RANDOM EFFECTS ESTIMATORS
WITH MANY INSTRUMENTAL VARIABLES

By Gary Chamberlain and Guido Imbens[1]

In this paper we propose a new estimator for a model with one endogenous regressor and many instrumental variables. Our motivation comes from the recent literature on the poor properties of standard instrumental variables estimators when the instrumental variables are weakly correlated with the endogenous regressor. Our proposed estimator puts a random coefficients structure on the relation between the endogenous regressor and the instruments. The variance of the random coefficients is modelled as an unknown parameter. In addition to proposing a new estimator, our analysis yields new insights into the properties of the standard two-stage least squares (TSLS) and limited-information maximum likelihood (LIML) estimators in the case with many weak instruments. We show that in some interesting cases, TSLS and LIML can be approximated by maximizing the random effects likelihood subject to particular constraints. We show that statistics based on comparisons of the unconstrained estimates of these parameters to the implicit TSLS and LIML restrictions can be used to identify settings when standard large sample approximations to the distributions of TSLS and LIML are likely to perform poorly. We also show that with many weak instruments, LIML confidence intervals are likely to have under-coverage, even though its finite sample distribution is approximately centered at the true value of the parameter. In an application with real data and simulations around this data set, the proposed estimator performs markedly better than TSLS and LIML, both in terms of coverage rate and in terms of risk.

KEYWORDS: Instrumental variables, random effects, quasi-likelihood, two-stage least squares, risk function, returns to schooling.

1. INTRODUCTION

IN THIS PAPER we propose a new estimator for a model with a single endogenous regressor and many instrumental variables. Our motivation comes from the recent literature on the poor properties of standard instrumental variables estimators when the instrumental variables are weakly correlated with the endogenous regressor. The starting point of this literature is a paper by Angrist and Krueger (1991), henceforth AK, who report two-stage least squares (TSLS) estimates from a data set characterized by a combination of a large number of observations and a large number of weak instrumental variables. Bound, Jaeger, and Baker (1995), henceforth BJB, show that even with such large samples TSLS can lead to extremely misleading confidence intervals. Subsequently researchers have focused on improving inference in these settings. Bekker (1994) and Staiger and Stock (1997) develop alternative asymptotic distributions for TSLS and the limited-information maximum likelihood (LIML) estimator. Bekker's approach is based on increasing the number of instrumental variables along with the sample size, whereas Staiger and Stock's approach is based on a vanishing correlation between the instrumental variables and the endogenous variables. Kleibergen (2002)

and Moreira (2003) develop new tests for the structural coefficients with improved finite sample properties.

Our proposed estimator puts a random coefficients structure on the relation between the endogenous regressor and the instruments. The variance of the random coefficients is modelled as an unknown parameter. We refer to this estimator as the random-effects quasi-maximum likelihood estimator (REQML).[2] Using standard asymptotics (with the number of observations going to infinity, keeping the number of instrumental variables fixed), the estimator has the same limiting normal distribution as LIML and TSLS. However, its finite sample properties will differ especially in settings with many weak instruments. Because REQML is a maximum likelihood estimator in a model with few parameters, inference based on it is likely to be accurate if the model is correctly specified. Even when the instruments are (almost) independent of the endogenous regressor, and the model is (close to) not identified, REQML confidence intervals are unlikely to be misleading. Although in that case asymptotic normality of the structural coefficient need not be a good approximation to the finite sample distribution, the likelihood function will be flat around the true value of the parameter, and thus the confidence intervals will be wide.[3]

In addition to proposing a new estimator, our analysis yields new insights into the properties of TSLS and LIML in the case with many weak instruments. We show that in some interesting cases TSLS and LIML can be approximated by maximizing the random effects likelihood subject to particular constraints. When the variance of the first stage coefficients is fixed at a large number, REQML and its standard error are approximately equal to TSLS and its standard error. When a particular transformation of this variance (involving the structural coefficients and the reduced-form error covariance matrix) is fixed at a large value, we recover LIML and its standard error. Since we argued that inference based on REQML is likely to be accurate in these settings, this approximation of TSLS and LIML by restricted REQML implies that any inaccuracies in asymptotic approximations of TSLS and LIML must be due to violations of the restrictions. Thus statistics based on comparisons of the unconstrained estimates of these parameters to the implicit TSLS and LIML restrictions can be used to identify settings when standard large sample approximations to the distributions of TSLS and LIML are likely to perform poorly. For TSLS this statistic is closely related to the concentration parameter suggested by BJB and Staiger and Stock (1997) as a valuable diagnostic. For LIML the diagnostic is new. Our results also imply that with many weak instruments LIML confidence intervals are likely to have under-coverage. For TSLS we provide a simple calculation for the approximate bias of its standard error.

We apply the proposed REQML estimator to data previously analyzed by AK in their study of the effect of years of schooling on earnings using quarter of birth, on its

---

[2]The estimator can also be interpreted as the posterior mode in a hierarchical Bayes set up with a hierarchical structure for the coefficients on the instrumental variables and a flat prior distribution on the hierarchical parameter (the variance of the coefficients on the instrumental variables). In that perspective, the problem with TSLS is that of a flat prior distribution in a high-dimensional parameter space being very informative.

[3]Note that this argument does not necessarily apply to LIML in our setting, because although LIML is maximum likelihood, it is so in a setting with many parameters.

own and interacted with year and state of birth, as an instrumental variable. Our analysis uses 162,515 observations and 505 instrumental variables. The LIML and REQML estimates of the returns to schooling are similar for the AK data, and somewhat larger than the TSLS estimate. The REQML confidence interval is somewhat wider than the LIML interval; both of these are considerably wider than the TSLS interval. Following BJB, we also apply the estimator to a modified version of the AK data with randomly generated instrumental variables. In that case the REQML confidence interval is extremely wide, correctly indicating the lack of information in the data. The LIML interval, although wider than the TSLS interval, is still misleadingly informative. Consistent with our theoretical discussion, the restricted REQML estimates and their standard errors are very close to TSLS and LIML in both settings. We also evaluate risk functions under data generated according to the fixed-effects model using parameter values similar to those estimated on the AK data. The risk for the REQML estimator is generally less than the risk of the LIML estimator. The risk improvement is substantial when the coefficients on the doubtful instrumental variables are all zero. The risk of the TSLS estimator is very large for some parameter values. Coverage rates of the REQML intervals are very close to the nominal value of .95 in all cases. The coverage rates of the TSLS intervals are extremely poor, as low as .00. The LIML intervals have better coverage rates than the TSLS intervals, but the coverage rate can still be as low as .40.

2. THEORY

2.1. *The Model*

In this section we set up the basic model and define TSLS and LIML. We shall work with the following reduced form model:

$$
(1) \quad
\begin{aligned}
Y_1 &= X_1\xi_1 + X_2\pi_1 + U_1, \\
Y_2 &= X_1\xi_2 + X_2\pi_2 + U_2, \\
\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \Big| X &\sim \mathcal{N}(0, \Sigma \otimes I_n),
\end{aligned}
$$

where Y_1 ($n \times 1$) is the endogenous regressor, Y_2 ($n \times 1$) is the primary outcome, X_1 ($n \times j$) is a set of exogenous regressors, X_2 ($n \times k$) is a set of instruments, and $X = (X_1 \quad X_2)$ is $n \times (j + k)$ with full column rank. All the analysis will be done conditional on X. The regression coefficients on the instrumental variables X_2 satisfy a proportionality restriction:

$$
\pi_1 = \beta, \quad \pi_2 = \beta\gamma,
$$

where β is $k \times 1$ and γ is scalar, so that $Y_2 = X_1\delta + Y_1\gamma + V$, with $\delta = \xi_2 - \gamma\xi_1$ and $V = U_2 - \gamma U_1$, is the structural equation of interest.

Consider the least-squares regression statistics:

$$
\hat{Y} = \begin{pmatrix} \hat{\xi}_1 & \hat{\xi}_2 \\ \hat{\pi}_1 & \hat{\pi}_2 \end{pmatrix} = (X'X)^{-1}X'Y, \quad S = (Y - X\hat{Y})'(Y - X\hat{Y}),
$$

where $Y = (Y_1 \quad Y_2)$. Rather than use the full likelihood function, we shall base the likelihood on the joint distribution of $\hat{\Pi} = (\hat{\pi}_1 \quad \hat{\pi}_2)$, which is $k \times 2$, and S. This simplifies the analysis, since the likelihood based on $\hat{\Pi}$ and S does not depend upon (ξ_1, ξ_2). $\hat{\pi}$ and S are independent with

$$\hat{\pi} = \begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix}, \quad \hat{\pi} \sim \mathcal{N}\big(\phi \otimes \beta, \Sigma \otimes (\tilde{X}_2'\tilde{X}_2)^{-1}\big), \quad S \sim \text{Wishart}(n - j - k, \Sigma),$$

where $\phi' = (1 \ \gamma)$ and $\tilde{X}_2 = (I_n - X_1(X_1'X_1)^{-1}X_1')X_2$.

We divide the set of instrumental variables into two subsets, one (small) set of instrumental variables that are a priori viewed as more important, and one (large) set of instrumental variables that are viewed as potentially weak. In applications such as ours this distinction may arise when there is a small set of basic instrumental variables, and a potentially large number of additional instrumental variables is generated by interacting this set of basic instrumental variables with functions of the exogenous covariates. However, this division merely adds generality, and it is not essential for our discussion that the first set is nonempty. We shall only impose the random coefficients structure on the second set of instrumental variables. Thus partition X_2 and β:

$$X_2\beta = X_{21}\beta_1 + X_{22}\beta_2,$$

where X_{21} is $n \times k_1$, X_{22} is $n \times p$, β_1 is $k_1 \times 1$, and β_2 is $p \times 1$. We are interested in the case where p is large relative to k_1. Before imposing a random coefficients structure, it is convenient to work with a one-to-one function of $\hat{\pi}$, denoted by $\hat{\pi}^*$, so that the elements of $\hat{\pi}^*$ corresponding to different instruments are independent:

$$\hat{\pi}^* = (I_2 \otimes F)\hat{\pi} \quad \text{where} \quad F = \begin{pmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{pmatrix}$$

is a nonsingular, $k \times k$ matrix with $F'F = \tilde{X}_2'\tilde{X}_2$. F is partitioned conformably with β. $\hat{\pi}^*$ and S are independent with

$$\hat{\pi}^* \sim \mathcal{N}(\phi \otimes \beta^*, \Sigma \otimes I_k), \quad \beta^* = \begin{pmatrix} \beta_1^* \\ \beta_2^* \end{pmatrix} = F\beta = \begin{pmatrix} F_{11}\beta_1 + F_{12}\beta_2 \\ F_{22}\beta_2 \end{pmatrix}.$$

Let $Z = (\hat{\pi}^*, S)$. We shall treat Z as the observation with

$$\mathcal{Z} = \big\{(z_1, z_2) : z_1 \in \mathcal{R}^{2k}, \ z_2 \in \mathcal{S}\big\}$$

as the sample space, where $\mathcal{S}$ is the set of 2×2 positive-definite matrices. The parameter space is

$$\Theta = \big\{(\gamma, \beta_1^*, \beta_2^*, \Sigma) : \gamma \in \mathcal{R}, \ \beta_1^* \in \mathcal{R}^{k_1}, \ \beta_2^* \in \mathcal{R}^p, \ \Sigma \in \mathcal{S}\big\}.$$

So our model is $\{P_\theta : \theta \in \Theta\}$ with $Z \sim P_\theta$ for some $\theta = (\gamma, \beta_1^*, \beta_2^*, \Sigma) \in \Theta$ where

$$P_\theta = \mathcal{N}(\phi \otimes \beta^*, \Sigma \otimes I_k) \times \text{Wishart}(n - j - k, \Sigma).$$

The (fixed effect) log likelihood function is

$$(2) \qquad l_{fe}(\gamma, \beta^*, \Sigma) = -\frac{1}{2}(n-j)\ln[\det(\Sigma)]$$

$$-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}\left(S + (\hat{\Pi}^* - \beta^*\phi')'(\hat{\Pi}^* - \beta^*\phi')\right)\right],$$

where $\hat{\Pi}^* = F\hat{\Pi}$, suppressing additive terms that do not depend upon $(\gamma, \beta^*, \Sigma)$. Maximizing this over θ gives LIML.[4] We shall also need the log likelihood function with β^* concentrated out:

$$(3) \qquad l^*_{fe}(\gamma, \Sigma) = -\frac{1}{2}\left[(n-j)\ln[\det(\Sigma)] + \operatorname{tr}\left[\Sigma^{-1}(S + \hat{\Pi}^{*'}\hat{\Pi}^*)\right] - \psi'\hat{\Pi}^{*'}\hat{\Pi}^*\psi/\psi'\Sigma\psi\right],$$

where $\psi = \Sigma^{-1}\phi$.[5]

The TSLS estimator of γ and its standard error are

$$(4) \qquad \hat{\gamma}_{TSLS} = \frac{\hat{\pi}_1^{*'}\hat{\pi}_2^*}{\hat{\pi}_1^{*'}\hat{\pi}_1^*}, \quad \kappa = \begin{pmatrix} -\hat{\gamma}_{TSLS} \\ 1 \end{pmatrix},$$

$$se(\hat{\gamma}_{TSLS}) = \left[\kappa'(S + \hat{\Pi}^{*'}\hat{\Pi}^*)\kappa/(n\hat{\pi}_1^{*'}\hat{\pi}_1^*)\right]^{1/2}.$$

2.2. Random Effects Quasi Maximum Likelihood

In this section we define the REQML estimator. We shall use the following family of distributions for the random coefficient β_2^*: $\{\mathcal{N}(0, \sigma_\beta^2 I_p) : \sigma_\beta \geq 0\}$. See an earlier version of this paper, Chamberlain and Imbens (2001), for a discussion of the invariance properties of this choice of distribution. Given a value for σ_β, the distribution for β_2^* implies a joint distribution for (Z, β_2^*), which implies a marginal distribution for Z, which is the basis for the random-effects quasi-likelihood function. Note that in large samples the random coefficients specification for β_2^* will not affect the maximum likelihood estimates for the remaining parameters (γ, Σ), so that the asymptotic distribution (for a fixed number of instruments) of the REQML estimator is identical to that of LIML and TSLS. Only in finite samples, and in particular with many instruments, will there be a difference between the three estimators.

Conditional on β_2^*, S is independent of $\hat{\Pi}^*$ and the distribution of S does not depend upon β_2^*. So $\hat{\Pi}^*$ and S are independent in the marginal distribution of Z, with

$$\hat{\pi}^* \sim \mathcal{N}\left(\phi \otimes \begin{pmatrix} \beta_1^* \\ 0 \end{pmatrix}, A\right), \quad S \sim \text{Wishart}(n-j-k, \Sigma),$$

where

$$A = \sigma_\beta^2 \phi\phi' \otimes \begin{pmatrix} 0 & 0 \\ 0 & I_p \end{pmatrix} + \Sigma \otimes I_k.$$

---

[4] Maximizing over ξ_1 and ξ_2 in the likelihood function for the full model gives a concentrated likelihood. The log likelihood in (2) differs from the log of this concentrated likelihood function only by replacing n by $(n-j)$.

[5] The derivation for this concentrated likelihood function is given in the working paper version, Chamberlain and Imbens (2001).

Define $\psi = \Sigma^{-1}\phi$, $\tau = \phi'\Sigma^{-1}\phi$, $\lambda = (\tau\sigma_\beta^2)^{-1}$. It is convenient to index the parameter space using $(\gamma, \beta_1^*, \lambda, \Sigma)$ instead of $(\gamma, \beta_1^*, \sigma_\beta, \Sigma)$. Then our log-likelihood function is

$$(5) \qquad l_{re}(\gamma, \beta_1^*, \lambda, \Sigma) = -\frac{1}{2}\Bigg[(n-j)\ln[\det(\Sigma)] - p\ln[\lambda/(\lambda+1)] + \operatorname{tr}(\Sigma^{-1}S)$$

$$+ \left[\hat{\pi}^* - \left(\phi \otimes \begin{pmatrix} I_{k_1} \\ 0 \end{pmatrix}\right)\beta_1^*\right]' A^{-1}\left[\hat{\pi}^* - \left(\phi \otimes \begin{pmatrix} I_{k_1} \\ 0 \end{pmatrix}\right)\beta_1^*\right]\Bigg].$$

Concentrating out β_1^*, the log-likelihood function is[6]

$$(6) \qquad l_{re}^*(\gamma, \lambda, \Sigma) = -\frac{1}{2}\Bigg[(n-j)\ln[\det(\Sigma)] - p\ln[\lambda/(\lambda+1)]$$

$$+ \operatorname{tr}\left[\Sigma^{-1}(S + \hat{\Pi}^{*'}\hat{\Pi}^*)\right] - \psi'\hat{\Pi}^{*'}P\hat{\Pi}^*\psi/\psi'\Sigma\psi\Bigg],$$

where

$$P = \begin{pmatrix} I_{k_1} & 0 \\ 0 & (\lambda+1)^{-1}I_p \end{pmatrix}.$$

The ML estimate of γ given (λ, Σ) is obtained from the corresponding eigenvector:

$$(7) \qquad b_1 = \arg\max_{b \neq 0} b'\hat{\Pi}^{*'}P\hat{\Pi}^*b/b'\Sigma b, \quad \hat{\gamma}(\lambda, \Sigma) = e_2'\Sigma b_1/e_1'\Sigma b_1,$$

where $e_1' = (1 \quad 0)$ and $e_2' = (0 \quad 1)$.

2.3. Comparison with LIML

Here we provide two comparisons with LIML that highlight the role of λ. In this discussion we assume that the reduced form covariance matrix Σ is known. Even though it is rare in practice that Σ is known, the precision of the reduced form covariance matrix is much higher than that of the structural coefficients in settings like ours with large samples and many weak instruments, and the following approximations will be seen to provide good guidance to finite sample biases in our application and simulations. First, consider the constrained REQML given λ (and Σ). As $\lambda \to 0$, $P \to I_k$, and the constrained REQML equals the LIML estimate of γ given Σ based on maximizing (3).

Next, consider the case in which the random-effects specification applies to all of β^*, so that $k_1 = 0$. Then $P = (\lambda+1)^{-1}I_k$. Comparing the fixed and random effects concentrated likelihood functions in (3) and (6) respectively, we have $l_{re}^*(\gamma, \lambda, \Sigma) = (\lambda+1)^{-1}l_{fe}^*(\gamma, \Sigma) + h(\lambda, \Sigma)$, where the additive term $h(\lambda, \Sigma)$ does not depend upon γ. So in this case, the REQML estimate of γ given (λ, Σ) is identical to the LIML estimate of γ given Σ. In addition, the second derivative with respect to γ of the random-effects log-likelihood function equals $(\lambda+1)^{-1}$ times the second derivative with respect to γ of the fixed-effects concentrated log-likelihood function. Suppose that the standard error reported for γ is the inverse of the square root of minus the second derivative of the

---

[6]The derivation for this concentrated likelihood function is given in the working paper version, Chamberlain and Imbens (2001).

concentrated log-likelihood function, evaluated at the ML estimate of γ. Then if λ and Σ are given, the REQML standard error is $(\lambda + 1)^{1/2}$ times the LIML standard error.

2.4. *Comparison with TSLS*

Here we discuss a link between REQML and TSLS. Consider the REQML estimate given (σ_β, Σ):

$$\hat{\gamma}(\sigma_\beta, \Sigma) = \arg\max_\gamma l_{re}^*(\gamma, (\tau\sigma_\beta^2)^{-1}, \Sigma).$$

As $\sigma_\beta \to \infty$, we have $P \to I_k$. So

$$(8) \qquad \lim_{\sigma_\beta \to \infty} \left[l_{re}^*(\gamma, (\tau\sigma_\beta^2)^{-1}, \Sigma) + p\ln(\sigma_\beta) \right]$$

$$= -\frac{1}{2}p\ln(\tau) + \frac{1}{2}\psi'\hat{\Pi}^{*'}\hat{\Pi}^*\psi/\psi'\Sigma\psi + c(\Sigma),$$

where the additive term $c(\Sigma)$ does not depend upon γ. The limiting objective function in (8) differs from the fixed-effects concentrated log-likelihood function in (3) by the term $-p\ln(\tau)/2$ (ignoring additive terms that do not depend upon γ). Consider the term $-p\ln(\tau)/2 = -p\ln(\phi'\Sigma^{-1}\phi)/2$. It is maximized as a function of γ at $\gamma = \sigma_{12}/\sigma_{11}$, which is the population least-squares value corresponding to the coefficient on Y_1 in the least-squares regression of Y_2 on $(Y_1 \quad X)$. The second derivative at this value is $-p\sigma_{11}^2/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$.

Now consider the case where all elements of β^* are equal to zero. Suppose furthermore that p is large. With $\beta^* = 0$ the probability limit of $\hat{\Pi}^{*'}\hat{\Pi}^*/p$ as $p \to \infty$ is equal to Σ, and the second term in (8) will therefore become flat as a function of γ. Thus the maximum likelihood estimator for γ will tend to the argmax of the first term, σ_{12}/σ_{11}. Moreover, its large sample variance will be approximated by minus the inverse of the second derivative, $(\sigma_{11}\sigma_{22} - \sigma_{12}^2)/(p\sigma_{11}^2)$.

In the same setting, with $\beta^* = 0$ and p large, consider TSLS. As p gets large, $p^{-1}\hat{\pi}_1^{*'}\hat{\pi}_2^*/p^{-1}\hat{\pi}_1^{*'}\hat{\pi}_1^*$ will converge to σ_{12}/σ_{11}. Now consider the estimator for the variance of the TSLS estimator: $\kappa'\Sigma\kappa/\hat{\pi}_1^{*'}\hat{\pi}_1^*$, which will be approximately equal to $(\sigma_{22} - \sigma_{12}^2/\sigma_{11})/(p\sigma_{11})$. Hence the TSLS estimator and variance will be close to the constrained REQML and its large sample variance with σ_β^2 fixed at a large value. Note that σ_β^2 is closely related to the concentration parameter $\beta'\tilde{X}_2'\tilde{X}_2\beta/\sigma_{11} = \beta^{*'}\beta^*/\sigma_{11}$, which in our random effects setup if $k_1 = 0$ has expectation equal to $p\sigma_\beta^2/\sigma_{11}$. BJB and Staiger and Stock (1997) discuss the value of the concentration parameter as a diagnostic for poor small sample properties of TSLS.

3. APPLICATION

3.1. *The AK Data*

Our data are a subset of the census data used by AK, containing males born in either the first or fourth quarters between 1930 and 1939.[7] The sample size is $n = $

---

[7]For convenience we only use the two quarters with the biggest contrast in the effect of compulsory schooling laws on age at entry. This is not essential for any of the analysis that follows.

TABLE I

RETURNS TO SCHOOLING USING QUARTER OF BIRTH AS INSTRUMENTAL VARIABLES

| | | | REQML | | |
| | | | $\sigma_\beta = 10^3$ | $\lambda = 10^{-6}$ | |
	TSLS	LIML	(Pseudo TSLS)	(Pseudo LIML)	unrestricted
Real QOB	.073	.094	.073	.094	.096
	(.057, .089)	(.061, .129)	(.057, .088)	(.061, .129)	(.056, .139)
Random QOB	.059	−.330	.059	−.330	.220
	(.042, .076)	(−.999, −.149)	(.041, .076)	(−.999, −.149)	(−1000, 1000)

Note: Point estimates with (nominal) .95 confidence intervals in parentheses. LIML = limited information maximum likelihood; TSLS = two-stage least squares; REQML = random-effects quasi-maximum-likelihood. The three versions of the REQML are (i) Pseudo TSLS with $\sigma_\beta = 10^3$ and $\Sigma = \hat{\Sigma}$, (ii) Pseudo LIML with $\lambda = 10^{-6}$ and $\Sigma = \hat{\Sigma}$, and (iii) unrestricted. The instrumental variables are quarter of birth (QOB) ($k_1 = 1$) and interactions of QOB with 504 indicator variables for state/year cells ($p = 504$). The data are for males born in either the first or fourth quarters between 1930 and 1939; $n = 162,515$.

162,515. The outcome variable Y_2 is the log of weekly earnings in 1979. The treatment variable Y_1 is years of school completed. The predictor variables X_1 consist of indicator variables based on the individual's state of birth and year of birth. There are 509 state/year cells (fifty states plus the District of Columbia and ten years minus one empty cell); $X_{1st} = 1$ if individual s is from state/year cell t, and equals 0 otherwise ($s = 1, \ldots, n; t = 1, \ldots, 509$). X_{21} consists of a single instrumental variable: X_{21s} equals 1 if individual s was born in the fourth quarter, and equals 0 otherwise, motivated by AK by compulsory schooling laws. The doubtful instrumental variables X_{22} consist of interactions between quarter of birth and the state/year indicators; $X_{22st} = X_{21s} \cdot X_{1st}$ ($t = 1, \ldots, 509$). Note that the columns of X_{22} sum to X_{21}. In addition, it turns out that four of the state/year cells have no variation in quarter of birth, so that the rank of X is 1014. We drop 5 columns of X_{22} so that the remaining columns give X of full rank.[8] Our specification generalizes the models used by AK, who only interacted quarter of birth with state *or* year dummies. This modification increases the effective number of instrumental variables in the fixed-effects model from 180 to 505.

Table I contains the results for our subset of the AK data. The table also contains the results for the same data, but with the actual quarter of birth replaced by randomly generated indicators, with probability .5 on the first quarter and .5 on the fourth quarter. The (nominal) .95 confidence intervals for the LIML and REQML estimators are obtained from the concentrated (profile) likelihood. We find lower and upper values for γ such that the concentrated log-likelihood function differs from its maximum value by $G^{-1}(.95)/2$, where G is the distribution function for a chi-square distribution with one degree of freedom.[9] The confidence interval for TSLS is formed from the point estimate plus or minus 1.96 times the standard error given in (4). With the real data, the LIML and REQML estimates of γ are somewhat larger than the TSLS estimate.

---

[8]In our application $k_1 = 1$ and $p = 504$.

[9]Using confidence intervals based on normal approximations with asymptotic standard errors based on the second derivative of the log likelihood function leads to very similar results. The confidence interval for the AK data for the LIML estimator is in that case (.061, .128) instead of the interval (.061, .129) based on the concentrated log-likelihood function.

For comparison with TSLS, we can form standard errors by dividing the width of the confidence interval by $2 \cdot 1.96$. This gives .022 for random effects and .017 for fixed effects; these standard errors are considerably larger than the .008 for TSLS.

Consider the restricted REQML estimator when σ_β and Σ are given. We present REQML estimates given $(\sigma_\beta = 10^3, \Sigma = \hat{\Sigma})$. ($\hat{\Sigma}$ is based on the residuals from least-squares regression of Y on X and is given below in (9).) Fixing σ_β at a large value gives an estimate and confidence interval that closely match the estimate and confidence interval for TSLS. Next, consider the REQML estimator when λ and Σ are given. We present REQML estimates given $(\lambda = 10^{-6}, \Sigma = \hat{\Sigma})$. Fixing λ at a small value gives an estimate and confidence interval that closely match the estimate and confidence interval for LIML.

Finally consider the results from using the randomly generated quarter of birth numbers. The TSLS estimate is .059 with a standard error of .009. The narrow confidence interval for TSLS, even though the instrumental variables are independent of the endogenous variables, is the result found by BJB. The LIML confidence interval is wider but still misleading, since it excludes positive values for γ. The random-effects interval is essentially the whole line.[10] The restricted REQML estimators again match the point estimate and confidence interval for TSLS and LIML very closely.

3.2. Risk Calculations

The results for the AK data suggest that the REQML estimates are more credible than those for the other two estimators in this setting with large samples and many instrumental variables. To further investigate this, we shall evaluate risk functions, under the fixed-effects model, for these estimators. We fix X at the values for the real data used in Table I. Given a value for $\theta = (\gamma, \beta^*, \Sigma)$, we generate J independent draws for the sufficient statistics: $Z^{(i)} = (\hat{\pi}^{*(i)}, S^{(i)})$, $i = 1, \ldots, J$. The Monte Carlo approximation is $\hat{R}(\theta, d) = \sum_{i=1}^{J} L(\theta, d(Z^{(i)}))/J$. We shall use bounded loss functions, based on truncated absolute error and squared error:

$$L_1(\theta, a) = \min\{|\gamma - a|, .2\}, \quad L_2(\theta, a) = [\min\{|\gamma - a|, .2\}]^2.$$

We set $\Sigma = \hat{\Sigma} = S/(n - j - k)$, where S is the matrix of sums of squares and cross products of the residuals from the least-squares regression of Y on X:

$$(9) \qquad \hat{\Sigma} = \begin{pmatrix} 10.133 & .678 \\ .678 & .448 \end{pmatrix}.$$

We consider several values for $(\gamma, \beta_1^*, \|\beta_2^*\|)$.[11] They are centered at REQML estimates using the AK data as in Table I. This gives $\gamma = .096$, $\beta_1^* = 30.4$, and $\|\beta_2^*\|/\sqrt{p} = .831$. We also double these values and set them to 0. In order to clearly display the bias in the TSLS estimate, when we double γ we also change its sign.

[10] The numerical procedure used to solve for the lower and upper values of γ searches over the interval $[-1000, 1000]$; but the concentrated log-likelihood was greater than the maximum value minus $G^{-1}(.95)/2$ throughout this interval.

[11] In the working paper version, Chamberlain and Imbens (2001) show that the risk only depends on β_2^* through $\|\beta_2^*\| = \sqrt{\beta_2^{*'}\beta_2^*}$.

TABLE II
RISK CALCULATIONS AND COVERAGE RATES

γ	β_1^*	$\frac{\|\beta_2^*\|}{\sqrt{p}}$	λ	Risk Based on Absolute Error Loss			Risk Based on Squared Error Loss			Coverage Rate Nominal 95% CI		
				REQML	LIML	TSLS	REQML	LIML	TSLS	REQML	LIML	TSLS
.096	$\hat{\beta}_1^*$	$\hat{\sigma}_\beta$	14.4	.017	.036	.024	.021	.049	.025	.95	.61	.16
.000	$\hat{\beta}_1^*$	$\hat{\sigma}_\beta$	13.2	.018	.035	.054	.022	.047	.054	.95	.63	.00
−.192	$\hat{\beta}_1^*$	$\hat{\sigma}_\beta$	5.5	.027	.038	.199	.035	.049	.199	.95	.79	.00
.096	$\hat{\beta}_1^*$	0	∞	.017	.050	.025	.022	.067	.026	.95	.54	.15
.000	$\hat{\beta}_1^*$	0	∞	.018	.049	.057	.023	.067	.057	.95	.57	.00
−.192	$\hat{\beta}_1^*$	0	∞	.028	.051	.200	.035	.067	.200	.95	.73	.00
.096	$\hat{\beta}_1^*$	$2\hat{\sigma}_\beta$	3.6	.015	.020	.020	.019	.025	.021	.95	.72	.23
.000	$\hat{\beta}_1^*$	$2\hat{\sigma}_\beta$	3.3	.015	.020	.046	.019	.025	.047	.95	.74	.00
−.192	$\hat{\beta}_1^*$	$2\hat{\sigma}_\beta$	1.4	.022	.024	.178	.028	.030	.179	.95	.85	.00
.096	0	$\hat{\sigma}_\beta$	14.4	.109	.101	.027	.131	.123	.028	.97	.42	.12
.000	0	$\hat{\sigma}_\beta$	13.2	.108	.101	.063	.129	.122	.063	.96	.44	.00
−.192	0	$\hat{\sigma}_\beta$	5.5	.103	.105	.200	.123	.125	.200	.96	.60	.00
.096	0	0	∞	na	na	na	na	na	na	.98	.38	.09
.000	0	0	∞	na	na	na	na	na	na	.98	.38	.00
−.192	0	0	∞	na	na	na	na	na	na	.95	.38	.00
.096	0	$2\hat{\sigma}_\beta$	3.6	.033	.032	.023	.044	.042	.024	.95	.64	.17
.000	0	$2\hat{\sigma}_\beta$	3.3	.033	.032	.053	.044	.042	.053	.95	.66	.00
−.192	0	$2\hat{\sigma}_\beta$	1.4	.035	.035	.198	.046	.046	.198	.95	.81	.00

Note: Absolute Error Loss: $R(\theta, d)$ with loss function $L_1(\theta, a) = \min\{|\gamma - a|, .2\}$. Squared Error Loss: $\sqrt{R(\theta, d)}$ with loss function $L_2(\theta, a) = [\min\{|\gamma - a|, .2\}]^2$. $\hat{\beta}_1^* = 30.4$, $\hat{\sigma}_\beta = .831$. We do not report the risk when the parameter is not identified (the case with $\beta_1^* = 0$ and $\|\beta_2^*\|/\sqrt{p} = 0$), as in that case the relative ranking by risk is not meaningful.

The risk values for estimating γ under absolute error loss, truncated at .2, are shown in Table II. For a given value of $\theta = (\gamma, \beta_1^*, \beta_2^*, \Sigma)$, the same $J = 8,000$ samples are used to evaluate the risk of each of the three estimators. The risk for the REQML estimator is generally less than the risk of the LIML estimator. The risk improvement is substantial when $\|\beta_2^*\| = 0$; for example, a mean absolute error of .017 versus .050. The TSLS risk becomes large as γ moves away from the population least-squares value of $\sigma_{12}/\sigma_{11} = .067$; when $\gamma = -.192$, the TSLS risk is close to the truncation value of .2. The square-root risk comparisons give similar results. Coverage rates correspond to $1 - R(\theta, d)$, with loss function $L(\theta, a)$ equal to 0 if $\gamma \in a$ and 1 otherwise, where a is an interval on the line. The decision rule $d(z)$ is an interval estimator. The three interval estimators are the ones used in Table I. Table II gives the coverage rates, for the same 8,000 Monte Carlo samples, of the nominal .95 confidence intervals. The coverage rate is the fraction of Monte Carlo samples for which the interval estimate contains the population value of γ. The coverage rates of the REQML intervals are always very

close to the nominal value of .95, and never below .95. The TSLS intervals have poor coverage rates, as low as .00 in some designs. The LIML intervals have better coverage rates than the TSLS intervals, but the rates for the LIML intervals can still be below .40.

It is interesting to interpret the difference between the LIML and REQML coverage rates in the light of the result in Section 2 that with no important instruments ($k_1 = 0$), and with Σ known, the LIML estimator is identical to the REQML estimator but the standard error of the latter is higher by a factor of $\sqrt{\lambda + 1}$. The designs with $\beta_1^* = 0$ correspond most closely to this case (although $k_1 = 1$, this instrument is irrelevant so its presence makes little difference). Consider the design with $\beta_1^* = 0$, $\gamma = 0.096$, $\|\beta_2^*\|/\sqrt{p} = \hat{\sigma}_\beta = 0.831$, implying that $\lambda = (\phi' \Sigma^{-1} \phi \sigma_\beta^2)^{-1} = 14.4$. Hence the standard error for LIML should be smaller than the REQML standard error by a factor of 0.255. If the REQML standard error is correct, and the distribution of LIML is approximately normal and centered at the truth, this should lead to a coverage rate for the LIML (nominal 95%) intervals of $\Phi(1.96 \times 0.255) - \Phi(-1.96 \times 0.255) = \Phi(.50) - \Phi(-.50) = 0.38$ (where Φ is the normal distribution function), close to the coverage rate of 0.42 we find. Taking the design with $\beta_1^* = 0$, $\gamma = 0.096$, $\|\beta_2^*\|/\sqrt{p} = 2\hat{\sigma}_\beta = 1.662$, we find $\lambda = 3.6$. Here the adjustment to the standard errors suggests that the coverage rate of the nominal 95% LIML confidence intervals should be 0.64, exactly what we find in our Monte Carlo study.[12]

4. CONCLUSION

We have developed a random-effects estimator for instrumental variables analyses with many instrumental variables. In addition to leading to a new estimator, this approach sheds light on the poor performance of LIML and TSLS with many weak instruments. We show that in interesting cases LIML and TSLS can be approximated by restricted versions of our REQML estimator. When these restrictions do not hold, LIML and TSLS confidence intervals are likely to have under-coverage. In the Monte Carlo analysis, we find that the coverage rates for the LIML intervals can be as low as .40, close to what is predicted by our theoretical calculations, whereas the coverage rates for the random-effects intervals are always very close to the nominal value of .95. The risk for the REQML estimator is generally less than the risk of the LIML estimator, with the risk improvement substantial when the coefficients on the doubtful instrumental variables are in fact all equal to zero. In agreement with the earlier literature, the TSLS estimator does very poorly in terms of maximum risk and coverage rates. An interpretation of this poor performance is that TSLS corresponds to REQML with σ_β fixed at a large value.

We offer two recommendations to researchers using instrumental variables methods in settings with many potentially weak instruments. First, in addition to considering the concentration parameter to assess the credibility of TSLS as recommended by BJB and Staiger and Stock (1997), one should calculate the λ-statistic to assess the potential bias in the LIML standard errors. Second, in case these calculations suggest that TSLS and

[12]Clearly this adjustment does not work for extreme values of λ. For example if $\sigma_\beta = 0$, then $\lambda = \infty$, and the adjustment would suggest zero coverage rates for the LIML intervals. Similarly, consistent with the derivation of the adjustment, it is not as accurate in the case where the first set of instruments is powerful. Nevertheless it shows the value of the theoretical calculations involving λ in understanding the finite sample behavior of the various estimators.

LIML confidence intervals are likely to have poor coverage rates, the random-effects estimator is a conceptually straightforward way of efficiently combining many weak instrumental variables.

Dept. of Economics, Harvard University, Littauer Center 123, North Yard, Cambridge, MA 02138, U.S.A.; gary_chamberlain@harvard.edu
and
Dept. of Economics and Dept. of Agricultural and Resource Economics, University of California Berkeley, Berkeley, CA 94720-3880, U.S.A; imbens@econ.berkeley.edu; http:// elsa.berkeley.edu/users/imbens/

REFERENCES

ANGRIST, J., AND A. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979–1014.

BEKKER, P. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

BOUND, J., D. JAEGER, AND R. BAKER (1995): "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

CHAMBERLAIN, G., AND G. IMBENS (2001): "Random Effects Estimators with Many Instrumental Variables," Manuscript, Harvard University; *http://elsa.berkeley.edu/users/imbens/*.

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

MOREIRA, M. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

STAIGER, D., AND J. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.