Bayesian Two-Stage Regression with Parametric Heteroscedasticity

Arto Luoma
arto.luoma@uta.fi
Department of Mathematics and Statistics
33014 University of Tampere, Finland

Jani Luoto
jani.luoto@econ.jyu.fi
School of Business and Economics
P.O. Box 35, FIN-40014 University of Jyväskylä, Finland

Abstract:

In this paper we expand Kleibergen and Zivot’s (2003) Bayesian Two Stage (B2S) model by allowing for unequal variances. Our choice for modelling heteroscedasticity is a fully Bayesian parametric approach. As an application we present a cross-country Cobb-Douglas production function estimation.

Keywords: parameter heteroscedasticity, Bayesian two-stage model, Cobb-Douglas production function
1. Introduction

After Anderson and Rubin (1949) developed their limited information maximum likelihood (LIML) and Theil (1953) his two-stage least squares (2SLS) technique, instrumental variables (IV) regression has became a standard textbook approach in classical econometrics. The development of Bayesian analysis of such models started two decades later, being initiated by Drèze (1976); see also e.g. Drèze and Morales (1976) Drèze and Richard (1983), and Bauwens and van Dijk (1989). Drèze’s idea was to equalize the classical and Bayesian analysis of IV models using suitable diffuse priors for the parameters. Unfortunately, his prior ignores important information concerning the near nonidentification of structural parameters due to weak instruments; see e.g. Kleibergen and Zivot (2003) for discussion.

Mainly due to this undesirable property of the Drèze prior, recent research on Bayesian analysis of IV models has started to address the above mentioned problem of local non-identification; see e.g. Geweke (1996), Kleibergen and van Dijk (1998) and Chao and Phillips (1998, 2002). Following this tradition, Kleibergen and Zivot (2003) developed a new Bayesian two-stage (B2S) approach. In order to mimic classical 2SLS techniques, which essentially handle the problem of local non-identification, they constructed a prior for the parameters of the restricted reduced form specification and thus functionalized the steps used to obtain the 2SLS estimator.

In this paper we expand Kleibergen and Zivot’s (2003) B2S model by allowing for unequal variances. In classical analysis, modelling heteroscedasticity improves the efficiency of estimation and enables the variance estimates to be consistent. Thus, not surprisingly, modelling heteroscedasticity has become standard in classical IV literature; see e.g. White (1982), Cumby et al. (1983), and Davidson and MacKinnon (1993). However, there is (to our knowledge) no single Bayesian study of IV models with unequal variances, although from the Bayesian point of view modelling heteroscedasticity should improve the precision of estimates and the quality of predictive inference. The latter follows from the fact that modelling heteroscedasticity allows predictive inferences to be more precise for some units and less so for other.

Our choice for modelling heteroscedasticity is a fully Bayesian parametric approach. Specifically, we assume that \( \text{var}(y_i) = \sigma^2 \zeta_i^\theta \), where \( y \) is the response variable and \( \zeta \) a variable explaining the variance. This specification requires only one unknown heteroscedasticity parameter (\( \theta \)) and, with a certain normalization, yields a simple constant Jeffreys' prior density for \( \theta \); see e.g. Ohtani (1982),
Boscarding and Gelman (1996) and Tanizaki and Zhang (2001). Alternatively, we could follow, for example, Geweke (1993) and model heteroscedasticity using a nonparametric approach. This, however, would require estimation of several unknown parameters, which might give rise to identification and estimation problems in our relatively complex nonlinear model.

To give an empirical illustration of the properties of the heteroscedastic B2S model, we follow Benhabib and Spiegel (1994) and Papageorgiou (2003) and construct a simple exercise of aggregate production function estimation as an application. We choose this example, since the problems of endogeneity and heteroscedasticity are well documented in the cross-country growth literature, see e.g. Benhabib and Spiegel (1994) and the surveys of Temple (1999a) and Durlauf et al. (2005).

The paper is organized as follows: In Section 2 we present a heteroscedastic Bayesian two-stage model (hereafter HB2S model). In Section 3 we give an example of estimating the empirical Cobb-Douglas aggregate production function. Section 4 concludes the paper.

2. The Bayesian Two-Stage Model with Parametric Heteroscedasticity

Consider the following limited information simultaneous equation model

\[ y_1 = Y_2 \beta + Z \gamma + \varepsilon_1, \]  

\[ Y_2 = X \Pi + Z \Gamma + V_2, \]  

where \( Y = (y_1, y_2) \) is an \( N \times m \) matrix of endogenous variables, \( Z \) an \( N \times k_1 \) matrix of included exogenous variables, \( X \) an \( N \times k_2 \) matrix of excluded exogenous variables, that is, instruments, and \( \varepsilon_1 \) an \( N \times 1 \) vector of errors and \( V_2 \) an \( N \times (m-1) \) matrix of errors. Vectors \( \beta \) and \( \gamma \) contain the structural parameters of interest. The matrices \( Z \) and \( X \) are assumed to be of full column rank, uncorrelated with \( \varepsilon_1 \) and \( V_2 \), and weakly exogenous for the structural parameter vector \( \beta \).

If the observation vectors \( y_i \) in the above simultaneous equation model have unequal covariance matrices, they are said to be heteroscedastic. In the following, we will model heteroscedasticity by

---

\(^1\) See also Barro (1999) and Temple (1999a, 1999b, 2001).
assuming that the elements $\varepsilon_{1i}$ of $\varepsilon$ and the rows $V_{2i}$ of $V_2$ are normally distributed with zero mean and the $m \times m$ covariance matrix

$$\Sigma_i = \text{var}(\varepsilon_{1i} \ V_{2i}) = f(\xi_i, \theta) \left( \begin{array}{cc} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{array} \right),$$

where heteroscedasticity is captured by the function $f(\xi_i, \theta)$, $\xi_1, \ldots, \xi_N$ being the known values of some positive-valued variable. Several alternative specifications of $f(\xi_i, \theta)$ have been suggested in the literature; see e.g. Judge et al. (1985), Greene (1990), Griffiths (1999) and Tanizaki and Zhang (2001). Here, we consider the following simple functional form

$$f(\xi_i, \theta) = \xi_i^{-\theta},$$

where $\theta \in [0, 1]$, and the extreme of $\theta = 0$ corresponds to homoscedastic errors; see e.g. Greene (1990) and Boscarding and Gelman (1996). If we substitute the reduced form Equation (2) into the structural form Equation (1) and reparametrize it slightly, we get the following nonlinearly restricted reduced form specification

$$y_1 = W\delta + v_1,$$

$$Y_2 = UB + V_2,$$

where $W = (UB \ Z)$, $\delta = (\beta' \ \gamma')'$, $U = (X \ Z)$, $B = (\Pi' \ \Gamma')'$ and $v_1 = \varepsilon_1 + V_2\beta$. Thus, $U$ is an $N \times k$ matrix, where $k = k_1 + k_2$, and $W$ is an $N \times (k_1 + m - 1)$ matrix. Denoting

$$\Omega_i = \text{var}(v_{1i} \ V_{2i}) = f(\xi_i, \theta) \left( \begin{array}{cc} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{array} \right); \ \omega_{11,2} = \omega_{11} - \Omega_{12}^{-1}\Omega_{21}^{-1} \omega_{22}; \ \phi = \Omega_{22}^{-1}\Omega_{21}^{-1},$$

we obtain that $e_1 = v_1 - V_2\phi$ is uncorrelated with $V_2$ and

$$\text{var}(e_{1i}) = f(\xi_i, \theta)\omega_{12}.$$
Since in the Bayesian two-stage approach suggested by Kleibergen and Zivot (2003) the prior distribution is so constructed that it explicitly incorporates this kind of knowledge, we will choose it as a starting-point for our heteroscedasticity-corrected limited information model.

The likelihood of our model is

\[ p(Y|X,Z,\eta) = p(Y_2|X,Z,\eta)p(y_1|X,Z,\eta), \]

where

\[ p(y_1|Y_2,X,Z,\eta) \propto \omega_{12}^{-0.5N}|\Lambda|^{-0.5} \exp\left(-0.5\omega_{12}^{-1}(y_1 - W\delta - V_2\phi)\Lambda^{-1}(y_1 - W\delta - V_2\phi)\right), \tag{7} \]

\[ p(Y_2|X,Z,\eta) \propto |\Omega_{22}|^{-0.5N}|\Lambda|^{-0.5(m-1)} \exp\left(-0.5\text{tr}\Omega_{22}^{-1}(Y_2 - UB)\Lambda^{-1}(Y_2 - UB)\right). \tag{8} \]

Here, \( \eta \) denotes the vector of all parameters and \( \Lambda = \text{diag}(f(\xi_1,0),\ldots,f(\xi_n,0)) \).

As a prior distribution we will use a modification of Jeffreys' uninformative prior distribution. Jeffreys' prior is defined as \( p(\eta) \propto |I(\eta)|^{1/2} \), where

\[ I(\eta) = -E\left[ \frac{\partial^2}{\partial \eta \partial \eta'} \log p(Y | \eta) \right] \]

is the Fisher information matrix for \( \eta \). Our modifications will be as follows: firstly, we calculate the second-order derivative with respect to \( \text{vec}(B) \) from the logarithm of the conditional density (8) instead of the full log likelihood. Secondly, we remove the prior dependence between \( B \) and \( \delta \) by replacing the corresponding non-diagonal blocks in \( I(\eta) \) with zero matrices. Then, if we assume that the geometric mean of \( \zeta_i, i = 1,\ldots,T \), is unity, the joint prior is given by

\[ p(\eta) \propto |\Omega_{22}|^{-0.5(m+k-1)}|\omega_{12}^{-0.5(2m+k)}|U'\Lambda^{-1}U|^{-0.5(m-3)}|W'\Lambda^{-1}W|^{0.5}. \tag{9} \]
One can normalize the weight variable $\zeta$ by dividing it by its geometric mean. This has two advantages: firstly, one need not adjust the prior distribution, and secondly, the dispersion parameters $\omega_{12}$ and $\Omega_{22}$ have a consistent meaning under different values of $\theta$. See also Boscardin and Gelman (1996), who discuss the issue in the context of one-stage regression models.

The presence of $\Pi$ in the prior (9) reflects the fact that the model is not informative regarding $\beta$ when $\Pi$ has reduced rank; see Kleibergen and Zivot (2003) for further discussion. Multiplying the likelihood function by the joint prior (13) yields, after some tedious algebra, the following conditional and marginal posteriors

$$p(\phi|Y, B, \theta, \omega_{12}, \Omega_{22}) \propto \omega_{12}^{-0.5(m-\delta_1)} |W^\top W|^{-0.5} \exp\left\{ -0.5\omega_{12}^{-1}(\delta - \hat{\delta})W^\top \Lambda^{-1}W(\delta - \hat{\delta}) \right\},$$

$$p(\phi|Y, B, \theta, \omega_{12}, \Omega_{22}) \propto \omega_{12}^{-0.5(m-\delta_2)} |V_2^\top \Lambda^{-1}M V_2|^{-0.5} \exp\left\{ -0.5\omega_{12}^{-1}(\phi - \hat{\phi})V_2^\top \Lambda^{-1}M V_2(\phi - \hat{\phi}) \right\},$$

$$p(\omega_{12}|Y, B, \theta, \Omega_{22}) \propto \omega_{12}^{-0.5(N+2)} (\nu^\top \Lambda^{-1}M \nu)^{0.5N} \exp\left\{ -0.5\omega_{12}^{-1} \nu^\top \Lambda^{-1}M \nu \right\},$$

$$p(\Omega_{22}|Y, B, \theta) \propto |\Omega_{22}|^{-0.5(N+m+1)} |V_2^\top \Lambda^{-1}V_2|^{0.5} \exp\left\{ -0.5tr \Omega_{22}^{-1} V_2^\top \Lambda^{-1}V_2 \right\},$$

$$p(B, \theta|Y) \propto (\nu^\top \Lambda^{-1}M \nu)^{-0.5N} |U^\top U|^{0.5(m-\delta_1)} |\Lambda|^{-0.5m} |V_2^\top \Lambda^{-1}M V_2|^{-0.5} |V_2^\top \Lambda^{-1}V_2|^{-0.5(N-k-1)},$$

where

$$\hat{\delta} = (W^\top \Lambda^{-1}W)^{-1} W^\top \Lambda^{-1}(y_1 - V_2 \phi), \quad \hat{\phi} = (V_2^\top \Lambda^{-1}M V_2)^{-1} V_2^\top \Lambda^{-1}M y_1, \quad M = I - W(W^\top \Lambda^{-1}W)^{-1} W^\top \Lambda^{-1},$$

$$\nu = y_1 - V_2 \hat{\phi} \quad \text{and} \quad V_2 = Y_2 - UB.$$
Kleibergen and Zivot (2003) discuss some properties of their B2S model and compare it to the original Drèze (1976) approach. We briefly review their discussion and make some comparison between our parametric heteroscedasticity-corrected model and their B2S model.

1. As with the Drèze and B2S approaches, the posteriors are not invariant to the ordering of the endogenous variables; that is, if \( y_1 \) is exchanged with some of the variables in \( Y_2 \), the results do not remain identical. See Drèze (1976) for the argument.

2. The mean of the conditional posterior of \( \beta \) in the B2S model is essentially \( \hat{\beta}_{2SLS} \). However, this is not true for the HB2S model, since heteroscedasticity correction gives more weight to ‘good’ observations, while \( \hat{\beta}_{2SLS} \) weights all observations equally. The difference between the heteroscedastic-corrected estimate of \( \beta \) and \( \hat{\beta}_{2SLS} \) depends, of course, on the degree of heteroscedasticity.

3. As with the B2S approach, the marginal posterior of \( \Pi \) does not have the non-integrable asymptote at \( \Pi = 0 \) which appears in the Drèze approach. The argument is similar to that used for the Kleibergen and Zivot (2003) and is omitted.

4. As with the B2S approach (without heteroscedasticity correction), the form of the posterior of \( B \) is closely related to the marginal posterior which results from a standard diffuse prior analysis of the reduced form regression of \( Y_2 \) on \( U \) with heteroscedasticity correction.

3. Empirical Example

3.1. Estimated Model

To illustrate some of the properties of the HB2S model we construct a simple exercise of aggregate production function estimation with cross-country data. We chose this example, since problems of endogeneity and heteroscedasticity are well documented in the cross-country growth literature; see e.g. Benhabib and Spiegel (1994), Papageorgiou (2003), and the surveys of Temple (1999) and Durlauf et al. (2005). For example, Benhabib and Spiegel (1994) analyse the biases of coefficient estimates which result from the correlation between the accumulated physical and human capital series and the error term, and find that there is likely to be an upward coefficient bias in the input share of capital and human capital estimates, and a downward bias in estimates of the input share of labour. Our analysis is close to that of Benhabib and Spiegel (1994) or Papageorgiou (2003). However, we do not separate aggregate labour and human capital stocks; rather we follow Bils and Klenow (2000) and assume that individual human capital stock is related to individuals, years of
schooling and years of experience. This implies that each individual has some degree of human capital and thus aggregate human-capital stock should be modelled as \( H_t = h_t L_t \), where \( h_t \) is average human-capital stock per person and \( L_t \) is labour force.

Taking the log differences of the assumed Romer type Cobb-Douglas production function \( Y_t = A_t^{\alpha} K_t^\beta H_{Y,t}^\gamma e_t \) (see Romer, 1990), we obtain the following equation for long-run growth,

\[
\log \left( \frac{Y_{it}}{Y_{i0}} \right) = (\alpha + \beta) \log \left( \frac{A_{it}}{A_{i0}} \right) + \alpha \log \left( \frac{K_{it}}{K_{i0}} \right) + \beta \log \left( \frac{H_{Y,it}}{H_{Y,i0}} \right) + \log \left( \frac{e_{it}}{e_{i0}} \right),
\]

where \( H_{Y,it} \) is the human capital engaged in final-goods production. In Equation (15), we assume that the resource constraint \( H = H_{A,it} + H_{Y,it} \) holds. One problem in estimating Equation (15) is that we should replace an unobservable \( \log(A_t / A_0) \) by some function of observables. Otherwise the estimates of factor shares will be biased; see e.g. Temple (1999). We follow Papageorgiou (2003) and propose the following specification for the growth rate of technology,

\[
\frac{A_{it} - A_{i0}}{A_{i0}} = \delta H_{A,it} + \mu H_{A,it} \left( \frac{\bar{A}_o}{A_{i0}} - 1 \right),
\]

where \( \bar{A}_o \) is the technology frontier, \( H_{A,it} \) is the human capital engaged in R&D activities and \( \delta \) and \( \mu \) are the innovation and imitation parameters, respectively. In Equation (16), human-capital speeds technology growth through innovation and imitation. Using Equation (16) we can write Equation (15) in the estimation form

\[
\log \left( \frac{Y_{it}}{Y_{i0}} \right) = \psi \cdot d_i + (\alpha + \beta) \left[ (\delta - \mu) H_{A,i0} + \mu \left( \frac{Y_{0,\max}}{Y_{i0}} \right) H_{A,i0} \right] + \alpha \log \left( \frac{K_{it}}{K_{i0}} \right) + \beta \log \left( \frac{H_{Y,it}}{H_{Y,i0}} \right) + u_{it},
\]

where \( d_i \) is a vector of deterministic components (constant and dummy variables), and \( u_{it} \) a normally distributed error term with zero mean and \( \sigma_{11} \) variance. We follow Benhabib and Spiegel (1994) and Papageorgiou (2003) in assuming that \( (Y_0 / L_0)^{\max} / (Y_{i0} / L_{i0}) \) approximates \( \bar{A}_0 / A_{i0} \).
Since human capital may also speed technology adoption and may be to some extent necessary for technology use, we propose the following two alternative specifications for technology growth with production technology, $Y_t = A_t a + \beta K_t \alpha H_t \beta e_t$ and $Y_t = A_t a + \beta K_t \alpha L_t \beta e_t$; see e.g. Benhabib and Spiegel (1994) and Bils and Klenow (2000). With similar steps we obtain the corresponding empirical specifications

\[
\log \left( \frac{Y_t}{Y_{i0}} \right) = \gamma_d + (\alpha + \beta) \left[ (\delta - \mu) H_{i0} + \mu \left( \frac{Y_{i0}}{Y_{i0}} \right) H_{i0} \right] + \alpha \log \left( \frac{K_{i0}}{K_{i0}} \right) + \beta \log \left( \frac{H_{i0}}{H_{i0}} \right) + u_t, \quad (18)
\]

\[
\log \left( \frac{Y_t}{Y_{i0}} \right) = \gamma_d + (\alpha + \beta) \left[ (\delta - \mu) H_{i0} + \mu \left( \frac{Y_{i0}}{Y_{i0}} \right) H_{i0} \right] + \alpha \log \left( \frac{K_{i0}}{K_{i0}} \right) + \beta \log \left( \frac{L_{i0}}{L_{i0}} \right) + u_t. \quad (19)
\]

We estimate Equations (17)-(19) using our HB2S model. As a weight variable we use $\zeta_i = y_{i0}$, where

\[
y_{i0} = \exp \left( \log \left( \frac{Y_{i0}}{L_{i0}} \right) - \frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{Y_{i0}}{L_{i0}} \right) \right).
\]

This corresponds to dividing the output per labour force ($Y_{i0}/L_{i0}$) by its geometric mean. We use the output per worker in the weight coefficients $f(\zeta,0)$, $i = 1, ..., N$, since we expect countries with higher initial income to have more stable growth paths due to developed institutional structures, which have the ability to reduce the overall risk in society. Alternatively, we could use some institutional indicator. However, since the choice of institutional indicators which approximate the true ‘level’ of institutional quality is somewhat difficult, and far from unique, we decided to abandon this approach.

### 3.2 Estimation Results

The data and choice of instruments are described in Appendix A. To generate a Monte Carlo sample from the joint posterior of $\theta$ and $B$ we used a version of the random walk Metropolis algorithm for Markov Chain Monte Carlo (MMCMC). The algorithm uses a multivariate normal distribution for the jump distribution on changes in $\theta$ and $B$. Our simulation procedure was as follows: We first simulated 10,000 draws using a diagonal covariance matrix with diagonal entries 0.0000001 in the
jump distribution. We then used these draws to estimate the posterior covariance matrix of $\theta$ and $B$ and scaled it by the factor $2.4^2/d$, where $d$ is the number of simulated parameters, to obtain an optimal covariance matrix for the jump distribution; see e.g. Gelman et al. (2004). We continued by simulating 10,000 draws and calculated a more accurate covariance matrix for $\theta$ and $B$. We repeated this 5-7 times. Finally, we ran 200,000-300,000 draws and picked out every 100th/200th draw after excluding the first 100,000-200,000.

Figure 1. Residual plots of the first- and second-stage regressions, corresponding to equations (8) and (7), respectively. The dotted lines are approximate 95% probability intervals, based on the normality assumption. The first row gives the residual plots against $y_{i0}$, when $\log(Y_{T}/Y_{0})$, $\log(H_{T}/H_{0})$ and $\log(K_{T}/K_{0})$ of model (17) are regressed on the instrumental variables. The second row gives the corresponding residual plots for $\log(Y_{T}/Y_{0})$, $\log(H_{T}/H_{0})$, and $\log(K_{T}/K_{0})$ when model (18) is used. Finally, the third row gives the residual plots for $\log(Y_{T}/Y_{0})$, $\log(L_{T}/L_{0})$ and $\log(K_{T}/K_{0})$, corresponding to model (19). The residuals are obtained when the unknown parameters are replaced by their posterior means.
In the B2S model we also used a multivariate normal distribution for the jump distribution on the changes in $B$. The covariance matrix of the classical first-stage regression (scaled by the factor $2.42/d$) was used to obtain the covariance matrix of the multivariate normal jump distribution. We simulated 300,000 rounds and used the first 100,000 rounds as a burn-in period. Finally, we picked out every 200th draw from those included.\footnote{The estimation was implemented using R, a statistical computing environment. R is freely available under the General Public Licence at www.R-project.org. The code and data sets will be available at http://mtl.uta.fi/codes/HB2S/.} Table 2 in the Appendix B shows the convergence diagnostics of the estimated models.

Figure 1 displays the residual plots for the first- and second-stage regressions, corresponding to Equations (8) and (7), respectively. The residuals have been obtained by replacing the unknown parameters by their posterior means and they have been plotted against the normalized initial output $y_0$. The approximate 95% probability belts, based on the normality assumption, are also shown. Table 1 shows the estimation results for Equations (17)-(19), obtained using the ordinary least squares (OLS) method and the Bayesian estimation of the B2S and HB2S models.

On the basis of the figures and the posterior summaries of the heteroscedasticity parameter $\theta$, we see that the data support heteroscedasticity in each model. Heteroscedasticity is especially obvious in the cases of output and physical capital growth, less so in human capital and only slight in labour growth. When we compare the estimated average discrepancies\footnote{The estimated average discrepancy approximates the expected deviance. In the limit of large sample size, the model with lowest expected deviance will have the highest posterior probability. We prefer using discrepancy between data and model to using Bayes factors in model comparisons. We agree with Gelman et al. (2004), who consider Bayes factors to be in most cases irrelevant, since they are used to compute the relative probabilities of the models conditional on one of them being true.} $\bar{D}_{avg}(Y)$ between the B2S and HB2S models, we see that the data lend strong support to the latter. It seems that the margin between the estimated average discrepancies of B2S and HB2S depends on the degree of heteroscedasticity, since the gap is around 20 for models (17) and (18), while for model (19) it is about 14 only. Note that model (19) has the lowest heteroscedasticity parameter due to the small amount of heteroscedasticity in the labour growth series (see the third row in Figure 1). Furthermore, the model defined by (19) has the smallest estimated average discrepancy (-432). However, in the ‘economic theory’ sense, this result does not necessarily indicate that Equation (19) is more preferable than the other models, since the first-stage regression could be more informative in this model, increasing the overall model fit.
We also find that the IV regression estimates of $\alpha$ are, in general, higher, and the estimates of $\beta$ lower, than the corresponding OLS estimates\(^4\). Thus our results confirm the finding of Benhabib and Spiegel (1994) that there is an upward coefficient bias in the OLS estimates of $\alpha$ and human capital share $\beta$ (Equations 17-18), and a downward bias in the OLS estimates of the labour share parameter $\beta$ (Equation 19)\(^5\); see also Griliches and Jacques (1995) for a discussion of endogeneity of regressors in the aggregate production function approach.

Table 1. Growth regressions for Equations 17-19

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Equation (17)</th>
<th>Equation (18)</th>
<th>Equation (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha + \beta$</td>
<td>OLS</td>
<td>B2S</td>
<td>HB2S</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$\beta$</td>
<td>$\alpha$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$\beta$</td>
<td>$\alpha$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$\beta$</td>
<td>$\alpha$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>Constant</td>
<td>0.275</td>
<td>0.335</td>
<td>0.305</td>
</tr>
<tr>
<td>dummyAfrica</td>
<td>-0.262</td>
<td>-0.348</td>
<td>-0.298</td>
</tr>
<tr>
<td>dummyLatinAmerica</td>
<td>-0.187</td>
<td>-0.261</td>
<td>-0.253</td>
</tr>
<tr>
<td>$\hat{D}_{avg}(Y)$</td>
<td>-378.7</td>
<td>-399.0</td>
<td>-373.0</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Note that multiplying $L_i$ by $h_i$ seems to reduce $\beta$ more in the IV models than in the ordinary regression model.

---

\(^4\) See also the OLS results, where the dummies are excluded from the analysis, in Table 3 (Appendix B). Our OLS results are, in general, quite similar to those in previous studies. Specifically, our estimates for physical capital share $\alpha$ lie between 0.432 and 0.53 and are positive at the 1% significance level. The estimates of human capital/labour shares are relatively low and positive at the 5% level.

\(^5\) Note that multiplying $L_i$ by $h_i$ seems to reduce $\beta$ more in the IV models than in the ordinary regression model.
Finally, based on the results reported in Table 1, the data are not consistent with the innovation parameter $\delta$ being positive; see e.g. Benhabib and Spiegel (1994), who obtained similar results in their analyses. On the other hand, there is weak (or moderate) support in the data for the imitation parameter $\mu$ being positive. Thus, contrary to Papageorgiou (2003), our results slightly favour catch-up progress over country-specific technological progress as the channel through which accumulation of human capital affects output growth. This is quite sensible, since only about 15 per cent of the countries in our sample have economically meaningful innovation activities; see also Benhabib and Spiegel (1994).

4. Conclusion

In this paper we have presented a relatively straightforward way to model unequal variances in Bayesian two-stage instrumental variable regression. We have done this using a fully Bayesian parametric approach. As noted, modelling heteroscedasticity is important also in the Bayesian instrumental variable context, since it improves the precision of estimates and the quality of predictive inference.

We used a simple production function approach as a tool to provide an empirical illustration of some properties of the heteroscedastic B2S model. On the basis of residual plots and estimated discrepancies between the data and the models, we have shown that the data lend strong support to the use of the HB2S model instead of the homoscedastic B2S model.

Because our modelling of heteroscedasticity is relatively limited, we suggest that future research on Bayesian IV regression under unequal variances should focus on multiplicative heteroscedasticity, which is flexible and includes most of the useful formulations for heteroscedasticity as special cases; see e.g. Tanizaki and Zhang (2001).
References:


APPENDIX A

Data and Instruments

Our estimation involves data on 85 countries (see Table 4 in Appendix B). The stock of physical capital is estimated using each country’s investment rates from Penn World Tables 6.1 and perpetual inventory methods. The capital stock in 1960 is estimated using $K_i = \frac{I_i}{(g_i + d + n_i)}$, where $I$ denotes the investments, $g$ the growth rate of GDP per worker, $d$ the depreciation rate and $n$ the growth rate of the population, calculated as the average growth rate from 1961 to 1970. The depreciation rate $d$ is assumed to be 0.076.

In the case of human capital we follow Bils and Klenow (2000), who approximate the human capital per person using the years of schooling per person and the experience of each age group. Specifically, we assume that the log of human capital stock of a worker of age $a$ is

$$\ln h(a) = f(s) + \gamma_1(a - s - 6) + \gamma_2(a - s - 6)^2, \quad (20)$$

where $\gamma_1$ and $\gamma_2$ are parameters of return to experience, $s$ is average years of schooling and $f(s) = \theta \cdot s^{1+\psi} / (1-\psi)$; $\psi > 0$, $\theta > 0$. Equation (20) is of the same form as that of Bils and Klenow (2000); however we assume that the influence of a teacher on human capital is zero. Using Equation (20) we calculate the average human capital stock for all age groups between 20 and 59 in 1970 and 2000 by weighting the human capital of the age group by its proportion of the country’s total population. In Equation (20) we set $\gamma_1 = 0.0512$ and $\gamma_2 = -0.00071$, which corresponds to the average estimates across 52 countries as reported in Bils and Klenow (2000). We set $\psi$ at 0.28 and set $\theta$ so that the mean of $f'(s) = \theta / s^{\psi}$ equals the mean Mincerian returns across 56 countries, which is 0.099; see Bils and Klenow (2000). Finally, $H_{A,t}$ is determined as $H_0$ times the percentage of the population aged 15 or over with some higher education (complete + incomplete).

The education series are from Cohen and Soto (2001). The population data are from the International Data Base of the U.S. Census Bureau (Population Division of the International

---

6 We have predicted the missing GDP and investment values for some countries in our data. The missing GDP values are predicted using the linear trend model while the investment share (I/GDP) values are predicted using the latest available data points. These countries (and missing years) are Angola (1997-2000), Central Africa (1999-2000), Cyprus (1997-2000), Fiji (2000), Guyana (2000), Haiti (1999-2000), Sierra Leone (1997-2000) and Singapore (1997-2000).
Programs Center (IPC)) and the United Nations population data (1995). The labour stock \( (L) \) in each country is obtained from World Development Indicators (2002). The output series have been taken from Penn World Tables 6.1.

Since physical and human capitals are accumulated factors, they are endogenous. This causes the simple OLS estimator to be inconsistent. A common means of dealing with the issue of endogeneity is to instrument for endogenous regressors with variables correlated with them but exogenous to them and the regressed variable. Moreover, the validity of an instrument requires that it cannot be a direct growth determinant or correlated with omitted growth determinants; see e.g. Durlauf et al. (2005). Therefore, we instrument the growth rates of aggregate human and physical capital using the distance from the equator (Gallup, Sachs and Mellinger 1998) and the following variables in 1970: age dependency ratio (dependents to working-age population), illiteracy rate (%) of people aged 15-24 from World Development Indicators (2002), and the level of physical capital per worker.

We make the assumption that the distance of a country from the equator, the initial (year 1970) values for age dependency ratio and youth illiteracy rate are not direct growth determinants; rather they influence the environment and investment culture, where individuals accumulate physical and human capital. For a more detailed discussion on these topics see e.g. Durlauf et al. (2005).

Since one may question the validity of our instruments, we check the consistency of the IV estimators using two specification tests. Firstly, Hansen's test for over-identification restrictions is used to see whether the model specification is correct and the instruments are uncorrelated with the error process. The second test is for weak instruments. We follow Stock and Yogo (2002), who propose quantitative definitions of weak instruments based on the maximum IV estimator bias or the maximum Wald test size distortion. The smallest p-value of the Hansen's test for over-identification restrictions for the regression models in this paper is 0.88 and the smallest test statistic of the Stock and Yogo test for weak instruments is 9.17. Thus, we can reject the null of weak instruments and can not reject the null of appropriate instruments at the 5% level. Note that we use classical tests here, since these are readily available and do not demand extra programming effort. Finally, we used African and Latin American country dummies since, based on the above test results, our instruments behave much more appropriately when these dummies are included in the analysis. The reason may be that these dummies approximate some omitted growth determinants which may be correlated with some of our instruments.
# APPENDIX B

## Table 2: Convergence diagnostics of the estimated models

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Equation (17)</th>
<th>Equation (18)</th>
<th>Equation (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB2S</td>
<td>Z-scores</td>
<td>HB2S</td>
<td>Z-scores</td>
</tr>
<tr>
<td>B2S</td>
<td></td>
<td>B2S</td>
<td></td>
</tr>
<tr>
<td>( \Theta )</td>
<td>1.48</td>
<td>-</td>
<td>0.17</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>-0.12</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>-1.71</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>( b_{13} )</td>
<td>-0.59</td>
<td>-1.74</td>
<td></td>
</tr>
<tr>
<td>( b_{14} )</td>
<td>-1.01</td>
<td>-1.80</td>
<td></td>
</tr>
<tr>
<td>( b_{15} )</td>
<td>-0.07</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>( b_{16} )</td>
<td>-0.70</td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td>( b_{17} )</td>
<td>0.94</td>
<td>-0.21</td>
<td></td>
</tr>
<tr>
<td>( b_{18} )</td>
<td>-0.17</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>( b_{19} )</td>
<td>0.03</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>( b_{21} )</td>
<td>0.25</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>-0.94</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>( b_{23} )</td>
<td>0.57</td>
<td>-0.87</td>
<td></td>
</tr>
<tr>
<td>( b_{24} )</td>
<td>-1.67</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>( b_{25} )</td>
<td>1.58</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>( b_{26} )</td>
<td>-0.04</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>( b_{27} )</td>
<td>0.27</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>( b_{28} )</td>
<td>-0.33</td>
<td>-0.19</td>
<td></td>
</tr>
<tr>
<td>( b_{29} )</td>
<td>-0.45</td>
<td>-0.61</td>
<td></td>
</tr>
</tbody>
</table>

The table gives Geweke statistic, posterior mean and standard error of the posterior mean (in parentheses) for each parameter. Geweke's statistic is the difference between the sample means of the first 10% and the last 50% of the simulated chain, divided by its estimated standard deviation. If the samples are drawn from the stationary distribution of the chain, the two means are equal and the Geweke statistic has an asymptotically standard normal distribution. Note that the standard error of the posterior mean depends on the number of simulations and should not be confused with the posterior standard deviation of the parameter.

Posterior means and standard errors of the posterior means (in parentheses):

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Equation (17)</th>
<th>Equation (18)</th>
<th>Equation (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB2S</td>
<td>-0.529 (0.004)</td>
<td>-0.519 (0.004)</td>
<td>-0.435 (0.004)</td>
</tr>
<tr>
<td>B2S</td>
<td>-0.175 (0.012)</td>
<td>-0.145 (0.014)</td>
<td>-0.265 (0.002)</td>
</tr>
<tr>
<td>( b_{11} )</td>
<td>-0.175 (0.003)</td>
<td>-0.245 (0.002)</td>
<td>-0.085 (0.002)</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>-0.138 (0.002)</td>
<td>-0.176 (0.002)</td>
<td>-0.060 (0.001)</td>
</tr>
<tr>
<td>( b_{13} )</td>
<td>0.0001 (1.4e-5)</td>
<td>7.4e-6 (3.3e-7)</td>
<td>1.155 (0.005)</td>
</tr>
<tr>
<td>( b_{14} )</td>
<td>0.0012 (3.3e-5)</td>
<td>-0.0002 (5.0e-6)</td>
<td>0.0006 (4.1e-5)</td>
</tr>
<tr>
<td>( b_{15} )</td>
<td>-0.591 (0.005)</td>
<td>-0.605 (0.004)</td>
<td>-0.465 (0.003)</td>
</tr>
<tr>
<td>( b_{16} )</td>
<td>1.412 (0.005)</td>
<td>1.383 (0.005)</td>
<td>1.200 (0.005)</td>
</tr>
<tr>
<td>( b_{17} )</td>
<td>-0.0004 (5.5e-5)</td>
<td>-0.0009 (6.1e-5)</td>
<td>0.0003 (3.1e-5)</td>
</tr>
<tr>
<td>( b_{18} )</td>
<td>0.031 (0.001)</td>
<td>0.042 (0.001)</td>
<td>0.053 (0.008)</td>
</tr>
<tr>
<td>( b_{19} )</td>
<td>4.811 (0.031)</td>
<td>4.736 (0.028)</td>
<td>4.765 (0.031)</td>
</tr>
<tr>
<td>( b_{21} )</td>
<td>-1.147 (0.006)</td>
<td>-1.140 (0.005)</td>
<td>-1.155 (0.006)</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>-0.672 (0.005)</td>
<td>-0.667 (0.005)</td>
<td>-0.651 (0.005)</td>
</tr>
<tr>
<td>( b_{23} )</td>
<td>-0.0001 (3.8e-5)</td>
<td>-1.1e-5 (7.3e-7)</td>
<td>1.475 (0.0258)</td>
</tr>
<tr>
<td>( b_{24} )</td>
<td>0.002 (8.1e-5)</td>
<td>0.0003 (1.1e-5)</td>
<td>0.0003 (1.3e-5)</td>
</tr>
<tr>
<td>( b_{25} )</td>
<td>-0.662 (0.009)</td>
<td>-0.674 (0.008)</td>
<td>-0.640 (0.009)</td>
</tr>
<tr>
<td>( b_{26} )</td>
<td>0.448 (0.015)</td>
<td>0.507 (0.013)</td>
<td>0.474 (0.016)</td>
</tr>
<tr>
<td>( b_{27} )</td>
<td>-0.0097 (1.0e-5)</td>
<td>0.010 (0.0001)</td>
<td>0.010 (0.0001)</td>
</tr>
<tr>
<td>( b_{28} )</td>
<td>-0.333 (0.002)</td>
<td>-0.335 (0.002)</td>
<td>-0.332 (0.003)</td>
</tr>
</tbody>
</table>

where \( \text{vec}(B) = (b_{11}, b_{12}, ..., b_{19}, b_{21}, b_{22}, ..., b_{29}) \).
Table 3. OLS results of growth regressions for Equations 17-19 (dummies excluded)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Equation (17)</th>
<th>Equation (18)</th>
<th>Equation (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>0.197</td>
<td>0.159</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.094)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>$(\alpha+\beta)(\delta-\mu)$</td>
<td>-0.03</td>
<td>0.005</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.017)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>$(\alpha+\beta)\mu$</td>
<td>0.057$^e$</td>
<td>0.001$^b$</td>
<td>0.001$^b$</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.231$^b$</td>
<td>0.267$^b$</td>
<td>0.202$^b$</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.124)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.513$^c$</td>
<td>0.500$^c$</td>
<td>0.534$^c$</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.064)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.587</td>
<td>0.592</td>
<td>0.577</td>
</tr>
<tr>
<td>OBS</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
</tbody>
</table>

a  p-value of one-sided hypothesis test < 0.10
b  p-value of one-sided hypothesis test < 0.05
c  p-value of one-sided hypothesis test < 0.01

White’s heteroscedasticity-corrected standard errors in parentheses.

Table 4. Sample of 85 countries

<table>
<thead>
<tr>
<th>Algeria</th>
<th>Malawi</th>
<th>Colombia</th>
<th>Fiji</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt</td>
<td>Mali</td>
<td>Costa Rica</td>
<td>Indonesia</td>
<td>Greece</td>
</tr>
<tr>
<td>Jordan</td>
<td>Mauritius</td>
<td>Dominican Rep</td>
<td>Korea South</td>
<td>Ireland</td>
</tr>
<tr>
<td>Morocco</td>
<td>Mozambique</td>
<td>Ecuador</td>
<td>Malaysia</td>
<td>Italy</td>
</tr>
<tr>
<td>Syria</td>
<td>Niger</td>
<td>El Salvador</td>
<td>Philippines</td>
<td>Japan</td>
</tr>
<tr>
<td>Tunisia</td>
<td>Nigeria</td>
<td>Guatemala</td>
<td>Thailand</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Angola</td>
<td>Senegal</td>
<td>Guyana</td>
<td>Bangladesh</td>
<td>New Zealand</td>
</tr>
<tr>
<td>Benin</td>
<td>Sierra Leone</td>
<td>Haiti</td>
<td>India</td>
<td>Portugal</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>South Africa</td>
<td>Honduras</td>
<td>Nepal</td>
<td>Singapore</td>
</tr>
<tr>
<td>Burundi</td>
<td>Tanzania</td>
<td>Jamaica</td>
<td>Australia</td>
<td>Spain</td>
</tr>
<tr>
<td>Cameroon</td>
<td>Uganda</td>
<td>Mexico</td>
<td>Austria</td>
<td>Sweden</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Zimbabwe</td>
<td>Panama</td>
<td>Canada</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Gabon</td>
<td>Argentina</td>
<td>Paraguay</td>
<td>Cyprus</td>
<td>United States</td>
</tr>
<tr>
<td>Ghana</td>
<td>Bolivia</td>
<td>Peru</td>
<td>Denmark</td>
<td>Hungary</td>
</tr>
<tr>
<td>Kenya</td>
<td>Brazil</td>
<td>Uruguay</td>
<td>Finland</td>
<td>Romania</td>
</tr>
<tr>
<td>Madagascar</td>
<td>Chile</td>
<td>China</td>
<td>France</td>
<td>Turkey</td>
</tr>
</tbody>
</table>