Latent Instrumental Variable Methods for Endogenous Regressors in Hierarchical Bayes Models with Dirichlet Process Priors

Peter Ebbes\textsuperscript{a}, Peter Lenk\textsuperscript{b}, Michel Wedel\textsuperscript{b},

\textsuperscript{a} The Pennsylvania State University, University Park, PA 16802

\textsuperscript{b} University of Michigan, Ann Arbor, MI 48109

March 27, 2006

\textsuperscript{*}Corresponding author. E-mail: pebbes@psu.edu
Latent Instrumental Variable Methods for Endogenous Regressors in Hierarchical Bayes Models with Dirichlet Process Priors

Abstract

A well-known and widespread inferential problem with non-experimental data, especially in the social sciences and economics, occurs when in linear models predictor variables are correlated with the error terms of the model; that is, these predictor variables are endogenous. Ignoring the endogeneity of the explanatory variables results in biased and inconsistent estimates. Instrumental Variable (IV) methods correct the estimates through the use of exogenous variables, or “instruments,” that are correlated with the endogenous predictors and uncorrelated with the error terms of the dependent variable. However, the search for appropriate instruments in any given application may be difficult and contentious. In addition, if the instruments are weak or not truly exogenous, IV methods may be misleading. An alternative to searching for observed instruments is to infer latent instruments from the data. We propose using mixtures of Dirichlet processes as the prior for the distribution of these latent instruments. This method is a generalization of previous work on latent instrumental variables methods in at least three ways: (1) it does not impose restrictions on the distribution of the latent instruments; (2) the inference is exact, and (3) it easily extends to multilevel and limited dependent variable models. Applications to simulated data sets reveal good performance of the method across several distributions of the unobserved instrument. An empirical application reveals and corrects for endogeneity in a hierarchical regression model.

KEYWORDS: Instrumental Variables, Endogeneity, Hierarchical Bayes, Dirichlet Process, Latent Variables
1 Introduction

In making causal inferences from non-experimental data, Instrumental Variables (IV) methods are essential, standard tools in economics and related fields (Sargan 1958 and 1959). Explanatory variables may be endogenous because of omitted variables, co-evolution of the predictor the dependent variables, or measurement error. Ignoring predictor endogeneity results in biased and inconsistent estimators. This problem can be corrected by identifying exogenous variables, or “instruments,” that are correlated with the endogenous predictors and independent of the dependent variable’s error terms. IV methods use the instrument to partition the variance of the endogenous regressors into endogenous and exogenous components and employ only the latter component in OLS, likelihood, or Bayesian estimation of the causal effect (Bowden and Turkington 1984, Angrist, Imbens and Rubin 1996, Kleibergen and Zivot, 2003).

The problem presented by endogenous regressors can easily be seen in simple regression with one, endogenous, stochastic predictor: $Y = \beta_0 + \beta_1 X + \epsilon$. The model for the stochastic regressor is $X = \mu_X + \eta$. The error terms are normally distributed with mean 0 and $E(\epsilon^2) = \sigma^2_\epsilon, E(\eta^2) = \sigma^2_\eta,$ and $E(\epsilon \eta) = \sigma_{\epsilon \eta}$. Because of the correlation between the equations for $X$ and $Y$, the conditional mean and variance of $Y$ given $X = x$ become:

$$
\mu_{Y|X=x} = \left( \beta_0 - \frac{\sigma_{\epsilon \eta}}{\sigma^2_\eta} \mu_x \right) + \left( \beta_1 + \frac{\sigma_{\epsilon \eta}}{\sigma^2_\eta} \right) x \quad \text{and} \quad \sigma^2_{Y|X=x} = \sigma^2_\epsilon - \frac{\sigma^2_{\epsilon \eta}}{\sigma^2_\eta}.
$$

If the endogeneity of $X$ is ignored, standard procedures estimate the reduced-form parameters $\beta^*_0 = \beta_0 - \sigma_{\epsilon \eta} \mu_x / \sigma^2_\eta$ and $\beta^*_1 = \beta_1 + \sigma_{\epsilon \eta} / \sigma^2_\eta$ and not the structural parameters $\beta_0$ and $\beta_1$. In addition, mean-squared residuals estimate $\sigma^2_\epsilon - \sigma^2_{\epsilon \eta} / \sigma^2_\eta$ and not $\sigma^2_\epsilon$. The crux of the problem is that $\mu_x$ is constant. If it varied, then the structural parameters could be identified.

IV methods assume that a third variable, $Z$, is available such that $X$ and $Z$ are corre-
lated and $E(Z\epsilon) = 0$. $Z$ is called the “instrument” and is used to identify the structural parameters of the model. (See Judge et al. 1985 for an overview). The equation for $X$ becomes $X = \alpha_0 + \alpha_1 Z + \eta$ where the $\eta$ has the same properties as before. Thus, $X$ is partitioned into a random variable $\alpha_0 + \alpha_1 Z$ that is exogenous (uncorrelated with $\epsilon$) and a random variable $\eta$ that is endogenous (correlated with $\epsilon$). The exogenous part is used to estimate $\beta_0$ and $\beta_1$ by projecting $X$ onto $Z$. Essentially, IV methods use the predicted value $\hat{X}$ given $Z = z$ instead of $X$ as the regressor. These projections lead to the well-known, instrumental variables estimator: $\beta^{IV} = (D_X' H_Z D_X)^{-1} D_X' H_Z Y$, where $D_X$ is the design matrix for regressing $Y$ onto $X$; $H_Z$ is the hat-matrix $D_Z (D_Z' D_Z)^{-1} D_Z'$ for projecting $X$ onto $Z$; and $D_Z$ is the design matrix for regressing $X$ onto $Z$.

The search for appropriate instruments in any given situation may be nontrivial. In addition, if the instruments are weak (low correlation between $X$ and $Z$) or not truly exogenous (non-zero correlation between $Z$ and $\epsilon$), IV methods are inaccurate and can produce results that are potentially worse than those obtained by simply ignoring the original endogeneity problem (Bound, Jaeger and Baker, 1995; Stock, Wright and Yogo, 2002; Hahn and Hausman, 2002). Instruments that are plausibly exogenous often prove to be weak. Finding strong instruments can be difficult: the higher the correlation among the endogenous predictors and the instruments, the less defensible the claim that the instruments are uncorrelated with the model’s disturbances. In other cases, valid instruments may not be readily available, due to theoretical or practical limitations. This situation becomes amplified in hierarchical models (Lindley and Smith, 1972) where endogenous regressors may occur at different levels; a problem that does not seem to have received much attention in the literature.

The lack of strong instruments has been addressed in two research streams. One focuses on improving IV methods for weak instruments (e.g. Staiger and Stock, 1997; Bekker, 1994), and the other proposes adjustment methods that do not require observed or external
instruments (Madansky, 1959; Erickson and Whited, 2002; Lewbel, 1997; Rigobon, 2003; Ebbes et al., 2005). The latter methods that use internal IVs exploit characteristics, such as skewness and heteroscedasticity, of the endogenous regressors’ distributions to construct “latent” instrumental variables (LIV).

To continue with our simple example, a LIV formulation of the endogenous predictor is \( X = \theta + \eta \) where \( \theta \) is stochastic and uncorrelated with \( \epsilon \), and \( \eta \) has the same properties as before: \( E(\eta) = 0, E(\eta^2) = \sigma^2_\eta \), and \( E(\epsilon \eta) = \sigma_{\epsilon \eta} \). The random \( \theta \) has distribution \( G \). The regression function of \( Y \) given \( X \) now becomes:

\[
\mu_{Y|X=x} = \beta_0 - \frac{\sigma_{\epsilon \eta}}{\sigma^2_\eta} \theta + \left( \beta_1 + \frac{\sigma_{\epsilon \eta}}{\sigma^2_\eta} \right) x.
\]

Because \( \theta \) varies among sample units, it is possible to identify \( \sigma_{\epsilon \eta} \), since one is able to decompose \( X \) into exogenous and endogenous components. Clearly, if \( \theta \) and \( \eta \) both have normal distributions, they are not identified in the equation for \( X \), and \( \left( \sigma_{\epsilon \eta}/\sigma^2_\eta \right) (\theta - \mu_\theta) \) can be absorbed by the error term in the model for \( Y \). Therefore, in LIV methods, the distribution of \( \theta \) is assumed to be non-normal (mostly either skewed or multi-modal).

Although previous LIV methods effectively address the endogeneity problem in many situations, they too have limitations. First, these procedures assume that the distribution of the endogenous regressors convolute the exogenous, latent instruments with endogenous, random disturbances. In our simple example, the density of \( X \) is \( f(x) = \int h(x - \theta) dG(\theta) \) where \( G \) is the distribution of \( \theta \), and \( h \) is the density, often taken to be normal, of \( \eta \). Thus, the distribution of the latent instruments plays two roles: it determines the estimator of the structural parameters \( \beta \) and the distribution of the stochastic, endogenous regressors. The distributional assumptions are difficult to check reliably in practice and may not hold, and incorrectly specifying \( G \) may lead to suboptimal estimates of \( \beta \). Second, the small sample performance of standard IV and LIV inference varies widely in applications. Third,
existing methods are not easily generalizable to multi-level models or other extensions of
the linear model including (hierarchical) limited dependent variable models.

Bayesian inference with Markov chain Monte Carlo (MCMC) provides a simple
approach to convolutions by decomposing the mixture with the use of latent variables (Tan-
ner and Wong 1987). Instead of computing the convolution with numerical integration,
through MCMC one may generate \( \{\theta_i\} \), which de-convolutes the integral. In this article
we employ MCMC to estimate latent instruments that partition the endogenous regressors
into unobserved exogenous and endogenous random variables. This framework is particu-
larly powerful when coupled with mixtures of Dirichlet processes (MDP) (Ferguson 1973;
Antoniak 1974, Escobar 1994) as the prior for the latent instruments’ distributions, thereby
relaxing their specification. The MDP avoids specifying the distribution of \( \theta \) to satisfy both
the convolution for the distribution of \( X \) and the assumptions for the LIV method, which
may otherwise not be mutually compatible.

In our simple example, the MDP-LIV model is given by

\[
\begin{align*}
  y_i &= \beta_0 + \beta_1 x_i + \epsilon_i; \quad x_i = \theta_i + \eta_i; \quad \text{and} \quad (\epsilon_i, \eta_i) \overset{\sim}{\sim} N_2(0, \Sigma) \text{ for } i = 1, \ldots, n, \\
  \theta_i &\sim G \text{ and } G \sim DP(G_0, \rho) : \text{Dirichlet process with base measure } G_0 \text{ and precision } \rho, \\
  G_0 &= \text{Normal c.d.f. with mean } \mu_G \text{ and variance } \sigma^2_G.
\end{align*}
\]

The Dirichlet process prior for \( G \) is a probability distribution on the space of distributions
for the unobserved instrument (Escobar 1994; Escobar and West 1995, and Damien 2005).
The parameters of the Dirichlet process are the base measure \( G_0 \) that centers the process
and the precision parameter \( \rho > 0 \) that describes the strength of prior belief in \( G_0 \). For
illustrative purposes, we assume that \( G_0 \) is a univariate normal distribution with unknown
mean \( \mu_G \) and variance \( \sigma^2_G \), although other specifications are easily implemented. We use
conditionally conjugate priors for the other parameters. Section 2 presents the details for
Simulation results for the simple model indicate the potential of using MDP-LIV. We generate data with two distributions for the instrument: (1) a Bernouli distribution, (2) a Gamma distribution. \( \sigma_{\epsilon \eta} \) was taken to be either 0, 0.36, and 0.79, representing situations with no, moderate and severe endogeneity, respectively. We use a sample size of \( n = 1000 \). Data was generated for \( y_i, x_i, \) and \( \theta_i, i = 1, ..., n \), but the latter was excluded in the estimation, which allows us to investigate the performance of the proposed method if the instrument is not observed. We set \( \beta_0 = 1, \beta_1 = 2, \sigma^2_\epsilon = \sigma^2_\eta = 1 \), and the mean and variance of \( \theta_i \) equal to 0 and 1.5, respectively. We generate 15 data sets, discard the first 5000 iterations of the MCMC chain and use 20000 target draws, using every 10th draw for each data set. The MCMC chains are stationary after burn-in. The results are presented in Table 1. We present the mean and standard deviations of the posterior means of the parameters computed across the 15 simulated data sets, and also present OLS estimates for comparison.

—INSERT TABLE 1 HERE—

It follows from Table 1 that with the MDP-LIV model one effectively estimates the true structural parameters, while OLS, which estimates the reduced form parameters, is biased for the structural parameters, which occurs when \( \sigma_{\epsilon \eta} \neq 0 \). When \( x_i \) is truly exogenous, OLS is unbiased and slightly more efficient than the nonparametric Bayes model, as expected. For nonzero values of \( \sigma_{\epsilon \eta} \), OLS is clearly biased but the nonparametric model is not, regardless of the distribution of the latent instrument. The performance of the nonparametric Bayesian LIV procedure in these synthetic data sets is encouraging.

The next section extends the simple regression model with one endogenous predictor to multivariate regression with multiple endogenous variables. The Bayesian analysis of this model is key to applying the procedure to hierarchical regressions in Section 3, as well
as more complex models, such as hierarchial probit or Tobit models. Section 4 presents empirical examples. Section 5 concludes by discussing limitations and topics for future research.

2 Endogeneity in Multivariate Regression

The multivariate regression model, which is interesting in its own right (Drèze and Richard 1983), is the basis of the hierarchical Bayes (HB) regression model and many other models, such as the multinomial-probit model, which makes it of wider interest to develop solutions to the endogeneity problem for this model. Here we present the MDP-LIV for the multivariate regression model, which in applications is easily extended to these other models. Section 2.1 presents the multivariate model with endogenous regressors, and Section 2.2 details the full conditional distributions for MCMC estimation.

2.1 Multivariate Normal Model

A random sample of \( n \) multivariate observations with endogenous, stochastic regressors follow the model given by:

\[
y_i = \beta_0 + B'_1 x_{1i} + B'_2 x_{2i} + \epsilon_i \quad \text{for } i = 1, \ldots, n, \tag{1}
\]

\[
x_{1i} = \theta_i + \Phi' z_i + \eta_i \quad \text{for } i = 1, \ldots, n, \tag{2}
\]

\[
(\epsilon'_i, \eta'_i)' \sim N_{m+p_1}(0, \Sigma), \tag{3}
\]

\[
\Sigma = \begin{bmatrix}
\Sigma_{\epsilon} & \Sigma_{\epsilon \eta} \\
\Sigma_{\epsilon \eta} & \Sigma_{\eta}
\end{bmatrix}
\]

where \( E[\epsilon' \epsilon] = \Sigma_{\epsilon}; \ E[\epsilon' \eta] = \Sigma_{\epsilon \eta}; \) and \( E[\eta' \eta] = \Sigma_{\eta} \).

Equation (1) provides the model for the \( m \)-vector of dependent observations \( y_i \) where \( x_{1i} \) is a \( p_1 \)-vector of endogenous regressors, which are correlated with \( \epsilon_i \); and \( x_{2i} \) is a \( p_2 \)-vector
of exogenous regressors, which are independent of \((\epsilon_i, \eta_i)\). \(\beta_0\) is a \(m\)-vector of intercepts; \(B_1\) is a \(p_1 \times m\) matrix of coefficients, and \(B_2\) is a \(p_2 \times m\) matrix of coefficients. Equation (2) provides the model for the endogenous, stochastic regressors \(x_{1i}\), where \(\theta_i\) is a \(p_1\)-vector of latent instruments; \(z_i\) is a \(q\)-vector of exogenous variables independent of \((\epsilon_i, \eta_i)\); and \(\Phi\) is a \(q \times p_1\) matrix of regressors. Wooldridge (2002) recommends that the \(z_i\) include all exogenous regressors \(x_{2i}\) from Equation (1). The primary situation that we are considering is when \(z_i = x_{2i}\), and observed instruments for \(x_{1i}\) do not exist. However, our analysis also extends to the situation where observed instruments are available as well. In Equation (3), the error terms are jointly normally distributed with zero mean and covariance \(\Sigma\).

The latent instruments \(\theta_i\) are independent of \(\epsilon_i\) and \(\eta_i\), and \(\{\theta_i\}\) are a random sample from distribution \(G\). Distribution \(G\) is unknown and has a Dirichlet process prior \(DP(G_0, \rho)\). We will assume that the base measure \(G_0\) is a \(p_1\)-variate normal distribution with mean \(\mu_G\) and variances \(\Sigma_G\) although this assumption is not central to our analysis. The base measure is the prior expectation of \(G\). The precision parameter \(\rho\) expresses the prior certainty about the base measure.

Equations (1) and (2) can be written as the structural model:

\[
\begin{bmatrix}
    I_m & -B_1' \\
    0 & I_{p_1}
\end{bmatrix}
\begin{bmatrix}
    y_i \\
    x_{1i}
\end{bmatrix}
= 
\begin{bmatrix}
    \beta_0 \\
    \theta_i
\end{bmatrix}
+ \begin{bmatrix}
    B_2' \\
    0
\end{bmatrix}
\begin{bmatrix}
    0 \\
    \Phi
\end{bmatrix}
\begin{bmatrix}
    x_{2i} \\
    z_i
\end{bmatrix}
+ 
\begin{bmatrix}
    \epsilon_i \\
    \eta_i
\end{bmatrix},
\]

where \(I_m\) is the \(m \times m\) identity matrix. The reduced form of the model for \(w_i = (y_i', x_{1i}')'\) is obtained by substituting Equation (2) into Equation (1), effectively integrating \(x_{1i}\) out.
of the mean for $y_i$:

\[
    w_i = (y_i', x_i')' \sim N_{m+p1}(u_i^*, \Lambda), \tag{4}
\]

\[
    u_i^* = \begin{bmatrix}
        \beta_0 + B_1'(\theta_1 + \Phi'z_i) + B_2'x_{2i} \\
        \theta_i + \Phi'z_i
    \end{bmatrix},
\]

\[
    \Lambda = \begin{bmatrix}
        \Sigma_e + \Sigma_{\eta\eta} B_1 + B_1'\Sigma_{\eta\eta} B_1 & \Sigma_{\eta\eta} + B_1'\Sigma_{\eta}\ \\
        \Sigma_{\eta\eta} + \Sigma_{\eta} B_1 & \Sigma_{\eta}
    \end{bmatrix}. \tag{5}
\]

Another representation, which is useful in computing full conditional distributions, expresses the joint distribution as the conditional distribution of $Y$ given $X_1$ times the marginal of $X_1$:

\[
    w_i = (y_i', x_{1i}')' \sim N_m(\mu_{y|x_1}, \Sigma_{Y|X_1})N_{p1}(\theta_i + \Phi'z_i, \Sigma_\eta), \tag{6}
\]

\[
    \mu_{y|x_1} = \beta_0 + B_1'x_{1i} + B_2'x_{2i} + \Sigma_{\eta\eta}^{-1}(x_{1i} - \theta_i - \Phi'z_i),
\]

\[
    \Sigma_{Y|X_1} = \Sigma_e - \Sigma_{\eta\eta}^{-1}\Sigma_{\eta\eta}. \tag{7}
\]

Both of these representations (4) and (6) of the model are algebraically equivalent to expressing the joint density of $w_i = (y_i', x_{1i}')'$ as a quadratic form:

\[
    f(w_i) \propto \exp \left[ -\frac{1}{2}(w_i - u_i^*)'\Sigma^{-1}(w_i - u_i^*) \right], \tag{8}
\]

\[
    u_i^* = \begin{bmatrix}
        \beta_0 + B_1'x_{1i} + B_2'x_{2i} \\
        \theta_i + \Phi'z_i
    \end{bmatrix}.
\]

The superscript $s$ on $u_i^*$ indicates the structural form of the mean. Equation (8), although not the standard form for a normal density, is often more convenient to use in deriving full conditional distributions.
The models in Equation (1) and (2) can be compactly written as:

\[ Y = XB + E_Y, \quad (9) \]
\[ X_1 = \Theta + Z\Phi + E_X, \quad (10) \]

where \( Y \) is an \( n \times m \) matrix of dependent observations with \( y'_i \) in the \( i^{th} \) row. \( X = [\overline{1}_n \ X_1 \ X_2] \) where \( \overline{1}_n \) is an \( n \)-vector of ones; \( X_1 \) is an \( n \times p_1 \) matrix of endogenous observations with \( x'_{1i} \) in the \( i^{th} \) row; and \( X_2 \) is an \( n \times p_2 \) matrix of exogenous observations with \( x'_{2i} \) in the \( i^{th} \) row. \( X \) is an \( n \times p \) matrix where \( p = 1 + p_1 + p_2 \). The \( p \times m \) matrix of coefficients is \( B = [\beta_0 \ B_1' \ B_2']' \).

The \( n \times m \) matrix of error terms \( E_Y \) has \( \epsilon'_i \) in the \( i^{th} \) row. \( \Theta \) is the \( n \times p_1 \) matrix of latent instruments with \( \theta'_i \) in the \( i^{th} \) row; \( Z \) is the \( n \times q \) matrix of exogenous variables with \( z'_i \) in the \( i^{th} \) row; and \( E_X \) is a \( n \times p_1 \) random matrix with \( \eta'_i \) in the \( i^{th} \) row.

Before analyzing the full conditional distributions, we first establish conventions for matrix normal distributions. The vectorization \( \vec{A} \) of a matrix \( A \) stacks its rows into a vector:

\[ \vec{E}_Y = (\epsilon'_1, \ldots, \epsilon'_n)' \] is an \( mn \)-vector, and \( \vec{E}_X = (\eta'_1, \ldots, \eta'_n) \) is a \( np_1 \)-vector. The variance of a random matrix is defined to be the variance of its vectorization:

\[ \text{var}(E_Y) = \text{var}(\vec{E}_Y) = I_n \otimes \Sigma_{\epsilon} \quad \text{and} \quad \text{var}(E_X) = I_n \otimes \Sigma_{\eta} \] where \( I_n \) is a \( n \times n \) identity matrix, and \( \otimes \) is the Kronecker product operator. The covariance between \( E_Y \) and \( E_X \) is

\[ \text{cov}(E_Y, E_X) = I_n \otimes \Sigma_{\epsilon \eta}. \]

If \( W \) is a \( J \times K \) matrix of normal random variables with mean \( U \), a \( J \times K \) matrix, and variance \( \Psi \otimes \Upsilon \) where \( \Psi \) is a \( J \times J \) positive definite matrix, and \( \Upsilon \) is a \( K \times K \) positive definite matrix, then it follows a matrix normal distribution:

\[ W \sim N_{J \times K}(U, \Psi, \Upsilon), \]
\[ f(W) = (2\pi)^{-rac{JK}{2}} |\Psi|^{-rac{1}{2}} |\Upsilon|^{-rac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Upsilon^{-1}(W - U)'\Psi^{-1}(W - U) \right] \right\}. \]

The equivalent distribution for its vectorization is \( \vec{W} \sim N_{JK}(\vec{U}, \Psi \otimes \Upsilon). \)
Now, define \( W = [Y \ X_1] \), a \( n \times (m + p_1) \) matrix with \( w_i' \) in the \( i^{th} \) row. The matrix normal equivalent of Equation (4), which is obtained by substituting Equation (10) into Equation (9), is:

\[
W \sim N_{n \times (m + p_1)}(U^r, I_n, \Lambda),
\]

\[
U^r = \left\{ \bar{I}_n\beta_0' + (\Theta + Z\Phi)B_1 + X_2 B_2 \right\} (\Theta + Z\Phi),
\]

where the superscript “\( r \)” refers to the reduced form of the mean, and \( \Lambda \) is given in Equation (5). The conditional distribution of \( Y \) given \( X_1 \) is:

\[
Y|X_1 = N_{n \times m}(U_{Y|X_1}, I_n, \Sigma_{Y|X_1}),
\]

\[
U_{Y|X_1} = XB + (X_1 - \Theta - Z\Phi)\Sigma^{-1}_\eta \Sigma_{\eta\epsilon},
\]

where \( \Sigma_{Y|X_1} \) is given in Equation (7). Both of the representations in Equations (11) and (12) are algebraically equivalent to writing the joint density of \( W \) as a quadratic form, analogous to Equation (8):

\[
f(W) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma^{-1}(W - U^*)'(W - U^*) \right] \right\},
\]

\[
U^* = [XB \ (\Theta + Z\Phi)],
\]

where the superscript “\( s \)” refers to the structural form.
To complete the model specification, the prior distributions are:

\[ \tilde{B} \sim N_{mp}(u_{B,0}, V_{B,0}) \text{ and } \tilde{\Phi} \sim N_{p_1q}(u_{\Phi,0}, V_{\Phi,0}), \]

\[ \Sigma \sim IW_{m+p_1}(r_0, S_0) : \text{inverted Wishart with } r_0 \text{ degrees of freedom and scale } S_0, \]

\[ \theta_i \sim G \text{ and } G \sim DP(G_0, \rho) : \text{Dirichlet process with measure } G_0 \text{ and precision } \rho, \]

\[ G_0 = N_{p_1}(\mu_G, \Sigma_G); \mu_G \sim N_{p_1}(u_{G,0}, V_{G,0}) \text{ and } \Sigma_G \sim IW_{p_1}(r_{G,0}, S_{G,0}), \]

\[ \rho \sim \Gamma(a_0, b_0) : \text{Gamma distribution with shape } a_0 \text{ and scale } b_0. \]

### 2.2 Full Conditionals for MCMC

The full conditional distributions for the MCMC modify the standard computations of a multivariate regression model to account for endogeneity of a multivariate regression model.

1. **Full conditional of \( B \).** In the model, \( B \) only appears in Equation (9) for \( Y \). The most direct method of computing the full conditional of \( B \) is to combine the conditional distribution of \( Y \) given \( X_1 \), which is given in (12), with the prior for \( B \):

\[
\begin{align*}
\tilde{B} | \text{Rest} & \sim N_{mp}(u_{B,n}, V_{B,n}), \\
V_{B,n} & = \left[ (X'X \otimes \Sigma^{(e)}) + V_{B,0}^{-1} \right]^{-1}, \\
u_{B,n} & = V_{B,n} \left[ (X' \otimes \Sigma^{(e)}) \tilde{Y} + (X' \otimes \Sigma^{(eq)}) \tilde{R}_{X_1} + V_{B,0}^{-1} u_{B,0} \right], \\
R_{X_1} & = X_1 - \Theta - Z\Phi, \\
\Sigma^{(e)} & = \left( \Sigma_e - \Sigma_{eq} \Sigma_{q}^{-1} \Sigma_{qe} \right)^{-1} \text{ and } \Sigma^{(eq)} = -\Sigma^{(e)} \Sigma_{eq} \Sigma_{q}^{-1}. 
\end{align*}
\]

The matrices \( \Sigma^{(e)} \) and \( \Sigma^{(eq)} \) are blocks of \( \Sigma^{-1} \).

2. **Full conditional of \( \Phi \).** The most direct method to derive this distribution is to express
the density of $W$ in its quadratic form of Equation (13):

$$
\tilde{\Phi} | \text{Rest} \sim N_p(u_{\Phi,n}, V_{\Phi,n}),
$$

$$
V_{\Phi,n} = \left[ (Z'Z \otimes \Sigma^{(\eta)}) + V_0^{-1} \right]^{-1},
$$

$$
u_{\Phi,n} = V_{\Phi,n} \left[ (Z'Z \otimes \Sigma^{(\eta)}) \left( \bar{X}_n - \bar{\Theta} \right) + (Z' \otimes \Sigma^{(\eta)}) \bar{R}_Y + V_0^{-1} u_{0,0} \right],
$$

$$
R_Y = Y - XB,
$$

$$
\Sigma^{(\eta)} = \left( \Sigma - \Sigma_{\eta} \Sigma_{\epsilon}^{-1} \Sigma_{\epsilon \eta} \right)^{-1} \text{ and } \Sigma^{(\eta \epsilon)} = -\Sigma^{(\eta)} \Sigma_{\eta} \Sigma_{\epsilon}^{-1}.
$$

3. Full conditional of $\Sigma$. We use the quadratic form in Equation (13) to obtain:

$$
\Sigma | \text{Rest} \sim IW(r_n, S_n),
$$

$$
r_n = r_0 + n \text{ and } S_n = S_0 + (W - U^*)'(W - U^*).
$$

4. Full conditional posterior distribution of $\theta_i$. The algorithm of Escobar (1994) and Escobar and West (1995) recursively generates the latent $\theta_i$ by using their predictive distribution, under the Dirichlet prior, given the other values of $\theta_j$ for $j \neq i$. The current values of $\{\theta_i\}$ reduce to $\tilde{n} < n$ distinct values or clusters (Antoniak 1974).

We denote these $\tilde{n}$ distinct values of $\theta_i$ by $\hat{\theta}_s$, $s = 1, ..., \tilde{n}$. $n_s^\leq n - 1$ is the number of observations in ‘cluster’ $s$ with common value $\hat{\theta}_s$ without observation $i$, and $\tilde{n}$ is the number of distinct clusters when observation $i$ is removed. The full conditional distribution of $\theta_i$ given $\{\hat{\theta}_s\}$ is:

$$
\theta_i | \text{Rest} \sim q_{0(i)}h(\theta_i) + \sum_{l=1}^{\tilde{n}} n_l^{-} \bar{q}_{(i)} \delta_{\bar{\theta}_l}(\theta_i),
$$

where $\delta_{\bar{\theta}_l}(\theta_i) = 1$ if $\theta_i = \bar{\theta}_l$ and zero otherwise. $q_{0(i)}$ is proportional to $\rho$ times the joint (marginal) density of $w_i = (y_i', x_{1i}')'$ after integrating out $\theta_i$ with respect to $G_0$, 


and $h$ is the full conditional distribution of $\theta_i$. This conditional density $h$ of $\theta_i$ given $G_0$ is:

$$\theta_i \mid \text{Rest} \sim N_{p_1}(\mu_{\theta_i}, V_{\theta_i}),$$

$$V_{\theta_i} = [\Sigma^{(\eta)} + \Sigma_G^{-1}]^{-1},$$

$$\mu_{\theta_i} = V_{\theta_i} \left[ \Sigma^{(\eta)} (x_{1i} - \Phi' z_i) + \Sigma^{(\eta e)} (y_i - B' x_i) + \Sigma_G^{-1} \mu_G \right],$$

$$\Sigma^{(\eta)} = (\Sigma_\eta - \Sigma_{\eta e} \Sigma_e^{-1} \Sigma_{\eta e})^{-1} \text{ and } \Sigma^{(\eta e)} = -\Sigma^{(\eta)} \Sigma_{\eta e} \Sigma_e^{-1}.$$

The distribution of $w_i$ after integrating $\theta_i$ out with respect to $G_0$ is normal with mean:

$$\begin{bmatrix}
\beta_0 + B_1' \mu_G + B_1' \Phi' z_i + B_2' x_{2i} \\
\mu_G + \Phi' z_i
\end{bmatrix}$$

and variance:

$$\begin{bmatrix}
\Sigma_e + \Sigma_{en} B_1 + B_1' (\Sigma_\eta + \Sigma_G) B_1 & \Sigma_{en} + B_1' (\Sigma_\eta + \Sigma_G) \\
\Sigma_{en} + (\Sigma_\eta + \Sigma_G) B_1 & \Sigma_\eta + \Sigma_G
\end{bmatrix}.$$}

Also, $\tilde{q}_{l(i)}$ is proportional to the density of $w_i$ conditional on $\theta_i = \tilde{\theta}_i$; and $q_{0(i)} + \sum_{l=1}^{n_{-}} \tilde{n}_l \tilde{q}_{l(i)} = 1$, i.e. $q_{0(i)}$ and $n_{-} \tilde{q}_{l(i)}$ are normalized to 1.

In a remixing step (West, Müller and Escobar 1994) the $\tilde{\theta}_i$’s are updated by replacing the current values by new values that are drawn from a normal distribution with variance and mean:

$$V_{\tilde{\theta}_i} = [\tilde{n}_l \Sigma^{(\eta)} + \Sigma_G^{-1}]^{-1},$$

$$\mu_{\tilde{\theta}_i} = V_{\tilde{\theta}_i} \left[ \Sigma^{(\eta)} \sum_{i \in J_l} (x_{1i} - \Phi' z_i) + \Sigma^{(\eta e)} \sum_{i \in J_l} (y_i - B' x_i) + \Sigma_G^{-1} \mu_G \right].$$

13
Here, \( J_l \) denotes the set of observations for which \( \theta_i = \tilde{\theta}_l, \ l = 1, ..., \tilde{n} \).

5. Full conditional distribution of \( \mu_G \):

\[
\mu_G | \text{Rest} \sim N_p(u_{G,\tilde{n}}, V_{G,\tilde{n}}),
\]
\[
V_{G,\tilde{n}} = [\tilde{n}\Sigma_G^{-1} + V_{G,0}^{-1}]^{-1},
\]
\[
u_{G,\tilde{n}} = V_{G,\tilde{n}} \left[ \Sigma_G^{-1} \sum_{l=1}^{\tilde{n}} \tilde{\theta}_l + V_{G,0}^{-1} u_{G,0} \right].
\]

6. Full conditional of \( \Sigma_G \):

\[
\Sigma_G | \text{Rest} \sim IW_p(r_{G,\tilde{n}}, S_{G,\tilde{n}}),
\]
\[
r_{G,\tilde{n}} = r_{G,0} + \tilde{n},
\]
\[
S_{G,\tilde{n}} = S_{G,0} + \sum_{l=1}^{\tilde{n}} (\tilde{\theta}_l - \mu_G) (\tilde{\theta}_l - \mu_G)'.
\]

7. Full conditional of \( \rho \). Following Escobar and West (1995), we sample \( \rho \) in two steps:

\[
\zeta | \text{Rest} \sim \text{Beta}(\rho + 1, n),
\]
\[
\rho | \text{Rest} \sim \pi \Gamma (a_0 + \tilde{n}, b_0 - \log(\zeta)) + (1 - \pi) \Gamma (a_0 + \tilde{n} - 1, b_0 - \log(\zeta)),
\]
\[
\pi / (1 - \pi) = (a_0 + \tilde{n} - 1) / (n(b_0 - \log(\zeta))).
\]

3 Endogeneity in a hierarchical model

As an illustration of how the results in the previous section are applicable to a wider range of models, we next consider as an illustration a linear hierarchical model with random coefficients. The heterogeneity in the random effects is explained by covariates (Lindley and Smith 1972, Albert 1988). These covariates are, however, potentially endogeneous.
We consider the repeated measures situation where \( n_i \) observations are made on the \( i^{th} \) unit, and \( n \) units are in the study. Each dependent observation \( y_{ij} \) is a \( m \)-vector. The model that we use to address that problem is given by

\[
y_{ij} = B_i' x_{ij} + \epsilon_{ij} \text{ for } j = 1, \ldots, n_i; \ i = 1, \ldots, N, \tag{15}
\]

\[
\bar{B}_i = \phi_0 + \Phi_1' z_{1i} + \Phi_2' z_{2i} + \eta_i, \tag{16}
\]

\[
z_{1i} = \theta_i + \Psi' z_{3i} + \xi_i. \tag{17}
\]

Equation (15) is the unit–level model that describes the variation in the dependent variables for unit \( i \), and Equation (16) is the population model that describes the heterogeneity in the unit–level parameters \( B_i \). This parameter heterogeneity is decomposed into two parts: one part is related to observed variables \( z_{1i} \) and \( z_{2i} \), and the other part \( \eta_i \) is random or unexplained variation across the population. The regressors \( x_{ij} \) and \( z_{2i} \) are exogenous or independent of the error terms \( \epsilon_{ij} \), \( \eta_i \), and \( \xi_i \). The regressor \( z_{1i} \) is endogenous with respect to \( \eta_i \): \( E(z_{1i} \eta_i) \neq 0 \). In our framework, other sources of endogeneity, say between \( x_{ij} \) and \( \epsilon_{ij} \) are possible with the appropriate modification of the analysis in Section 2. However, we focus on endogeneity in the population model because it has not been addressed in the literature. Equation (17) is the latent instrumental variable equation for the endogenous regressor \( z_{1i} \).

To fix notation for the unit–level model in Equation (15), \( B_i \) is a \( p \times m \) random coefficient matrix; \( x_{ij} \) is a \( p \)-vector of regressor variables for the unit level model, and \( \epsilon_{ij} \) is \( N_m(0, \Sigma_\epsilon) \). In Equation (16), \( z_{1i} \) is a \( q_1 \)-vector of endogenous regressors, and \( z_{2i} \) is a \( q_2 \)-vector of exogenous regressors. \( \phi_0 \) is a \( mp \)-vector of intercepts; \( \Phi_1 \) is a \( q_1 \times mp \) matrix of coefficients; and \( \Phi_2 \) is a \( q_2 \times mp \) matrix of coefficients. The random terms \( \eta_i \) are \( N_{mp}(0, \Sigma_\eta) \). In Equation (17), \( \theta_i \) is a \( q_1 \)-vector of latent instruments; \( z_{3i} \) is a \( q_3 \)-vector of exogenous regressors that include \( z_{2i} \); \( \Psi \) is a \( q_3 \times q_1 \) matrix of coefficients; and the error terms \( \xi_i \) are \( N_{q_1}(0, \Sigma_\xi) \). We
will assume that \( E(\eta_i\xi'_i) = \Sigma_{\eta\xi} \) is nonzero, while \( E(\epsilon_i\eta'_i) = 0 \) and \( E(\epsilon_i\xi'_i) = 0 \). We use the unscripted \( \Sigma \) for the variance of \((\eta'_i, \xi'_i)\).

As in Section 2, the model can be expressed in matrix notation:

\[
\begin{align*}
Y_i &= X_iB_i + E_Y \text{ for } i = 1, \ldots, N \quad (18) \\
B &= Z\Phi + E_B \quad (19) \\
Z_1 &= \Theta + Z_3\Psi + E_Z \quad (20)
\end{align*}
\]

where \( Y_i \) is a \( n_i \times m \) matrix with \( y'_{ij} \) in the \( j \)th row; \( X_i \) is a \( n_i \times p \) with \( x'_{ij} \) in the \( j \)th row; and \( E_Y \) has \( e'_{ij} \) in the \( j \)th row. \( B \) is a \( N \times mp \) matrix with \( \vec{B}'_i \) in the \( i \)th row; \( Z = [\bar{1}_n Z_1 Z_2] \) and \( \Phi = [\phi_0 \Phi_1' \Phi_2']' \). \( Z \) is a \( N \times q \) matrix with \( q = 1 + q_1 + q_2 \); and \( \Phi \) is a \( q \times mp \) matrix. \( E_B \) has \( \eta'_i \) in the \( i \)th row. \( \Theta \) is a \( N \times q_1 \) matrix of latent instruments with \( \theta'_i \) in the \( i \)th row; \( Z_3 \) is a \( N \times q_3 \) matrix of exogenous regressors with \( z'_{3i} \) in the \( i \)th row; and \( E_Z \) is a \( N \times q_1 \) random matrix with \( \xi'_i \) in the \( i \)th row. As in Section 2, we define \( W = [B \ Z_1] \) a \( N \times (mp + q_1) \) matrix.

The prior specification for the model is:

\[
\begin{align*}
\Sigma_e &\sim IW_m(r_{r,0}, S_{r,0}) \text{ and } \Sigma \sim IW_{mp+q_1}(r_0, S_0), \\
\Phi &\sim N_{mpq}(u_{\Phi,0}, V_{\Phi,0}) \text{ and } \Psi \sim N_{q_1q_3}(u_{\Psi,0}, V_{\Psi,0}), \\
\theta_i &\sim DP(G_0, \rho); \ G_0 = N_{q_1}(\mu_G, \Sigma_G); \text{ and } \rho \sim \Gamma(a_0, b_0), \\
\mu_G &\sim N_{q_1}(u_{G,0}, V_{G,0}) \text{ and } \Sigma_G \sim IW_{q_1}(r_{G,0}, S_{G,0}).
\end{align*}
\]

The MDP-LIV analysis of the HB multivariate regression model follows from the multivariate regression model in Section 2 by identifying Equation (19) with Equation (9) and Equation (20) with Equation (10), after adding (18) for the observational equation. The prior specification in this section follows that in Section 2, with the appropriate change...
in notation. All posteriors take the forms described in the previous section, and the only remaining technical details are the full conditional distributions for $\vec{B}_i$ and $\Sigma_\epsilon$. The full conditional for $\vec{B}_i$ can obtained via the conditional distribution $\vec{B}_i|z_{1i}$ in (16) and (17). It follows that

$$\vec{B}_i|\text{Rest} \sim N_{mp}(u_{B_i}, V_{B_i}) \text{ for } i = 1, \ldots, N,$$

$$V_{B_i} = \left( (X_i'X_i \otimes \Sigma^{-1}_\epsilon) + \Sigma^{(\eta)} \right)^{-1},$$

$$u_{B_i} = V_{B_i} \left\{ (X_i' \otimes \Sigma^{-1}_\epsilon) \bar{Y}_i + \Sigma^{(\eta)} \Phi' z_i + \Sigma^{(\eta)} \Sigma_{\eta \xi} \Sigma^{-1}_\xi \left[ z_{1i} - (\theta_i + \Psi' z_i) \right] \right\},$$

$$\Sigma^{(\eta)} = (\Sigma_\eta - \Sigma_{\eta \xi} \Sigma^{-1}_\xi \Sigma_{\xi \eta})^{-1},$$

and

$$\Sigma_\epsilon|\text{Rest} \sim IW_m(r_{\epsilon,N}, S_{\epsilon,N}),$$

$$r_{\epsilon,N} = r_{\epsilon,0} + \sum_{i=1}^N n_i \text{ and } S_{\epsilon,N} = S_{\epsilon,0} + \sum_{i=1}^N (Y_i - X_iB_i)'(Y_i - X_iB_i),$$

which is the standard full conditional for the HB, multivariate normal regression problem.

## 4 Applications

We present empirical studies using the HB regression model with endogenous variables in the population level model that is presented in Section 3. We first demonstrate with synthetic data that the MDP prior for the latent instruments yields the correct estimates with and without endogeneity, then we apply the model to an empirical data set.
4.1 Synthetic Data sets

We generated 15 synthetic datasets according to a hierarchical model in Equations (15) to (17). We assume that the dependent $y_{ij}$’s are scalars ($m = 1$) and that Equation (15) is a simple regression ($p = 2$). In Equation (16), we assume the presence of one ($q_1 = 1$) endogenous covariate and one exogenous covariate ($q_2 = q_3 = 1$). The model specification for the simulated data is:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij} \text{ for } j = 1, \ldots, 15 \text{ and } i = 1, \ldots, 500,$$

$$\beta_0 = \phi_0 + \phi_0 z_{1i} + \phi_2 z_{2i} + \eta_{1i}; \quad \beta_1 = \phi_{10} + \phi_{11} z_{1i} + \phi_{12} z_{2i} + \eta_{2i} \text{ and }$$

$$z_{1i} = \theta_i + \psi_1 z_{2i} + \xi_i, \text{ where }$$

$$\phi_{00} = \phi_{01} = \phi_{11} = 1, \phi_{10} = 2, \phi_{02} = -\phi_{12} = 0.5, \text{ and } \psi_1 = -0.25$$

$$\text{var}(\epsilon_{ij}) = 2, \text{var}(\eta_{1i}) = \text{var}(\eta_{2i}) = 1 \text{ and } E(\eta_{1i}\eta_{2i}) = 0.5,$$

$$E(\eta_{1i}\xi_i) = E(\eta_{2i}\xi_i) = \sigma_{\eta \xi},$$

where each of $N = 500$ sampling units have $n_i = 15$ repeated measurements. $\{x_{ij}\}$ and $\{z_i\}$ are random samples from standard normal distributions. We consider three situations: no, moderate and severe endogeneity, with $\sigma_{\eta \xi}$ being 0, 0.36 and 0.79, respectively. We present the results for a (univariate) Bernoulli distribution of $\theta_i$ and a skewed Gamma distribution. We set the variance of $\theta_i$ equal to 1.5 and set its mean to zero. We discard the first 5000 iterations of the MCMC chain and use 20000 target draws, retaining every 10th draw for each data set. The MCMC chains are stationary after burn-in. The results are presented in Table 2. We present the mean and standard deviations of the posterior means computed across the 15 simulated data sets, focusing on the relevant estimates from the population level regressions, and also present results from a standard hierarchical Bayes model for comparison.
Three results are apparent. First, the MDP-LIV estimators and the standard HB estimates are essentially equal when $z_i$ is exogenous in Case A. Second, the estimates for $\phi_{00}$ and $\phi_{10}$, which are the intercepts in the models for $\beta_{0i}$ and $\beta_{1i}$, are nearly the same for the two methods, regardless of whether $z_i$ is endogenous or not. Third, as expected, when endogeneity is present, the standard HB estimates of $\phi_{10}$ and $\phi_{11}$, the slopes for $z_i$, are biased. In addition, the standard HB estimates of the $\text{var}(\eta_1)$ and $\text{var}(\eta_2)$ underestimate in the presence of an endogenous covariate, but not so for the MDP-LIV model.

It can be seen that the bias due to endogeneity in standard hierarchical Bayes model estimates of $\Phi$ and $\Sigma_\eta$ may be large, while for both distributions of the unobserved instrument the results of the MDP–LIV model are quite encouraging.

—INSERT TABLE 2 HERE—

4.2 Empirical data set

We next analyze data from a conjoint experiment on the market for new personal computers (Lenk et al. 1996). Conjoint analysis (Green and Rao 1971; Green and Srinivasan 1978 and 1990; Cattin and Wittink 1982) is a popular method in marketing research to measure subjects’ preferences for product and service attributes. It involves product alternatives constructed from pre-specified product characteristics at a number of discrete levels, using fractional factorial designs. The preference ratings of a sample of consumers of these hypothetical products (called product profiles) are commonly analyzed with hierarchical Bayes (HB) models to account for observed (through covariates) and unobserved differences in the effect parameters ($\beta_i$, called part-worths) of these product characteristics across individuals. The data set that we use pertains to 174 individuals that rated the likelihood of purchasing each of 20 profiles of personal computers on an 11-point scale (whereas Lenk et al. 1996, use 16 profiles, we use all 20 and removed four outliers from the data).
The twenty profiles are defined on 13 characteristics each having two levels, on an 11-point scale: (1) Telephone line (yes/no), (2) RAM (8/16Mb), (3) Screen Size (14”/17”), (4) CPU (50/100Hz), (5) Hard Disk (340/630Mb), (6) CDRom (no/yes), (7) Cache (128/256Kb), (8) Color (Beige/Black), (9) Availability (Mail/Store), (10) Warranty (1/3 year), (11) Software (no/yes), (12) Money Back Guarantee (no/30 days), (13) Price (2000/3500).

Information on six demographic and usage variables and one psychographic variable is available: Gender, Work Experience, Own/Lease, Professional, Software Applications, and Expertise. The latter psychographic variable, expertise, is constructed as the sum of two self-evaluations of knowledge about the microcomputer market, on five-point scales. It is likely to be correlated with the error term due to measurement error, which is expected to be pervasive in these measurements. In addition, omitted psychographic constructs, such as ability, may cause endogeneity of the expertise measure. Thus, we conjecture that the expertise-variable is correlated with the error term.

The following model was brought to these data:

\[ y_{ij} = x_{ij}'\beta_i + \epsilon_{ij}; \beta_i = \phi_0 + \phi_1 z_{1i} + \Phi_2 z_{2i} + \eta_i; \text{ and } z_{1i} = \theta_i + \Psi z_{2i} + \xi_i, \]

\[ \theta_i \sim G; G \sim DP (G_0, \rho); G_0 = N (\mu_G, \sigma_G^2); \text{ and } \rho \sim \Gamma(a_0, b_0), \]

\[ (\eta_i', \xi_i')' \sim N_{pq} (0, \Sigma) \text{ and } \epsilon_{ij} \sim N (0, \sigma^2_\epsilon), \]

with \( i = 1, \ldots, 174 \) and \( j = 1, \ldots, 20 \). \( y_{ij} \) is a scalar \( (m=1) \), \( x_{ij} \) a 13-vector of explanatory variables, and \( z_{1i} \) the potentially endogenous covariate expertise, and \( z_{2i} \) a 6 \times 1 vector with the exogenous covariates (which were standardized). We used a burn in of the MCMC chain of 15,000 iterations, use 50,000 target draws, and thin 1 in 50 draws. We focus on posterior inference for \( (\phi_0, \phi_1, \Phi_2) \), using standard conjugate flat priors as specified in the previous section, and a Normal baseline distribution \( G_0 \). We report the posterior means and standard deviations, and compare the estimates in Table 4 with those of a standard
hierarchical model in Table 3.

—INSERT TABLE 3 AND 4 HERE—

Our main conclusions are that in particular the effects of the unit-level covariates on the price coefficient are biased due to endogeneity. The price coefficients themselves ($\beta_{13i}$, not shown), however, are not biased, but the estimate of their variance is inflated in the standard HB model.

Table 4 shows that as compared to the HB model in Table 3, more estimates of $\Phi$ have at least 97.5 percent of their posterior distribution away from zero, when taking endogeneity into account (in particular “Years”, “Tech” and “Soft”). Respondents with more work-experience (“Years”) tend to find CPU less important, place more emphasis on warranty service and less on bundled productivity software. In addition, the computer professionals (“Tech”) and respondents that run more software applications (“Soft”) find memory more important and are more price sensitive. All coefficients reflecting the effect of individual level variables on price, except for that of years of work experience (“Years”), are larger in magnitude than in the standard HB model. In particular, it can be seen that respondents that consider themselves an expert (“Expert”) are much less price sensitive.

The variances of the distribution of the part-worths across individuals ($\text{diag}(\Sigma_\eta)$), also shown in Tables 3 and 4, are in general much smaller for the MDP-LIV model, ranging from 0.06 for “Distribution” to 0.25 for “Multimedia”, than for the HB model, where they range from 0.23 for “Distribution” to 0.41 for “Multimedia.” This does not hold for the price variable, however, for which the estimates are $\Sigma_{13,13} = 1.48$, respectively $\Sigma_{13,13} = 0.79$. The covariance between the error of price and the error of the endogenous regressor “Expert” is about $-0.77$ (0.20), which provides strong evidence of endogeneity.

For model-diagnostic purposes, we investigate the distributions of $\tilde{\eta}_i = \beta_i - (\phi + \Phi z_i)$, $i = 1, \ldots, n$, by computing the residuals for each MCMC draw, and computing skewness and
kurtosis across $i$. The skewness’ of the error distributions in the HB model range from -0.5 to 0.5, for the nonparametric IV model from -0.1 to 0.5. Estimates of kurtosis are in the 3-4 range. We do not consider these values alarming, especially because the distribution of the price-effect residuals seems to be close to normal for both the standard HB model and the nonparametric LIV model (skewness of -0.55 kurtosis 3.20, and skewness of -0.03, kurtosis 3.40, respectively).

5 Conclusion

We derived a nonparametric approach to problems of endogeneity in simple and hierarchical regression models. The approach alleviates problems of identifying observable instruments in such situations, as has been customary in the economics literature, as well as problems related to them being weak once they are identified. The proposed model introduces an unobserved instrument that allows one to partition the variance of the endogenous regressor in endogenous and exogenous components, similar to when using exogenous instruments. A Dirichlet process prior is placed on the space of distribution functions of the unobserved instrument, which renders the model robust to specific assumptions on its distribution. Synthetic data analysis for both the simple and hierarchical regression models shows that the approach performs satisfactorily. In an empirical application to conjoint data collected in a marketing context, endogeneity was shown to exist for hierarchical effects of a psychometric variable operating on the distribution of part-worths across individuals, and the estimates from the standard hierarchical model were shown to be strongly biased.

The MDP–LIV multivariate regression model and its analysis can be easily extended to a wide variety of models. In this paper, we illustrated such an extension to hierarchical Bayes regression with endogenous regressors in the population–level model for parameter heterogeneity. A useful extension for panel choice data prevalent in psychology, economics
and business, is the HB multinomial probit model where the observed dependent variables are 0/1 indicators. HB probit models are analyzed by linking the indicators $D_{ijk}$ for subject $i$, choice occasion $j$, and alternative $k$ to latent, multivariate normal random variables $Y_{ij}$ (Albert and Chib 1993, and McCulloch and Rossi 1994). The latent $Y$’s are generated from their full conditionals, which are truncated normal distributions with $y_{i,j,k^*} \geq y_{i,j,k}$ when alternative $k^*$ was selected over the other alternatives. Once the latent $Y$’s are generated, the analysis follows the hierarchial Bayes, MDP–LIV multivariate regression model provided above. The common theme is that if a part of any model corresponds to a multivariate regression and in applications the regressors maybe endogenous, then the techniques described in this paper can be used as the core of MCMC sampling schemes for correct identification of parameter values.

In spite of a larger computational burden imposed by MDP priors than standard IV methods, we therefore believe that our approach may have applications in areas of economics, sociology, psychology and business where endogeneity problems in a wide range of statistical models may have been prevalent, but have remained heretofore uninvestigated because of the lack of proper instruments or appropriate models.

6 References


Table 1: Linear model parameter estimates for synthetic data with bimodal and gamma distributions of the unobserved instrument, and three levels of endogeneity for $\sigma_{\epsilon \eta}$.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Endogeneity $\sigma_{\epsilon \eta}$</th>
<th>Method</th>
<th>$\beta_0 = 1$</th>
<th>$\beta_1 = 2$</th>
<th>$\sigma_{\epsilon \eta}^2 = 1$</th>
<th>$\sigma_{\epsilon \nu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bimodal</td>
<td>$\sigma_{\epsilon \eta} = 0$</td>
<td>OLS</td>
<td>1.00</td>
<td>1.99</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.041)</td>
<td>(0.020)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayes LIV</td>
<td>1.00</td>
<td>2.00</td>
<td>1.01</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.041)</td>
<td>(0.043)</td>
<td>(0.029)</td>
<td>(0.127)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{\epsilon \eta} = 0.36$</td>
<td>OLS</td>
<td>1.00</td>
<td>2.15</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.041)</td>
<td>(0.012)</td>
<td>(0.046)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayes LIV</td>
<td>1.00</td>
<td>2.01</td>
<td>1.00</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.040)</td>
<td>(0.030)</td>
<td>(0.050)</td>
<td>(0.074)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{\epsilon \eta} = 0.79$</td>
<td>OLS</td>
<td>1.00</td>
<td>2.31</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.027)</td>
<td>(0.015)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayes LIV</td>
<td>1.00</td>
<td>2.00</td>
<td>1.01</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.034)</td>
<td>(0.028)</td>
<td>(0.049)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Skewed</td>
<td>$\sigma_{\epsilon \eta} = 0$</td>
<td>OLS</td>
<td>1.00</td>
<td>2.00</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.028)</td>
<td>(0.023)</td>
<td>(0.053)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayes LIV</td>
<td>1.00</td>
<td>2.00</td>
<td>1.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.028)</td>
<td>(0.049)</td>
<td>(0.053)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{\epsilon \eta} = 0.36$</td>
<td>OLS</td>
<td>1.01</td>
<td>2.14</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.027)</td>
<td>(0.018)</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayes LIV</td>
<td>1.01</td>
<td>2.00</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.025)</td>
<td>(0.021)</td>
<td>(0.066)</td>
<td>(0.128)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{\epsilon \eta} = 0.79$</td>
<td>OLS</td>
<td>1.01</td>
<td>2.32</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
<td>(0.028)</td>
<td>(0.043)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayes LIV</td>
<td>1.02</td>
<td>1.99</td>
<td>1.01</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.030)</td>
<td>(0.026)</td>
<td>(0.082)</td>
<td>(0.075)</td>
</tr>
</tbody>
</table>
Table 2: Hierarchical model parameter estimates for synthetic data with bimodal and gamma distributions of the unobserved instrument, and three levels of endogeneity, $\sigma_{\eta\xi}$.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Parameter</th>
<th>True $\sigma_{\eta\xi} = 0$</th>
<th>A: $\sigma_{\eta\xi} = 0.36$</th>
<th>B: $\sigma_{\eta\xi} = 0.79$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bimodal</td>
<td>$\phi_{00} = 1$</td>
<td>1.01 (0.10)</td>
<td>1.01 (0.07)</td>
<td>1.01 (0.10)</td>
</tr>
<tr>
<td></td>
<td>$\phi_{10} = 2$</td>
<td>1.99 (0.08)</td>
<td>1.99 (0.07)</td>
<td>2.02 (0.08)</td>
</tr>
<tr>
<td></td>
<td>$\phi_{01} = 1$</td>
<td>0.97 (0.06)</td>
<td>1.00 (0.06)</td>
<td>0.99 (0.04)</td>
</tr>
<tr>
<td></td>
<td>$\phi_{11} = 1$</td>
<td>1.01 (0.06)</td>
<td>1.00 (0.06)</td>
<td>1.02 (0.04)</td>
</tr>
<tr>
<td></td>
<td>$\text{var}(\eta_1) = 1$</td>
<td>0.98 (0.06)</td>
<td>1.01 (0.08)</td>
<td>1.00 (0.03)</td>
</tr>
<tr>
<td></td>
<td>$\text{var}(\eta_2) = 1$</td>
<td>1.00 (0.07)</td>
<td>1.01 (0.07)</td>
<td>0.95 (0.07)</td>
</tr>
<tr>
<td></td>
<td>$E(\eta_1 \xi)$</td>
<td>0.04 (0.11)</td>
<td>0.34 (0.15)</td>
<td>0.70 (0.10)</td>
</tr>
<tr>
<td></td>
<td>$E(\eta_2 \xi)$</td>
<td>-0.02 (0.14)</td>
<td>0.37 (0.20)</td>
<td>0.68 (0.08)</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\phi_{00} = 1$</td>
<td>1.01 (0.10)</td>
<td>1.00 (0.09)</td>
<td>0.97 (0.11)</td>
</tr>
<tr>
<td></td>
<td>$\phi_{10} = 2$</td>
<td>1.99 (0.10)</td>
<td>1.98 (0.15)</td>
<td>1.94 (0.11)</td>
</tr>
<tr>
<td></td>
<td>$\phi_{01} = 1$</td>
<td>1.02 (0.07)</td>
<td>1.01 (0.05)</td>
<td>1.00 (0.04)</td>
</tr>
<tr>
<td></td>
<td>$\phi_{11} = 1$</td>
<td>0.98 (0.07)</td>
<td>1.00 (0.03)</td>
<td>1.01 (0.04)</td>
</tr>
<tr>
<td></td>
<td>$\text{var}(\eta_1) = 1$</td>
<td>1.01 (0.07)</td>
<td>1.01 (0.07)</td>
<td>0.98 (0.07)</td>
</tr>
<tr>
<td></td>
<td>$\text{var}(\eta_2) = 1$</td>
<td>0.97 (0.07)</td>
<td>1.03 (0.06)</td>
<td>0.98 (0.09)</td>
</tr>
<tr>
<td></td>
<td>$E(\eta_1 \xi)$</td>
<td>-0.02 (0.12)</td>
<td>0.30 (0.09)</td>
<td>0.71 (0.07)</td>
</tr>
<tr>
<td></td>
<td>$E(\eta_2 \xi)$</td>
<td>0.04 (0.12)</td>
<td>0.34 (0.07)</td>
<td>0.70 (0.07)</td>
</tr>
</tbody>
</table>
Table 3: Estimates of $\Phi$ from the standard HB model for the empirical application. Here bold indicate estimates for which zero is not in the 95% posterior interval; italics not in the 90% interval.

<table>
<thead>
<tr>
<th></th>
<th>CNST</th>
<th>HotLine</th>
<th>RAM</th>
<th>ScrnSz</th>
<th>CPU</th>
<th>HrdDsk</th>
<th>MultMd</th>
<th>Cache</th>
<th>Color</th>
<th>Dstrbtn</th>
<th>Wrrnty</th>
<th>Sftwr</th>
<th>Grntee</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNST</td>
<td>4.76</td>
<td>0.10</td>
<td>0.32</td>
<td>0.18</td>
<td>0.41</td>
<td>0.13</td>
<td>0.50</td>
<td>0.04</td>
<td>0.00</td>
<td>0.10</td>
<td>0.14</td>
<td>0.19</td>
<td>0.12</td>
<td>-1.17</td>
</tr>
<tr>
<td>(0.11)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.03</td>
<td>0.16</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.07</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.03</td>
<td>0.06</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>(0.11)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>YEARS</td>
<td>-0.27</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.09</td>
<td>-0.04</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.07</td>
<td>-0.08</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>(0.11)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>OWN</td>
<td>-0.04</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td>0.07</td>
<td>0.00</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>TECH</td>
<td>-0.12</td>
<td>-0.02</td>
<td>0.07</td>
<td>0.02</td>
<td>0.07</td>
<td>0.02</td>
<td>-0.03</td>
<td>0.00</td>
<td>-0.06</td>
<td>-0.05</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>SOFT</td>
<td>0.17</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>EXPERT</td>
<td>0.33</td>
<td>0.08</td>
<td>-0.10</td>
<td>0.03</td>
<td>0.07</td>
<td>0.05</td>
<td>-0.02</td>
<td>0.06</td>
<td>0.00</td>
<td>-0.04</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>(0.13)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{\eta\eta}$</td>
<td>1.88</td>
<td>0.31</td>
<td>0.28</td>
<td>0.26</td>
<td>0.35</td>
<td>0.26</td>
<td>0.41</td>
<td>0.29</td>
<td>0.23</td>
<td>0.23</td>
<td>0.25</td>
<td>0.27</td>
<td>0.29</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Table 4: Estimates of $\Phi$ for the DP-IV model for the empirical application. Here bold indicate estimates for which zero is not in the 95% posterior interval; italics not in the 90% interval.

<table>
<thead>
<tr>
<th></th>
<th>CNST</th>
<th>HotLine</th>
<th>RAM</th>
<th>ScrnSz</th>
<th>CPU</th>
<th>HrdDsk</th>
<th>MultMd</th>
<th>Cache</th>
<th>Color</th>
<th>Dstrbnn</th>
<th>Wrrnty</th>
<th>Sftwr</th>
<th>Grntee</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNST</td>
<td>4.77</td>
<td>0.10</td>
<td>0.32</td>
<td>0.18</td>
<td>0.41</td>
<td>0.13</td>
<td>0.50</td>
<td>0.05</td>
<td>0.00</td>
<td>0.10</td>
<td>0.14</td>
<td>0.19</td>
<td>0.12</td>
<td>-1.16</td>
</tr>
<tr>
<td>(0.11)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>0.03</td>
<td>0.16</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.42</td>
</tr>
<tr>
<td>(0.13)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.12)</td>
<td></td>
</tr>
<tr>
<td>YEARS</td>
<td>-0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.11</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.09</td>
<td>-0.08</td>
<td>0.05</td>
<td>-0.02</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.10)</td>
<td></td>
</tr>
<tr>
<td>OWN</td>
<td>-0.07</td>
<td>0.00</td>
<td>0.08</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td>0.06</td>
<td>0.01</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.16</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.10)</td>
<td></td>
</tr>
<tr>
<td>TECH</td>
<td>-0.17</td>
<td>-0.01</td>
<td>0.08</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.07</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>-0.02</td>
<td>-0.29</td>
</tr>
<tr>
<td>(0.13)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td>SOFT</td>
<td>0.12</td>
<td>0.00</td>
<td>0.09</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.25</td>
</tr>
<tr>
<td>(0.14)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.12)</td>
<td></td>
</tr>
<tr>
<td>EXPERT</td>
<td>0.54</td>
<td>0.08</td>
<td>-0.16</td>
<td>0.04</td>
<td>0.24</td>
<td>0.06</td>
<td>0.02</td>
<td>0.09</td>
<td>0.04</td>
<td>-0.02</td>
<td>-0.10</td>
<td>0.03</td>
<td>-0.01</td>
<td>1.08</td>
</tr>
<tr>
<td>(0.28)</td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.11)</td>
<td>(0.08)</td>
<td>(0.12)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.10)</td>
<td>(0.26)</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{\eta\eta}$</td>
<td>1.87</td>
<td>0.13</td>
<td>0.10</td>
<td>0.07</td>
<td>0.21</td>
<td>0.08</td>
<td>0.25</td>
<td>0.10</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
<td>0.11</td>
<td>1.48</td>
</tr>
</tbody>
</table>