.002	0.026	0.002	0.002	0.001
100	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.013	0.001	0.001	0.000
200	0.001	0.001	−0.001	0.000	0.001	0.001	−0.001	0.000	0.007	0.000	0.001	0.000
500	−0.001	0.000	0.000	0.000	−0.001	0.000	0.000	0.000	0.003	0.000	0.000	0.000
1000	−0.000	0.000	0.000	0.000	−0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
2000	0.000	0.000	−0.000	0.000	0.000	0.000	−0.000	0.000	0.001	0.000	0.000	0.000
n	$\gamma = 0.25$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.031	0.003	0.025	0.003	−0.000	0.002	0.002	0.002	0.028	0.003	0.001	0.001
100	0.029	0.002	0.021	0.001	−0.001	0.001	0.002	0.001	0.014	0.001	0.001	0.000
200	0.030	0.002	0.021	0.001	0.001	0.001	0.000	0.001	0.007	0.000	0.001	0.000
500	0.028	0.001	0.021	0.001	−0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
1000	0.028	0.001	0.021	0.001	−0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000
2000	0.029	0.001	0.021	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
n	$\gamma = 0.5$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.060	0.006	0.044	0.004	−0.001	0.003	−0.001	0.002	0.025	0.002	0.003	0.001
100	0.057	0.005	0.042	0.003	0.001	0.001	−0.000	0.001	0.014	0.001	0.001	0.000
200	0.055	0.004	0.042	0.002	−0.001	0.001	0.000	0.001	0.007	0.000	0.000	0.000
500	0.055	0.003	0.042	0.002	−0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
1000	0.055	0.003	0.041	0.002	−0.000	0.000	−0.000	0.000	0.001	0.000	0.000	0.000
2000	0.054	0.003	0.041	0.002	−0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000
n	$\gamma = 1$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.100	0.012	0.078	0.008	−0.004	0.003	−0.000	0.002	0.024	0.002	0.004	0.001
100	0.100	0.011	0.074	0.006	−0.000	0.001	0.000	0.001	0.013	0.001	0.001	0.001
200	0.098	0.010	0.072	0.006	−0.000	0.001	−0.001	0.001	0.007	0.000	0.001	0.000
500	0.096	0.010	0.071	0.005	−0.000	0.000	−0.001	0.000	0.003	0.000	0.000	0.000
1000	0.096	0.009	0.072	0.005	−0.001	0.000	−0.000	0.000	0.001	0.000	0.000	0.000
2000	0.095	0.009	0.072	0.005	−0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000

The σ_i are scaling constants which express the possible heteroskedasticity. The map $h(\cdot)$ allows for a heteroskedastic regressor. Condition (iii) is quite general and allows for various standard specifications, including abrupt breaks or smooth transitions such as $\sigma_i = \sqrt{\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\frac{i}{n}}$. If σ_i is not a function of n, then the convergence of $\hat{\beta}_n$ is almost surely.

3.2. Multiple regressors

Let $y_i = \sum_{j=1}^{p} \beta_j x_{ji} + u_i$ $(i = 1, \ldots, n)$ and, along the lines of Feigin and Resnick (1994), let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$ be the solution to the linear programming problem of maximizing the objective function $f(\beta_1, \ldots, \beta_p) = \sum_{j=1}^{p} \beta_j$ subject to the n constraints $y_i - \sum_{j=1}^{p} \beta_j x_{ji} \geq 0$. Note that (1) is the solution to the above problem for the special case when $p = 1$. The finite-sample and asymptotic properties of the extended LPE $\hat{\boldsymbol{\beta}}$ is the subject of further research. We conjecture that the extended LPE consistently estimates $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ under conditions analogous to Assumption 2. The proposed estimator is easily computable using standard numerical computing environments such as MATLAB. Our simulations indicate that the extended LPE can have very reasonable finite-sample properties, also in the presence of heteroskedastic or serially correlated errors.[5]

4. Monte Carlo results

In this section we report simulation results concerning the estimation of the slope parameters β_1 and β_2 in the regression

$$\begin{cases} y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i u_i \\ x_{1i} = v_{1i} + \gamma u_i \\ x_{2i} = v_{2i} + \gamma u_i, \quad i = 1, \ldots, n \end{cases} \tag{6}$$

where v_{1i} is a chi-square RV with three degrees of freedom, $v_{1i} \sim \chi^2(3)$, and v_{2i} is a chi-square RV with four degrees of

[5] Sample MATLAB code can be downloaded from http://www.mysmu.edu/staff/danielpreve.

Table 6
Simulation results: univariate regression with serially correlated errors—specification (v). Each table entry, based on 1000 Monte Carlo replications, reports the empirical bias/mean squared error (MSE) of different estimators for the slope parameter $\beta_1 = 2.5$ in the univariate regression $y_i = 2.5x_{1i} + u_i$, where $x_{1i} = v_{1i} + \gamma u_i$, $v_{1i} \sim \chi^2(3)$ and $u_i = w_i(1 + 0.8w_{i-1})$ with i.i.d. noise $w_i \sim U(0, 10)$. The following estimators are considered: the ordinary least squares estimator (LSE), the instrumental variable estimator (IVE) and the linear programming estimator (LPE). For the IVE, the variable v_{1i} is used as an instrument. Finally, for both the LSE and IVE an intercept is included in the regression equation. Different sample sizes (n) and levels of endogeneity (γ) are considered.

n	$\gamma = 0$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.052	1.419	–	–	0.052	1.419	–	–	0.263	0.136	–	–
100	0.006	0.646	–	–	0.006	0.646	–	–	0.119	0.029	–	–
200	0.019	0.335	–	–	0.019	0.335	–	–	0.060	0.007	–	–
500	−0.012	0.131	–	–	−0.012	0.131	–	–	0.024	0.001	–	–
1000	−0.007	0.064	–	–	−0.007	0.064	–	–	0.012	0.000	–	–
2000	0.001	0.032	–	–	0.001	0.032	–	–	0.006	0.000	–	–
n	$\gamma = 0.25$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	3.199	10.323	–	–	−0.593	4.342	–	–	0.232	0.102	–	–
100	3.193	10.247	–	–	−0.180	1.108	–	–	0.117	0.027	–	–
200	3.196	10.242	–	–	−0.088	0.380	–	–	0.062	0.007	–	–
500	3.196	10.225	–	–	−0.029	0.131	–	–	0.024	0.001	–	–
1000	3.204	10.275	–	–	−0.012	0.071	–	–	0.011	0.000	–	–
2000	3.204	10.271	–	–	−0.008	0.031	–	–	0.006	0.000	–	–
n	$\gamma = 0.5$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	1.883	3.555	–	–	−0.614	33.279	–	–	0.195	0.067	–	–
100	1.879	3.536	–	–	−0.369	8.628	–	–	0.108	0.021	–	–
200	1.882	3.546	–	–	−0.241	1.411	–	–	0.057	0.006	–	–
500	1.882	3.543	–	–	−0.061	0.186	–	–	0.023	0.001	–	–
1000	1.882	3.545	–	–	−0.034	0.078	–	–	0.012	0.000	–	–
2000	1.882	3.544	–	–	−0.014	0.032	–	–	0.006	0.000	–	–
n	$\gamma = 1$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.983	0.968	–	–	0.212	18.330	–	–	0.177	0.049	–	–
100	0.984	0.969	–	–	0.002	16.321	–	–	0.100	0.017	–	–
200	0.984	0.969	–	–	−0.145	22.566	–	–	0.056	0.005	–	–
500	0.984	0.969	–	–	−0.252	3.497	–	–	0.023	0.001	–	–
1000	0.984	0.969	–	–	−0.071	0.162	–	–	0.011	0.000	–	–
2000	0.984	0.969	–	–	−0.032	0.044	–	–	0.006	0.000	–	–

freedom, $v_{2i} \sim \chi^2(4)$. The sequences $\{v_{1i}\}$ and $\{v_{2i}\}$ are mutually independent. We write $u_i \sim U(0, b)$ to indicate that u_i is uniformly distributed on the interval $[0, b]$ and consider different specifications of (6):

(i) $\beta_1 = 2.5$, $\beta_2 = 0$, $\sigma_i = 1$ and $u_i \sim U(0, 10)$.[6]
(ii) Same specification as in (i) but with $\beta_2 = -1.5$.
(iii) $\beta_1 = 2.5$, $\beta_2 = 0$, $\sigma_i = \sqrt{0.25 + 0.75\frac{i}{n}}$ and $u_i \sim U(0, \sqrt{12})$ with $\mathrm{Var}(u_i) = 1$.
(iv) Same specification as in (iii) but with $\beta_2 = -1.5$.
(v) $\beta_1 = 2.5$, $\beta_2 = 0$, $\sigma_i = 1$ and $u_i = w_i(1 + 0.8w_{i-1})$ with i.i.d. noise $w_i \sim U(0, 10)$.
(vi) Same specification as in (v) but with $\beta_2 = -1.5$.

For the odd-numbered specifications, which are all simple regressions, we use the LPE in (1). For the even-numbered specifications we use the extended LPE described in Section 3 and compute it using the MATLAB function `linprog`. We report the empirical bias and mean squared error (MSE) over 1000 Monte Carlo replications and consider the following estimators: the LSE, IVE and LPE. We consider different sample sizes and levels of endogeneity. The simulation results are shown in Tables 2–7.

In general, the results indicate that the LPE has a higher bias than the IVE but a substantially lower MSE, suggesting that the LPE has a considerably smaller variance than the IVE. For example, for specification (v) with $\gamma = 0.5$ and $n = 200$ the MSE of the IVE and LPE is 1.411 and 0.006, respectively. Similarly, the results for the extended LPE indicate that it can be consistent in the presence of heteroskedastic or serially correlated errors and that its variability is much lower than that of the IVE.

5. Conclusions

In this paper we have established the exact finite-sample distribution of a LPE for the slope parameter in a constrained simple linear regression model when (1) the regressor is non-stochastic, and (2) the regressor is stochastic and potentially

[6] Hence, $\alpha = 5$ and $\varepsilon_i \sim U(-5, 5)$ in this specification.

Table 7
Simulation results: bivariate regression with serially correlated errors—specification (vi). Each table entry, based on 1000 Monte Carlo replications, reports the empirical bias/mean squared error (MSE) of different estimators for the slope parameters $\beta_1 = 2.5$ and $\beta_2 = -1.5$ in the bivariate regression $y_i = 2.5x_{1i} - 1.5x_{2i} + u_i$, where $x_{1i} = v_{1i} + \gamma u_i$, $x_{2i} = v_{2i} + \gamma u_i$, with $v_{1i} \sim \chi^2(3)$ and $v_{2i} \sim \chi^2(4)$, and $u_i = w_i(1 + 0.8w_{i-1})$ with i.i.d. noise $w_i \sim U(0, 10)$. The following estimators are considered: the ordinary least squares estimator (LSE), the instrumental variable estimator (IVE) and the extended linear programming estimator (LPE). For the IVE, the variables v_{1i} and v_{2i} are used as instruments. Finally, for both the LSE and IVE an intercept is included in the regression equation. Different sample sizes (n) and levels of endogeneity (γ) are considered.

n	$\gamma = 0$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.047	1.533	−0.037	1.055	0.047	1.533	−0.037	1.055	0.362	0.469	0.038	0.240
100	0.037	0.746	−0.019	0.532	0.037	0.746	−0.019	0.532	0.198	0.145	0.011	0.064
200	0.012	0.330	0.023	0.236	0.012	0.330	0.023	0.236	0.093	0.031	0.012	0.016
500	0.010	0.129	−0.009	0.093	0.010	0.129	−0.009	0.093	0.036	0.005	0.002	0.002
1000	0.007	0.066	0.003	0.049	0.007	0.066	0.003	0.049	0.017	0.001	0.003	0.000
2000	−0.001	0.032	0.003	0.025	−0.001	0.032	0.003	0.025	0.009	0.000	0.000	0.000
n	$\gamma = 0.25$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	1.992	4.146	1.493	2.389	−0.342	8.125	−0.291	3.731	0.340	0.427	0.023	0.241
100	2.018	4.172	1.482	2.287	−0.164	1.195	−0.139	0.758	0.187	0.113	0.007	0.055
200	2.009	4.083	1.494	2.276	−0.099	0.430	−0.088	0.320	0.096	0.029	0.005	0.013
500	2.002	4.028	1.499	2.265	−0.032	0.142	−0.015	0.100	0.037	0.004	0.002	0.002
1000	2.005	4.030	1.499	2.255	−0.022	0.062	−0.008	0.048	0.018	0.001	0.001	0.001
2000	1.999	4.003	1.502	2.263	−0.008	0.033	−0.016	0.025	0.009	0.000	0.001	0.000
n	$\gamma = 0.5$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	1.095	1.250	0.837	0.748	−0.004	24.035	0.186	19.156	0.262	0.243	0.057	0.149
100	1.094	1.220	0.836	0.719	−0.312	17.710	−0.289	14.171	0.158	0.098	0.023	0.056
200	1.102	1.228	0.829	0.701	−0.270	1.757	−0.212	1.211	0.085	0.025	0.010	0.012
500	1.099	1.215	0.832	0.698	−0.074	0.224	−0.076	0.150	0.041	0.006	−0.001	0.003
1000	1.104	1.222	0.827	0.686	−0.029	0.080	−0.033	0.055	0.017	0.001	0.002	0.001
2000	1.104	1.220	0.827	0.686	−0.021	0.033	−0.009	0.024	0.009	0.000	0.001	0.000
n	$\gamma = 1$											
	LSE				IVE				LPE			
	β_1		β_2		β_1		β_2		β_1		β_2	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
50	0.560	0.326	0.430	0.197	0.322	13.875	0.430	10.342	0.245	0.216	0.015	0.129
100	0.568	0.329	0.423	0.185	0.269	10.957	−0.030	14.681	0.128	0.058	0.025	0.034
200	0.566	0.324	0.424	0.183	0.091	13.125	0.151	6.518	0.082	0.022	0.006	0.012
500	0.566	0.322	0.424	0.181	−0.161	3.937	−0.101	2.023	0.034	0.004	0.004	0.002
1000	0.566	0.320	0.425	0.181	−0.133	0.423	−0.108	0.290	0.019	0.001	0.000	0.001
2000	0.566	0.321	0.424	0.180	−0.042	0.057	−0.033	0.039	0.010	0.000	0.000	0.000

endogenous. The exact distribution may be used for statistical inference or to bias-correct the LPE. In addition, we have shown that the LPE is strongly consistent under fairly general conditions on the related distributions. In particular, the LPE is robust to various heavy-tailed specifications and its functional form indicates that it can be insensitive to outliers in y_i or x_i. We have also identified a number of cases where the LPE is superconsistent. In contrast to existing results for the LPE, in a time series setting, our results in a cross-sectional setting are valid also in the case when the slope parameter is negative.

We provided conditions under which the LPE is consistent in the presence of serially correlated, heteroskedastic errors and described how the LPE can be extended to the case with multiple regressors. Our simulation results indicated that the LPE and extended LPE can have very reasonable finite-sample properties compared to the LSE and IVE, also in the presence of heteroskedastic or serially correlated errors. Clearly, one advantage of the LPE is that, in contrast to the IVE, it does not require an instrument.

Acknowledgments

We are grateful to conference participants at the Singapore Econometrics Study Group Meeting (2009, Singapore) for their comments and suggestions. We also wish to thank Michael McAleer and two referees for very helpful comments. All remaining errors are our own. The first author gratefully acknowledges partial research support from the Jan Wallander and Tom Hedelius Research Foundation (grant P 2006-0166:1) and the Sim Kee Boon Institute for Financial Economics at Singapore Management University. The second author wish to thank CNPq/Brazil for partial financial support.

Appendix

Proof of Corollary 1. Note that $\hat{\beta}_n \overset{\text{a.s.}}{\to} \beta$ iff $R_n \overset{\text{a.s.}}{\to} 0$. By condition (ii), 0 is not a limit point of S, hence, there exists a $\delta > 0$ such that the two sets $(-\delta, \delta)$ and S are disjoint. Let $c = \delta/2$ and let

$\epsilon > 0$ be arbitrary. Then, in view of Proposition 1,

$$P(|R_n| > \epsilon) = \prod_{i=1}^{n}[1 - F_u(x_i\epsilon)] \overset{\text{(i)–(ii)}}{\leq} [1 - F_u(c\epsilon)]^n.$$

By (iv), $F_u(u) > 0$ for every $u > 0$. Hence, $R_n \xrightarrow{p} 0$ as $n \to \infty$. Finally, since $R_1, \ldots, R_n$ forms a stochastically decreasing sequence, it follows that $R_n \xrightarrow{\text{a.s.}} 0$. □

Proof of Proposition 2.

$$\begin{aligned} P(\hat{\beta}_n - \beta \leq z) &\overset{\text{(ii)}}{=} 1 - P(u_1/x_1 > z)^n = 1 - [1 - P(u_1/x_1 \leq z)]^n \\ &= 1 - [1 - F_z(z)]^n, \end{aligned}$$

where $F_z(z)$ is the cdf of $z = u_1/x_1$. Let $f_{u_1,x_1}(u, x)$ denote the joint pdf of u_1 and $x_1 = v_1 + \gamma u_1$, and $f_{u_1}(u)$ the marginal pdf of u_1. Denote by $f_{x_1|u_1=u}(x)$ the conditional pdf of x_1 given that $u_1 = u$. Then, for $u > 0$

$$f_{u_1,x_1}(u, x) = f_{x_1|u_1=u}(x)f_{u_1}(u) = f_{v_1}(x - \gamma u)f_{u_1}(u), \tag{7}$$

where $f_{v_1}(v)$ is the pdf of v_1. By Theorem 3.1 in Curtiss (1941),

$$\begin{aligned} f_z(z) &= F_z'(z) \\ &= \int_{-\infty}^{\infty} |x| f_{u_1,x_1}(zx, x)\mathrm{d}x \overset{\text{(ii)}}{=} \int_0^{\infty} x f_{u_1,x_1}(zx, x)\mathrm{d}x. \end{aligned} \tag{8}$$

Now consider the case when $\gamma > 0$. By (7) and (8),

$$f_z(z) = \int_0^{\infty} x f_{v_1}(x - \gamma zx) f_{u_1}(zx)\mathrm{d}x,$$

for $0 < z < 1/\gamma$ and zero otherwise. Hence,

$$F_z(z) = \mathbf{1}_{\{0<z<1/\gamma\}} \int_0^z \int_0^{\infty} x f_{v_1}(x - \gamma tx) f_{u_1}(tx)\mathrm{d}x\mathrm{d}t + \mathbf{1}_{\{z \geq 1/\gamma\}}.$$

The proof when $\gamma = 0$ is analogous. □

Proof of Proposition 3. Let $\epsilon > 0$ be arbitrary. Then,

$$\begin{aligned} P(|R_n| > \epsilon) &\overset{\text{(i)}}{=} P[u_1 > \epsilon(\gamma u_1 + v_1), \ldots, u_n > \epsilon(\gamma u_n + v_n)] \\ &\overset{\text{(ii)}}{\leq} P(u_1 > \epsilon v_1, \ldots, u_n > \epsilon v_n) \overset{\text{(ii)}}{=} P(u_1 > \epsilon v_1)^n. \end{aligned}$$

A simple proof by contradiction, based on a geometric argument, shows that $P(u_1 > \epsilon v_1) < 1$. Hence, $R_n \xrightarrow{p} 0$ as $n \to \infty$ and once more the strong convergence of $\hat{\beta}_n = \beta + R_n$ follows. □

References

Curtiss, J.H., 1941. On the distribution of the quotient of two chance variables. The Annals of Mathematical Statistics 12 (4), 409–421.

Datta, S., Mathew, G., McCormick, W.P., 1998. Nonlinear autoregression with positive innovations. Australian & New Zealand Journal of Statistics 40 (2), 229–239.

Datta, S., McCormick, W.P., 1995. Bootstrap inference for a first-order autoregression with positive innovations. Journal of the American Statistical Association 90, 1289–1300.

Davis, R.A., McCormick, W.P., 1989. Estimation for first-order autoregressive processes with positive or bounded innovations. Stochastic Processes and their Applications 31, 237–250.

Feigin, P.D., Kratz, M.F., Resnick, S.I., 1996. Parameter estimation for moving averages with positive innovations. The Annals of Applied Probability 12 (4), 1157–1190.

Feigin, P.D., Resnick, S.I., 1994. Limit distributions for linear programming time series estimators. Stochastic Processes and their Applications 51, 135–165.

Nielsen, B., Shephard, N., 2003. Likelihood analysis of a first-order autoregressive model with exponential innovations. Journal of Time Series Analysis 24 (3), 337–344.

Preve, D., 2011. Linear programming-based estimators in nonnegative autoregression. Manuscript.