

January 2013

Bayesian Estimation of Panel Data Fractional Response Models with Endogeneity: An Application to Standardized Test Rates

Lawrence Kessler

University of South Florida, lkess23@gmail.com

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Economics Commons](#), [Education Commons](#), and the [Statistics and Probability Commons](#)

Scholar Commons Citation

Kessler, Lawrence, "Bayesian Estimation of Panel Data Fractional Response Models with Endogeneity: An Application to Standardized Test Rates" (2013). *Graduate Theses and Dissertations*.
<http://scholarcommons.usf.edu/etd/4518>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Bayesian Estimation of Panel Data Fractional Response Models with Endogeneity:
An Application to Standardized Test Rates

by

Lawrence M. Kessler

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Economics
College of Arts and Sciences
University of South Florida

Major Professor: Murat Munkin, Ph.D.
Benjamin Craig, Ph.D.
Beom Lee, Ph.D.
Gabriel Picone, Ph.D.
Philip Porter, Ph.D.

Date of Approval:
March 7, 2013

Keywords: Bayesian Inference, Economics of Education, Fractional Probit,
Longitudinal Modeling, Instrumental Variables

Copyright © 2013, Lawrence M. Kessler

Acknowledgments

This paper would not have been possible without the support of numerous people, although it is only possible to name a particular few. First, I'd like to thank my advisor and mentor, Dr. Murat Munkin, for his excellent guidance, patience, and advice. Without his teachings none of these MCMC algorithms would exist for your reading pleasure. He also introduced me to a second mentor, Dr. Benjamin Craig, who has been extremely generous with his time and expertise and has been invaluable on both an academic and professional level, for which I am extremely grateful. I'd like to thank my committee members Dr. Benjamin Craig, Dr. Beom Lee, Dr. Murat Munkin, Dr. Gabriel Picone, and Dr. Philip Porter. They provided a tremendous amount of support, thoughtful criticisms, and helpful comments throughout this process.

I'd also like to thank my family and friends. To my parents, a one-page acknowledgement section is not nearly enough space to express my gratitude, but it is all that I have been allotted. Thank you for all of your love and support throughout my formidable years as well as the more recent (mildly) productive ones. I would not be the person I am today without your guidance and wisdom. To my sister, who is also a great friend, thank you for providing me with incredible insight and advice throughout my academic career. I am grateful for all of your help and encouragement. Finally to my wife-to-be, thank you for all of your patience, kindness, love, and support. Without you, I would not be getting married.

Contents

List of Tables	ii
List of Figures	iii
Abstract	iv
1 Introduction	1
2 Model specification under strict exogeneity (baseline model)	7
2.1 Prior distributions	11
2.2 Sampling from the posterior and estimation of marginal effects	13
2.3 Linear model with correlated random effects	18
3 Model specification with an endogenous explanatory variable (IV model)	20
3.1 Prior distributions	24
3.2 Sampling from the posterior and estimation of marginal effects	25
4 Application: the effect of school spending on student achievement	30
4.1 Literature	30
4.2 Data summary	34
4.2.1 Standardized test pass rates and state assigned school grades	38
4.2.2 Expenditures per pupil and instrumental variables	42
4.2.3 Additional control variables	45
4.3 Results	50
5 Model comparison	64
6 Conclusion	69
References	72
Appendices	85
Appendix 1: MCMC calculations for fractional probit baseline model	85
Appendix 2: MCMC steps for linear model with correlated random effects	91
Appendix 3: MCMC calculations for fractional probit IV model	93

List of Tables

1	Variable definitions and summary statistics (N = 1,138 schools; T = 7 years)	40
2	Pass rates on 4th grade reading FCAT. Percentiles, 1999-2005	42
3	Pass rates on 5th grade math FCAT. Percentiles, 1999-2005	42
4	FCAT scores (by state assigned school grades) 4th grade reading exam	43
5	FCAT scores (by state assigned school grade) 5th grade math exam .	43
6	School level expenditures per pupil in 1999 dollars. Percentiles, 1999-2005	46
7	Average expenditures per pupil by school grade and year in 1999 dollars	47
8	Total taxes levied per capita (district level) in 1999 dollars. Percentiles, 1999-2005	47
9	Posterior estimates for 4th grade reading. Baseline model assuming school expenditures is exogenously determined	53
10	Posterior estimates for 5th grade math. Baseline model assuming school expenditures is exogenously determined	54
11	Posterior estimates for 4th grade reading. Expenditure equation . . .	58
12	Posterior estimates for 4th grade reading. IV model with endogenous spending - structural equation	61

13	Posterior estimates for 5th grade math. IV model with endogenous spending - structural equation	62
14	Special education classification rates, 1997-2005	63

List of Figures

1	Posterior distribution of spending – linear model	56
2	Posterior distribution of spending – fractional probit model	56
3	Posterior distribution of spending – linear IV model	63
4	Posterior distribution of spending – fractional probit IV model	63
5	Posterior distribution of deltaUE – linear IV model	67
6	Posterior distribution of deltaUE – fractional probit IV model	67
7	Marginal effects (of school spending) at different percentiles of spending fractional probit models	68

Abstract

In this paper I propose Bayesian estimation of a nonlinear panel data model with a fractional dependent variable (bounded between 0 and 1). Specifically, I estimate a panel data fractional probit model which takes into account the bounded nature of the fractional response variable. I outline estimation under the assumption of strict exogeneity as well as when allowing for potential endogeneity. Furthermore, I illustrate how transitioning from the strictly exogenous case to the case of endogeneity only requires slight adjustments. For comparative purposes I also estimate linear specifications of these models and show how quantities of interest such as marginal effects can be calculated and compared across models. Using data from the state of Florida, I examine the relationship between school spending and student achievement, and find that increased spending has a positive and statistically significant effect on student achievement. Furthermore, this effect is roughly 50% larger in the model which allows for endogenous spending. Specifically, a \$1,000 increase in per-pupil spending is associated with an increase in standardized test pass rates ranging from 6.2-10.1%.

1 Introduction

This paper proposes Bayesian estimation of a panel data model with a fractional dependent variable (bounded between zero and one) and an endogenous explanatory variable. The model is used to analyze the relationship between public school spending and student achievement among Florida elementary schools from 1999 through 2005. Due to a wave of education reforms implemented in the late-90's such as the *A-plus Plan for Education* (A+ plan) and the *No Child Left Behind Act* (NCLB), school spending may be determined in part by student achievement, and therefore spending is modeled as potentially endogenous through the use of simultaneous equation modeling (SEM) and instrumental variables (IV).

The outcome variable of interest, student achievement, is measured as the proportion of each school's students passing Florida's standardized test, the FCAT (Florida Comprehensive Assessment Test). Since pass rates are bounded between zero and one, the model is presented as a nonlinear fractional response model. Traditionally, fractional response data has been handled using a linear probability model or a log-odds transformation, however, these specifications both have limitations. The log-odds model cannot handle y values equal to zero or one without the use of additional adjustments, and parameters of primary interest such as partial or marginal effects can be difficult to interpret (Wooldridge, 2002). While the linear probability

model assumes constant marginal effects, such that a one unit change in spending will always change pass rates by the same amount regardless of the initial level of school spending. If taken literally, this can lead to predicted pass rates of less than 0% or greater than 100%, which would not make sense. Furthermore, it seems more realistic to assume that, if spending has an effect on student achievement, this effect would likely diminish as spending increases. These limitations can be overcome by specifying a nonlinear fractional response model which will bound the relationship between spending and student achievement to the $(0, 1)$ interval and allow for diminishing marginal returns (Papke and Wooldridge, 1996, 2008). When working with panel data however, additional complexities arise due to the presence of unobserved heterogeneity.

Empirically, unobserved heterogeneity can be incorporated through a series of indicator variables (FE, fixed effects estimator) or as random variables fixed at the school level (RE, random effects estimator). However, consistency of the RE estimator hinges on a strong assumption of independence between the random effects and the covariates, which is often unrealistic, whereas nonlinear FE models generally suffer from an incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000).¹ Therefore, to control for unobserved heterogeneity in nonlinear models, a standard approach is to either place restrictions on the distribution of the unobserved effects or rely on a semiparametric approach. The main advantage of a semiparametric approach is that, by design, no restrictions need to be placed on the distribution of

¹One special circumstance is the conditional logit model (also known as fixed effects logit), in which the unobserved effects can be eliminated through the use of a conditional density. However, this procedure is not applicable when the outcome variable is fractional (Wooldridge, 2002).

the unobserved effects. However, a major limitation is that quantities of interest such as marginal effects and average partial effects generally cannot be identified.

As an alternative, I employ a correlated random effects approach (Mundlak, 1978; Chamberlain, 1982, 1984), in which dependence between the unobserved effects and explanatory variables is allowed but is restricted through a distributional assumption. This method is particularly attractive in the nonlinear case because it provides a simple way of avoiding the incidental parameters problem associated with the FE model while avoiding the strong assumption of independence in the RE model. Furthermore, by making this additional assumption on the unobserved effects, quantities of interest such as marginal effects can easily be identified (Altonji and Matzkin, 2005; Wooldridge, 2005).

Using the correlated random effects approach, I consider Bayesian estimation of a fractional response panel probit model, and provide an extension in order to allow for an endogenous explanatory variable. For comparative purposes I also estimate a linear specification with correlated random effects.² From a frequentist’s standpoint, the fractional response panel probit model was first introduced by Papke and Wooldridge (2008), who use the model to estimate the relationship between school spending and student achievement among Michigan elementary schools. Papke and Wooldridge initially assume that all explanatory variables are strictly exogenous, and estimate the model parameters using a “generalized estimating equation” (GEE) (Liang and Zeger, 1986)³ with the mean and variance of the fractional response variable and a

²In the linear case I consider the correlated random effects estimator instead of fixed effects because the linear fixed effects model cannot identify covariates which are time-invariant, and variables with little time-variation, while identifiable, are often estimated imprecisely.

³GEE is similar to, and asymptotically equivalent to weighted multivariate nonlinear least squares

“working correlation matrix” for GEE estimation. This approach does not rely on the joint likelihood distribution of the fractional response, which can be computationally burdensome to evaluate using frequentist methods. Once endogeneity is introduced, however, the GEE estimation method used by Papke and Wooldridge is no longer consistent. Therefore, as an alternative, they estimate a less-efficient two-step “pooled fractional probit quasi maximum likelihood estimation” (QMLE) method,⁴ and afterwards they must employ simulation methods (bootstrapping in particular) in order to obtain asymptotic standard errors, adjusted for their two-stage approach.

In recent years the fractional probit model with endogeneity has been applied in a variety of panel data settings. A few examples include Hanna (2010) who examined whether multinational firms increased production overseas in response to heavier environmental regulations imposed domestically. Nguyen (2010), who analyzed the effect of fertility on the female labor supply as measured by the fraction of hours worked per week. Gardeazabal (2010), who examined the effect of economic fluctuations on political party vote shares in Spanish general elections, and, McCabe and Snyder (2011) who examined whether online access to academic journals increased the citation rates of published articles.

In contrast to the Frequentist methods used in the aforementioned articles, my approach is a likelihood based estimation method which becomes feasible through (WMNLS). However, in many cases it is not possible to find the true variance matrix $var(y_i | x_i)$, which is needed for WMNLS. Therefore, in the GEE procedure a “working” version of $var(y_i | x_i)$ is specified based on distributional assumptions (see Imbens & Wooldridge, 2007; and Papke & Wooldridge, 2008).

⁴In step one, Papke and Wooldridge use a control function to estimate the equation for endogenous school spending (the endogenous explanatory variable). Then in step two they obtain the residuals from the spending equation and plug them into the main outcome equation to estimate the effect of endogenous school spending on student achievement (using the pooled probit QMLE method).

the use of a Bayesian data augmentation technique. This allows me to work directly with the joint likelihood distribution of the fractional response, and create a fully efficient estimation method in which all parameters and standard errors can be estimated simultaneously through the use of Bayesian Markov Chain Monte Carlo (MCMC) simulation methods. Using this approach, likelihood based estimation can be performed in the case of strict exogeneity as well as in the case of an endogenous explanatory variable, and transitioning from one model to the other only requires slight adjustments, which are rather straightforward. Conversely, in the Frequentist framework, this transition from a strictly exogenous model to one which allows for endogeneity requires a more drastic change in estimation methods, which nevertheless remains inefficient.

A secondary motivation for the Bayesian approach proposed here is from a policy perspective, as quantities of interest such as marginal effects can be calculated directly and can also be interpreted as probabilities, whereas those calculated in the frequentist's framework can only be identified up to a scale factor (see Imbens and Wooldridge, 2007; Papke and Wooldridge, 2008).

The analysis is related to several previous Bayesian contributions, including Albert and Chib (1993), Li (1998), Chib and Carlin (1999), and Bacolod and Tobias (2006). Albert and Chib (1993) introduce a Bayesian treatment of the discrete binary response model using the data augmentation method. Then Li (1998) uses this data augmentation technique to estimate a simultaneous equation model with a limited dependent variable and an endogenous explanatory variable. Chib and Carlin (1999) extend these methods to the longitudinal setting with the panel probit MCMC algo-

rithm, and finally, Bacolod and Tobias (2006) employ Bayesian panel data methods in order to analyze the relationship between school inputs and student achievement. In this paper I provide further extensions to the panel probit model in order to allow for a fractional outcome variable as well as a continuous endogenous covariate.

The remainder of the paper is organized as follows: In the next section I define the fractional response panel probit model under strict exogeneity, which I refer to as the baseline model. A Bayesian estimation method is then presented using MCMC simulation methods, and the calculation of marginal effects is addressed. In Section 3 a respecification of the model is proposed in order to allow for an endogenous explanatory variable, in which identification of the structural parameters of interest is performed through the use of an instrumental variables (IV) technique and simultaneous equation modeling (SEM). In Section 4 the proposed algorithms are used to analyze the relationship between school spending and student achievement among Florida elementary schools. In Section 5 I outline specification tests which can be implemented using Bayes factors as the criteria for model comparison. Finally, in Section 6 I conclude with a summary and brief discussion.

2 Model specification under strict exogeneity (baseline model)

Unlike in the case of the frequentist approach, introduction of endogeneity in the Bayesian approach leads to an estimation method which is just a slight modification of that for the strictly exogenous model. To simplify the exposition of the model, and estimation, I start with the baseline model under strict exogeneity. Assuming a probit specification, I express the conditional mean of the fractional response as

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{g}_i, c_i) = \Phi(c_i + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{g}_i\boldsymbol{\phi}), \quad (1)$$

where there are N independent institutions observed over T periods, such that i and t index institutions and time respectively, and the sample consists of NT observations; $\Phi(\cdot)$ represents the standard normal cumulative distribution function; y_{it} is an outcome variable, $0 \leq y_{it} \leq 1$; c_i represents time-constant unobserved difference across institutions; \mathbf{x}_{it} denotes a $1 \times k$ vector of explanatory variables which vary across institutions and time; and \mathbf{g}_i denotes a $1 \times h$ vector of time-invariant regressors.⁵

Following Chamberlain (1982, 1984), I formalize a relationship between the indi-

⁵The probit function, $\Phi(\cdot)$, is specified because we will be making use of the normal distribution to allow for correlation between c_i and \mathbf{x}_{it} in (1). Therefore, specifying a probit function which also makes use of the normal distribution is a convenient choice. Alternatively, one could choose to specify the logistic function $\Lambda(\cdot)$, however, computationally this would be more demanding.

vidual effects and time-varying explanatory variables such that c_i is a function of all lagged, present, and future values of \mathbf{x}_{it} :

$$\begin{aligned} c_i &= \psi + \mathbf{x}_{i1}\boldsymbol{\lambda}_1 + \mathbf{x}_{i2}\boldsymbol{\lambda}_2 + \dots + \mathbf{x}_{iT}\boldsymbol{\lambda}_T + a_i, \\ a_i &| \quad \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT} \sim N(0, \sigma_a^2), \end{aligned} \tag{2}$$

where ψ is an intercept term, $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_T$ are $k \times 1$ parameter vectors, and a_i is a normally distributed error term with zero mean and conditional variance σ_a^2 .

Combining equations (1) and (2) yields

$$E(y_{it} | \mathbf{x}_{it}, \mathbf{g}_i, a_i) = \Phi(\psi + \mathbf{x}_i\boldsymbol{\lambda} + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{g}_i\boldsymbol{\phi} + a_i), \tag{3}$$

where $\mathbf{x}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}]$, and $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_T]$.

To simplify the notation I denote $\mathbf{W}_{it} = [1, \mathbf{x}_i, \mathbf{x}_{it}, \mathbf{g}_i]$ as a vector of all observable data, $\boldsymbol{\Omega} = [\psi, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\phi}]$ as a parameter vector, and $\boldsymbol{\iota}_T$ as a T -vector of ones. Then for institution i at all time periods (3) can be written as

$$E(\mathbf{y}_i | \mathbf{W}_i, a_i) = \Phi(\mathbf{W}_i\boldsymbol{\Omega} + \boldsymbol{\iota}_T a_i). \tag{4}$$

The model above is specified in a semiparametric way such that the conditional mean of y_{it} is defined, but the distribution of y_{it} is not. However, in this Bayesian approach a set of augmented data is introduced in the parameter set which makes it possible to obtain fully efficient likelihood based estimates for the parameters in (3), even though the distribution of y_{it} is not explicitly specified.

The augmented data is created by introducing a dummy vector \mathbf{d}_{it} ($S \times 1$) for each observation in the sample. These dummy vectors are constructed such that each element, d_{its} ($s = 1, \dots, S$), takes a value of either one or zero such that the proportion of ones in \mathbf{d}_{it} is equal to the outcome variable, y_{it} (the proportion of students passing the FCAT exam at school i time t). In the application section that follows all FCAT scores were rounded to the nearest hundredth by the FLDOE, therefore, in such a construction $S = 100$.⁶ Thus, vector \mathbf{d}_{it} is treated as fully observed for each observation, and by construction each \mathbf{d}_{it} vector consists of ones and zeros such that the first $100 \times y_{it}$ elements are ones and the following $100 \times (1 - y_{it})$ elements are zeros.

The dummy vector \mathbf{d}_{it} is defined by S latent random normal variables:

$$\begin{aligned} y_{its}^* &= \mathbf{W}_{it}\boldsymbol{\Omega} + a_i + u_{its}, \\ u_{its} &\sim N(0, 1), \end{aligned} \tag{5}$$

such that

$$\begin{aligned} d_{its} &= 1, \text{ if } y_{its}^* = \mathbf{W}_{it}\boldsymbol{\Omega} + a_i + u_{its} > 0, \\ d_{its} &= 0, \text{ if } y_{its}^* = \mathbf{W}_{it}\boldsymbol{\Omega} + a_i + u_{its} \leq 0. \end{aligned}$$

This process defines the random variable d_{its} , for which the likelihood function can

⁶As an example, if $y_{it} = 0.73$, indicating that 73% of the students in school i time t passed the FCAT exam, then the first 73 ($s = 1, \dots, 73$) elements of d_{it} will be equal to one and the remaining 27 ($s = 74, \dots, 100$) elements will be equal to zero.

be written as

$$\Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i)^{d_{its}} [1 - \Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i)]^{1-d_{its}},$$

such that

$$E(d_{its}|\mathbf{W}_{it}, a_i) = \Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i).$$

The joint likelihood for the entire \mathbf{d}_{it} is

$$\Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i)^{\sum_{s=1}^S d_{its}} [1 - \Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i)]^{\sum_{s=1}^S (1-d_{its})}. \quad (6)$$

This likelihood will produce the same point estimates as the more familiar probit likelihood function:

$$\Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i)^{y_{it}} [1 - \Phi(\mathbf{W}_{it}\boldsymbol{\Omega} + a_i)]^{(1-y_{it})},$$

since it is a monotonic retransformation of (6). Specifically, raising (6) to the power of $\frac{1}{S}$ results in the probit likelihood function but with fractional response variable y_{it} , $0 \leq y_{it} \leq 1$, rather than a binary outcome. Thus, this allows me to specify a fully parametric model with data augmentation in which the moment condition, (4), is satisfied since by construction

$$y_{it} = \frac{1}{S} \sum_{s=1}^S d_{its}.$$

Using the data augmentation technique, I include the augmented data, y_{its}^* , directly into the likelihood function (Tanner and Wong, 1987; Albert and Chib, 1993), and the augmented data density for observation i, t can be written as

$$\begin{aligned}
 p(\mathbf{d}_{it}, y_{it}^* \mid \boldsymbol{\Omega}, a_i, \mathbf{W}_{it}) = & \\
 \frac{1}{\sqrt{2\pi}} \exp \left(-.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)] \right) \times & \\
 \sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)], & \quad (7)
 \end{aligned}$$

where I is simply an indicator function which takes the value 1 if the statement in the parenthesis is true and 0 otherwise.

2.1 Prior distributions

In Bayesian statistics, inference is made based upon a posterior distribution formed by combining information provided by data and prior knowledge about the parameters of interest. The data are summarized in terms of the likelihood function or data density, while the prior, which can be viewed as information provided by specialists or findings from previous research is incorporated through a probability density function.

Bayesian estimation follows by first assigning prior distributions to all parameters in the model. However, in many instances there is no reliable prior information available. In this case one can proceed by using flat or “diffuse” priors so that, relative to the likelihood function, the prior contributes very little information to the posterior. This enables the posterior distribution to be dominated by the data (i.e.

likelihood function), which Gelman et al. (1995) explain allows ‘the data to speak for itself.’⁷

The parameters ψ , $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}$ are all assigned commonly used conjugate-normal priors:

$$\psi \sim N(\underline{\psi}, \underline{\mathbf{H}}_{\psi}^{-1}), \quad \boldsymbol{\lambda} \sim N(\underline{\boldsymbol{\lambda}}, \underline{\mathbf{H}}_{\boldsymbol{\lambda}}^{-1}), \quad \boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1}), \quad \boldsymbol{\phi} \sim N(\underline{\boldsymbol{\phi}}, \underline{\mathbf{H}}_{\boldsymbol{\phi}}^{-1}),$$

which are centered at zero mean and made diffuse by choosing a large variance equal to 10. Following Chib and Carlin (1999), these parameters are then drawn together in one block as

$$\boldsymbol{\Omega} = [\psi, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\phi}] \sim N(\underline{\boldsymbol{\Omega}}, \underline{\mathbf{H}}_{\boldsymbol{\Omega}}).$$

where the variance term, $\underline{\mathbf{H}}_{\boldsymbol{\Omega}} = 10\mathbf{I}_{1+T+k+k+h}$. The second stage variance parameter is also assigned a commonly used conjugate-inverse gamma prior:

$$\sigma_a^2 \sim IG(a_a, b_a),$$

where the hyperparameters a_a and $1/b_a$ represent the shape and scale parameters.⁸

In hierarchical models care must be taken in choosing values for the hyperparam-

⁷Gelman et al. (2004) also note that when using a diffuse prior, the mean of the posterior distribution will be a weighted average of the prior mean values and the standard maximum likelihood estimates.

⁸Such that the mean equals $\frac{1/b_a}{a_a-1}$ and variance is $\frac{1/b_a^2}{(a_a-1)^2(a_a-2)}$.

ters, as noninformative second stage priors can lead to improper posteriors (Carlin, 1996; Hobert and Casella, 1996). Furthermore, Chib and Carlin (1999) show that if the variance prior is “overly vague” the Markov chain will likely suffer from slow convergence. To avoid this, the hyperparameters are set to $a_a = 3$ and $b_a = 0.025$, so that the mean and standard deviation are both equal to 20, a proper, but rather vague prior specification (see Koop, Poirier, and Tobias, 2007).

2.2 Sampling from the posterior and estimation of marginal effects

Let Δ equal a vector of all parameters in the model such that $\Delta = (\Omega, a_i, \sigma_a^2)$. Then the augmented joint posterior density, which is proportional to the product of the augmented data density (7) and the prior distributions of the parameters Ω , a_i , and σ_a^2 can be written as

$$\begin{aligned}
p(\Delta, \mathbf{y}^* | \mathbf{y}, \mathbf{W}) &\propto \\
&\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-0.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\Omega - a_i)' (y_{its}^* - \mathbf{W}_{it}\Omega - a_i) \right] \times \\
&\prod_{i=1}^N \prod_{t=1}^T \left[\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right] \times \\
&(2\pi)^{-(1+Tk+k+h)/2} |\underline{\mathbf{H}}_\Omega|^{1/2} \exp [-0.5(\Omega - \underline{\Omega})'\underline{\mathbf{H}}_\Omega(\Omega - \underline{\Omega})] \times \\
&\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp (-0.5a_i'\sigma_a^{-2}a_i) \times \frac{1}{\Gamma(a_a)b_a^{a_a}} (\sigma_a^2)^{-(a_a+1)} \exp \left(-\frac{1}{b_a\sigma_a^2} \right). \tag{8}
\end{aligned}$$

The model parameters can then be estimated via Gibbs sampling (Geman and Geman, 1984). The basic idea behind the Gibbs sampler is to partition the joint posterior into smaller blocks known as *full conditional posterior densities*. If analytically tractable, these full conditional densities can then be drawn from directly, from which successive and repeated draws will create an ergodic Markov chain – a sequence of draws which eventually converges to some target density.⁹ After a number of iterations, this joint sequence will converge to the joint posterior of interest (8). Functions of the posterior such as the mean and variance can then be estimated based on the simulated draws, and will satisfy a central-limit theorem as the length of the simulation tends to infinity (Chib and Greenberg, 1996; Gamerman and Lopes, 2006). The parameter estimates are then based on the following full conditional distributions (calculations are presented in Appendix 1):

1. The conditional posterior kernel for the latent variable y_{its}^* is normally distributed as $y_{its}^* \mid y_{it}, \mathbf{W}_{it}, a_i, \boldsymbol{\Omega}, \sigma_a^2 \sim N[\mathbf{W}_{it}\boldsymbol{\Omega} + a_i, 1]$, and is truncated at zero such that

$$y_{its}^* > 0 \text{ if } d_{its} = 1,$$

$$y_{its}^* \leq 0 \text{ if } d_{its} = 0.$$

⁹In terms of Markov Chains, ergodicity is equivalent to the strong law of large numbers (Gill, 2002). It states that if $\pi(\theta)$ is the target distribution and θ_i and θ_j are random draws from the chain such that $p(\theta_i, \theta_j)$ measures the probability that the chain will move from θ_i to θ_j then $\lim_{n \rightarrow \infty} p^n(\theta_i, \theta_j) = \pi(\theta_j)$.

2. The full conditional density for $a_i \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, \boldsymbol{\Omega}, \sigma_a^2$ is normally distributed as $a_i \sim N \left[\bar{a}_i, \bar{H}_a^{-1} \right]$, where

$$\begin{aligned} \bar{H}_a &= S \times T + \sigma_a^{-2} \\ \bar{a}_i &= \bar{H}_a^{-1} \left[\sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega}) \right) \right]. \end{aligned}$$

3. The full joint conditional density of block $\boldsymbol{\Omega} = [\boldsymbol{\psi}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\phi}]$ is normally distributed as $\boldsymbol{\Omega} \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, a_i, \sigma_a^2 \sim N \left[\bar{\boldsymbol{\Omega}}, \bar{\mathbf{H}}_{\boldsymbol{\Omega}}^{-1} \right]$, where

$$\begin{aligned} \bar{\mathbf{H}}_{\boldsymbol{\Omega}} &= \underline{\mathbf{H}}_{\boldsymbol{\Omega}} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}_{it}' \mathbf{W}_{it} \\ \bar{\boldsymbol{\Omega}} &= \bar{\mathbf{H}}_{\boldsymbol{\Omega}}^{-1} \left[\underline{\mathbf{H}}_{\boldsymbol{\Omega}} \boldsymbol{\Omega} + \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}_{it}' \left(\sum_{s=1}^S (y_{its}^* - a_i) \right) \right]. \end{aligned}$$

4. Finally, the full conditional density of the variance parameter σ_a^2 is inverse gamma, i.e.

$$\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, \boldsymbol{\Omega}, \sigma_a^2 \sim IG \left(\frac{N}{2} + a_a, \left[b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a_i' a_i \right]^{-1} \right).$$

Steps of the MCMC algorithm are as follows:

Algorithm 1

1. Sample $y_{its}^{*(1)}$ from $p(y_{its}^* | y_{it}, \mathbf{W}_{it}, a_i^{(0)}, \boldsymbol{\Omega}^{(0)}, \sigma_a^{2(0)})$
2. Sample $a_i^{(1)}$ from $p(a_i | y_{it}, \mathbf{W}_{it}, y_{its}^{*(1)}, \boldsymbol{\Omega}^{(0)}, \sigma_a^{2(0)})$
3. Sample $\boldsymbol{\Omega}^{(1)}$ from $p(\boldsymbol{\Omega} | y_{it}, \mathbf{W}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \sigma_a^{2(0)})$
 where $\boldsymbol{\Omega}^{(1)} = [\psi^{(1)}, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\phi}^{(1)}]$
4. Sample $\sigma_a^{2(1)}$ from $p(\sigma_a^2 | y_{it}, \mathbf{W}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)})$
5. Repeat steps 1-4 R times and at each step update the conditioning variables with their most recent values.

The Gibbs sampling process begins by assigning initial values to the parameters of interest $(a_i^{(0)}, \boldsymbol{\Omega}^{(0)}, \sigma_a^{2(0)})$, where the superscripts represent the current iteration step. These initial values can be drawn from their corresponding prior distributions. Each parameter is then sampled successively from their respective conditional distributions, and at each step the parameter values are updated. The first iteration is completed after sampling from all four conditional densities. After this process is repeated M times, convergence to the target distribution will take place and the subsequent $R - M$ draws will come directly from the joint posterior of interest (8).

The model parameters were all assigned conjugate priors to ensure that each full conditional (steps 1-4) is of known form and can be directly drawn from. For instance the conditional posterior of $\boldsymbol{\Omega}$ will be normally distributed, while the conditional posterior of σ_a^2 will be inverse gamma.

An added complication of the probit model (along with most nonlinear models) is that the estimated slope coefficients, $\mathbf{\Omega}$, do not have a direct interpretation. Rather they only indicate the sign (positive or negative) and statistical significance of the estimated effect. To estimate the magnitude one must obtain the marginal effects, which for a continuous explanatory variable, Ω_j , can be calculated as

$$\frac{\partial E(y_{it} | \mathbf{W}_{it}, a_i)}{\partial \mathbf{W}_j} = \Omega_j \phi(\mathbf{W}_{it} \mathbf{\Omega} + a_i), \quad (9)$$

where ϕ is the standard normal probability density function (pdf). Equation (9) shows that the marginal effects depend on the data, \mathbf{W}_{it} , and on the random component, a_i . In the classical framework, estimating (9) is not particularly straightforward because the error term, a_i , is not observed. As a solution, Papke and Wooldridge (2008) have outlined a procedure in which they eliminate a_i from (9) by dividing the remaining observed parameters ($\psi, \boldsymbol{\lambda}, \boldsymbol{\beta}, \phi$) by a “scale factor” of $(1 + \sigma_a^2)^{1/2}$. They then obtain scaled versions of the average partial effects¹⁰ of Ω_j by differentiating the adjusted equation with respect to W_j and plugging in values for \mathbf{W}_{it} such as its average over T or over N and T .

Conversely, with the Bayesian methods proposed here, I not only obtain posterior means for all parameters in (9) (including a_i), but through Gibbs sampling I can also obtain the entire distribution of each parameter. Therefore, the whole posterior

¹⁰Marginal effects with a_i integrated or “averaged” out.

distribution of the marginal effect is available as

$$\frac{\partial E(y_{it} \mid \mathbf{W}_{it}, a_i)}{\partial \mathbf{W}_j} = \widehat{\boldsymbol{\Omega}}_j^{(r)} \phi(\mathbf{W}_{it} \widehat{\boldsymbol{\Omega}}^{(r)} + \widehat{a}_i^{(r)}), \quad (10)$$

where the superscript r denotes the r^{th} draw of the Gibbs sampler. The marginal effects can then be calculated by simply plugging in interesting values of \mathbf{W}_{it} such as minimum or maximum values, or by averaging \mathbf{W} over N and T . The posterior mean of the marginal effect can then be calculated by averaging (10) over MCMC draws:

$$R^{-1} \sum_{r=1}^R \widehat{\boldsymbol{\Omega}}_j^{(r)} \phi(\mathbf{W}_{it} \widehat{\boldsymbol{\Omega}}^{(r)} + \widehat{a}_i^{(r)}). \quad (11)$$

Straightforward calculations of the posterior standard deviations and highest posterior density intervals (HPDIs) are available as well.

2.3 Linear model with correlated random effects

MCMC estimation of the linear model will be very similar to the probit model outlined above, with a few minor adjustments. Introduction of the latent variable, y_{its}^* , is no longer required so the full conditionals will be based on the actual joint posterior density with likelihood function:

$$p(y_{it} \mid \boldsymbol{\Omega}, a_i, \sigma_a^2, \sigma_u^2, \mathbf{W}_{it}) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp \left[-0.5 (y_{it} - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i)' \sigma_u^{-2} (y_{it} - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i) \right], \quad (12)$$

rather than the augmented posterior (8). Therefore the full conditionals will contain the actual data y_{it} rather than y_{its}^* . Furthermore, identification of the linear model does not require any restrictions being placed on the variance parameter, σ_u^2 , (in the probit model I must assume that $\sigma_u^2 = 1$ for identification). Consequently an additional inverse-gamma prior for σ_u^2 will be introduced, and an MCMC step will be added to the algorithm in order to estimate σ_u^2 . Details for the baseline linear MCMC are provided in Appendix 2.

3 Model specification with an endogenous explanatory variable (IV model)

In this section an extension to the baseline model is proposed in order to allow for potential endogeneity of a continuous explanatory variable, denoted q_{it} . To account for q_{it} , equation (5) can be rewritten as

$$y_{its}^* = q_{it}\delta + \psi + \mathbf{x}_i\boldsymbol{\lambda} + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{g}_{i,1}\boldsymbol{\phi}_1 + a_i + u_{its}, \quad (13)$$

where all I have done in (13) is added q_{it} to the right hand side.¹¹ The potentially endogenous explanatory variable, q_{it} , can then be estimated through an instrumental variables regression (of linear form):

$$q_{it} = \tau_i + \mathbf{z}_{it}\boldsymbol{\gamma} + \mathbf{g}_{i,2}\boldsymbol{\phi}_2 + \varepsilon_{it}, \quad (14)$$

where τ_i represents the time-constant unobserved effects; \mathbf{z}_{it} is a $1 \times L$ vector of time-varying instruments, including \mathbf{x}_{it} and a set of m exclusion restrictions which enter (14) only; $\mathbf{g}_{i,2}$ represents a vector of time-invariant instruments, including $\mathbf{g}_{i,1}$ and any time-invariant exclusion restrictions; and ε_{it} is an idiosyncratic error term.

¹¹In the baseline model, variable q_{it} would have been blocked together with the other explanatory variables in either \mathbf{x}_{it} or \mathbf{g}_i .

In practice, proper exclusion restrictions may or may not be available as they must be correlated with the endogenous explanatory variable q_{it} , but unrelated to the unobservables ε_{it} and u_{its} . However, in the theoretical section that follows I assume that valid exclusion restrictions are available in (14).

The unobserved effects, τ_i , are defined (using Chamberlain's correlated random effects assumption) as a linear function of all time-varying explanatory variables in all time periods:

$$\begin{aligned} \tau_i &= \eta + \mathbf{z}_{i1}\boldsymbol{\mu}_1 + \mathbf{z}_{i2}\boldsymbol{\mu}_2 + \dots + \mathbf{z}_{iT}\boldsymbol{\mu}_T + b_i \\ b_i &| \quad \mathbf{z}_{i1}, \dots, \mathbf{z}_{iT} \sim N(0, \sigma_b^2), \end{aligned} \tag{15}$$

where η is an intercept term, and $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T$ are $L \times 1$ parameter vectors. Plugging this auxiliary equation into (14):

$$q_{it} = \eta + \mathbf{z}_i\boldsymbol{\mu} + \mathbf{z}_{it}\boldsymbol{\gamma} + \mathbf{g}_{i,2}\boldsymbol{\phi}_2 + b_i + \varepsilon_{it}, \tag{16}$$

where $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T]$.

The nature of endogeneity in q_{it} is accounted for through the relationship between the idiosyncratic error terms u_{its} and ε_{it} ; which for individual i at time t , I assume is constant across the augmented data, s . Following the usual treatment of Bayesian IV models, the error terms are assumed a joint distribution which is bivariate normal with zero mean and variance-covariance matrix Σ (see Geweke, 1996; Chao and Phillips, 1998; Kleibergen and van Dijk, 1998; Kleibergen and Zivot, 2003; Hoogerheide et al.,

2007) where

$$\boldsymbol{\Sigma} = \text{var} \begin{pmatrix} u_{its} \\ \varepsilon_{it} \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & \sigma_{u\varepsilon} \\ \sigma_{u\varepsilon} & \sigma_\varepsilon^2 \end{pmatrix},$$

where $\text{var}(u_{its}) = \sigma_u^2$, $\text{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$ and $\text{cov}(u_{its}, \varepsilon_{it}) = \sigma_{u\varepsilon}$.

To simplify notation I denote $\mathbf{W}_{it} = [q_{it}, 1, \mathbf{x}_i, \mathbf{x}_{it}, \mathbf{g}_{i1}]$ as a matrix of all observable data in equation (13) and $\boldsymbol{\Omega} = [\delta, \psi, \boldsymbol{\lambda}, \boldsymbol{\beta}, \phi_1]$ as the vector of all parameters. Similarly $\mathbf{Q}_{it} = [1, \mathbf{z}_i, \mathbf{z}_{it}, \mathbf{g}_{i2}]$ represents a matrix of observable data in equation (15) and $\boldsymbol{\Upsilon}$ as the vector of parameters $[\eta, \boldsymbol{\mu}, \boldsymbol{\gamma}, \phi_2]$. The model can then be written as a system of equations such that

$$\begin{aligned} y_{its}^* &= \mathbf{W}_{it}\boldsymbol{\Omega} + a_i + u_{its}, \\ q_{it} &= \mathbf{Q}_{it}\boldsymbol{\Upsilon} + b_i + \varepsilon_{it}. \end{aligned} \tag{17}$$

Given the formulation above, the augmented joint likelihood function is commonly expressed as the joint bivariate normal distribution of $(u_{its}, \varepsilon_{it})$. However, for computational convenience the augmented likelihood function can also be written as the product of the conditional distribution of u_{its} given ε_{it} and the marginal distribution of ε_{it} (Li, 1998). This allows us to write the augmented data density in a one-dimensional form using partitioned elements of $\boldsymbol{\Sigma}$. Let $\Delta = [\boldsymbol{\Omega}, a_i, \sigma_a^2, \sigma_u^2, \sigma_{u\varepsilon}, \mathbf{W}_{it}, \boldsymbol{\Upsilon}, b_i, \sigma_b^2, \sigma_\varepsilon^2, \mathbf{Q}_{it}]$, then, the augmented data density for observation i, t can be written as

$$\begin{aligned}
p(\mathbf{d}_{it}, y_{it}^*, q_{it} \mid \Delta) = & \\
& \frac{1}{\sqrt{2\pi}\delta_u} \exp \left(-.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})' \delta_u^{-1} (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})] \right) \times \\
& \left(\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right) \times \\
& \frac{1}{\sqrt{2\pi}\delta_\varepsilon} \exp [-.5(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)' \delta_\varepsilon^{-1} (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)], \tag{18}
\end{aligned}$$

where

$$\delta_u = \sigma_u^2 - \frac{\sigma_{u\varepsilon}^2}{\sigma_\varepsilon^2}, \quad \delta_\varepsilon = \sigma_\varepsilon^2, \quad \text{and} \quad \delta_{u\varepsilon} = \frac{\sigma_{u\varepsilon}}{\sigma_\varepsilon^2}, \tag{19}$$

and

$$\varepsilon_{it} = q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i.$$

The degree of endogeneity between q_{it} and y_{it} is captured through the covariance parameter $\delta_{u\varepsilon}$. Identification of the probit model requires restrictions on one of the variance parameters, therefore δ_u is set to one.

3.1 Prior distributions

Given specification (18), all parameters can be assigned priors similar to those in Section 2. The parameter vectors $\underline{\Omega}$ and $\underline{\Upsilon}$ are assigned conjugate normal priors:

$$\underline{\Omega} = [\delta, \psi, \boldsymbol{\lambda}, \boldsymbol{\beta}, \phi_1] \sim N(\underline{\Omega}, \underline{\mathbf{H}}_{\Omega}^{-1}), \quad \underline{\Upsilon} = [\eta, \boldsymbol{\mu}, \boldsymbol{\gamma}, \phi_2] \sim N(\underline{\Upsilon}, \underline{\mathbf{H}}_{\Upsilon}^{-1}),$$

and are specified as diffuse, while the second stage variance parameters are assigned conjugate inverse-gamma priors:

$$\delta_{\varepsilon} \sim IG(a_{\delta_{\varepsilon}}, b_{\delta_{\varepsilon}}), \quad \sigma_a^2 \sim IG(a_a, b_a), \quad \sigma_b^2 \sim IG(a_b, b_b).$$

Finally, the covariance parameter (between the errors u_{its} and ε_{it}) is assigned a diffuse normal prior as

$$\delta_{u\varepsilon} \sim N(\underline{\delta}_{u\varepsilon}, \underline{\mathbf{H}}_{\delta_{u\varepsilon}}^{-1}),$$

with $\underline{\delta}_{u\varepsilon} = 0$ and $\underline{\mathbf{H}}_{\delta_{u\varepsilon}}^{-1} = 10$.

3.2 Sampling from the posterior and estimation of marginal effects

The full augmented posterior density is proportional to the product of the augmented data density (18) and the prior densities of all parameters. This can be written as

$$\begin{aligned}
p(\mathbf{y}^*, \boldsymbol{\Omega}, a_i, \sigma_a^2, \delta_{u\varepsilon}, \boldsymbol{\Upsilon}, b_i, \sigma_b^2, \delta_\varepsilon \mid y, \mathbf{W}, q, \mathbf{Q}) \propto & \\
\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})] \times & \\
\prod_{i=1}^N \prod_{t=1}^T \left(\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right) \times & \\
\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\delta_\varepsilon} \exp[-.5(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)] \times & \\
\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp[-.5a_i'\sigma_a^{-2}a_i] \times \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_b} \exp[-.5b_i'\sigma_b^{-2}b_i] \times & \\
(2\pi)^{-\frac{1+Tk+k+h+1}{2}} |\mathbf{H}_\Omega|^{-\frac{1}{2}} \exp[-.5(\boldsymbol{\Omega} - \underline{\boldsymbol{\Omega}})'\mathbf{H}_\Omega(\boldsymbol{\Omega} - \underline{\boldsymbol{\Omega}})] \times & \\
(2\pi)^{-\frac{1+Tk+L+h}{2}} |\mathbf{H}_\Upsilon|^{-\frac{1}{2}} \exp[-.5(\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})'\mathbf{H}_\Upsilon(\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})] \times & \\
(2\pi)^{-\frac{1}{2}} |\mathbf{H}_{\delta_{u\varepsilon}}|^{-\frac{1}{2}} \exp[-.5(\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})'\mathbf{H}_{\delta_{u\varepsilon}}(\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})] \times & \\
\frac{1}{\Gamma(a_a)b_a^{a_a}} (\sigma_a^2)^{-(a_a+1)} \exp\left(-\frac{1}{b_a\sigma_a^2}\right) \times \frac{1}{\Gamma(a_b)b_b^{a_b}} (\sigma_b^2)^{-(a_b+1)} \exp\left(-\frac{1}{b_b\sigma_b^2}\right) \times & \\
\frac{1}{\Gamma(a_{\delta_\varepsilon})b_{\delta_\varepsilon}^{a_{\delta_\varepsilon}}} (\delta_\varepsilon)^{-(a_{\delta_\varepsilon}+1)} \exp\left(-\frac{1}{b_{\delta_\varepsilon}\delta_\varepsilon}\right). & \tag{20}
\end{aligned}$$

The parameter estimates are then based on the following full conditional distributions (calculations are presented in Appendix 3):

1. The conditional posterior kernel of the latent data y_{its}^* is normally distributed as $y_{its}^* \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N[\mathbf{W}_{it}\mathbf{\Omega} + a_i + \delta_{u\varepsilon}\varepsilon_{it}, 1]$, and is truncated at zero such that

$$y_{its}^* > 0 \text{ if } d_{its} = 1 \text{ and}$$

$$y_{its}^* \leq 0 \text{ if } d_{its} = 0.$$

2. The full conditional density for a_i is normally distributed as

$$a_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N\left[\bar{a}_i, \bar{H}_a^{-1}\right] \text{ where}$$

$$\bar{H}_a = S \times T + \sigma_a^{-2}$$

$$\bar{a}_i = \bar{H}_a^{-1} \left[\sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\mathbf{\Omega} - \delta_{u\varepsilon}\varepsilon_{it}) \right) \right].$$

3. The full conditional of the parameter vector $\mathbf{\Omega} = [\delta, \psi, \boldsymbol{\lambda}, \boldsymbol{\beta}, \phi_1]$ is normally distributed as $\mathbf{\Omega} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N\left[\bar{\mathbf{\Omega}}, \bar{\mathbf{H}}_\Omega^{-1}\right]$ with

$$\bar{\mathbf{H}}_\Omega = \mathbf{H}_\Omega + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}'_{it} \mathbf{W}_{it}$$

$$\bar{\mathbf{\Omega}} = \bar{\mathbf{H}}_\Omega^{-1} \left[\mathbf{H}_\Omega \mathbf{\Omega} + \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}'_{it} \left(\sum_{s=1}^S (y_{its}^* - a_i - \delta_{u\varepsilon}\varepsilon_{it}) \right) \right].$$

4. The posterior distribution of the variance parameter σ_a^2 is inverse gamma, i.e.

$$\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_b^2, \delta_\varepsilon \sim IG \left[\frac{N}{2} + a_a, \left(b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a'_i a_i \right)^{-1} \right].$$

5. The conditional distribution of b_i is normally distributed as

$$b_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{b}_i, \bar{H}_b^{-1} \right] \text{ where}$$

$$\bar{H}_b = T \times \delta_\varepsilon^{-1} + S \times T \times \delta_{u\varepsilon}^2 + \sigma_b^{-2}$$

$$\bar{b}_i = \bar{H}_b^{-1} \left[\delta_\varepsilon^{-1} \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it} \boldsymbol{\Upsilon}) - \delta_{u\varepsilon} \sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i - \delta_{u\varepsilon} (q_{it} - \mathbf{Q}_{it} \boldsymbol{\Upsilon})) \right) \right].$$

6. The full conditional of the parameter vector $\boldsymbol{\Upsilon} = [\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\phi}_2]$ is normally

$$\text{distributed as } \boldsymbol{\Upsilon} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Omega}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\boldsymbol{\Upsilon}}, \bar{\mathbf{H}}_\boldsymbol{\Upsilon}^{-1} \right] \text{ where}$$

$$\bar{\mathbf{H}}_\boldsymbol{\Upsilon} = \underline{\mathbf{H}}_\boldsymbol{\Upsilon} + \sum_{i=1}^N \sum_{t=1}^T \mathbf{Q}'_{it} \mathbf{Q}_{it} \delta_\varepsilon^{-1} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{Q}'_{it} \mathbf{Q}_{it} \delta_{u\varepsilon}^2$$

$$\bar{\boldsymbol{\Upsilon}} = \bar{\mathbf{H}}_\boldsymbol{\Upsilon}^{-1} \left[\underline{\mathbf{H}}_\boldsymbol{\Upsilon} \boldsymbol{\Upsilon} + \sum_{i=1}^N \sum_{t=1}^T \left(\delta_\varepsilon^{-1} \mathbf{Q}'_{it} (q_{it} - b_i) - \delta_{u\varepsilon} \mathbf{Q}'_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i - \delta_{u\varepsilon} q_{it} + \delta_{u\varepsilon} b_i) \right) \right]$$

7. The conditional distribution of the covariance parameter $\delta_{u\varepsilon}$ is normally

$$\text{distributed as } \delta_{u\varepsilon} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\delta}_{u\varepsilon}, \bar{H}_{\delta_{u\varepsilon}}^{-1} \right] \text{ where}$$

$$\bar{H}_{\delta_{u\varepsilon}} = \underline{H}_{\delta_{u\varepsilon}} + S \times \sum_{i=1}^N \sum_{t=1}^T \varepsilon'_{it} \varepsilon_{it}$$

$$\bar{\delta}_{u\varepsilon} = \bar{H}_{\delta_{u\varepsilon}}^{-1} \left[\underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} + \sum_{i=1}^N \sum_{t=1}^T \varepsilon'_{it} \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i) \right) \right].$$

8. The posterior distribution of the variance parameter σ_b^2 is inverse gamma, i.e.

$$\sigma_b^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \delta_\varepsilon \sim IG \left[\frac{N}{2} + a_b, \left(b_b^{-1} + \frac{1}{2} \sum_{i=1}^N b'_i b_i \right)^{-1} \right].$$

9. Finally, the posterior distribution of the variance parameter δ_ε is inverse gamma, i.e.

$$\delta_\varepsilon \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2 \sim IG \left[\frac{NT}{2} + a_{\delta_\varepsilon}, \left(b_{\delta_\varepsilon}^{-1} + \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it} \boldsymbol{\Upsilon} - b_i)^2 \right)^{-1} \right].$$

Steps of the MCMC algorithm are as follows:

Algorithm 2

1. Sample $y_{its}^{*(1)}$ from $p(y_{its}^* \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, a_i^{(0)}, \boldsymbol{\Omega}^{(0)}, \sigma_a^{2(0)}, b_i^{(0)}, \boldsymbol{\Upsilon}^{(0)}, \delta_{u\varepsilon}^{(0)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$
2. Sample $a_i^{(1)}$ from $p(a_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, \boldsymbol{\Omega}^{(0)}, \sigma_a^{2(0)}, b_i^{(0)}, \boldsymbol{\Upsilon}^{(0)}, \delta_{u\varepsilon}^{(0)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$
3. Sample $\boldsymbol{\Omega}^{(1)}$ from $p(\boldsymbol{\Omega} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \sigma_a^{2(0)}, b_i^{(0)}, \boldsymbol{\Upsilon}^{(0)}, \delta_{u\varepsilon}^{(0)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$

where $\boldsymbol{\Omega}^{(1)} = [\delta^{(1)}, \psi^{(1)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\lambda}^{(1)}, \phi_1^{(1)}]$

4. Sample $\sigma_a^{2(1)}$ from $p(\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)}, b_i^{(0)}, \boldsymbol{\Upsilon}^{(0)}, \delta_{u\varepsilon}^{(0)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$
5. Sample $b_i^{(1)}$ from $p(b_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)}, \sigma_a^{2(1)}, \boldsymbol{\Upsilon}^{(0)}, \delta_{u\varepsilon}^{(0)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$
6. Sample $\boldsymbol{\Upsilon}^{(1)}$ from $p(\boldsymbol{\Upsilon} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)}, \sigma_a^{2(1)}, b_i^{(1)}, \delta_{u\varepsilon}^{(0)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$

where $\boldsymbol{\Upsilon}^{(1)} = [\eta^{(1)}, \boldsymbol{\gamma}^{(1)}, \boldsymbol{\mu}^{(1)}, \phi_2^{(1)}]$

7. Sample $\delta_{u\varepsilon}^{(1)}$ from $p(\delta_{u\varepsilon} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)}, \sigma_a^{2(1)}, b_i^{(1)}, \boldsymbol{\Upsilon}^{(1)}, \sigma_b^{2(0)}, \delta_\varepsilon^{(0)})$
8. Sample $\sigma_b^{2(1)}$ from $p(\sigma_b^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)}, \sigma_a^{2(1)}, b_i^{(1)}, \boldsymbol{\Upsilon}^{(1)}, \delta_{u\varepsilon}^{(1)}, \delta_\varepsilon^{(0)})$
9. Sample $\delta_\varepsilon^{(1)}$ from $p(\delta_\varepsilon \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^{*(1)}, a_i^{(1)}, \boldsymbol{\Omega}^{(1)}, \sigma_a^{2(1)}, b_i^{(1)}, \boldsymbol{\Upsilon}^{(1)}, \delta_{u\varepsilon}^{(1)}, \sigma_b^{2(1)})$

10. Repeat steps 1-9 R times and at each step update the conditioning variables with their most recent values.

Successive draws from the conditional densities one through nine via Gibbs sampling will create a sequence of draws that will eventually converge to the full posterior (20).

Given MCMC estimates of $\delta_\varepsilon^{(r)}$ and $\delta_{u\varepsilon}^{(r)}$ I then obtain estimates of the variance parameters σ_ε^2 , σ_u^2 , and $\sigma_{u\varepsilon}$ by solving the equations:

$$\sigma_\varepsilon^{2(r)} = \delta_\varepsilon^{(r)}, \quad \sigma_{u\varepsilon}^{(r)} = \delta_{u\varepsilon}^{(r)} \times \sigma_\varepsilon^{2(r)} \quad \text{and} \quad \sigma_u^{2(r)} = 1 + \frac{\sigma_{u\varepsilon}^{2(r)}}{\sigma_\varepsilon^{2(r)}}$$

Where the superscript r indicates that I obtain values for σ_u^2 , σ_ε^2 , and $\sigma_{u\varepsilon}$ during each MCMC draw ($r = 1, \dots, R$).

For a continuous explanatory variable, Ω_j , straightforward calculations of the posterior distribution for the marginal effects follow as

$$\widehat{\Omega}_j^{(r)} \phi(\mathbf{W}_{it} \widehat{\Omega}^{(r)} + \widehat{a}_i^{(r)} + \widehat{\delta}_{u\varepsilon}^{(r)} \widehat{\varepsilon}_{it}^{(r)}). \quad (21)$$

4 Application: the effect of school spending on student achievement

4.1 Literature

Examining the relationship between school spending and student achievement has been the focus of a large body of work, and much debate, dating back to the Coleman Report (Coleman et al. 1966). The Coleman Report gained national attention primarily because of its findings that the main determinants of student achievement were not related to school inputs, but rather the characteristics of the student's family and friends. In response to these findings, numerous researchers have attempted to analyze the relationship between school inputs (including school spending) and student achievement using the "education production function" (EPF) – a model which studies the relationship between inputs into the learning process and some measure of educational output. Inputs generally include school spending or other school resources, student and school characteristics, and family background, while output is measured using some degree of student achievement, typically in the form of standardized test scores (Hanushek, 2003).

Although a large number of researchers have used the EPF to examine the relationship between school spending and student achievement, little, if any consensus

has been reached. In a series of often cited literature reviews, Hanushek (1986, 1996, and 2003) has summarized the results of hundreds of EPFs studies dating from the early-1970's to mid-1990's. Using a statistical technique known as vote counting, he categorized the regression results of each study based on statistical significance and direction (positive or negative). In his recent review, Hanushek (2003) showed that out of 163 EPF studies, 66% found a statistically insignificant relationship between school spending and student achievement, 27% found a positive (and significant) relationship, and 7% were negative (and significant). Based on similar findings from his previous reviews, Hanushek (1986) wrote "there appears to be no strong or systematic relationship between school expenditures and student performance" (Hanushek, p. 1162). This statement has however been challenged by several researchers. In particular, Hedges et al. (1994) and Greenwald et al. (1996) criticized Hanushek's vote counting methodology, and reanalyzed the literature through meta-analysis. Both of these studies found that a strong and positive relationship between school spending and student achievement did in fact exist.

Other researchers have argued that the mixed conclusions found in the literature are not due to a lacking relationship between school spending and student achievement, but are the product of modeling differences and misspecifications. In particular, many of the traditional EPF studies may fail to account for endogeneity, arising because school spending might be correlated with unobservable determinants of student achievement (Ferguson & Ladd, 1996; Ludwig & Bassi, 1999; Guryan, 2001; Webbink, 2005). This could occur if relevant explanatory variables, such as family inputs, are excluded from the model (omitted variables), but are related to both school spending

and student test scores. For example, if highly motivated parents spend additional time helping their children with their homework, and also choose to send their children to schools with more resources, then a researcher who does not observe parental motivation may find that school resources have a positive effect on student achievement. However, in reality the higher achievement is due in part by parental motivation (Tiebout, 1956; Mayer, 1997; Webbink, 2005). In this case, traditional EPF models which assume strict exogeneity may overstate the true relationship between school spending and student achievement. Ludwig and Bassi (1999) explained that if this were the only reason to expect estimation bias then traditional EPF studies could be viewed as “upper bound estimates.” However, there is also a possibility that some schools are funded in a compensatory manner, whereby lower achieving schools increase spending in an effort to raise the student achievement level. If this is true then traditional EPF estimates may actually underestimate the true relationship between school spending and student achievement (Heckman et al., 1996).

In an effort to account for these endogenous changes in school spending numerous researchers have applied instrumental variable (IV) regression techniques, and have generally found a larger positive relationship between school spending and student achievement as compared to the traditional studies which assume strict exogeneity (see Ferguson and Ladd, 1996; Ludwig & Bassi, 1999; Dewey et al., 2000; Roy, 2003; Levacic et al., 2005; Papke, 2005; Webbink, 2005; Jenkins et al., 2006; Papke and Wooldridge, 2008). In practice however, finding a proper instrument can be a challenging task as it must be correlated with school spending but have no relationship with student achievement otherwise.

In much of the literature, researchers have tried to solve this problem by creatively exploiting some change in “nature,” often brought about by a new government policy that leads to (arguably) exogenous variations in school resources. For instance, Roy (2003), Papke (2005), and Papke and Wooldridge (2008) all used longitudinal data to analyze the relationship between school spending and student achievement among Michigan elementary schools in the mid 1990’s. During this time period a new law (Proposal A) was passed which changed Michigan’s school funding scheme to one that relied more heavily on state funding (through sales tax revenues) and less on local property taxes. This led to large changes in school spending (see Papke, 2005), which the researchers assumed was unrelated to student achievement. As another example, Guryan (2001) took advantage of changes in state funding among Massachusetts school districts caused by the Massachusetts Education Reform Act of 1993. The policy was implemented in an effort to equalize spending across Massachusetts school districts, which Guryan argued led to exogenous changes in district spending levels.

In many instances however, there are school systems of interest but no natural experiments to exploit. In these cases the issue of endogeneity can still be addressed if the researcher can locate an external instrument. For example, Dewey et al. (2000) proposed using political variables such as whether the democratic or republican party had control of the state government (both legislative and executive). However, this data will not have much variation as it is can only be measured at the state level, and therefore can only be used to compare schools across states. As another alternative Ferguson and Ladd (1996) analyzed district-level data on Alabama schools and used per-capita income and property values as instruments. However, this instrument

might not be valid, as previous research has indicated that student achievement may have a positive impact on housing prices (Hayes & Taylor, 1996; Bogart and Cromwell, 1997; Black, 1999; Weimer and Wolkoff, 2001; Figlio & Lucas, 2004).

4.2 Data summary

The purpose of the empirical work in this section is to further investigate the relationship between school spending and student achievement by analyzing data on a large subset of Florida elementary schools, observed over a seven year period. The paper attempts to address the endogeneity problem in a number of ways. First, data are collected from a variety of sources in order to control for as many relevant school, student, and family characteristics as possible. In addition, the panel data methods outlined in Sections 2 and 3 are applied in order to allow for individual school effects which control for any time-constant unobserved difference across schools such as (but not limited to): school administration, school policy differences, school structural differences, geographical differences, and other historical differences. Finally, the model will attempt to identify a causal relationship between school spending and student achievement through the use of SEM and IV methods in order to capture exogenous changes in school spending.

The data are comprised of 1138 public elementary schools, located in 28 of the larger Florida school districts. (For the purposes of this paper a large school district is defined as one which contains at least 10 elementary schools.) The data set was constructed using multiple sources. School-level data were collected for a seven year

period between 1999 and 2005 from the Florida School Indicators Report (FSIR) and the Common Core of Data (CCD), provided by the Florida Department of Education and National Center of Education Statistics respectively. The year 1999 refers to the school year of fall 1998 - spring 1999, and 2005 refers to the school year of fall 2004 - spring 2005. Additional city-level data were collected for the year 2000 using the U.S. Census Bureau's database, and district-level data were gathered from the property valuation and tax data spreadsheets (1999-2005) supplied by the Florida Department of Revenue.

The FSIR database provided detailed school level information on standardized test scores as well as school and teacher characteristics. School and teacher characteristics included variables such as the school size (number of students), teachers' education, the proportion of staff devoted to instruction, and school spending per-pupil (which was converted into 1999 dollars using the Southeast CPI data.) Student composition measures included variables such as the percentage of students classified into gifted, special education, and English as a second language (ESOL) programs, as well as the percentage of students absent for more than 21 days in a school year. The CCD provided information on each school's racial and ethnic composition and physical location (city and zip code.) These data were then matched with data from the 2000 Census to derive adult education levels for each neighborhood. Finally, the Florida Department of Revenue data was used to gather county level property taxes.¹²

The sample time period is of particular interest because it coincides with the 1999

¹²In Florida, a school district's boundary is defined by the county boundary. Therefore the terms school district and county can be used interchangeably.

implementation of Florida's A-plus Program for Education, a school accountability reform aimed at increasing student achievement throughout the state. At the center of the reform was a school grading system used by the state government to rank each Florida public school on a scale of A through F. Grades were based mainly on each school's overall student performance on the FCAT exam – a high stakes standardized test administered annually to Florida's public school students. Schools that received a D or an F grade were provided with assistance and intervention plans, which if necessary included access to additional resources and the reassignment or even replacement of school staff members. For a detailed summary of assistance plans available for F and D schools see Chakrabarti (2007). The commissioner of education (responsible for budget development and school assessment/accountability) was also allowed to give preference to these schools when allocating Federal and State grants designed to improve student achievement (FLDOE, Rule 6A-1.09981, 1999). As a result, I believe that school spending may be determined, in part, by student achievement, and therefore treating spending as an exogenous covariate may lead to biased estimates.

In order to control for endogenous school spending I used district-level property taxes as an instrumental variable, however, it is important to note that student achievement was measured at the school level. The chosen instrument was measured as the total dollar value of property taxes levied per capita at the district-level. (These figures were calculated by the Florida Legislative Committee on Intergovernmental Relations (LCIR) staff as the total property taxes levied per county divided by county population estimates. The Southeast Consumer Price Index (CPI) data,

collected from the Bureau of Labor Statistics, was then used to transform these figures into 1999 dollars.) Florida public schools are financed with a mixture of state, local and federal funding. Specifically, in 2007 Florida school districts received approximately 40% of their finances from state sources, supplied primarily from legislative appropriations; 50% from local funds, provided mainly by property taxes; and 10% from the federal government. Since school districts are funded in large part by property taxes, this instrument should be highly correlated with school spending. Descriptive statistics indicate that school spending per student was larger in districts with higher property taxes per capita, as the correlation coefficient between these two variables is 0.346. This indicates that the chosen instrument is moderately correlated with the endogenous explanatory variable. This instrument has been used previously in the literature by Ferguson and Ladd (1996), however, one issue is that student achievement may have a positive effect on housing prices, which could invalidate the chosen instrument (Black, 1999; Figlio and Lucas, 2004). Therefore, to dissolve any relationship that may exist between school-level student achievement and district-level property taxes, the sample is comprised only of those schools located within the larger Florida school districts where there will likely be a mixture of high performing and low performing schools. Thus, even if one school displays very high student achievement, this should not boost all property values in the entire district, merely those located in close proximity to the high performing school. Furthermore, Florida saw a large increase in housing prices during this time period due to increased real estate investment (in what is commonly referred to as the housing bubble). Thus there was an increase in property taxes and in turn an increase in school spending

which was unrelated to student achievement. These unique circumstances provide additional justification for the validity of the chosen instrument.

The variables used in the analysis are defined and summarized in Table 1. The explanatory variables can be separated into three subcategories: time varying regressors, x_{it} , (including GIFT, DISAB, ABSENT, and DEGREE) which will be treated with the Chamberlain-device, regressors with little or no variation across time, g_{it} , (including PARENT EDUC, BLACK, HISPANIC, ENGLISH, and INSTRUCT), and the potentially endogenous explanatory variable, q_{it} , (EXPEND), which was measured as real spending in 1999 dollars, and was calculated using the Southeast CPI data.

4.2.1 Standardized test pass rates and state assigned school grades

In line with the *A+ plan* the outcome variable of interest, student achievement, was measured using each school's overall FCAT performance. The FCAT exams were graded on a scale of 1 (lowest) through 5 (highest) where levels 1 and 2 represent "below basic" achievement and "basic" achievement respectively, while levels 3 and above correspond to "proficiency." The analysis focused on fourth grade reading and fifth grade math outcomes as these were the only primary grade levels tested continuously throughout the sample period.¹³

The FSIR database provided detailed information on the percentage of students scoring in each of the five FCAT achievement levels. From this I created measures

¹³Currently the FCAT is administered to all students in grades 3 through 10, testing their knowledge in math, reading, writing, and science. Though, when first implemented in 1998, the reading exams were only given to students in grades 4, 8, and 10 while the math exams were given to students in grades 5, 8, and 10. It was not until 2001 that the FCAT exams were extended to all grades between third and tenth.

indicating the percentage of fourth graders at school i time t that passed the reading FCAT exam and the percentage of fifth graders that passed the math FCAT exam (achieved a level 3 or higher). These outcome measures, denoted READ_PASS and MATH_PASS, were chosen for their policy relevance as they were the main measures used by the state to calculate school grades.

Overall, the average pass rate for the 4th grade reading exam was 57.8% and ranged from 7% up to 99%, while the average pass rate for the 5th grade math exam was 47.6% and ranged from 0% to 95%. Table 2 displays more descriptive summary statistics on pass rates for the fourth grade reading FCAT. The table indicates that for every percentile, the proportion of students passing increased in every year with the exception of the year 2000. In the beginning of the sample period FCAT pass rates increased slightly from one year to the next, but towards the end of the sample period, these yearly gains were much more prominent. For example, in 1999 the reading pass rate among the lowest 10th percentile of schools was only 27%, in 2002 it had risen to 33%, and by 2005 this had risen to 52%. For schools in the 50th percentile, the pass rate rose from 53% in 1999 to 55% in 2002, and dramatically up to 71% in 2005. Finally, for those in the top 90th percentile, the pass rate rose from 74% in 1999, to 75% in 2002, and 87% in 2005.

Table 1: Variable definitions and summary statistics (N = 1,138 schools; T = 7 years)

Variable	Definition	Mean	S.D.	Min	Max
Dependent Variables					
READ_PASS	Percentage of the school's 4th grade students that achieved a level 3 or higher on the reading FCAT	57.8%	17.3%	7%	99%
MATH_PASS	Percentage of the school's 5th grade students that achieved a level 3 or higher on the math FCAT	47.6%	17.2%	0%	95%
Explanatory Variables					
EXPEND	School expenditures per pupil (real in 1999 dollars)	\$4,695.32	\$1,063	\$722	\$12,808
PARENT_EDUC	Percentage of neighborhood population age 25+ with a Bachelor's degree or higher (year 2000)	22.9%	9.7%	0.6%	58.3%
BLACK	Percentage of the school's students that are Black	27%	25%	0.1%	100%
HISPANIC	Percentage of the school's students that are Hispanic	20.4%	22.9%	0.1%	100%
ENGLISH	Percentage of the school's students in English for speakers of other languages (ESOL) programs	9.8%	11.6%	0%	69.8%
INSTRUCT	Percentage of the school's staff devoted to instruction	65.1%	7.2%	35.2%	88.7%
GIFT	Percentage of the school's students classified as gifted	4%	5%	0%	54.9%
DISAB	Percentage of the school's students classified as disabled/special education	15.6%	5.8%	0%	45.8%
ABSENT	Percentage of the school's students absent from school for 21 days or more in a school year	6.5%	3.3%	0%	47.5%
DEGREE	Percentage of the school's teachers with advanced degrees (Masters or Ph.D.)	32.1%	11%	0%	71.5%
TAXES	Total taxes levied per capita (district level; real in 1999 dollars)	\$403.93	\$102.29	\$159.76	\$859.27

Table 3 reports similar summary statistics for the fifth grade math FCAT exam. The pass rates associated with the math exam were almost always lower than those of the reading exam. However, the table reveals a similar pattern of increased improvement over time. Papke (2005) and Papke and Wooldridge (2008) found a similar trend (of increased achievement) to exist among public school students in the state of Michigan. They attributed this to “a ‘teaching to the test phenomenon,’ making the tests easier over time, increased real spending, or some combination of these” (Papke and Wooldridge, p. 127). In this paper, I try to determine how much of this increase in student achievement among Florida students, if any, can be attributed to an increase in real spending.

In order to put these figures into perspective, Tables 4 and 5 illustrate how FCAT performance varied across the state assigned school grades. For the average A school roughly 70% of the fourth grade students passed the reading test, and 60% of the fifth graders passed the math test. Among these students about 8% scored a perfect 5 out of 5, and more than 25% scored a level 4, while less than 17% of the school’s students received a level 1 (the lowest possible score). Conversely, for the average F school, the passing rates for the reading and math exams were 22% and 15% respectively, and for both exams, less than 1% of the student population had achieved a level 5, less than 5% had achieved a level 4, and over 55% of the school’s students had received a level 1.

Table 2: Pass rates on 4th grade reading FCAT. Percentiles, 1999-2005

Year	10 th	25 th	50 th	75 th	90 th
1999	27%	39%	53%	66%	74%
2000	28%	39%	51%	63%	72%
2001	31%	41%	52%	64%	73%
2002	33%	43%	55%	66%	75%
2003	37%	48%	60%	71%	80%
2004	49%	59%	69%	78%	85%
2005	52%	61%	71%	79%	87%

Table 3: Pass rates on 5th grade math FCAT. Percentiles, 1999-2005

Year	10 th	25 th	50 th	75 th	90 th
1999	15%	25%	38%	50%	62%
2000	23%	32%	44%	56%	65%
2001	25%	35%	46%	58%	68%
2002	27%	36%	48%	61%	70%
2003	28%	38%	50%	61%	73%
2004	29%	39%	51%	63%	73%
2005	33%	44%	56%	67%	77%

4.2.2 Expenditures per pupil and instrumental variables

The average school in the sample spent roughly \$4,700 per student in a given year (in 1999 dollars) though this ranged from \$722 up to \$12,808 per student. Three schools in the sample had a budget of less than \$1,000 per student (all of which were during the 2000 school year) and seven had a budget of over \$10,000 per student (all in 2005). Interestingly, the three lowest spending schools were all located in Broward county and were all higher performing schools as two of the three had been assigned an A grade by the state and the third had received a B (in 2000). Conversely, the seven highest spending schools were all located in Miami-Dade county, and among these

Table 4: FCAT scores (by state assigned school grades) 4th grade reading exam

School Grade	% Pass	Percentage of Students at Each Achievement Level				
		Level 1	Level 2	Level 3	Level 4	Level 5
A	70.7%	16.2%	12.9%	33.0%	29.1%	8.6%
B	60.3%	23.3%	16.1%	32.9%	22.1%	5.3%
C	48.9%	32.7%	18.3%	29.8%	16.1%	3.0%
D	31.7%	50.7%	17.5%	22.2%	8.4%	1.1%
F	21.8%	61.5%	16.4%	16.9%	4.4%	0.5%

Table 5: FCAT scores (by state assigned school grade) 5th grade math exam

School Grade	% Pass	Percentage of Students at Each Achievement Level				
		Level 1	Level 2	Level 3	Level4	Level 5
A	60.5%	14.9%	24.3%	26.4%	25.7%	8.4%
B	50.3%	21.3%	28.1%	25.1%	20.1%	5.1%
C	37.4%	30.2%	32.2%	21.6%	13.3%	2.5%
D	23.9%	45.2%	30.7%	15.7%	7.1%	1.1%
F	15.03%	56.7%	28.0%	11.0%	3.7%	0.33%

schools only one was awarded an A grade, two had received C's, and the remaining four had all been assigned D grades (in 2005).¹⁴

Table 6 contains percentiles of spending per pupil from 1999 through 2005. For each percentile, average expenditures per pupil rose every year except between the years 2001 and 2002. For example, in 1999 the lowest 10th percentile of schools spent an average of \$3,371 per student, and in 2005 they spent an average of \$4,282 per student, an increase of 27% in 7 years. Among schools spending in the top 90th percentile, average per pupil spending rose from \$5,342 in 1999 to \$6,954 in 2005, an

¹⁴To determine whether these outliers had any impact on the subsequent analysis, additional MCMC simulations were performed without such outliers included. The results were unaffected.

increase of 30% in 7 years. These figures also indicate that there was a large spending gap between the higher spending schools and the lower spending schools during the sample period. In each year, schools in the 50th spending percentile spent roughly 20% more than those in the 10th percentile, and schools in the top 90th spending percentile spent roughly 60% more than those in the 10th percentile.

Table 7 reports average school spending per-student across state assigned school grades and by year. The table indicates that, throughout the grade distribution, school spending had increased over time. Furthermore, regardless of the year, the lower performing D and F schools spent a considerable amount more per student than the higher performing A, B, and C schools. For example, the average A school only spent \$3,773 per student in 1999 compared to \$4,642 per student among D schools and \$5,144 among F schools. While, in 2005 average per-pupil spending among A schools was \$5,338, but over \$6,000 for the lower performing D and F schools. This indicates that Florida schools were likely funded in a compensatory manner; perhaps due in part by the assistance and intervention plans assigned to the lower performing schools, as mandated by the *A+ plan* for education. This suggests that variations in school spending were not exogenous; rather they were determined in part by school grades and ultimately by student achievement. I therefore adopt an instrumental variables approach whereby the selection equation models school spending as a function of school level characteristics as well as an additional district level exclusion restriction.

Table 8 contains summary statistics for the instrumental variable. The table indicates that for every percentile, district level property taxes per capita increased during every year of the sample period. For example, in 1999 median property taxes

was roughly \$374 per capita and in the year 2005 median property taxes was \$492 per capita, an increase of 31%. Much of this can be attributed to the rise in housing prices caused by the real estate boom of the late 90s and early-2000's.

4.2.3 Additional control variables

While the main goal was to analyze the impact of school spending on student achievement, I also included additional explanatory variables for family background, student composition, race/ethnicity, and teacher characteristics which have previously been shown to affect student achievement.

Family background has consistently been found to be a strong indicator of student achievement (Coleman et al., 1966; Jencks et al., 1972; Mayer, 1997; Davis-Keane, 2005). To try and account for this, much of the research attempts to include information regarding parental education and family income. However, in two separate studies Gyimah-Brempong and Gyapong (1991) and Dewey et al. (2000) have shown that including family income measures (or proxies for family income) as an input in the EPF may lead to confounding results due to endogeneity and/or multicollinearity because richer families are more likely to send their children to better schools (leading to endogeneity), or perhaps to schools that spend more money (leading to multicollinearity). As an alternative, both Gyimah-Brempong and Gyapong (1991) and Dewey et al. (2000) suggest using parental education as the only measure of family background. Therefore in this analysis family background was measured using the adult education rate (percentage of adults age 25 or older with a Bachelor's degree or higher) associated with each school's neighborhood, according to the US Census

Table 6: School level expenditures per pupil in 1999 dollars. Percentiles, 1999-2005

Year	10 th	25 th	50 th	75 th	90 th
1999	\$3,371	\$3,655	\$4,027	\$4,567	\$5,342
2000	\$3,380	\$3,690	\$4,169	\$4,753	\$5,392
2001	\$3,549	\$3,907	\$4,375	\$5,031	\$5,740
2002	\$3,464	\$3,878	\$4,348	5,019	\$5,693
2003	\$3,641	\$4,004	\$4,577	\$5,188	\$5,883
2004	\$3,875	\$4,317	\$4,903	\$5,610	\$6,341
2005	\$4,282	\$4,750	\$5,377	\$6,140	\$6,954

(2000). The average adult education rate was 22.9%, though this ranged from 0.6% up to 58.3%.

Student composition was controlled for with the inclusion of the percentage of students classified into gifted, special education, and English as a second language (ESOL) programs (Variables GIFT, DISAB, and ENGLISH respectively). On average, roughly 4% of a school's students were classified as gifted, 15.6% as special education, and 9.8% as limited English proficient. Gifted programs usually provide a more advanced curriculum than those of an average student's whereas the converse is true for programs focusing on students with learning disabilities. Therefore I assume, *a priori*, that schools with a higher percentage of gifted students were likely to perform better on standardized tests, while schools with a larger proportion of special education students may have performed worse. This latter assumption must be considered carefully however, as many special education students were legally exempt from taking the FCAT exams.¹⁵ Some researchers have proposed that schools

¹⁵State law mandates that all students in the appropriate grade levels must participate in state assessments like the FCATs. However, the inclusion of special education students is determined by each student's Individual Education Plan (IEP).

Table 7: Average expenditures per pupil by school grade and year in 1999 dollars

School Grade	1999	2000	2001	2002	2003	2004	2005
A	\$3,773	\$4,034	\$4,265	\$4,226	\$4,442	\$4,790	\$5,338
B	\$3,882	\$3,896	\$4,322	\$4,431	\$4,704	\$5,100	\$5,543
C	\$4,035	\$4,305	\$4,546	\$4,747	\$5,196	\$5,612	\$5,973
D	\$4,642	\$5,082	\$5,490	\$5,300	\$6,194	\$6,637	\$6,622
F	\$5,144	\$4,944	-	\$5,811	\$5,249	-	\$6,036

Table 8: Total taxes levied per capita (district level) in 1999 dollars. Percentiles, 1999-2005

Year	10 th	25 th	50 th	75 th	90 th
1999	\$229.78	\$267.82	\$374.30	\$396.79	\$402.49
2000	\$234.04	\$300.11	\$379.98	\$419.88	\$452.68
2001	\$241.76	\$308.45	\$387.19	\$420.42	\$471.04
2002	\$274.30	\$335.41	\$403.42	\$460.55	\$508.04
2003	\$291.68	\$359.01	\$429.89	\$489.89	\$523.75
2004	\$306.52	\$380.66	\$456.77	\$527.43	\$555.80
2005	\$352.52	\$430.48	\$492.17	\$567.19	\$604.23

may have even used this law to their advantage by classifying some of their poorer test takers into special education programs in an attempt to raise test pass rates (Figlio and Getzler, 2002; Jacob, 2005; Cullen and Reback, 2006). If this is true, then DISAB may have no association or even a positive effect on FCAT pass rates. Finally, I assume *a priori* that ENGLISH has a negative effect on the FCAT exams since both exams require a comprehensive understanding of the English language.¹⁶

Among the racial/ethnicity measures, the mean percentage of Black and Hispanic students in attendance were 27% and 20.4% respectively. Previous studies have

¹⁶The negative effect of ENGLISH may be negligible because ESOL students are not required to take the FCAT exam during their first two years in the program.

consistently found a negative relationship between BLACK and student achievement (Rivkin, 1995; Fryer and Levitt, 2004; Neal, 2006; Hanushek, Kain and Rivkin, 2009). However, the effect of HISPANIC might not be as straightforward, especially in the state of Florida. At the national level, studies have typically found that Hispanic students score lower on standardized test scores than do White students (Ingels et al., 1994; National Center for Education Statistics, 1998; Phillips, 2000; U.S. Department of Education, 2005), which many have attributed to “language acquisition barriers” (Wojtkiewicz and Donato, 1995; National Center for Education Statistics, 1995; Bali and Alvarez, 2004). However, in Florida the Hispanic population consists of a large proportion of Cuban and Puerto Rican families who place a strong emphasis on education and tend to speak better English relative to Hispanics living elsewhere in the United States (EDRFL, 2005). Furthermore, since I am also controlling for the proportion of students with limited English proficiency, the coefficient associated with Hispanic may only capture the achievement effect for those Hispanic students that can speak English well. Therefore, *a priori*, I assume that BLACK will have a negative effect on student achievement but make no such assumptions with respect to the HISPANIC coefficients.

To control for the variation in school and teacher quality I also included measures for teachers’ education and the percentage of each school’s staff devoted to instructional purposes. Teachers’ education, as measured by the percentage of teachers with advanced degrees, is of great interest because most states associate advanced degrees with teacher quality and award higher salaries to those holding a master’s or doctorate degree (Goldhaber and Brewer 1998). For example in 2005 the average salary

for a Florida school teacher with a Bachelor's degree was \$38,516, while those with a Master's or Doctorate degree earned an average salary of \$45,678 and \$52,047 respectively (Florida Department of Education, 2005). Those in favor of this increased pay scale argue that teachers with advanced degrees have a higher level of knowledge and expertise which they can pass on to their students. While a few studies have suggested that this may be true (Ferguson, 1991; Greenwald et al., 1996), the majority of research has concluded that no such effect exists; teachers' education has no significant impact on student achievement (Hanushek, 1986; Hanushek, 1997; Jepsen and Rivkin, 2002; Rowan et al., 2002; Buddin and Zamarro, 2009). Kelly Henson, the director of Georgia's Professional Standards Commission, stated that one possible explanation for this relationship, or lack thereof, is that some teachers might be "taking the path of least resistance to get a pay raise." Suggesting that some teachers are obtaining their advanced degrees from online colleges that offer less demanding curricula or are receiving their degrees in fields unrelated to the subjects they teach because the coursework might be easier. In the sample the mean proportion of teachers with a master's degree or higher was 32.1%, and approximately 65% of the school's staff was employed for instructional purposes. I assume *a priori* that the percentage of a school's staff devoted to instructional purposes (INSTRUCT) will be positively associated with student achievement but make no such assumption with regards to the percentage of teachers with an advanced degree (DEGREE.) Finally, I controlled for the percentage of students absent for more than 21 days in a school year, as these students miss more of the curriculum than their cohorts and therefore it is likely that they would perform worse on the FCAT exams than the average student.

4.3 Results

To initiate the Gibbs sampler, starting values for the model parameters were randomly drawn from the uniform distribution bounded between 0 and 1, and values for the projection error terms a_i and b_i were randomly drawn from the normal distribution with mean 0 and standard deviation 0.1.¹⁷ The reason for why the diffuse priors were not used to draw the initial values was that when I did so it took considerably more time for the Markov chain of the multi-parameter model to converge to the stationary distribution. I fit the baseline models (both the linear and fractional probit models, assuming strict exogeneity) using the algorithm outlined in Section 2, running the Markov chains for 20,000 iterations, following an initial 5,000 replication burn-in phase. The IV models were then estimated using the algorithm outlined in Section 3. After some preliminary simulations, I found that the sequential draws of the IV models (both linear and fractional probit) displayed higher degrees of autocorrelation, indicating that the simulations may be slower to converge. This was addressed by running longer MCMC simulations of 35,000 iterations with an initial 20,000 replication burn-in phase. Separate models were fit for each FCAT exam (4th grade reading and 5th grade math).

In Table 9 I present the posterior estimates for the baseline models (linear and fractional probit) with 4th grade reading pass rates as the dependent variable. The table reports posterior means, posterior standard deviations (SD) and 95% highest posterior density intervals (HPDI) for all parameters listed in Table 1. The results found for the 5th grade math exam (reported in Table 10) were very similar to those

¹⁷For the linear model a_i and b_i are drawn from the standard normal distribution $N(0, 1)$.

presented in Table 9, and therefore I limit my discussion to the estimates associated with 4th grade reading.

The baseline results suggest that an increase in per pupil expenditures does in fact have a positive and significant effect on reading pass rates among 4th grade students. This holds true for both the linear specification and the fractional probit model. However, in the probit model the estimated spending effect is slightly smaller. This is consistent with the notion that spending may exhibit diminishing marginal returns which can be accounted for in the probit model. Conversely, the linear model assumes a constant relationship between the explanatory variables and the outcome variable, and therefore may slightly overestimate the true spending effect. Furthermore, the standard errors of the probit model are much smaller than those from the linear specification resulting in much stronger effects. Thus, the effect of expenditure is sharpened up with the more robust model specification. One advantage for using the linear model, however, is that the slope coefficients are easy to interpret since they have a one-to-one correspondence with the marginal effects. For instance, in the linear model the estimated posterior mean of spending is 0.066. This implies that a \$1,000 increase in per student spending would increase reading pass rates by 6.6%. Conversely, in the probit model, the slope coefficient is 0.173, and is statistically significant, however, for interpretation the marginal effects must be calculated. Therefore, I also estimate the posterior distribution for the marginal effects of all parameters (using sample average values of W_{it} , over N and T) and report the posterior means, standard deviations, and highest posterior density intervals of the marginal effects in the last three columns of Table 9. For the marginal effect of spending, I

find a posterior mean of 0.062, indicating that a \$1,000 increase in spending would increase pass rates by 6.2%; a nontrivial effect, almost identical to that of the linear model.

The posterior distributions of the marginal effects of spending are plotted below in Figures 1 and 2. For the linear model (Figure 1), the spending effect ranges from 0.0097 to 0.1202, while in the probit model (Figure 2), the posterior distribution of the marginal effect ranges from 0.0603 to 0.0643. These plots show that the spending effect is always positive as the entire posterior distribution is above zero in both models.

The posterior estimates also indicate that PARENT EDUC, INSTRUCT and GIFT all have large positive effects on reading FCAT pass rates.¹⁸ While DEGREE has a small positive effect, and BLACK has a small negative effect. The posterior mean for HISPANIC and ENGLISH are statistically insignificant in the linear model, but display a positive effect in the nonlinear model, and DISAB has a positive and significant posterior mean in both models. The latter could imply that schools were trying to “game the system” by classifying some of their poorer test takers into test-exempt special education programs. However, it is also possible that these estimates are a by-product of model misspecification since the baseline models do not account for endogenous spending. Thus, if Florida schools follow a compensatory funding scheme these results may be biased

¹⁸According to the estimated marginal effects of the fractional probit model (reported in table 9) a 10% increase in the percentage of staff devoted for instructional purposes would, on average lead to a 4.8% increase in pass rates, while a 10% increase in the percentage of gifted students would increase pass rates by 6.4%. Parental education has close to a one-to-one correspondence with reading pass rates, (a 10% increase in the adult education rate would lead to a 9.2% increase in pass rates).

Table 9: Posterior estimates for 4th grade reading. Baseline model assuming school expenditures is exogenously determined

Variable	Linear Model			Fractional Probit Model							
	(1)			(2)			(3)				
	Mean	SD	HPDI	Coefficient	Mean	SD	HPDI	Coefficient	Mean	SD	HPDI
EXPEND (thousands)	0.066	(0.015)	(0.036, 0.095)		0.173	(0.001)			0.062	(0.001)	(0.061, 0.063)
PARENT EDUC	0.659	(0.368)	(-0.049, 1.400)		2.561	(0.133)			0.921	(0.048)	(0.830, 1.018)
BLACK	-0.134	(0.156)	(-0.433, 0.174)		-0.090	(0.024)			-0.032	(0.009)	(-0.050, -0.016)
HISPANIC	-0.262	(0.241)	(-0.742, 0.210)		1.482	(0.031)			0.533	(0.011)	(0.510, 0.554)
ENGLISH	0.047	(0.34)	(-0.603, 0.724)		0.199	(0.031)			0.072	(0.011)	(0.050, 0.093)
INSTRUCT	0.667	(0.31)	(0.061, 1.280)		1.358	(0.028)			0.488	(0.010)	(0.469, 0.509)
GIFT	0.546	(0.13)	(0.309, 0.797)		1.793	(0.062)			0.644	(0.022)	(0.600, 0.688)
DISAB	0.373	(0.125)	(0.144, 0.616)		0.324	(0.04)			0.116	(0.014)	(0.088, 0.145)
ABSENT	0.313	(0.185)	(0.004, 0.660)		0.149	(0.039)			0.054	(0.014)	(0.026, 0.081)
DEGREE	0.098	(0.058)	(-0.006, 0.206)		0.169	(0.014)			0.061	(0.005)	(0.051, 0.071)
Variance Parameters											
σ_u^2	0.38	(0.10)			-				-		
σ_a^2	0.55	(0.16)			2.30	(0.19)					

Table 10: Posterior estimates for 5th grade math. Baseline model assuming school expenditures is exogenously determined

Variable	Linear Model			Fractional Probit Model				
	Mean	SD	HPDI	Mean	SD	HPDI		
			(1)			(3)		
			Coefficient			Marginal Effects		
			(2)					
			Mean			Mean		
			SD			SD		
			HPDI			HPDI		
EXPEND (thousands)	0.048	(0.015)	(0.018, 0.078)	0.126	(0.001)	0.046	(0.001)	(0.045, 0.047)
PARENT EDUC	0.663	(0.368)	(-0.040, 1.408)	2.455	(0.136)	0.903	(0.050)	(0.825, 1.026)
BLACK	-0.107	(0.154)	(-0.409, 0.195)	-0.263	(0.023)	-0.097	(0.008)	(-0.114, -0.081)
HISPANIC	-0.268	(0.244)	(-0.752, 0.211)	0.950	(0.030)	0.349	(0.011)	(0.329, 0.372)
ENGLISH	0.149	(0.338)	(-0.502, 0.826)	0.198	(0.031)	0.073	(0.011)	(0.051, 0.095)
INSTRUCT	0.468	(0.310)	(-0.143, 1.082)	0.685	(0.028)	0.252	(0.010)	(0.232, 0.273)
GIFT	0.668	(0.129)	(0.432, 0.912)	1.925	(0.059)	0.708	(0.022)	(0.665, 0.750)
DISAB	0.379	(0.125)	(0.146, 0.620)	-0.021	(0.014)	0.138	(0.015)	(0.108, 0.167)
ABSENT	0.092	(0.187)	(-0.217, 0.443)	-0.521	(0.038)	-0.192	(0.014)	(-0.220, -0.164)
DEGREE	0.029	(0.058)	(-0.074, 0.138)	0.375	(0.040)	-0.008	(0.005)	(-0.018, 0.003)
Variance Parameters								
σ_u^2	0.378	(0.104)		-	-	-	-	-
σ_a^2	0.552	(0.157)		2.302	(0.194)			

In order to allow for endogenous spending I apply the IV method outlined in Section 3. The selection equation, used to estimate school spending, includes all explanatory variables from the original outcome equation as well as the instrument, real property taxes per-capita measured at the district level. In Table 11 I display the posterior results of the spending equation. Since spending is measured as a continuous variable the selection equation is always linear, therefore the slope coefficients can be treated as the marginal effects in both the linear model and the fractional probit model.

The results reported in Table 11 indicate that the chosen instrument, property taxes per-capita, has a strong positive effect on school expenditures in both the linear and nonlinear model; this helps justify the choice in instruments. The table also indicates that the proportion of students classified into gifted, special education, and English as a second language programs all have a positive impact on school spending. This is expected since these specialized classes are generally smaller in size and taught by specialists or teachers with additional certifications who are typically paid a higher salary. The proportion of teachers with advanced degrees also displays a positive effect which is consistent with school spending patterns, as the average salary was roughly 18.5% greater for teachers with a Master's compared to a Bachelor's degree, and 35% higher for those with a Doctorate versus a Bachelor's. Interestingly, I find that parental education has a negative effect on school spending. This is consistent with previous research which has indicated that, on average, higher educated parents place a greater weight on their children's education (Coleman, et al., 1966; Hanushek, 1996; Davis-Keane, 2005). Therefore, it is more likely that higher educated parents

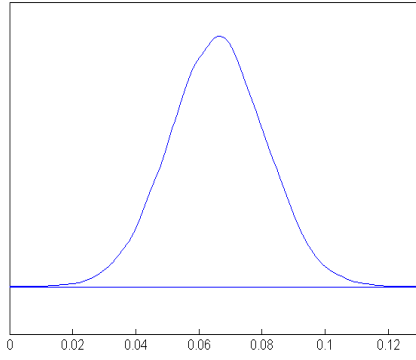


Figure 1: Posterior distribution of spending – linear model

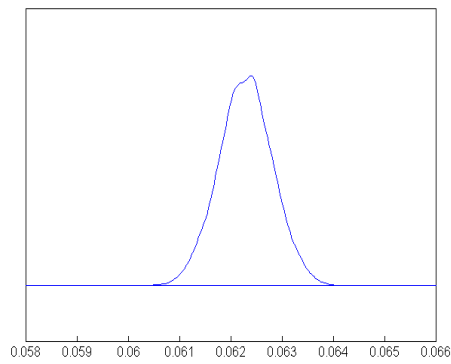


Figure 2: Posterior distribution of spending – fractional probit model

would enroll their children into higher performing A or B schools, which as illustrated in Table 7, spend less money on average than their lower performing counterparts. Thus a higher adult education rate may be associated with lower school spending but higher school quality.

Finally, Table 12 presents the estimated results of the structural equation with endogenous spending and 4th grade reading pass rates as the dependent variable. Once again the discussion is limited to the estimates associated with the 4th grade reading exam. Estimates for the 5th grade math exam are reported in Table 13 and are very similar to those reported in Table 12.

Similar to the baseline results, I find a spending effect that is positive and statistically significant; however, compared to the baseline model, this effect is much larger in magnitude. In the linear IV model, the posterior mean for EXPEND is 0.101, indicating that a \$1,000 increase in per pupil spending would lead to a 10.1% increase in reading pass rates. This is more than 3 percentage points larger than the estimated effect from the baseline linear model.

In the fractional probit IV model the posterior mean for school spending is 0.268, which is more than 1.5 times larger than the spending coefficient in the baseline probit model. To interpret this coefficient I calculate the marginal effects (reported in the last set of columns of Table 12), and find an estimated spending effect (posterior mean of the marginal effect of spending) of 9.6%. Thus, I find that the estimated effects of both the linear IV and fractional probit IV models are similar, although the probit estimates are slightly smaller, as are their standard errors.

Table 11: Posterior estimates for 4th grade reading. Expenditure equation

Variable	Linear Model			Fractional Probit Model		
	Coefficient			Coefficient		
	Mean	SD	HPDI	Mean	SD	HPDI
TAXES	7.24	(0.20)	(6.75, 7.54)	8.45	(0.26)	(8.04, 9.04)
PARENT EDUC	-1.55	(0.37)	(-2.25, -0.82)	-1.64	(0.35)	(-2.40, -0.99)
BLACK	1.03	(0.16)	(0.70, 1.34)	1.12	(0.17)	(0.75, 1.40)
HISPANIC	-0.44	(0.25)	(-0.92, 0.04)	-0.38	(0.26)	(-0.89, 0.16)
ENGLISH	1.22	(0.19)	(0.857, 1.59)	1.59	(0.24)	(1.15, 2.09)
INSTRUCT	-0.56	(0.18)	(-0.91, -0.21)	-1.16	(0.18)	(-1.52, -0.83)
GIFT	1.25	(0.15)	(0.95, 1.54)	1.40	(0.23)	(0.94, 1.84)
DISAB	1.34	(0.13)	(1.096, 1.61)	0.94	(0.18)	(0.58, 1.26)
ABSENT	-0.022	(0.22)	(-0.45, 0.41)	-0.20	(0.23)	(-0.66, 0.23)
DEGREE	0.37	(0.07)	(0.23, 0.49)	0.33	(0.07)	(0.18, 0.47)

For the IV models, the posterior distributions of the marginal effects of spending are plotted in Figures 3 and 4. For the linear IV model (Figure 3), the spending effect ranges from 0.015 to 0.192, while for the fractional probit IV model (Figure 4), the posterior distribution of the marginal effect ranges from 0.0748 to 0.1095. Similar to the baseline models, these plots indicate that the spending effect is always positive as the entire posterior distribution is above zero in both models; though the estimated marginal effects are more than 50% larger for the IV models. Observing IV estimates above the baseline estimates is consistent with previous literature (Ferguson and Ladd, 1996; Dewey et al., 2000; Roy, 2003; Levacic et al., 2005; Papke and Wooldridge, 2008) and is also consistent with the idea that Florida schools were funded in a compensatory manner.

The IV results presented in Table 12 also indicate that parental education has a positive effect on student achievement. According to the fractional probit IV model, a 10% increase in the adult education rate is associated with a 3.6% increase in

reading pass rates. While this effect is nontrivial, it would be very difficult for a community to increase the adult education rate by 10%, therefore its relevance from a policy perspective is negligible. Of greater interest are the posterior estimates associated with INSTRUCT and DEGREE, two inputs that schools do have control over. I find that INSTRUCT has a strong positive effect on student outcomes, while DEGREE does not. Focusing on the fractional probit IV estimates, a 10% increase in the percentage of staff devoted to instruction (INSTRUCT) would, on average, increase reading pass rates by 4.8%, a relatively large effect; while a 10% increase in the percentage of teachers with advanced degrees would only increase pass rates by 0.35%, an almost nonexistent effect. It is important to note here that this does not imply that teacher quality is irrelevant; in fact numerous studies have found that variations in teacher quality have a large impact on student achievement (Murnane and Phillips, 1981; Hanushek, 1992; Rivkin et al., 2005). Rather this simply implies that teacher's education may not be a very strong indicator for teacher quality as was discussed earlier.

The IV results also indicate that GIFT has a large positive effect on reading pass rates, while the posterior mean of DISAB is now insignificant. This implies that schools were probably not classifying the poorer test takers into test-exempt special education programs just to boost FCAT scores, as was suggested by the baseline results. To investigate this further I look at the special education classification rate for all Florida elementary schools in the sample between 1997 and 2005. The average classification rates are reported in Table 14. Two extra years of data are included (1997 and 1998) in order to determine whether the introduction of either the FCAT exam-

inations (in 1998) or the implementation of school accountability reforms (in 1999) had any impact on the classification rate. If schools were “gaming the system” and classifying their poorer test takers into test-exempt special education programs there would likely be a large spike in the classification rate shortly after the introduction of the FCAT exams and/or after the A+ plan. Instead the classification rate was fairly stable throughout the sample period, thus providing little evidence to indicate that schools were “gaming the system.” The posterior mean of BLACK is still negative, which is expected, while HISPANIC is now positive. This is consistent with previous research indicating that the Hispanic population in Florida places a very strong emphasis on education (EDRFL, 2005), and also because I am already accounting for limited English proficiency. Furthermore, Hanushek and Raymond (2004) found that, while school accountability reforms such as the *A+ plan for education* led to wider achievement gaps between Black and White students, they actually narrowed the Hispanic-White achievement gap.

Table 12: Posterior estimates for 4th grade reading. IV model with endogenous spending - structural equation

Variable	Linear Model			Fractional Probit Model			
	Mean	SD	HPDI	Coefficient	Mean	SD	HPDI
EXPEND (thousands)	0.101	(0.015)	(0.071, 0.131)	0.268	(0.012)	(0.005)	(0.086, 0.104)
PARENT EDUC	0.561	(0.406)	(-0.270, 1.490)	1.009	(0.101)	(0.037)	(0.293, 0.434)
BLACK	-0.162	(0.054)	(-0.270, -0.062)	-0.700	(0.066)	(0.024)	(-0.294, -0.198)
HISPANIC	0.454	(0.072)	(0.303, 0.592)	0.628	(0.129)	(0.046)	(0.149, 0.338)
ENGLISH	0.070	(0.040)	(-0.008, 0.147)	-0.035	(0.066)	(0.024)	(-0.056, 0.038)
INSTRUCT	0.404	(0.041)	(0.322, 0.487)	1.342	(0.056)	(0.022)	(0.435, 0.522)
GIFT	0.516	(0.038)	(0.439, 0.587)	1.450	(0.079)	(0.030)	(0.462, 0.579)
DISAB	-0.023	(0.057)	(-0.142, 0.087)	0.046	(0.063)	(0.022)	(-0.026, 0.062)
ABSENT	-0.002	(0.044)	(-0.091, 0.084)	0.005	(0.063)	(0.023)	(-0.042, 0.046)
DEGREE	0.042	(0.015)	(0.012, 0.071)	0.099	(0.021)	(0.008)	(0.021, 0.050)
Variance Parameters							
σ_a^2	0.012	(0.004)		1.01	(0.01)		
σ_a^2	0.707	(0.606)		5.48	(1.57)		
σ_ε^2	0.336	(0.108)		0.36	(0.17)		
σ_b^2	1.019	(0.416)		0.90	(0.49)		
$\sigma_{u\varepsilon}$	-0.022	(0.01)		-0.06	(0.01)		

Table 13: Posterior estimates for 5th grade math. IV model with endogenous spending - structural equation

Variable	Linear Model				Fractional Probit Model				
	(1)		(2)		(3)				
	Mean	SD	HPDI	Coefficients	Mean	SD	Mean	SD	HPDI
EXPEND (thousands)	0.068	(0.015)	(0.038, 0.10)		0.189	(0.009)	0.069	(0.003)	(0.062, 0.076)
PARENT EDUC	0.532	(0.409)	(-0.288, 1.484)		0.908	(0.082)	0.333	(0.030)	(0.277, 0.393)
BLACK	-0.205	(0.054)	(-0.312, -0.104)		-0.764	(0.052)	-0.280	(0.019)	(-0.315, -0.235)
HISPANIC	0.329	(0.071)	(0.187, 0.467)		0.317	(0.107)	0.116	(0.039)	(0.055, 0.226)
ENGLISH	0.047	(0.039)	(-0.028, 0.122)		0.003	(0.050)	0.001	(0.018)	(-0.031, 0.042)
INSTRUCT	0.195	(0.041)	(0.113, 0.277)		0.710	(0.047)	0.261	(0.017)	(0.223, 0.292)
GIFT	0.626	(0.037)	(0.551, 0.698)		1.644	(0.072)	0.603	(0.027)	(0.553, 0.657)
DISAB	0.043	(0.057)	(-0.078, 0.156)		0.189	(0.057)	0.069	(0.021)	(0.029, 0.111)
ABSENT	-0.222	(0.043)	(-0.307, -0.138)		-0.630	(0.058)	-0.231	(0.021)	(-0.273, -0.189)
DEGREE	-0.020	(0.015)	(-0.049, 0.010)		-0.068	(0.020)	-0.025	(0.007)	(-0.039, -0.011)
Variance Parameters									
σ_u^2	0.011	(0.003)			1.00	0.002			
σ_a^2	0.702	(0.613)			8.32	2.45			
σ_ε^2	0.335	(0.107)			0.33	0.10			
σ_b^2	1.006	(0.412)			0.82	0.57			
$\sigma_{u\varepsilon}$	-0.013	(0.014)			-0.03	0.01			

Table 14: Special education classification rates, 1997-2005

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005
Classification Rate	14.5%	14.9%	15.1%	15.4%	15.6%	15.5%	15.9%	16.0%	15.8%

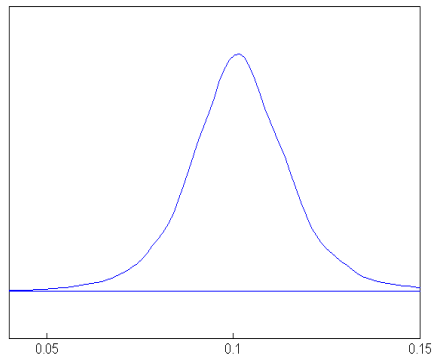


Figure 3: Posterior distribution of spending – linear IV model

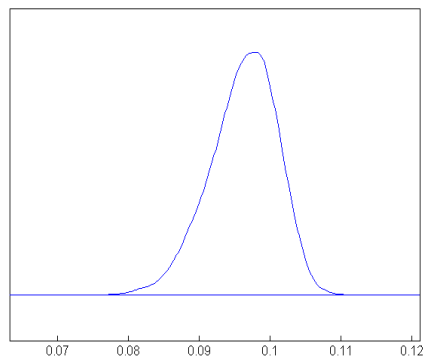


Figure 4: Posterior distribution of spending – fractional probit IV model

5 Model comparison

In this paper I consider two fractional probit models with correlated random effects: the single equation baseline model which assumes that all explanatory variables are exogenous and the simultaneous equation IV model which allows for endogeneity. The single equation model is obtained by restricting the covariance parameter, $\delta_{u\varepsilon}$, to zero while the IV model leaves $\delta_{u\varepsilon}$ unconstrained. Thus I can perform a formal test of endogeneity by setting $\delta_{u\varepsilon}$ to zero and testing the null hypothesis $H_0 : \delta_{u\varepsilon} = 0$ against the alternative $H_1 : \delta_{u\varepsilon} \neq 0$.

Denoting M_1 as the constrained model and M_2 as the unconstrained model, the Bayes factor for the null hypothesis is defined as

$$B_{1,2} = \frac{m(y | M_1)}{m(y | M_2)},$$

where $m(y|M_j)$ is the marginal likelihood of the model specification M_j , $j = 1, 2$. Since M_1 is simply a nested form of M_2 The Bayes factor, $BF_{1,2}$, can be calculated by using the Savage-Dickey density ratio approach (Verdinelli and Wasserman, 1995). Specifically,

$$B_{1,2} = \frac{p(\delta_{u\varepsilon}^* | data)}{p(\delta_{u\varepsilon}^*)},$$

where $p(\delta_{ue}^* | data)$ is the posterior density of the covariance parameter, δ_{ue} , and $p(\delta_{ue}^*)$ is the prior of δ_{ue} calculated at the points $\delta_{ue}^* = 0$. Estimating the prior density at δ_{ue}^* is straightforward, however, the unconditional posterior density $p(\delta_{ue}^* | data)$ is unknown and must be estimated using the output from the MCMC simulation. The posterior density of δ_{ue} can be estimated by averaging the full conditional densities over the number of MCMC draws and conditioning on the model parameters and augmented data (Deb, Munkin and Trivedi, 2009).¹⁹ This can be written as:

$$\tilde{p}(\delta_{ue}|data) = \frac{1}{R} \sum_{r=1}^R p(\delta_{ue} | y_{its}^{*(R)}, a_i^{(R)}, b_i^{(R)}, \boldsymbol{\Omega}^{(R)}, \boldsymbol{\Upsilon}^{(R)}, \sigma_a^{2(R)}, \sigma_b^{2(R)}, \delta_\varepsilon^R),$$

which should be evaluated at δ_{ue}^* .

Similar tests can be imposed in order to test whether the correlated random effects specification is appropriate by restricting $\boldsymbol{\lambda}$ to zero in the baseline models, and both $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ to zero in the IV models. In total there are three possible specification tests that need to be implemented. Each test compares a nested model with a non-nested model, and therefore the Savage-Dickey density ratio approach can be applied to all three.

First, I test whether the correlated random effects specification is appropriate in the baseline models by testing the joint null hypothesis $H_0 : \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_T = 0$ against the alternative which leaves these parameters unconstrained. The test for the joint null hypothesis $H_0 : \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_T = 0$ is strongly rejected for both models as the posterior mean and standard deviation of the estimated Bayes factor is 2.94×10^{-21} ($2.21 \times$

¹⁹Note: In the linear model, the posterior of δ_{ue}^* would be conditioned with respect to all of the same model parameters except y_{its}^* .

10^{-19}) for the linear model and 3.34×10^{-51} (4.78×10^{-52}) for the probit model. Thus providing overwhelming support in favor of the CRE specification.

For the IV models, I test whether the correlated random effects specification is appropriate in the outcome equation by testing the joint null hypothesis $H_0 : \lambda_1, \dots, \lambda_T = 0$ against the alternative which leaves these parameters unconstrained, and for the expenditure equation I test the joint null hypothesis of $H_0 : \mu_1, \dots, \mu_T = 0$. In both the linear model and the fractional probit model, the null hypotheses are strongly rejected with posterior means and standard deviations of less than 0.001, again providing evidence in favor of the CRE specifications.

To determine whether spending is endogenous I focus on the covariance parameter, $\delta_{u\varepsilon}$, which captures dependence between the error term in the outcome equation, u , and the error term in the expenditure equation ε . In Figures 5 and 6 I plot the posterior distribution of $\delta_{u\varepsilon}$. If spending were truly exogenous, the posterior distribution of $\delta_{u\varepsilon}$ would be centered at zero. However, in the linear model (Figure 5) the posterior is centered at -0.0761 and is separated from zero by more than two standard deviations (0.029), and in the probit model (Figure 6) $\delta_{u\varepsilon}$ is centered at -0.178 which is more than 4 standard deviations (0.037) away from zero. This provides some evidence of endogeneity. As a formal test I calculate the Bayes factor with null hypothesis $H_0 : \delta_{u\varepsilon} = 0$ against $H_A : \delta_{u\varepsilon} \neq 0$. The posterior mean and standard deviation of the estimated Bayes factor is 0.0014 (0.00023) for the linear model and 0.00136 (0.000195) for the fractional probit, both of which reject the null hypothesis of no endogeneity, thus providing evidence in favor of using the IV models over the baseline methods.

Finally, in Figure 7 I plot the marginal effects of spending at different spending

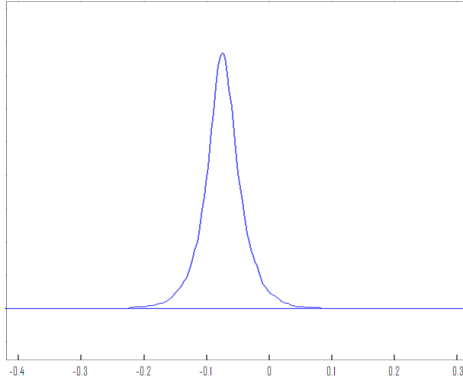


Figure 5: Posterior distribution of δaUE – linear IV model

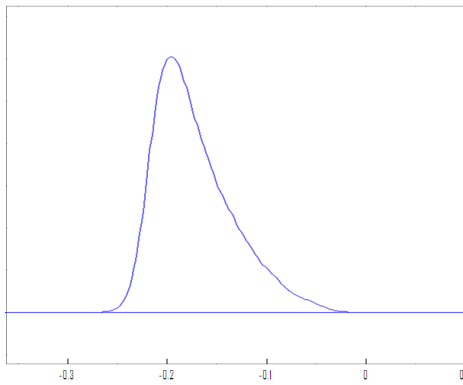


Figure 6: Posterior distribution of δaUE – fractional probit IV model

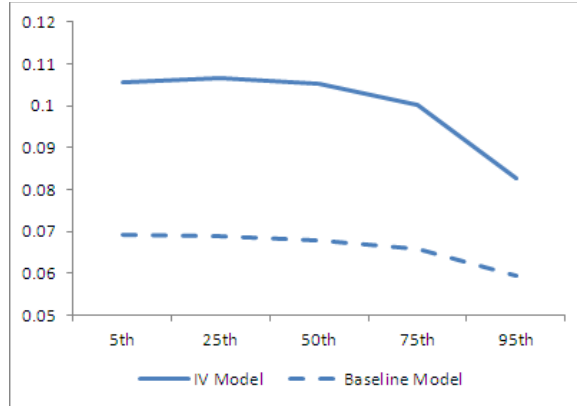


Figure 7: Marginal effects (of school spending) at different percentiles of spending fractional probit models

levels, using the fractional probit models. This is done to assess the importance of using a nonlinear model to allow for diminishing marginal returns of spending. The estimated marginal effects are calculated at the 5th, 25th, 50th, 75th, and 95th percentiles of the spending distribution, while all other explanatory variables are averaged over N and T . For the baseline model, the estimated marginal effects are calculated using Equation (10), while the IV model estimates are calculated using Equation (21). For both models I find that the marginal effects are larger among schools with below median spending levels, and as spending increases above the 50th percentile the marginal effects decrease. This indicates that a \$1,000 increase in spending will have a larger impact on pass rates among lower spending schools than it would among the higher spending schools. Therefore, to capture this diminishing marginal effect of spending on test pass rates, a nonlinear specification seems more appropriate than the traditional linear specification.

6 Conclusion

In this paper I used various models to examine the relationship between school spending and test pass rates among Florida elementary schools. For all models, Bayesian estimation methods were proposed through the use of Gibbs sampling (and data augmentation in the fractional probit models), which allowed for efficient estimation of all parameters of interest.

In the empirical analysis I did not examine the effects of school accountability programs such as the A+ plan for education and the NCLB nor did I attempt to identify any “teaching to the test” phenomenon. Rather, the main focus of the analysis was to quantify a causal relationship between school spending and student achievement.

In all model specifications I found that real school spending had a positive and statistically significant effect on student achievement. When estimating the average effect of spending, I found that the linear estimates and nonlinear fractional probit estimates were very similar, although the nonlinear estimates were slightly smaller in magnitude and more precise. When estimating the marginal effects of spending at various spending levels, I found evidence of diminishing marginal effects, giving greater motivation for the nonlinear models which allow for diminishing returns. Furthermore, the standard errors of all variables were much smaller in magnitude in

the nonlinear specifications than in the linear models and therefore specifying the fractional probit led to large gains in efficiency.

Using the MCMC algorithms proposed in this paper, it was rather straightforward to obtain slope coefficients and marginal effects for the nonlinear fractional probit models, however, estimation required the introduction of an augmented dataset which increased the dimensions of the parameter space by $S \times T$ observations per school. This can greatly increase computational time. However, if one needs to obtain more precise estimates of the average effects and/or estimates beyond the average effects, such as marginal effects across the distribution of a particular variable or variables, then the fractional probit models may be better suited.

Finally, in both the linear and nonlinear specifications, allowing for potential endogeneity of spending led to estimated spending effects that were roughly 50% larger than those found in the models which assume spending is strictly exogenous. For instance, in the single-equation baseline models the estimated effect of a \$1,000 increase in per pupil spending was an average increase in pass rates ranging from 6.2% (fractional probit model) to 6.6% (linear model). Whereas, in the simultaneous equation IV models the estimated spending effect increased to 9.6-10.1%.

Based on the formal Bayes factor specification tests, I found strong evidence in favor of the conclusion that school spending was endogenously related to student achievement, and in this case, failure to account for endogeneity could lead to estimated spending effects which were biased downwards. Of course, the IV results relied on the validity of the chosen instrument, and I cannot dismiss the possibility that variations in district-level property taxes were not entirely exogenous. Furthermore, in

an effort to capture exogenous changes in the instrument, I measured property taxes at the more aggregate district level, but test pass rates at the school level. While, this may dissolve the relationship between property taxes and student achievement, it left me with data that had much less variation than would data at the school level. Therefore, leading to weaker identification of the estimated spending effects. As a result the IV estimates were less precise than were the estimates from the baseline specification. However, this is a typical occurrence in any IV model regardless of data aggregation.

As a final note, while I did find that increased spending had a fairly large positive impact on student achievement, these estimated effects were based on a \$1,000 increase in per-pupil spending. For the average school in the sample, this equates to a 20% increase in spending which is rather substantial. Therefore, even though a positive relationship between school expenditures and student achievement exists, this does not suggest that increasing school spending would be the most effective way to increase student achievement, only that it could be part of a more comprehensive solution.

References

- Albert, J., & Chib, S. (1993), Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88, 669-679.
- Abrevaya, J., & Dahl, C. (2008). The Effects of Birth Inputs on Birthweight: Evidence From Quantile Estimation on Panel Data. *Journal of Business & Economic Statistics*, 379-397.
- Altonji, J. G., & Matzkin, R. L. (2005). Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors. *Econometrica*, 73(4), 1053-1102.
- Bacolod, M. P., & Tobias, J. L. (2006), Schools, School Quality, and Achievement Growth: Evidence from the Philippines. *Economics of Education Review*, 25, 619-632.
- Bali, V. A., & Alvarez, M. R. (2004). The Race Gap in Student Achievement Scores: Longitudinal Evidence from a Racially Diverse School District. *Policy Studies Journal*, 32:3. 393-415.
- Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 577-599.

- Bogart, W. T., & Cromwell, B. A. (1997). How Much More is a Good School District Worth? *National Tax Journal*, 50, 215-232.
- Buddin, R., & Zamarro, G. (2009). Teacher Qualifications and Student Achievement in Urban Elementary Schools. *Journal of Urban Economics* 66:2, 103-15.
- Carey, K. (1997). A Panel Data Design for Estimation of Hospital Cost Functions. *Review of Economics and Statistics*, 443-453.
- Carey, K. (2000). Hospital Cost Containment and Length of Stay: An Econometric Analysis. *Southern Economic Journal*, 363-380.
- Carlin, B. P. (1996). Hierarchical Longitudinal Modeling. In *Markov Chain Monte Carlo in Practice*. W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.), 303-319.
- Chakrabarti, R. (2007). Vouchers, Public School Response and the Role of Incentives: Evidence from Florida. Staff Report No. 306. Federal Reserve Bank of New York.
- Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, 5-46.
- Chamberlain, G. (1984). Panel data. In *Handbook of Econometrics*, Vol 2. Z. Griliches and M. D. Intriligator.
- Chao, J., & Phillips, P. (1998). Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys Prior. *Journal of Econometrics*, 49-86.
- Chiang, H. (2009). How Accountability Pressure on Failing Schools Affects Student Achievement. *Journal of Public Economics*, 93, 1045-57.

- Chib, S., & Carlin, B. (1999). On MCMC Sampling in Hierarchical Longitudinal Models. *Statistics and Computing*, 17-26.
- Chib, S., & Greenberg, E. (1996). Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometric Theory*, 409-431.
- Coleman, J., Campbell, E., Hobson, D., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health Education and Welfare.
- Cullen, J., & Reback, R. (2006). Tinkering Toward Accolades: School Gaming Under a Performance Accountability System. In *Advances in Applied Microeconomics*. T. J. Gronberg & D. W. Jansen.
- Davis-Keane, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, 19:2, 294-304.
- Deb, P., Munkin, M., & Trivedi, P. (2006). Bayesian Analysis of the Two-Part Model with Endogeneity: Application to Health Care Expenditure. *Journal of Applied Econometrics*, 1081-1099.
- Dewey, J., Husted, T. A., & Kenny, L. W. (2000). The Ineffectiveness of School Inputs: A Product of Misspecification? *Economics of Education Review*, 19, 27-45.
- EDRFL (2005). *Characteristics of Students by Place of Birth and Language Spoken in the Home: Florida Public Schools, Grades PK-12, 2003-04 School Year*. Tallahassee: Office of Economic and Demographic Research.

- Ferguson, R. F. (1991). Paying for Public Education: New Evidence on How and Why Money Matters. *Harvard Journal on Legislation* 28, 465-488.
- Ferguson, R. F., & Ladd, H. F. (1996). How and Why Money Matters: An Analysis of Alabama Schools. In *Holding Schools Accountable: Performance Based Reform in Education*. H. F. Ladd. Washington, DC: Brookings Institution.
- Figlio, D., & Getzler, L. (2002). Accountability, Ability and Disability: Gaming the System? National Bureau of Economic Research, W9307, Cambridge, MA.
- Figlio, D., & Lucas, M. (2004). What's in a grade? School Report Cards and the Housing Market. *American Economic Review*, 591-604.
- Florida Department of Education (FLDOE) Teacher Salary, Experience, and Degree Level 2004-05.
- Fryer, R., & Levitt, S. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *Review of Economics and Statistics*, 86:2, 447-464.
- Gamerman, D., & Lopes, H. F. (2006). Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd ed.). Boca Raton: Taylor & Francis.
- Gardeazabal, J. (2010), Vote Shares in Spanish General Elections as a Fractional Response to the Economy and Conflict. *Economics of Security*. Working Paper Series 33, DIW Berlin, German Institute for Economic Research.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., & Raton, B. (2004). *Bayesian Data Analysis* (2nd ed.). Florida: Chapman & Hall.

- Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 721-741.
- Geweke, J. (1996). Bayesian Reduced Rank Regression in Econometrics. *Journal of Econometrics*, 121-146.
- Gill, J. (2002). *Bayesian Methods: a Social and Behavioral Sciences Approach*. Boca Raton: Chapman & Hall.
- Goldhaber, D., & Brewer, D. (1998). When Should We Award Degrees for Teachers? *Phi Delta Kappan*, 80:2 134-38.
- Greenwald, R., Hedges, L., & Laine, R. (1996). The Effect of School Resources on Student Achievement. *Review of Educational Research*, 361-396.
- Guryan, J. (2001). Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts. NBER Working Paper 8269.
- Gyimah-Brempong, K., & Gyapong, A. (1991). Characteristics of Education Production Functions: An Application of Canonical Regression Analysis. *Economics of Education Review* 10: 7-17.
- Hanna, R. (2010). US Environmental Regulation and FDI: Evidence from a Panel of US-Based Multinational Firms. *American Economic Journal: Applied Economics*, 2:3, 158-89.
- Hanushek, E. (1986). The Economics of Schooling - Production and Efficiency in Public Schools. *Journal of Economic Literature*, 1141-1177.
- Hanushek, E. (1992). The Trade-Off Between Child Quantity and Quality. *Journal of Political Economy* vol. 100:1, 84-117.

- Hanushek, E. (1996). Measuring Investment in Education. *Journal of Economic Perspectives*, 9-30.
- Hanushek, E. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, 141-164.
- Hanushek, E. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, F64-F98.
- Hanushek, E., Kain, J., & Rivkin, S. (2009). New Evidence About Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement. *Journal of Labor Economics*, University of Chicago Press, 27:3, 349-383.
- Hanushek, E. A., & Raymond, M. E. (2004). The Effect of School Accountability Systems on the Level and Distribution of Student Achievement. *Journal of the European Economic Association*, 2(2-3), 406-415.
- Hayes, K. J. & Taylor, L. L. (1996). Neighborhood School Characteristics: What Signals Quality to Homebuyers? *Federal Reserve Bank of Dallas Economic Review*, 3, 2-9.
- Heckman, J., Layne-Farrar, A., & Todd, P. (1996). Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings. *Review of Economics and Statistics*, 562-610.
- Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does Money Matter? A Metaanalysis of Studies of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher*, 23(3).

- Hobert, J., & Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association*, 1461-1473.
- Hoogerheide, L., Kleibergen, F., & van Dijk, H. (2007). Natural Conjugate Priors for the Instrumental Variables Regression Model Applied to the Angrist-Krueger Data. *Journal of Econometrics*, 63-103.
- Hujer, R., Grammig, J., & Schnabel, R. (1994). A Comparative Empirical Analysis of Labor Supply and Wages of Married Women in the FRG and the USA - A Microeconomic Study Using SEP and PSID Panel Data. *Jahrbucher Fur Nationalokonomie Und Statistik*, 129-147.
- Imbens, G. W. & Wooldridge, J. M. (2007). What's New in Econometrics? NBER Research Summer Institute, Cambridge, July/August, 2007.
- Implementation of Florida's System of School Improvement and Accountability. (1999). Florida Department of Education Rule 6A-1.09981.
- Ingels, S. J., Dowd, K. L., Baldrige, J. D., Stipe, J. L., Bartot, V. H., & Frankel, M. R. (1994). *National Education Longitudinal Study of 1988: Second follow-up*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.
- Islam, N. (1995). Growth Empirics - A Panel Data Approach. *Quarterly Journal of Economics*, 1127-1170.
- Jacob, B. A. (2005). Accountability, Incentives and Behavior: Evidence from School Reform in Chicago. *Journal of Public Economics*, 89(5-6), 761-796.

- Jakubson, G. (1988). The Sensitivity of Labor Supply Parameter Estimates to Unobserved Individual Effects - Fixed Effects and Random Effects Estimates in a Nonlinear Model Using Panel Data. *Journal of Labor Economics*, 302-329.
- Jencks, C., Smith, M., Ackland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A Reassessment of the Effects of Family and Schooling in America*. New York: Basic Books.
- Jenkins, A., Levacic, R., & Vignoles, A. (2006). Estimating the Relationship Between School Resources and Pupil Attainment at GCSE. Department for Education and Skills.
- Jepsen, C., & Rivkin, S. (2002). *Class Size Reduction, Teacher Quality, and Academic Achievement in California Public Elementary Schools*. San Francisco: Public Policy Institute of California.
- Kane, T., Staiger, D., Samms, G. (2003). *School Accountability Ratings and Housing Values*. Brookings-Wharton Papers on Urban Affairs.
- Kleibergen, F., & van Dijk, H. (1998). Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures. *Econometric Theory*, 701-743.
- Kleibergen, F., & Zivot, E. (2003). Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics*, 29-72.
- Knight, M., Loayza, N., & Villanueva, D. (1993). Testing the Neoclassical Theory of Economic Growth - A Panel Data Approach. *International Monetary Fund Staff Papers*, 512-541.
- Koop, G., Poirier, D., & Tobias, J. (2007). *Bayesian Econometric Methods*. Cambridge: Cambridge University Press.

- Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114:2, 497-532.
- Lancaster, T. (2000). The Incidental Parameter Problem Since 1948. *Journal of Econometrics*, 391-413.
- Levacic, R., Jenkins, A., Vignoles, A., & Allen, R. (2005). The Effect of School Resources on Student Attainment in English Secondary Schools. Institute of Education and Centre for Economics of Education, Institute of Education.
- Li, K.(1998). Bayesian Inference in a Simultaneous Equation Model with Limited Dependent Variables. *Journal of Econometrics*, 85, 387-400.
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* 73, 13-22.
- Lindley, D., & Smith, A. (1972). Bayes Estimates for Linear Model. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 34:1, 1-41.
- Ludwig, J., & Bassi, L. J. (1999). The Puzzling Case of School Resources and Student Achievement. *Educational Evaluation and Policy analysis*, 21:4, 385-403.
- Mayer, S. (1997). *What Money Can't Buy*. Cambridge, MA: Harvard University Press.
- McCabe, M. J., & Snyder, C. M. (2011), 'Did Online Access to Journals Change the Economics Literature?' (January 23, 2011). Available at SSRN: <http://ssrn.com/abstract=1746243> or <http://dx.doi.org/10.2139/ssrn.1746243>.

- Mensah, Y. M., Schoderbek, M. P. & Werner, R. H. (2005). Public School Spending, Functional Cost Classifications, and Student Performance: A Simultaneous Equations Approach. Rutgers University Department of Accounting and Information Systems.
- Mundlak, Y. (1978). Pooling of Time-Series and Cross-Section Data. *Econometrica*, 69-85.
- Murnane, R. J., & Phillips, B. (1981). What do Effective Teachers of Inner-City Children Have in Common? *Social Science Research*, 10:1, 83-100.
- National Center for Education Statistics. (1995). The Condition of Education 1995: The Educational Progress of Hispanic Students. [Report 95767]. Washington DC: Author.
- National Center for Education Statistics. (1998). The Condition of Education 1998: The Educational Progress of Hispanic Students. [Report 98013]. Washington DC: Author.
- Neal, D. (2006). Why has Black-White Skill Convergence Stopped? In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: Elsevier.
- Neyman, J., & Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 1-32.
- Nguyen, H. B. (2010), Estimating a Fractional Response Model with a Count Endogenous Regressor and an Application to Female Labor Supply. In W. Greene and R. C. Hil (eds, Emerald Group Publishing Limited), *Maximum Simulated Likelihood Methods and Applications Advances in Econometrics*, Vol 26, 253-298.

- Papke, L. (2005). The Effects of Spending on Test Pass Rates: Evidence from Michigan. *Journal of Public Economics*, 821-839.
- Papke, L., & Wooldridge, J. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics*, 11, 619-632.
- Papke, L., & Wooldridge, J. (2008). Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates. *Journal of Econometrics*, 121-133.
- Phillips, M. (2000). Understanding Ethnic Differences in Academic Achievement: Empirical Lessons from National Data. In *Analytic Issues in the Assessment of Student Achievement*. D. Grissmer, & M. Ross. Washington DC: Department of Education, National Center for Education Statistics. 103–132.
- Raftery, A. E., and Lewis, S. M. (1992). How Many Iterations in the Gibbs Sampler? in *Bayesian Statistics*, Vol. 4. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith eds. Oxford University Press: Oxford, 763-773.
- Rivkin, S. G. (1995). Black/White Differences in Schooling and Employment. *Journal of Human Resources*, 30:4, 826-852.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73:2, 417-458.
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What Large-Scale Survey Research Tells us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record*, 104, 1525–1567.

- Roy, J. (2003). Impact of School Finance Reform on Resource Equalization and Academic Performance: Evidence from Michigan. Princeton University, Education Research Section Working Paper No. 8.
- Tanner, M. A., & Wong, W. (1987), The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*. 82, 528-550.
- Tiebout, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy* 64, 416-24.
- U.S. Department of Education. (2005). National Assessment of Educational Progress (NAEP). National Center for Education Statistics. <http://nces.ed.gov/nationsreportcard>.
- Verdinelli, I., & Wasserman, L. (1995), Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, 90, 614-618.
- Webbink, D. (2005). Causal Effects in Education. *Journal of Economic Surveys*, 535-560.
- Weimer, D., & Wolkoff, M. (2001). School Performance and Housing Values: Using Non-Contiguous District and Incorporation Boundaries to Identify School Effects. *National Tax Journal*, 231-253.
- Wojtkiewicz, R. A., & M. Donato, K. M. (1995). Hispanic Educational Attainment: The Effects of Family Background and Nativity. *Social Forces*, 74, 559-574.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Wooldridge, J.M. (2005). Unobserved Heterogeneity and Estimation of Average Partial Effects. In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, ed. D.W.K. Andrews and J.H. Stock. Cambridge: Cambridge University Press, 27-55.

Appendices

Appendix 1: MCMC calculations for fractional probit baseline model

The full augmented joint posterior can be written as

$$\begin{aligned}
& \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i) \right] \times \\
& \prod_{i=1}^N \prod_{t=1}^T \left(\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right) \times \\
& (2\pi)^{-(1+Tk+k+h)/2} |\mathbf{H}_{\boldsymbol{\Omega}}|^{1/2} \exp[-.5(\boldsymbol{\Omega} - \underline{\boldsymbol{\Omega}})' \mathbf{H}_{\boldsymbol{\Omega}}(\boldsymbol{\Omega} - \underline{\boldsymbol{\Omega}})] \times \\
& \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp(-.5a_i'\sigma_a^{-2}a_i) \times \frac{1}{\Gamma(a_a)b_a^{a_a}} (\sigma_a^2)^{-(a_a+1)} \exp\left(-\frac{1}{b_a\sigma_a^2}\right). \quad (22)
\end{aligned}$$

The steps of the MCMC algorithm are as follows:

1. The conditional posterior kernel for the latent variable y_{its}^* is normally distributed as $y_{its}^* \mid y_{it}, \mathbf{W}_{it}, a_i, \boldsymbol{\Omega}, \sigma_a^2 \sim N[\mathbf{W}_{it}\boldsymbol{\Omega} + a_i, 1]$, and is truncated at zero such that

$$y_{its}^* > 0 \text{ if } d_{its} = 1 \text{ and}$$

$$y_{its}^* \leq 0 \text{ if } d_{its} = 0.$$

This can be shown by collecting all terms related to y_{its}^* in (22):

$$\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i) \right] \times$$

$$\prod_{i=1}^N \prod_{t=1}^T \left(\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right)$$

which immediately implies that the latent variables y_{its}^* are conditionally independent with normal distribution $y_{its}^* \mid y_{it}, \mathbf{W}_{it}, a_i, \boldsymbol{\Omega}, \sigma_a^2 \sim N[\mathbf{W}_{it}\boldsymbol{\Omega} + a_i, 1]$ and each variable is truncated at zero.

2. The full conditional density for a_i is normally distributed as, $a_i \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, \boldsymbol{\Omega}, \sigma_a^2 \sim N[\bar{a}_i, \bar{H}_a^{-1}]$, where

$$\bar{H}_a = S \times T + \sigma_a^{-2}$$

$$\bar{a}_i = \bar{H}_a^{-1} \left[\sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega}) \right) \right].$$

To derive this full conditional density one must use the full posterior density kernel (22), dropping any terms that are free of a_i :

$$\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i) \right] \times$$

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp [-.5a_i'\sigma_a^{-2}a_i].$$

Focusing only on the terms in the brackets, and once again dropping terms free of a_i ,

I am left with

$$\exp \left(-0.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i) + a_i' \sigma_a^{-2} a_i] \right).$$

Then I can rewrite the term in the bracket as

$$\begin{aligned} & [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i) + a_i' \sigma_a^{-2} a_i] \\ = & y_{its}^* y_{its}^* - y_{its}^* \mathbf{W}_{it}\boldsymbol{\Omega} - y_{its}^* a_i - \boldsymbol{\Omega}' \mathbf{W}_{it}' y_{its}^* + \boldsymbol{\Omega}' \mathbf{W}_{it}' \mathbf{W}_{it}\boldsymbol{\Omega} \\ & + \boldsymbol{\Omega}' \mathbf{W}_{it}' a_i - a_i' y_{its}^* + a_i' \mathbf{W}_{it}\boldsymbol{\Omega} + a_i' a_i + a_i' \sigma_a^{-2} a_i. \end{aligned}$$

Again dropping any terms multiplicatively unrelated to a_i :

$$\begin{aligned} & - y_{its}^* a_i + \boldsymbol{\Omega}' \mathbf{W}_{it}' a_i - a_i' y_{its}^* + a_i' \mathbf{W}_{it}\boldsymbol{\Omega} + a_i' a_i + a_i' \sigma_a^{-2} a_i \\ = & a_i' (1 + \sigma_a^{-2}) a_i - a_i' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega}) - (y_{its}^* - \boldsymbol{\Omega}' \mathbf{W}_{it}') a_i \end{aligned}$$

Adding over s and t :

$$\begin{aligned} & a_i' (S \times T + \sigma_a^{-2}) a_i - a_i' \left[\sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega}) \right) \right] \\ & - \left[\sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \boldsymbol{\Omega}' \mathbf{W}_{it}') \right) \right] a_i \\ = & a_i' \bar{H}_a a_i - a_i' \bar{H}_a \bar{a}_i - \bar{a}_i \bar{H}_a a_i. \end{aligned}$$

Then I complete the square, by adding and subtracting the term $\bar{a}_i' \bar{H}_a \bar{a}_i$:

$$\begin{aligned}
&= a_i' \bar{H}_a a_i - a_i' \bar{H}_a \bar{a}_i - \bar{a}_i' \bar{H}_a a_i + \bar{a}_i' \bar{H}_a \bar{a}_i - \bar{a}_i' \bar{H}_a \bar{a}_i \\
&= (a_i - \bar{a}_i)' \bar{H}_a (a_i - \bar{a}_i) - \bar{a}_i' \bar{H}_a \bar{a}_i.
\end{aligned}$$

Since the last term is free of a_i , this simplifies to

$$(a_i - \bar{a}_i)' \bar{H}_a (a_i - \bar{a}_i).$$

From here it is straightforward to see that $a_i \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, \boldsymbol{\Omega}, \sigma_a^2 \sim N \left[\bar{a}_i, \bar{H}_a^{-1} \right]$.

3. The full conditional of parameter vector $\boldsymbol{\Omega}$ can be shown to be $\boldsymbol{\Omega} \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, a_i, \sigma_a^2 \sim N \left[\bar{\boldsymbol{\Omega}}, \bar{\mathbf{H}}_{\boldsymbol{\Omega}}^{-1} \right]$, where

$$\begin{aligned}
\bar{\mathbf{H}}_{\boldsymbol{\Omega}} &= \mathbf{H}_{\boldsymbol{\Omega}} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}_{it}' \mathbf{W}_{it} \\
\bar{\boldsymbol{\Omega}} &= \bar{\mathbf{H}}_{\boldsymbol{\Omega}}^{-1} \left[\mathbf{H}_{\boldsymbol{\Omega}} \boldsymbol{\Omega} + \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}_{it}' \left(\sum_{s=1}^S (y_{its}^* - a_i) \right) \right].
\end{aligned}$$

To derive this I collect the terms of the full posterior density kernel that are multiplicatively related to $\boldsymbol{\Omega}$:

$$\begin{aligned}
&\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i)' (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i) \right] \times \\
&(2\pi)^{-(1+T^k+k+h)/2} |\mathbf{H}_{\boldsymbol{\Omega}}|^{1/2} \exp \left[-.5 (\boldsymbol{\Omega} - \bar{\boldsymbol{\Omega}})' \bar{\mathbf{H}}_{\boldsymbol{\Omega}} (\boldsymbol{\Omega} - \bar{\boldsymbol{\Omega}}) \right].
\end{aligned}$$

Dropping terms unrelated to Ω :

$$\exp \left(-.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it}\Omega - a_i)' (y_{its}^* - \mathbf{W}_{it}\Omega - a_i) + (\Omega - \underline{\Omega})' \underline{\mathbf{H}}_{\Omega} (\Omega - \underline{\Omega})] \right).$$

Then I can rewrite the term in the bracket as

$$\begin{aligned} & [(y_{its}^* - \mathbf{W}_{it}\Omega - a_i)' (y_{its}^* - \mathbf{W}_{it}\Omega - a_i) + (\Omega - \underline{\Omega})' \underline{\mathbf{H}}_{\Omega} (\Omega - \underline{\Omega})] \\ = & y_{its}^* y_{its}^* - y_{its}^* \mathbf{W}_{it}\Omega - y_{its}^* a_i - \Omega' \mathbf{W}'_{it} y_{its}^* + \Omega' \mathbf{W}'_{it} \mathbf{W}_{it}\Omega + \Omega' \mathbf{W}'_{it} a_i \\ & - a_i' y_{its}^* + a_i' \mathbf{W}_{it}\Omega + a_i' a_i + \Omega' \underline{\mathbf{H}}_{\Omega} \Omega - \Omega' \underline{\mathbf{H}}_{\Omega} \underline{\Omega} + \underline{\Omega}' \underline{\mathbf{H}}_{\Omega} \underline{\Omega}. \end{aligned}$$

Dropping terms unrelated to Ω :

$$\begin{aligned} & - y_{its}^* \mathbf{W}_{it}\Omega - \Omega' \mathbf{W}'_{it} y_{its}^* + \Omega' \mathbf{W}'_{it} \mathbf{W}_{it}\Omega + \Omega' \mathbf{W}'_{it} a_i \\ & + a_i' \mathbf{W}_{it}\Omega + \Omega' \underline{\mathbf{H}}_{\Omega} \Omega - \Omega' \underline{\mathbf{H}}_{\Omega} \underline{\Omega} - \underline{\Omega}' \underline{\mathbf{H}}_{\Omega} \underline{\Omega} \\ = & \Omega' (\underline{\mathbf{H}}_{\Omega} + \mathbf{W}'_{it} \mathbf{W}_{it}) \Omega - \Omega' (\underline{\mathbf{H}}_{\Omega} \underline{\Omega} + \mathbf{W}'_{it} y_{its}^* + \mathbf{W}'_{it} a_i) \\ & - (\underline{\Omega}' \underline{\mathbf{H}}_{\Omega} + y_{its}^* \mathbf{W}_{it} + a_i' \mathbf{W}_{it}) \Omega. \end{aligned}$$

Adding over s , i , and t :

$$\begin{aligned} & \Omega' \left(\underline{\mathbf{H}}_{\Omega} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}'_{it} \mathbf{W}_{it} \right) \Omega - \Omega' \left(\underline{\mathbf{H}}_{\Omega} \underline{\Omega} + \sum_{i=1}^N \sum_{t=1}^T \left[\mathbf{W}'_{it} \sum_{s=1}^S (y_{its}^* + a_i) \right] \right) \\ & - \left(\underline{\Omega}' \underline{\mathbf{H}}_{\Omega} + \sum_{i=1}^N \sum_{t=1}^T \left[\sum_{s=1}^S (y_{its}^* + a_i) \mathbf{W}_{it} \right] \right) \Omega \\ = & \Omega' \overline{\mathbf{H}}_{\Omega} \Omega - \Omega' \overline{\mathbf{H}}_{\Omega} \overline{\Omega} - \overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \overline{\Omega}. \end{aligned}$$

To complete the square I add and subtract the term $\bar{\Omega}'\bar{\mathbf{H}}_\Omega\bar{\Omega}$:

$$\begin{aligned} &= \Omega'\bar{\mathbf{H}}_\Omega\Omega - \Omega'\bar{\mathbf{H}}_\Omega\bar{\Omega} - \bar{\Omega}'\bar{\mathbf{H}}_\Omega\Omega + \bar{\Omega}'\bar{\mathbf{H}}_\Omega\bar{\Omega} - \bar{\Omega}'\bar{\mathbf{H}}_\Omega\bar{\Omega} \\ &= (\Omega - \bar{\Omega})'\bar{\mathbf{H}}_\Omega(\Omega - \bar{\Omega}) - \bar{\Omega}'\bar{\mathbf{H}}_\Omega\bar{\Omega}. \end{aligned}$$

Since the last term is free of Ω , I am left with

$$(\Omega - \bar{\Omega})'\bar{\mathbf{H}}_\Omega(\Omega - \bar{\Omega}).$$

This shows that $\Omega \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, a_i, \sigma_a^2 \sim N \left[\bar{\Omega}, \bar{\mathbf{H}}_\Omega^{-1} \right]$.

4. The posterior distribution for σ_a^2 is inverse gamma:

$$\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, a_i, \Omega \sim IG \left(\frac{N}{2} + a_a, \left[b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a_i' a_i \right]^{-1} \right).$$

Collecting all terms related to σ_a^2 :

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp(-.5a_i'\sigma_a^{-2}a_i) \times \frac{1}{\Gamma(a_a)b_a^{a_a}} (\sigma_a^2)^{-(a_a+1)} \exp\left(-\frac{1}{b_a\sigma_a^2}\right).$$

Grouping like terms together, and ignoring terms unrelated to σ_a^2 I am left with

$$(\sigma_a^2)^{-\left(\frac{N}{2}+a_a+1\right)} \exp\left[-.5\sigma_a^{-2} (2b_a^{-1} + a_i' a_i)\right].$$

where, adding over i , I find that $\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, y_{its}^*, a_i, \boldsymbol{\Omega} \sim IG[\bar{a}_a, \bar{b}_a]$, where

$$\bar{a}_a = \frac{N}{2} + a_a \quad \text{and} \quad \bar{b}_a = \left[b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a_i' a_i \right]^{-1}.$$

This concludes the MCMC algorithm.

Appendix 2: MCMC steps for linear model with correlated random effects

1. The full conditional density for a_i is normally distributed as

$$a_i \mid y_{it}, \mathbf{W}_{it}, \boldsymbol{\Omega}, \sigma_a^2, \sigma_u^2 \sim N \left[\bar{a}_i, \bar{H}_a^{-1} \right], \text{ where}$$

$$\bar{H}_a = T \times \sigma_u^{-2} + \sigma_a^{-2}$$

$$\bar{a}_i = \bar{H}_a^{-1} \left[\sigma_u^{-2} \sum_{t=1}^T (y_{it} - \mathbf{W}_{it} \boldsymbol{\Omega}) \right].$$

2. The full joint conditional density of block $\boldsymbol{\Omega} = [\psi, \boldsymbol{\lambda}_t, \boldsymbol{\beta}, \phi]$ is normally distributed

as $\boldsymbol{\Omega} \mid y_{it}, \mathbf{W}_{it}, a_i, \sigma_a^2, \sigma_u^2 \sim N \left[\bar{\boldsymbol{\Omega}}, \bar{\mathbf{H}}_{\boldsymbol{\Omega}}^{-1} \right]$, where

$$\bar{\mathbf{H}}_{\boldsymbol{\Omega}} = \mathbf{H}_{\boldsymbol{\Omega}} + \sum_{i=1}^N \sum_{t=1}^T \sigma_u^{-2} \mathbf{W}_{it}' \mathbf{W}_{it}$$

$$\bar{\boldsymbol{\Omega}} = \bar{\mathbf{H}}_{\boldsymbol{\Omega}}^{-1} \left[\mathbf{H}_{\boldsymbol{\Omega}} \boldsymbol{\Omega} + \sum_{i=1}^N \sum_{t=1}^T \sigma_u^{-2} \mathbf{W}_{it}' (y_{it} - a_i) \right].$$

3. The conditional posterior distribution for σ_u^2 is inverse gamma:

$$\sigma_u^2 \mid y_{it}, \mathbf{W}_{it}, a_i, \mathbf{\Omega}, \sigma_a^2 \sim IG \left(\frac{NT}{2} + a_u, \left[b_u^{-1} + \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{W}_{it} \mathbf{\Omega} - a_i)^2 \right]^{-1} \right).$$

4. The conditional for σ_a^2 is inverse gamma:

$$\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, a_i, \mathbf{\Omega}, \sigma_u^2 \sim IG \left(\frac{N}{2} + a_a, \left[b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a_i' a_i \right]^{-1} \right).$$

This concludes the MCMC algorithm.

Appendix 3: MCMC calculations for fractional probit IV model

The full augmented posterior distribution of the IV model is proportional to

$$\begin{aligned}
& \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it}) \right] \times \\
& \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\delta_\varepsilon} \exp \left[-.5 (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)' \delta_\varepsilon^{-1} (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i) \right] \times \\
& \prod_{i=1}^N \prod_{t=1}^T \left(\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right) \times \\
& \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp[-.5a_i'\sigma_a^{-2}a_i] \times \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_b} \exp[-.5b_i'\sigma_b^{-2}b_i] \times \\
& (2\pi)^{-\frac{1+Tk+k+h+1}{2}} |\underline{\mathbf{H}}_\Omega|^{\frac{1}{2}} \exp[-.5(\boldsymbol{\Omega} - \underline{\boldsymbol{\Omega}})' \underline{\mathbf{H}}_\Omega (\boldsymbol{\Omega} - \underline{\boldsymbol{\Omega}})] \times \\
& (2\pi)^{-\frac{1+TL+L+h}{2}} |\underline{\mathbf{H}}_\Upsilon|^{\frac{1}{2}} \exp[-.5(\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})' \underline{\mathbf{H}}_\Upsilon (\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})] \times \\
& (2\pi)^{-\frac{1}{2}} |\underline{\mathbf{H}}_{\delta_{u\varepsilon}}|^{\frac{1}{2}} \exp[-.5(\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})' \underline{\mathbf{H}}_{\delta_{u\varepsilon}} (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})] \times \\
& \frac{1}{\Gamma(a_a)b_a^{a_a}} (\sigma_a^2)^{-(a_a+1)} \exp\left(-\frac{1}{b_a\sigma_a^2}\right) \times \frac{1}{\Gamma(a_b)b_b^{a_b}} (\sigma_b^2)^{-(a_b+1)} \exp\left(-\frac{1}{b_b\sigma_b^2}\right) \times \\
& \frac{1}{\Gamma(a_{\delta_\varepsilon})b_{\delta_\varepsilon}^{a_{\delta_\varepsilon}}} (\delta_\varepsilon)^{-(a_{\delta_\varepsilon}+1)} \exp\left(-\frac{1}{b_{\delta_\varepsilon}\delta_\varepsilon}\right). \tag{23}
\end{aligned}$$

The steps of this MCMC algorithm are as follows:

1. The posterior distribution of the latent dependent variable y_{its}^* is normally distributed as

$y_{its}^* \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N [y_{its}^* + \mathbf{W}_{it}\boldsymbol{\Omega} + a_i + \delta_{u\varepsilon}\varepsilon_{it}, 1]$, and is truncated at zero such that

$$y_{its}^* > 0 \text{ if } d_{its} = 1 \text{ and}$$

$$y_{its}^* \leq 0 \text{ if } d_{its} = 0,$$

which is based on the y_{its}^* terms of the augmented posterior distribution

$$\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it}) \right] \times \prod_{i=1}^N \prod_{t=1}^T \left(\sum_{s=1}^S [I(d_{its} = 1)I(y_{its}^* > 0) + I(d_{its} = 0)I(y_{its}^* \leq 0)] \right).$$

2. The full conditional density for a_i is normally distributed as

$a_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N [\bar{a}_i, \bar{H}_a^{-1}]$ where

$$\bar{H}_a = S \times T + \sigma_a^{-2}$$

$$\bar{a}_i = \bar{H}_a^{-1} \left[\sum_{t=1}^T \left(\sum_{c=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - \delta_{u\varepsilon}\varepsilon_{it}) \right) \right].$$

To derive this, I drop any terms from equation (23) that are multiplicatively unrelated

to a_i :

$$\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it}) \right] \times$$

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp[-.5a_i'\sigma_a^{-2}a_i].$$

Once again dropping terms free of a_i :

$$\exp \left(-.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it}) + a_i'\sigma_a^{-2}a_i] \right)$$

Then I can rewrite the term in the bracket as

$$\begin{aligned} & [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it}) + a_i'\sigma_a^{-2}a_i] \\ = & y_{its}^*y_{its}^* - y_{its}^*\mathbf{W}_{it}\boldsymbol{\Omega} - y_{its}^*a_i - y_{its}^*\delta_{u\varepsilon}\varepsilon_{it} - \boldsymbol{\Omega}'\mathbf{W}_{it}'y_{its}^* + \boldsymbol{\Omega}'\mathbf{W}_{it}'\mathbf{W}_{it}\boldsymbol{\Omega} \\ & + \boldsymbol{\Omega}'\mathbf{W}_{it}'a_i + \boldsymbol{\Omega}'\mathbf{W}_{it}'\delta_{u\varepsilon}\varepsilon_{it} - a_i'y_{its}^* + a_i'\mathbf{W}_{it}\boldsymbol{\Omega} + a_i'a_i + a_i'a_i + a_i'\delta_{u\varepsilon}\varepsilon_{it} \\ & - \varepsilon_{it}'\delta_{u\varepsilon}'y_{its}^* + \varepsilon_{it}'\delta_{u\varepsilon}'\mathbf{W}_{it}\boldsymbol{\Omega} + \varepsilon_{it}'\delta_{u\varepsilon}'a_i + \varepsilon_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}\varepsilon_{it} + a_i'\sigma_a^{-2}a_i. \end{aligned}$$

Again dropping terms free of a_i :

$$\begin{aligned} & - y_{its}^*a_i + \boldsymbol{\Omega}'\mathbf{W}_{it}'a_i - a_i'y_{its}^* + a_i'\mathbf{W}_{it}\boldsymbol{\Omega} + a_i'a_i + a_i'a_i + a_i'\delta_{u\varepsilon}\varepsilon_{it}' + \varepsilon_{it}'\delta_{u\varepsilon}'a_i + a_i'\sigma_a^{-2}a_i \\ = & a_i'(1 + \sigma_a^{-2})a_i - a_i'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - \delta_{u\varepsilon}\varepsilon_{it}') - (y_{its}^* - \boldsymbol{\Omega}'\mathbf{W}_{it}' - \varepsilon_{it}'\delta_{u\varepsilon}')a_i. \end{aligned}$$

Adding over s and t :

$$\begin{aligned}
& a_i'(S \times T + \sigma_a^{-2})a_i - a_i' \left(\sum_{t=1}^T \left[\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - \delta_{u\varepsilon}\varepsilon'_{it}) \right] \right) \\
& - \left(\sum_{t=1}^T \left[\sum_{s=1}^S (y_{its}' - \boldsymbol{\Omega}'\mathbf{W}'_{it} - \varepsilon'_{it}\delta'_{u\varepsilon}) \right] \right) a_i \\
& = a_i'\overline{H}_a a_i - a_i'\overline{H}_a \overline{a}_i - \overline{a}_i'\overline{H}_a a_i.
\end{aligned}$$

Then I complete the square, by adding and subtracting the term $\overline{a}_i'\overline{H}_a \overline{a}_i$:

$$\begin{aligned}
& = a_i'\overline{H}_a a_i - a_i'\overline{H}_a \overline{a}_i - \overline{a}_i'\overline{H}_a a_i + \overline{a}_i'\overline{H}_a \overline{a}_i - \overline{a}_i'\overline{H}_a \overline{a}_i \\
& = (a_i - \overline{a}_i)'\overline{H}_a (a_i - \overline{a}_i) - \overline{a}_i'\overline{H}_a \overline{a}_i.
\end{aligned}$$

Since the last term is free of a_i , this simplifies to

$$(a_i - \overline{a}_i)'\overline{H}_a (a_i - \overline{a}_i).$$

From here it is straightforward that

$$a_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\overline{a}_i, \overline{H}_a^{-1} \right].$$

3. The full conditional density of the parameter vector Ω is normally distributed as

$$\Omega \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \Upsilon, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\Omega}, \bar{\mathbf{H}}_\Omega^{-1} \right] \text{ where}$$

$$\bar{\mathbf{H}}_\Omega = \mathbf{H}_\Omega + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}'_{it} \mathbf{W}_{it}$$

$$\bar{\Omega} = \bar{\mathbf{H}}_\Omega^{-1} \left[\mathbf{H}_\Omega \Omega + \sum_{i=1}^N \sum_{t=1}^T \left(\mathbf{W}'_{it} \sum_{s=1}^S (y_{its}^* - a_i - \delta_{u\varepsilon} \varepsilon_{it}) \right) \right].$$

To show this I collect all terms from equation (23) that are related to Ω :

$$\begin{aligned} & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-0.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \Omega - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \Omega - a_i - \delta_{u\varepsilon} \varepsilon_{it}) \right] \\ & \times (2\pi)^{-(1+Tk+k+h/2)} |\mathbf{H}_\Omega|^{\frac{1}{2}} \exp \left[-0.5 (\Omega - \underline{\Omega})' \mathbf{H}_\Omega (\Omega - \underline{\Omega}) \right]. \end{aligned}$$

Once again dropping terms free of Ω I am left with

$$\begin{aligned} & \exp \left(-0.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it} \Omega - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \Omega - a_i - \delta_{u\varepsilon} \varepsilon_{it}) \right. \\ & \left. + (\Omega - \underline{\Omega})' \mathbf{H}_\Omega (\Omega - \underline{\Omega}) \right]. \end{aligned}$$

Then I can rewrite the term in the bracket as

$$\begin{aligned} & [(y_{its}^* - \mathbf{W}_{it} \Omega - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \Omega - a_i - \delta_{u\varepsilon} \varepsilon_{it}) + (\Omega - \underline{\Omega})' \mathbf{H}_\Omega (\Omega - \underline{\Omega})] \\ = & y_{its}^* y_{its}^* - y_{its}^* \mathbf{W}_{it} \Omega - y_{its}^* a_i - y_{its}^* \delta_{u\varepsilon} \varepsilon_{it} - \Omega' \mathbf{W}'_{it} y_{its}^* + \Omega' \mathbf{W}'_{it} \mathbf{W}_{it} \Omega + \Omega' \mathbf{W}'_{it} a_i \\ & + \Omega' \mathbf{W}'_{it} \delta_{u\varepsilon} \varepsilon_{it} - a_i' y_{its}^* + a_i' \mathbf{W}_{it} \Omega + a_i' a_i + a_i' a_i + a_i' \delta_{u\varepsilon} \varepsilon_{it} - \varepsilon'_{it} \delta'_{u\varepsilon} y_{its}^* + \varepsilon'_{it} \delta'_{u\varepsilon} \mathbf{W}_{it} \Omega \\ & + \varepsilon'_{it} \delta'_{u\varepsilon} a_i + \varepsilon'_{it} \delta'_{u\varepsilon} \delta_{u\varepsilon} \varepsilon_{it} + \Omega' \mathbf{H}_\Omega \Omega - \Omega' \mathbf{H}_\Omega \underline{\Omega} - \underline{\Omega}' \mathbf{H}_\Omega \Omega + \underline{\Omega}' \mathbf{H}_\Omega \underline{\Omega}. \end{aligned}$$

Again dropping terms unrelated to Ω :

$$\begin{aligned}
& - y_{its}^* \mathbf{W}_{it} \Omega - \Omega' \mathbf{W}'_{it} y_{its}^* + \Omega' \mathbf{W}'_{it} \mathbf{W}_{it} \Omega + \Omega' \mathbf{W}'_{it} a_i + \Omega' \mathbf{W}'_{it} \delta_{u\varepsilon} \varepsilon_{it} \\
& + a_i' \mathbf{W}_{it} \Omega + \varepsilon'_{it} \delta'_{u\varepsilon} \mathbf{W}_{it} \Omega + \Omega' \underline{\mathbf{H}}_{\Omega} \Omega - \Omega' \underline{\mathbf{H}}_{\Omega} \underline{\Omega} - \underline{\Omega}' \underline{\mathbf{H}}_{\Omega} \Omega \\
= & \Omega' (\underline{\mathbf{H}}_{\Omega} + \mathbf{W}'_{it} \mathbf{W}_{it}) \Omega - (\underline{\Omega}' \underline{\mathbf{H}}_{\Omega} + y_{its}^* \mathbf{W}_{it} - a_i' \mathbf{W}_{it} - \varepsilon'_{it} \delta'_{u\varepsilon} \mathbf{W}_{it}) \Omega \\
& - \Omega' (\underline{\mathbf{H}}_{\Omega} \underline{\Omega} + \mathbf{W}'_{it} y_{its}^* - \mathbf{W}'_{it} a_i - \mathbf{W}'_{it} \delta_{u\varepsilon} \varepsilon_{it}).
\end{aligned}$$

Adding over s , i , and t :

$$\begin{aligned}
& \Omega' \left(\underline{\mathbf{H}}_{\Omega} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}'_{it} \mathbf{W}_{it} \right) \Omega \\
& - \left(\underline{\Omega}' \underline{\mathbf{H}}_{\Omega} + \sum_{t=1}^T \sum_{i=1}^N \left[\mathbf{W}_{it} \sum_{s=1}^S (y_{its}^* - a_i' - \varepsilon'_{it}) \right] \right) \Omega \\
& - \Omega' \left(\underline{\mathbf{H}}_{\Omega} \underline{\Omega} + \sum_{t=1}^T \sum_{i=1}^N \left[\mathbf{W}'_{it} \sum_{s=1}^S (y_{its}^* - a_i - \delta_{u\varepsilon} \varepsilon_{it}) \right] \right) \\
= & \Omega' \overline{\mathbf{H}}_{\Omega} \Omega - \overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \Omega - \Omega' \overline{\mathbf{H}}_{\Omega} \overline{\Omega}.
\end{aligned}$$

To complete the square, I add and subtract $\overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \overline{\Omega}$:

$$\begin{aligned}
& = \Omega' \overline{\mathbf{H}}_{\Omega} \Omega - \overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \Omega - \Omega' \overline{\mathbf{H}}_{\Omega} \overline{\Omega} + \overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \overline{\Omega} - \overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \overline{\Omega} \\
& = (\Omega - \overline{\Omega})' \overline{\mathbf{H}}_{\Omega} (\Omega - \overline{\Omega}) - \overline{\Omega}' \overline{\mathbf{H}}_{\Omega} \overline{\Omega}.
\end{aligned}$$

Since the last term is free of Ω , I am left with

$$(\Omega - \overline{\Omega})' \overline{\mathbf{H}}_{\Omega} (\Omega - \overline{\Omega}).$$

Which shows that $\boldsymbol{\Omega} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\boldsymbol{\Omega}}, \bar{\mathbf{H}}_\Omega^{-1} \right]$.

4. The posterior distribution of the variance parameter σ_a^2 is inverse gamma, i.e.

$$\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_b^2, \delta_\varepsilon \sim IG \left[\frac{N}{2} + a_a, \left(b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a_i' a_i \right)^{-1} \right].$$

This is derived by collecting all terms in (20) that are multiplicatively related to σ_a^2 :

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_a} \exp[-.5a_i' \sigma_a^{-2} a_i] \times \frac{1}{\Gamma(a_a) b_a^{a_a}} (\sigma_a^2)^{-(a_a+1)} \exp \left(-\frac{1}{b_a \sigma_a^2} \right).$$

Grouping like terms together, and ignoring terms unrelated to σ_a^2 I am left with

$$(\sigma_a^2)^{-\left(\frac{N}{2} + a_a + 1\right)} \exp \left[-.5 \sigma_a^{-2} (2b_a^{-1} + a_i' a_i) \right].$$

where, adding over i , I find that $\sigma_a^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_b^2, \delta_\varepsilon \sim IG \left[\bar{a}_a, \bar{b}_a \right]$ where

$$\bar{a}_a = \frac{N}{2} + a_a \quad \text{and} \quad \bar{b}_a = \left[b_a^{-1} + \frac{1}{2} \sum_{i=1}^N a_i' a_i \right]^{-1}.$$

5. The full conditional of b_i is distributed normally as

$$b_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{b}_i, \bar{H}_b^{-1} \right] \text{ where}$$

$$\bar{H}_b = T \times \delta_\varepsilon^{-1} + S \times T \delta_{u\varepsilon}^2 + \sigma_b^{-2}$$

$$\bar{b}_i = \bar{H}_b^{-1} \left[\delta_\varepsilon^{-1} \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it} \boldsymbol{\Upsilon}) - \delta_{u\varepsilon} \sum_{t=1}^T \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \boldsymbol{\Omega} - a_i - \delta_{u\varepsilon} (q_{it} - \mathbf{Q}_{it} \boldsymbol{\Upsilon})) \right) \right].$$

To show this I collect all terms from equation (23) that are related to b_i :

$$\begin{aligned} & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp\left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})'(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}\varepsilon_{it})\right] \times \\ & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\delta_\varepsilon} \exp\left[-.5(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)\right] \\ & \times \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[-.5b_i'\sigma_b^{-2}b_i\right], \end{aligned}$$

where $\varepsilon_{it} = q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i$. This can be written as

$$\begin{aligned} & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp\left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i))'\right. \\ & \left. (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i))\right] \times \\ & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\delta_\varepsilon} \exp\left[-.5(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)\right] \\ & \times \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[-.5b_i'\sigma_b^{-2}b_i\right] \end{aligned}$$

Again dropping terms unrelated to b_i I am left with

$$\begin{aligned} & \exp\left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i))'\right. \\ & \left. \times (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i))\right] \\ & \times \exp\left[-.5(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)\right] \times \exp\left[-.5b_i'\sigma_b^{-2}b_i\right] \\ = & \exp\left(-.5 \sum_{s=1}^S [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}q_{it} + \delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_{u\varepsilon}b_i)'\right. \\ & \left. \times (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}q_{it} + \delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_{u\varepsilon}b_i) \right. \\ & \left. + (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i) + b_i'\sigma_b^{-2}b_i\right]. \end{aligned}$$

Then I can re-write the term in the bracket as

$$\begin{aligned}
& [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\epsilon}q_{it} + \delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_{u\epsilon}b_i)' \times \\
& (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\epsilon}q_{it} + \delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_{u\epsilon}b_i) \\
& + (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\epsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i) + b_i'\sigma_b^{-2}b_i] \\
= & y_{its}^*y_{its}^* - y_{its}^*\mathbf{W}_{it}\boldsymbol{\Omega} - y_{its}^*a_i - y_{its}^*\delta_{u\epsilon}q_{it} + y_{its}^*\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + y_{its}^*\delta_{u\epsilon}b_i \\
& - \boldsymbol{\Omega}'\mathbf{W}'_{it}y_{its}^* + \boldsymbol{\Omega}'\mathbf{W}'_{it}\mathbf{W}_{it}\boldsymbol{\Omega} + \boldsymbol{\Omega}'\mathbf{W}'_{it}a_i + \boldsymbol{\Omega}'\mathbf{W}'_{it}\delta_{u\epsilon}q_{it} - \boldsymbol{\Omega}'\mathbf{W}'_{it}\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\
& - \boldsymbol{\Omega}'\mathbf{W}'_{it}\delta_{u\epsilon}b_i - a_i'y_{its}^* + a_i'\mathbf{W}_{it}\boldsymbol{\Omega} + a_i'a_i + a_i'\delta_{u\epsilon}q_{it} - a_i'\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + a_i'\delta_{u\epsilon}b_i \\
& - q'_{it}\delta'_{u\epsilon}y_{its}^* + q'_{it}\delta'_{u\epsilon}\mathbf{W}_{it}\boldsymbol{\Omega} + q'_{it}\delta'_{u\epsilon}a_i + q'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}q_{it} - q'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} - q'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}b_i \\
& + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}y_{its}^* - \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}\mathbf{W}_{it}\boldsymbol{\Omega} - \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}a_i - \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}q_{it} + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\
& + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}b_i + b_i'\delta'_{u\epsilon}y_{its}^* - b_i'\delta'_{u\epsilon}\mathbf{W}_{it}\boldsymbol{\Omega} - b_i'\delta'_{u\epsilon}a_i - b_i'\delta'_{u\epsilon}\delta_{u\epsilon}q_{it} + b_i'\delta'_{u\epsilon}\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\
& + b_i'\delta'_{u\epsilon}\delta_{u\epsilon}b_i + q'_{it}\delta_\epsilon^{-1}q_{it} - q'_{it}\delta_\epsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} - q'_{it}\delta_\epsilon^{-1}b_i - \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta_\epsilon^{-1}q_{it} + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta_\epsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\
& + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta_\epsilon^{-1}b_i - b_i'\delta_\epsilon^{-1}q_{it} + b_i'\delta_\epsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + b_i'\delta_\epsilon^{-1}b_i + b_i'\sigma_b^{-2}b_i.
\end{aligned}$$

Again I drop terms unrelated to b_i to get

$$\begin{aligned}
& y_{its}^*\delta_{u\epsilon}b_i - \boldsymbol{\Omega}'\mathbf{W}'_{it}\delta_{u\epsilon}b_i + a_i'\delta_{u\epsilon}b_i - q'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}b_i + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}\delta_{u\epsilon}b_i \\
& + b_i'\delta'_{u\epsilon}y_{its}^* - b_i'\delta'_{u\epsilon}\mathbf{W}_{it}\boldsymbol{\Omega} - b_i'\delta'_{u\epsilon}a_i - b_i'\delta'_{u\epsilon}\delta_{u\epsilon}q_{it} + b_i'\delta'_{u\epsilon}\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + b_i'\delta'_{u\epsilon}\delta_{u\epsilon}b_i \\
& - q'_{it}\delta_\epsilon^{-1}b_i + \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta_\epsilon^{-1}b_i - b_i'\delta_\epsilon^{-1}q_{it} + b_i'\delta_\epsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + b_i'\delta_\epsilon^{-1}b_i + b_i'\sigma_b^{-2}b_i \\
= & b_i'(\delta'_{u\epsilon}\delta_{u\epsilon} + \delta_\epsilon^{-1} + \sigma_b^{-2})b_i \\
& - (y_{its}^*\delta_{u\epsilon} + \boldsymbol{\Omega}'\mathbf{W}'_{it}\delta_{u\epsilon} - a_i'\delta_{u\epsilon} + q'_{it}\delta'_{u\epsilon}\delta_{u\epsilon} - \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta'_{u\epsilon}\delta_{u\epsilon} + q'_{it}\delta_\epsilon^{-1} - \boldsymbol{\Upsilon}'\mathbf{Q}'_{it}\delta_\epsilon^{-1})b_i \\
& - b_i'(\delta'_{u\epsilon}y_{its}^* + \delta'_{u\epsilon}\mathbf{W}_{it}\boldsymbol{\Omega} + \delta'_{u\epsilon}a_i + \delta'_{u\epsilon}\delta_{u\epsilon}q_{it} - \delta'_{u\epsilon}\delta_{u\epsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_\epsilon^{-1}q_{it} - \delta_\epsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon}).
\end{aligned}$$

Grouping terms together and adding over s and t :

$$\begin{aligned}
& b_i' (S \times T \delta_{u\varepsilon}^2 + T \times \delta_\varepsilon^{-1} + \sigma_b^{-2}) b_i \\
& - \left(\delta_\varepsilon^{-1} \sum_{t=1}^T (q'_{it} - \mathbf{r}' \mathbf{Q}'_{it}) - \delta_{u\varepsilon} \sum_{t=1}^T \left(\sum_{s=1}^S [y_{its}^* - \mathbf{\Omega}' \mathbf{W}'_{it} - a'_i - \delta_{u\varepsilon} (q'_{it} - \mathbf{r}' \mathbf{Q}'_{it})] \right) \right) b_i \\
& - b_i' \left(\delta_\varepsilon^{-1} \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it} \mathbf{r}) - \delta_{u\varepsilon} \sum_{t=1}^T \left(\sum_{s=1}^S [y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} (q_{it} - \mathbf{Q}_{it} \mathbf{r})] \right) \right) \\
& = b_i' \overline{H}_b b_i - \overline{b}_i' \overline{H}_b b_i - b_i' \overline{H}_b \overline{b}_i.
\end{aligned}$$

Then I complete the square by adding and subtracting the term $\overline{b}_i' \overline{H}_b \overline{b}_i$:

$$\begin{aligned}
& = b_i' \overline{H}_b b_i - \overline{b}_i' \overline{H}_b b_i - b_i' \overline{H}_b \overline{b}_i + \overline{b}_i' \overline{H}_b \overline{b}_i - \overline{b}_i' \overline{H}_b \overline{b}_i \\
& = (b_i - \overline{b}_i)' \overline{H}_b (b_i - \overline{b}_i) - \overline{b}_i' \overline{H}_b \overline{b}_i.
\end{aligned}$$

Since the last term is free of b_i this simplifies to

$$(b_i - \overline{b}_i)' \overline{H}_b (b_i - \overline{b}_i).$$

From here it is straightforward that $b_i \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, \mathbf{\Omega}, \mathbf{r}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\overline{b}_i, \overline{H}_b^{-1} \right]$.

6. The full conditional density of the parameter vector Υ is normally distributed as

$$\Upsilon \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\Upsilon}, \bar{\mathbf{H}}_\Upsilon^{-1} \right] \text{ where}$$

$$\bar{\mathbf{H}}_\Upsilon = \mathbf{H}_\Upsilon + \sum_{i=1}^N \sum_{t=1}^T \mathbf{Q}'_{it} \mathbf{Q}_{it} \delta_\varepsilon^{-1} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{Q}'_{it} \mathbf{Q}_{it} \delta_{u\varepsilon}^2$$

$$\bar{\Upsilon} = \bar{\mathbf{H}}_\Upsilon^{-1} \left[\mathbf{H}_\Upsilon \underline{\Upsilon} + \sum_{i=1}^N \sum_{t=1}^T \left(\delta_\varepsilon^{-1} \mathbf{Q}'_{it} (q_{it} - b_i) - \delta_{u\varepsilon} \mathbf{Q}'_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} q_{it} + \delta_{u\varepsilon} b_i) \right) \right].$$

To show this I collect all terms from equation (23) that are related to Υ :

$$\begin{aligned} & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it}) \right] \times \\ & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi} \delta_\varepsilon} \exp \left[-.5 (q_{it} - \mathbf{Q}_{it} \Upsilon - b_i)' \delta_\varepsilon^{-1} (q_{it} - \mathbf{Q}_{it} \Upsilon - b_i) \right] \times \\ & (2\pi)^{-(1+Tk+L+h)/2} |\mathbf{H}_\Upsilon|^{\frac{1}{2}} \exp \left[-.5 (\Upsilon - \underline{\Upsilon})' \mathbf{H}_\Upsilon (\Upsilon - \underline{\Upsilon}) \right]. \end{aligned}$$

where $\varepsilon_{it} = q_{it} - \mathbf{Q}_{it} \Upsilon - b_i$, so this can be written as

$$\begin{aligned} & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} (q_{it} - \mathbf{Q}_{it} \Upsilon - b_i))' \right. \\ & \left. \times (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} (q_{it} - \mathbf{Q}_{it} \Upsilon - b_i)) \right] \times \\ & \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi} \delta_\varepsilon} \exp \left[-.5 (q_{it} - \mathbf{Q}_{it} \Upsilon - b_i)' \delta_\varepsilon^{-1} (q_{it} - \mathbf{Q}_{it} \Upsilon - b_i) \right] \times \\ & (2\pi)^{-(1+Tk+L+h)/2} |\mathbf{H}_\Upsilon|^{\frac{1}{2}} \exp \left[-.5 (\Upsilon - \underline{\Upsilon})' \mathbf{H}_\Upsilon (\Upsilon - \underline{\Upsilon}) \right]. \end{aligned}$$

Again dropping terms unrelated to Υ , I have

$$\begin{aligned} & \exp\left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i))'\right. \\ & \times (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)) \\ & \left. + (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i) + (\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})'\underline{\mathbf{H}}_\Upsilon(\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})\right]. \end{aligned}$$

Then I can re-write the term in the bracket as

$$\begin{aligned} & [(y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}q_{it} + \delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_{u\varepsilon}b_i)'] \times \\ & (y_{its}^* - \mathbf{W}_{it}\boldsymbol{\Omega} - a_i - \delta_{u\varepsilon}q_{it} + \delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + \delta_{u\varepsilon}b_i) \\ & + (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)'\delta_\varepsilon^{-1}(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i) + (\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})'\underline{\mathbf{H}}_\Upsilon(\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}})] \\ & = y_{its}^*y_{its} - y_{its}^*\mathbf{W}_{it}\boldsymbol{\Omega} - y_{its}^*a_i - y_{its}^*\delta_{u\varepsilon}q_{it} + y_{its}^*\delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + y_{its}^*\delta_{u\varepsilon}b_i \\ & - \boldsymbol{\Omega}'\mathbf{W}_{it}'y_{its}^* + \boldsymbol{\Omega}'\mathbf{W}_{it}'\mathbf{W}_{it}\boldsymbol{\Omega} + \boldsymbol{\Omega}'\mathbf{W}_{it}'a_i + \boldsymbol{\Omega}'\mathbf{W}_{it}'\delta_{u\varepsilon}q_{it} - \boldsymbol{\Omega}'\mathbf{W}_{it}'\delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\ & - \boldsymbol{\Omega}'\mathbf{W}_{it}'\delta_{u\varepsilon}b_i - a_i'y_{its}^* + a_i'\mathbf{W}_{it}\boldsymbol{\Omega} + a_i'a_i + a_i'\delta_{u\varepsilon}q_{it} - a_i'\delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + a_i'\delta_{u\varepsilon}b_i \\ & - q_{it}'\delta_{u\varepsilon}'y_{its}^* + q_{it}'\delta_{u\varepsilon}'\mathbf{W}_{it}\boldsymbol{\Omega} + q_{it}'\delta_{u\varepsilon}'a_i + q_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}q_{it} - q_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} - q_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}b_i \\ & + \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_{u\varepsilon}'y_{its}^* - \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_{u\varepsilon}'\mathbf{W}_{it}\boldsymbol{\Omega} - \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_{u\varepsilon}'a_i - \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}q_{it} + \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\ & + \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_{u\varepsilon}'\delta_{u\varepsilon}b_i + b_i'\delta_{u\varepsilon}'y_{its}^* - b_i'\delta_{u\varepsilon}'\mathbf{W}_{it}\boldsymbol{\Omega} - b_i'\delta_{u\varepsilon}'a_i - b_i'\delta_{u\varepsilon}'\delta_{u\varepsilon}q_{it} + b_i'\delta_{u\varepsilon}'\delta_{u\varepsilon}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\ & + b_i'\delta_{u\varepsilon}'\delta_{u\varepsilon}b_i + q_{it}'\delta_\varepsilon^{-1}q_{it} - q_{it}'\delta_\varepsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} - q_{it}'\delta_\varepsilon^{-1}b_i - \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_\varepsilon^{-1}q_{it} + \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_\varepsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} \\ & + \boldsymbol{\Upsilon}'\mathbf{Q}_{it}'\delta_\varepsilon^{-1}b_i - b_i'\delta_\varepsilon^{-1}q_{it} + b_i'\delta_\varepsilon^{-1}\mathbf{Q}_{it}\boldsymbol{\Upsilon} + b_i'\delta_\varepsilon^{-1}b_i + \boldsymbol{\Upsilon}'\underline{\mathbf{H}}_\Upsilon\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}\underline{\mathbf{H}}_\Upsilon\boldsymbol{\Upsilon} - \underline{\boldsymbol{\Upsilon}}'\underline{\mathbf{H}}_\Upsilon\boldsymbol{\Upsilon} \\ & + \underline{\boldsymbol{\Upsilon}}'\underline{\mathbf{H}}_\Upsilon\underline{\boldsymbol{\Upsilon}}. \end{aligned}$$

Collecting terms multiplicatively related to Υ :

$$\begin{aligned}
& y_{its}^* \delta_{u\epsilon} \mathbf{Q}_{it} \Upsilon - \Omega' \mathbf{W}'_{it} \delta_{u\epsilon} \mathbf{Q}_{it} \Upsilon - a'_i \delta_{u\epsilon} \mathbf{Q}_{it} \Upsilon - q'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} \mathbf{Q}_{it} \Upsilon \\
& + \Upsilon' \mathbf{Q}'_{it} \delta'_{u\epsilon} y_{its}^* - \Upsilon' \mathbf{Q}'_{it} \delta'_{u\epsilon} \mathbf{W}_{it} \Omega - \Upsilon' \mathbf{Q}'_{it} \delta'_{u\epsilon} a_i - \Upsilon' \mathbf{Q}'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} q_{it} \\
& + \Upsilon' \mathbf{Q}'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} \mathbf{Q}_{it} \Upsilon + \Upsilon' \mathbf{Q}'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} b_i + b'_i \delta'_{u\epsilon} \delta_{u\epsilon} \mathbf{Q}_{it} \Upsilon \\
& - q'_{it} \delta_{\epsilon}^{-1} \mathbf{Q}_{it} \Upsilon - \Upsilon' \mathbf{Q}'_{it} \delta_{\epsilon}^{-1} q_{it} + \Upsilon' \mathbf{Q}'_{it} \delta_{\epsilon}^{-1} \mathbf{Q}_{it} \Upsilon + \Upsilon' \mathbf{Q}'_{it} \delta_{\epsilon}^{-1} b_i \\
& + b'_i \delta_{\epsilon}^{-1} \mathbf{Q}_{it} \Upsilon + \Upsilon' \underline{\mathbf{H}}_{\Upsilon} \Upsilon - \Upsilon' \underline{\mathbf{H}}_{\Upsilon} \underline{\Upsilon} - \underline{\Upsilon}' \underline{\mathbf{H}}_{\Upsilon} \Upsilon + \underline{\Upsilon}' \underline{\mathbf{H}}_{\Upsilon} \underline{\Upsilon} \\
= & \Upsilon' (\underline{\mathbf{H}}_{\Upsilon} + \mathbf{Q}'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} \mathbf{Q}_{it} + \mathbf{Q}'_{it} \delta_{\epsilon}^{-1} \mathbf{Q}_{it}) \Upsilon \\
& - (\underline{\Upsilon}' \underline{\mathbf{H}}_{\Upsilon} + q'_{it} \delta_{\epsilon}^{-1} \mathbf{Q}_{it} - b'_i \delta_{\epsilon}^{-1} \mathbf{Q}_{it} - y_{its}^* \delta_{u\epsilon} \mathbf{Q}_{it} + \Omega' \mathbf{W}'_{it} \delta_{u\epsilon} \mathbf{Q}_{it} \\
& + a'_i \delta_{u\epsilon} \mathbf{Q}_{it} + q'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} \mathbf{Q}_{it} - b'_i \delta'_{u\epsilon} \delta_{u\epsilon} \mathbf{Q}_{it}) \Upsilon \\
& - \Upsilon' (\underline{\mathbf{H}}_{\Upsilon} \underline{\Upsilon} + \mathbf{Q}'_{it} \delta_{\epsilon}^{-1} q_{it} - \mathbf{Q}'_{it} \delta_{\epsilon}^{-1} b_i - \mathbf{Q}'_{it} \delta'_{u\epsilon} y_{its}^* + \mathbf{Q}'_{it} \delta'_{u\epsilon} \mathbf{W}_{it} \Omega \\
& + \mathbf{Q}'_{it} \delta'_{u\epsilon} a_i + \mathbf{Q}'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} q_{it} - \mathbf{Q}'_{it} \delta'_{u\epsilon} \delta_{u\epsilon} b_i).
\end{aligned}$$

Summing over s , i , and t , I get

$$\begin{aligned}
& \Upsilon' \left(\underline{\mathbf{H}}_{\Upsilon} + S \times \sum_{i=1}^N \sum_{t=1}^T \mathbf{Q}'_{it} \mathbf{Q}_{it} \delta_{u\varepsilon}^2 + \sum_{i=1}^N \sum_{t=1}^T \mathbf{Q}'_{it} \mathbf{Q}_{it} \delta_{\varepsilon}^{-1} \right) \Upsilon \\
& - (\underline{\Upsilon}' \underline{\mathbf{H}}_{\Upsilon} + \sum_{i=1}^N \sum_{t=1}^T [\delta_{\varepsilon}^{-1} \mathbf{Q}'_{it} (q'_{it} - b'_i) \\
& - \delta_{u\varepsilon} \mathbf{Q}'_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{\Omega}' \mathbf{W}'_{it} - a'_i - q'_{it} \delta_{u\varepsilon} + b'_i \delta_{u\varepsilon})]) \Upsilon \\
& - \Upsilon' (\underline{\mathbf{H}}_{\Upsilon} \underline{\Upsilon} + \sum_{i=1}^N \sum_{t=1}^T [\delta_{\varepsilon}^{-1} \mathbf{Q}'_{it} (q_{it} - b_i) - \delta_{u\varepsilon} \mathbf{Q}'_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} q_{it} + \delta_{u\varepsilon} b_i) \\
& \delta_{u\varepsilon} \mathbf{Q}'_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} q_{it} + \delta_{u\varepsilon} b_i)]) \\
& = \Upsilon' \overline{\mathbf{H}}_{\Upsilon} \Upsilon - \overline{\Upsilon}' \overline{\mathbf{H}}_{\Upsilon} \Upsilon - \Upsilon' \overline{\mathbf{H}}_{\Upsilon} \overline{\Upsilon}.
\end{aligned}$$

To complete the square I add and subtract $\overline{\Upsilon}' \overline{\mathbf{H}}_{\Upsilon} \overline{\Upsilon}$:

$$\begin{aligned}
& \Upsilon' \overline{\mathbf{H}}_{\Upsilon} \Upsilon - \overline{\Upsilon}' \overline{\mathbf{H}}_{\Upsilon} \Upsilon - \Upsilon' \overline{\mathbf{H}}_{\Upsilon} \overline{\Upsilon} + \overline{\Upsilon}' \overline{\mathbf{H}}_{\Upsilon} \overline{\Upsilon} - \overline{\Upsilon}' \overline{\mathbf{H}}_{\Upsilon} \overline{\Upsilon} \\
& = (\Upsilon - \overline{\Upsilon})' \overline{\mathbf{H}}_{\Upsilon} (\Upsilon - \overline{\Upsilon}) - \overline{\Upsilon}' \overline{\mathbf{H}}_{\Upsilon} \overline{\Upsilon}.
\end{aligned}$$

Since this last term is free of Υ I am left with

$$(\Upsilon - \overline{\Upsilon})' \overline{\mathbf{H}}_{\Upsilon} (\Upsilon - \overline{\Upsilon}).$$

From here you can see that $\Upsilon \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2, \delta_{\varepsilon} \sim N[\overline{\Upsilon}, \overline{\mathbf{H}}_{\Upsilon}^{-1}]$.

7. The full conditional density of the covariance parameter $\delta_{u\varepsilon}$ is normally distributed

as $\delta_{u\varepsilon} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\delta}_{u\varepsilon}, \bar{H}_{\delta_{u\varepsilon}}^{-1} \right]$ where

$$\bar{H}_{\delta_{u\varepsilon}} = \underline{H}_{\delta_{u\varepsilon}} + S \times \sum_{i=1}^N \sum_{t=1}^T \varepsilon'_{it} \varepsilon_{it}$$

$$\bar{\delta}_{u\varepsilon} = \bar{H}_{\delta_{u\varepsilon}}^{-1} \left[\underline{H}_{\delta_{u\varepsilon}} \underline{\delta}_{u\varepsilon} + \sum_{i=1}^N \sum_{t=1}^T \varepsilon'_{it} \left(\sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i) \right) \right].$$

To show this I collect all terms from equation (21) that are related to $\delta_{u\varepsilon}$:

$$\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it}) \right] \times$$

$$(2\pi)^{-\frac{1}{2}} \left| \underline{H}_{\delta_{u\varepsilon}} \right|^{\frac{1}{2}} \exp \left[-.5 (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})' \underline{H}_{\delta_{u\varepsilon}} (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon}) \right].$$

Dropping any terms unrelated to $\delta_{u\varepsilon}$:

$$\exp \left[-.5 \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it}) + (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})' \underline{H}_{\delta_{u\varepsilon}} (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon}) \right].$$

Multiplying out the terms within the brackets:

$$\begin{aligned} & \left[(y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it})' (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i - \delta_{u\varepsilon} \varepsilon_{it}) + (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon})' \underline{H}_{\delta_{u\varepsilon}} (\delta_{u\varepsilon} - \underline{\delta}_{u\varepsilon}) \right] \\ = & y_{its}^* y_{its}^* - y_{its}^* \mathbf{W}_{it} \mathbf{\Omega} - y_{its}^* a_i - y_{its}^* \delta_{u\varepsilon} \varepsilon_{it} - \mathbf{\Omega}' \mathbf{W}'_{it} y_{its}^* + \mathbf{\Omega}' \mathbf{W}'_{it} \mathbf{W}_{it} \mathbf{\Omega} + \mathbf{\Omega}' \mathbf{W}'_{it} a_i \\ & + \mathbf{\Omega}' \mathbf{W}'_{it} \delta_{u\varepsilon} \varepsilon_{it} - a_i' y_{its}^* + a_i' \mathbf{W}_{it} \mathbf{\Omega} + a_i' a_i + a_i' a_i + a_i' \delta_{u\varepsilon} \varepsilon_{it} - \varepsilon'_{it} \delta'_{u\varepsilon} y_{its}^* + \varepsilon'_{it} \delta'_{u\varepsilon} \mathbf{W}_{it} \mathbf{\Omega} \\ & + \varepsilon'_{it} \delta'_{u\varepsilon} a_i + \varepsilon'_{it} \delta'_{u\varepsilon} \delta_{u\varepsilon} \varepsilon_{it} + \delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \underline{\delta}_{u\varepsilon} - \underline{\delta}'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} + \underline{\delta}'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \underline{\delta}_{u\varepsilon}. \end{aligned}$$

Again dropping any terms multiplicatively unrelated to $\delta_{u\varepsilon}$:

$$\begin{aligned}
& - y_{its}^* \delta_{u\varepsilon} \varepsilon_{it} + \mathbf{\Omega}' \mathbf{W}'_{it} \delta_{u\varepsilon} \varepsilon_{it} + a'_i \delta_{u\varepsilon} \varepsilon_{it} - \varepsilon'_{it} \delta'_{u\varepsilon} y_{its}^* + \varepsilon'_{it} \delta'_{u\varepsilon} \mathbf{W}_{it} \mathbf{\Omega} \\
& + \varepsilon'_{it} \delta'_{u\varepsilon} a_i + \varepsilon'_{it} \delta'_{u\varepsilon} \delta_{u\varepsilon} \varepsilon_{it} + \delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} \\
= & \delta'_{u\varepsilon} (\underline{H}_{\delta_{u\varepsilon}} + \varepsilon'_{it} \varepsilon_{it}) \delta_{u\varepsilon} - (\delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} + y_{its}^* \varepsilon_{it} - \mathbf{\Omega}' \mathbf{W}'_{it} \varepsilon_{it} - a'_i \varepsilon_{it}) \delta_{u\varepsilon} \\
& - \delta'_{u\varepsilon} (\underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} + \varepsilon'_{it} y_{its}^* - \varepsilon'_{it} \mathbf{W}_{it} \mathbf{\Omega} - \varepsilon'_{it} a_i).
\end{aligned}$$

Summing over s , i , and t

$$\begin{aligned}
= & \delta'_{u\varepsilon} \left(\underline{H}_{\delta_{u\varepsilon}} + S \times \sum_{i=1}^N \sum_{t=1}^T \varepsilon'_{it} \varepsilon_{it} \right) \delta_{u\varepsilon} \\
& - \left(\delta'_{u\varepsilon} \underline{H}_{\delta_{u\varepsilon}} + \sum_{i=1}^N \sum_{t=1}^T \left[\varepsilon_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{\Omega}' \mathbf{W}'_{it} - a'_i) \right] \right) \delta_{u\varepsilon} \\
& - \delta'_{u\varepsilon} \left(\underline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} + \sum_{i=1}^N \sum_{t=1}^T \left[\varepsilon'_{it} \sum_{s=1}^S (y_{its}^* - \mathbf{W}_{it} \mathbf{\Omega} - a_i) \right] \right) \\
= & \delta'_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \overline{\delta'}_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \delta'_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \overline{\delta}_{u\varepsilon}.
\end{aligned}$$

To complete the square I add and subtract $\overline{\delta'}_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \overline{\delta}_{u\varepsilon}$:

$$\begin{aligned}
& \delta'_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \overline{\delta'}_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \delta_{u\varepsilon} - \delta'_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \overline{\delta}_{u\varepsilon} + \overline{\delta'}_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \overline{\delta}_{u\varepsilon} - \overline{\delta'}_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \overline{\delta}_{u\varepsilon} \\
= & (\delta_{u\varepsilon} - \overline{\delta}_{u\varepsilon})' \overline{H}_{\delta_{u\varepsilon}} (\delta_{u\varepsilon} - \overline{\delta}_{u\varepsilon}) - \overline{\delta'}_{u\varepsilon} \overline{H}_{\delta_{u\varepsilon}} \overline{\delta}_{u\varepsilon}.
\end{aligned}$$

Since this last term is free of $\delta_{u\varepsilon}$ I am left with

$$(\delta_{u\varepsilon} - \overline{\delta}_{u\varepsilon})' \overline{H}_{\delta_{u\varepsilon}} (\delta_{u\varepsilon} - \overline{\delta}_{u\varepsilon}).$$

From here it is clear that $\delta_{u\varepsilon} \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \sigma_a^2, \sigma_b^2, \delta_\varepsilon \sim N \left[\bar{\delta}_{u\varepsilon}, \bar{H}_{\delta_{u\varepsilon}}^{-1} \right]$.

8. The posterior distribution of the variance parameter σ_b^2 is inverse gamma, i.e.

$$\begin{aligned} \sigma_b^2 & \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \delta_\varepsilon \\ & \sim IG \left[\frac{N}{2} + a_b, \left(b_b^{-1} + \frac{1}{2} \sum_{i=1}^N b'_i b_i \right)^{-1} \right]. \end{aligned}$$

This is derived by collecting all terms in (23) that are multiplicatively related to σ_b^2 .

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_b} \exp[-.5b'_i \sigma_b^{-2} b_i] \times \frac{1}{\Gamma(a_b) b_b^{a_b}} (\sigma_b^2)^{-(a_b+1)} \exp \left(-\frac{1}{b_b \sigma_b^2} \right).$$

Dropping terms unrelated to σ_b^2 :

$$(\sigma_b^2)^{-\left(\frac{N}{2} + a_b + 1\right)} \exp \left[-.5\sigma_b^{-2} \left(2b_b^{-1} + \sum_{i=1}^N b'_i b_i \right) \right].$$

From here I can see that $\sigma_b^2 \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \delta_\varepsilon \sim IG \left[\bar{a}_b, \bar{b}_b \right]$

where

$$\bar{a}_b = \frac{N}{2} + a_b \quad \text{and} \quad \bar{b}_b = \left[b_b^{-1} + \frac{1}{2} \sum_{i=1}^N b'_i b_i \right]^{-1}.$$

9. The posterior distribution of the variance parameter δ_ε is inverse gamma, i.e.

$$\begin{aligned} \delta_\varepsilon & \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{its}^*, a_i, b_i, \mathbf{\Omega}, \mathbf{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2 \\ & \sim IG \left[\frac{NT}{2} + a_{\delta_\varepsilon}, \left(b_{\delta_\varepsilon}^{-1} + \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it} \mathbf{\Upsilon} - b_i)^2 \right)^{-1} \right]. \end{aligned}$$

To show this I collect all terms related to δ_ε and am left with

$$\prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\delta_\varepsilon} \exp \left[-.5(q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)' \delta_\varepsilon^{-1} (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i) \right] \times \\ \frac{1}{\Gamma(a_{\delta_\varepsilon}) b_{\delta_\varepsilon}^{a_{\delta_\varepsilon}}} (\delta_\varepsilon)^{-(a_{\delta_\varepsilon}+1)} \exp \left(-\frac{1}{b_{\delta_\varepsilon} \delta_\varepsilon} \right).$$

Dropping terms unrelated to δ_ε , and simplifying I get

$$(\delta_\varepsilon)^{-\left(\frac{NT}{2} + a_{\delta_\varepsilon} + 1\right)} \exp \left[-.5\delta_\varepsilon^{-1} \left(2b_{\delta_\varepsilon}^{-1} + \sum_{i=1}^N \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)^2 \right) \right].$$

Which shows that $\delta_\varepsilon \mid y_{it}, \mathbf{W}_{it}, q_{it}, \mathbf{Q}_{it}, y_{it}^*, a_i, b_i, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}, \delta_{u\varepsilon}, \sigma_a^2, \sigma_b^2 \sim IG [\bar{a}_{\delta_\varepsilon}, \bar{b}_{\delta_\varepsilon}]$

where

$$\bar{a}_{\delta_\varepsilon} = \frac{NT}{2} + a_{\delta_\varepsilon} \quad \text{and} \quad \bar{b}_{\delta_\varepsilon} = \left[b_{\delta_\varepsilon}^{-1} + \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T (q_{it} - \mathbf{Q}_{it}\boldsymbol{\Upsilon} - b_i)^2 \right]^{-1}.$$

This concludes the MCMC algorithm.