The Impact of Community Level Variables on Individual Level Outcomes: Theoretical Results and Demographic Applications

Gustavo Angeles, David K. Guilkey, Thomas A. Mroz

April 2002

MEASURE Evaluation

Carolina Population Center
University of North Carolina
at Chapel Hill
123 W. Franklin Street
Chapel Hill, NC 27516
Phone: 919-966-7482
Fax: 919-966-2391
measure@unc.edu
www.cpc.unc.edu/measure

Collaborating Partners:

Macro International Inc.
11785 Beltsville Drive
Suite 300
Calverton, MD 20705-3119
Phone: 301-572-0200
Fax: 301-572-0999
measure@macroint.com

John Snow Research and Training Institute
1616 N. Ft. Myer Drive
11th Floor
Arlington, VA 22209
Phone: 703-528-7474
Fax: 703-528-7480
measure_project@jsi.com

Tulane University
1440 Canal Street
Suite 2200
New Orleans, LA 70112
Phone: 504-584-3655
Fax: 504-584-3653
measure2@tulane.edu

Funding Agency:

Center for Population, Health
and Nutrition
U.S. Agency for
International Development
Washington, DC 20523-3600
Phone: 202-712-4959

WP-02-50

The research upon which this paper is based was sponsored by the MEASURE *Evaluation* Project with support from the United States Agency for International Development (USAID) under Contract No. HRN-A-00-97-00018-00.

The working paper series is made possible by support from USAID under the terms of Cooperative Agreement HRN-A-00-97-00018-00. The opinions expressed are those of the authors, and do not necessarily reflect the views of USAID.

The working papers in this series are produced by the MEASURE *Evaluation* Project in order to speed the dissemination of information from research studies. Most working papers currently are under review or are awaiting journal publication at a later date. Reprints of published papers are substituted for preliminary versions as they become available. The working papers are distributed as received from the authors. Adjustments are made to a standard format with no further editing.

A listing and copies of working papers published to date may be obtained from the MEASURE *Evaluation* Project at the address listed on the back cover.

Other MEASURE *Evaluation* Working Papers

WP-02-49 The Impact of a Reproductive Health Project Interventions on Contraceptive Use in Uganda (Katende C, Gupta N, Bessinger)

WP-02-48 Decentralization in Tanzania: the View of District Health Management Teams (Paul Hutchinson)

WP-02-47 Community effects on the risk of HIV infection in rural Tanzania (Shelah S. Bloom, Mark Urassa, Raphael Isingo, Japheth Ng'weshemi, J. Ties Boerma)

WP-02-46 The Determinants of Fertility in Rural Peru: Program Effects in the Early Years of the National Family Planning Program (Gustavo Angeles, David K. Guilkey and Thomas A. Mroz)

WP-02-45 Cost and Efficiency of Reproductive Health Service Provision at the Facility Level in Paraguay (Gustavo Angeles, Ruben Gaete and John F. Stewart)

WP-02-44 Decentralization, Allocative Efficiency and Health Service Outcomes in the Philippines (J. Brad Schwartz , David K. Guilkey and Rachel Racelis)

WP-01-43 Changes in Use of Health Services During Indonesia's Economic Crisis (Elizabeth Frankenberg, Bondan Sikoki, Wayan Suriastini, Duncan Thomas)

WP-01-42 Contraceptive Use in a Changing Service Environment: Evidence from the First Year of Indonesia's Economic Crisis (Elizabeth Frankenberg, Bondan Sikoki, Wayan Suriastini, DuncanThomas)

WP-01-41 Access as a Factor in Differential Contraceptive Use between Mayans and Ladinos in Guatemala (Eric Seiber and Jane T. Bertrand)

WP-01-40 Dimensions of Ratings of Maternal and Neonatal Health Services: A Factor Analysis (Rodolfo A. Bulatao and John A. Ross)

WP-01-39 Do Health Services Reduce Maternal Mortality? Evidence from Ratings of Maternal Health Programs (Rudolfo A. Bulatao and John A. Ross)

WP-01-38 Economic Status Proxies in Studies of Fertility in Developing Countries: Does the Measure Matter? (Kenneth A. Bollen, Jennifer L. Glanville, and Guy Stecklov)

WP-01-37 A Pilot Study of a Rapid Assessment Method to Identify Areas for AIDS Prevention in Cape Town, South Africa (Sharon S. Weir, Chelsea Morroni, Nicol Coetzee, John Spencer, and J. Ties Boerma)

WP-01-36 Decentralization and Local Government Health Expenditures in the Phillippines (J. Brad Schwartz, Rachel Racelis, and David K. Guilkey)

WP-01-35 Decentralization and Government Provision of Public Goods: The Public Health Sector in Uganda (John Akin, Paul Hutchinson and Koleman Strumpf)

WP-01-34 Appropriate Methods for Analyzing the Effect of Method Choice on Contraceptive Discontinuation (Fiona Steele and Siân L. Curtis)

WP-01-33 A Simple Guide to Using Multilevel Models for the Evaluation of Program Impacts (Gustavo Angeles and Thomas A.Mroz)

WP-01-32 The Effect of Structural Characteristics on Family Planning Program Performance in Côte d'Ivoire and Nigeria (Dominic Mancini, Guy Stecklov and John F. Stewart)

WP-01-31 Socio-Demographic Context of the AIDS Epidemic in a Rural Area in Tanzania with a Focus on People's Mobility and Marriage (J. Ties Boerma, Mark Urassa, Soori Nnko, Japheth Ng'weshemi, Raphael Isingo, Basia Zaba, and Gabriel Mwaluko)

WP-01-30 A Meta-Analysis of the Impact of Family Planning Programs on Fertility Preferences, Contraceptive Method Choice and Fertility (Gustavo Angeles, Jason Dietrich, David Guilkey, Dominic Mancini, Thomas Mroz, Amy Tsui and Feng Yu Zhang)

WP-01-29 Evaluation of Midwifery Care: A Case Study of Rural Guatemala (Noreen Goldman and Dana A. Glei)

WP-01-28 Effort Scores for Family Planning Programs: An Alternative Approach (John A. Ross and Katharine Cooper-Arnold)

WP-00-27 Monitoring Quality of Care in Family Planning Programs: A Comparison of Observation and Client Exit Interviews (Ruth E. Bessinger and Jane T. Bertrand)

WP-00-26 Rating Maternal and Neonatal Health Programs in Developing Countries (Rodolfo A. Bulatao and John A. Ross)

WP-00-25 Abortion and Contraceptive Use in Turkey (Pinar Senlet, Jill Mathis, Siân L. Curtis, and Han Raggers)

WP-00-24 Contraceptive Dynamics among the Mayan Population of Guatemala: 1978-1998 (Jane T. Bertrand, Eric Seiber and Gabriela Escudero)

WP-00-23 Skewed Method Mix: a Measure of Quality in Family Planning Programs (Jane T. Bertrand, Janet Rice, Tara M. Sullivan & James Shelton)

WP-00-21 The Impact of Health Facilities on Child Health (Eric R. Jensen and John F. Stewart)

WP-00-20 Effort Indices for National Family Planning Programs, 1999 Cycle (John Ross and John Stover)

WP-00-19 Evaluating Malaria Interventions in Africa: A Review and Assessment of Recent Research (Thom Eisele, Kate Macintyre, Erin Eckert, John Beier, and Gerard Killeen)

WP-00-18 Monitoring the AIDS epidemic using HIV prevalence data among young women attending antenatal clinics: prospects and problems (Basia Zaba, Ties Boerma and

Richard White)

WP-99-17 Framework for the Evaluation of National AIDS Programmes (Ties Boerma, Elizabeth Pisani, Bernhard Schwartländer, Thierry Mertens)

WP-99-16 National trends in AIDS knowledge and sexual behaviour in Zambia 1996-98 (Charles Banda, Shelah S. Bloom, Gloria Songolo, Samantha Mulendema, Amy E. Cunningham, J. Ties Boerma)

WP-99-15 The Determinants of Contraceptive Discontinuation in Northern India: A Multilevel Analysis of Calendar Data (Fengyu Zhang, Amy O. Tsui, C. M. Suchindran)

WP-99-14 Does Contraceptive Discontinuation Matter?: Quality of Care and Fertility Consequences (Ann Blanc, Siân Curtis, Trevor Croft)

WP-99-13 Socioeconomic Status and Class in Studies of Fertility and Health in Developing Countries (Kenneth A. Bollen, Jennifer L. Glanville, Guy Stecklov)

WP-99-12 Monitoring and Evaluation Indicators Reported by Cooperating Agencies in the Family Planning Services and Communication, Management and Training Divisions of the USAID Office of Population (Catherine Elkins)

WP-98-11 Household Health Expenditures in Morocco: Implications for Health Care Reform (David R. Hotchkiss, Zine Eddine el Idriss, Jilali Hazim, and Amparo Gordillo)

WP-98-10 Report of a Technical Meeting on the Use of Lot Quality Assurance Sampling (LQAS) in Polio Eradication Programs

WP-98-09 How Well Do Perceptions of Family Planning Service Quality Correspond to Objective Measures? Evidence from Tanzania (Ilene S. Speizer)

WP-98-08 Family Planning Program Effects on Contraceptive Use in Morocco, 1992-1995 (David R. Hotchkiss)

WP-98-07 Do Family Planning Service Providers in Tanzania Unnecessarily Restrict Access to Contraceptive Methods? (Ilene S. Speizer)

WP-98-06 Contraceptive Intentions and Subsequent Use: Family Planning Program Effects in Morocco (Robert J. Magnani)

WP-98-05 Estimating the Health Impact of Industry Infant Food Marketing Practices in the Philippines (John F. Stewart)

WP-98-03 Testing Indicators for Use in Monitoring Interventions to Improve Women's Nutritional Status (Linda Adair)

WP-98-02 Obstacles to Quality of Care in Family Planning and Reproductive Health Services in Tanzania (Lisa Richey)

WP-98-01 Family Planning, Maternal/Child Health, and Sexually-Transmitted Diseases in Tanzania:

Multivariate Results using Data from the 1996 Demographic and Health Survey and Service Availability Survey (Jason Dietrich)

The Impact of Community Level Variables on Individual Level Outcomes: Theoretical Results and Demographic Applications

Gustavo Angeles

David K. Guilkey

Thomas A. Mroz

April 2002

Abstract

We study alternative estimators of the impacts of higher level variables in multilevel models. This is important since many of the important variables in demographic research, such as community level access to family planning facilities, prices for services, and media campaigns are higher level factors having impacts on lower level outcomes such as contraceptive use. We present theoretical and Monte Carlo evidence about point estimation and standard error estimation for both two and three level models for continuous dependent variables, and we discuss the extension of these results to models with discrete dependent variables. A major conclusion of the paper is that readily available commercial software can be used to obtain both reliable point estimates and coefficient standard errors in models with two or more levels as long as appropriate corrections are made for possible error correlations at the highest level.

I. Introduction

Multilevel models have found widespread use in demography and related disciplines in recent years. These types of models are used when the outcome of interest, and its observed and unobserved determinants, have an hierarchical structure. By an hierarchical structure, we mean that there are important factors influencing decisions and outcomes that arise from a variety of levels of aggregation or observation. For example, whether individuals use contraception might depend on whether there are easily accessible clinics that provide family planning in the community where they live. The presence of a clinic in a community, of course, could influence the contraceptive choice of many individuals living there, and this gives rise to the multilevel structure of observed determinants of contraceptive use. Kreft and de Leeuw (1998) provide an excellent introductions to multilevel models, and Goldstein (1995) and Byrk and Raudenbush (1992) present more advanced treatments of these modeling approaches.

Typically the outcome of interest takes place at an individual level, and this usually is referred to as the lower- or micro-level outcome. In analyses with more than two levels, this is called the level-one outcome. These lower level, individual outcomes are usually influenced in part by individual, micro-level characteristics. In the family planning literature, for example, a woman's age and education and measures of her wealth have all been shown to have important effects on her use of contraceptives (Gertler and Molyneaux, 1994; Guilkey and Cochrane, 1995; Guilkey and Jayne, 1997).

What distinguishes the hierarchy in these types of analyses is the fact that some characteristics from a higher level also influence the lower-level outcomes. Researchers have found that food prices, for example, can influence whether a couple practices contraception (Stewart et al., 1991; Rous, 2001). High prices might indicate food shortages, or that it would be expensive to raise children. Couples might tend to be more likely to attempt to limit fertility when food prices are high than when food prices are low. Food prices, like many other contraceptive determinants, vary across communities but individuals within a single community all face the same level of food prices. In addition, it has been found that family planning clinics located within communities, and the availability of these sources for contraceptives and of contraceptive knowledge can affect whether individuals adopt family planning (Tsui, 1985; Bollen,

Guilkey and Mroz, 1995; Thomas and Maluccio, 1995; Guilkey and Jayne, 1997; Angeles et al., 2001). Often such higher level determinants of contraceptive use are observed, and researchers usually include these measures as explanatory variables in their empirical analyses when they are available.

At a basic level, there is nothing special about these higher level determinants that distinguishes them from individual characteristics like age and education. One can readily incorporate observed community-level characteristics along with observed individual-level characteristics as determinants of individual-level behaviors. The fact that these higher level characteristics do not differ within groups of individuals is, for the most part, irrelevant in the interpretation of impacts of observed covariates on individual-level outcomes. However, there can also be unobserved or unmeasured factors at the higher level that influence the lower-level outcomes. Such unmeasured factors give rise to multilevel error structure. These factors give rise to the important statistical considerations discussed extensively in the following sections.

The statistical properties of various estimators for multilevel models have been studied in detail (see Matyas, 1992, for a review), but the focus has been on the properties of the estimated coefficients of individual level variables in models with a two-level error structure. The impacts of community level (level two) variables, however, are frequently of considerable substantive interest because these are often policy variables than can be modified to effect individual level outcomes. In addition, longitudinal data sets with individuals sorted into communities, or data at a province, municipality, and individual levels, depend on three-levels of unobserved determinants. The purpose of this paper is to fill these gaps in the literature by focusing on the correct measurement and statistical testing of the impact of community level variables on individual level outcomes in multilevel models. We do so by presenting new theoretical results based on analytical derivations and Monte Carlo simulations. We then illustrate the methods in an analysis of children's weight using a data set from Cebu, Philippines with three levels of observed and unobserved determinants: the community, the individual, and over time for the individual.

While the theoretical and simulation results we present are specific to models with continuous dependent variables, our experiences with a wide variety of statistical models indicate that the major

conclusions of the paper should carry over to models with limited dependent variables. We demonstrate this in an empirical model of the determinants of contraception at the province, municipality and individual level. The data for this second application come from a different data set covering all of the Philippines.

II. Theoretical Results

This section summarizes the results presented in Angeles and Mroz (2001). We present analytical results for the two level model followed by Monte Carlo evidence for both two and three level models.

A. Analytical Results

Consider the following model:

$$Y_{ic} = \beta_0 + \beta_C X_c^C + \beta_{ic} X_{ic}^{IC} + \beta_I X_{ic}^I + \mu_c + \varepsilon_{ic} \quad (1)$$

where Y_{ic} is a continuous outcome variable for respondent i ($i=1,2,...,N_c$) from community c ($c=1,2,...,C$). The β's are unknown regression coefficients, X_c^C is a community level variable, X_{ic}^{IC} is an individual level variable with covariance τ with the community level variable ($\text{Cov}(X_c^C, X_{ic}^{IC})=\tau$), and X_{ic}^I is an individual level variable that is not correlated with either the community level variable or the other individual level variable. All the explanatory variables are independent of the error terms.

The error term is assumed to have two components. The first is a term that is specific to each individual, ε_{ic}; it is assumed to have mean zero and variance σ_ε^2 and to be independent for all i and c. The second unobserved component, μ_c, affects the outcome Y for all individuals within each community; it is assumed to have mean zero, variance σ_μ^2, and to be independent across communities. We define the total

error variance as $\sigma^2 = \sigma_\varepsilon^2 + \sigma_\mu^2$. The fraction of the total error variance due to the community level component of the error term is frequently referred to as the intra-class error correlation, and it is defined by $\rho = \sigma_\mu^2/\sigma^2$. Let $N_{TOT} = \sum_{c=1}^{C} N_c$ be the total number of individual level (level one) outcomes. After ordering observations by communities the (N_{TOT} x N_{TOT}) covariance matrix takes on the following form:

$$\Omega = \begin{vmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & 0 & 0 & \cdots & A_C \end{vmatrix}$$

A_c is the (N_c x N_c) covariance matrix for community c and it is given by

$$A_c = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

The two-level error structure results in a block diagonal covariance matrix for the disturbance term where the non-zero blocks (i.e. cases where the error term has non-zero correlation) correspond to individuals within the same community. This parsimonious specification assumes that the level of correlation is constant for all individuals within the same community and all communities have the same level of correlation.

Three estimators for the β's and their standard errors are commonly used: ordinary least squares (OLS), ordinary least squares with a corrected covariance matrix (OLS-C), and the maximum likelihood estimator (MLE) which is identical to the generalized least squares (GLS) estimator if Ω is known and the errors are normally distributed. To define these estimators, let $X_{ic} = [1 \quad X_c^C \quad X_{ic}^{IC} \quad X_{ic}^I]$ be the (N_c x 4) matrix of explanatory variables for community c and X be the (N_{TOT} x 4) matrix formed by stacking all the community observations. Define a similar vector Y for the ordered N_{TOT} observed outcomes. The OLS estimator is:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

The naive estimate of the covariance matrix generated by a standard statistical package would be:

$$Cov(\hat{\beta})_{OLS} = \sigma^2(X'X)^{-1} \quad (2)$$

The correct covariance matrix of the OLS estimator, however, would take account of the multilevel error structure. It takes on the following form:

$$Cov(\hat{\beta})_{OLS-C} = (X'X)^{-1}X'\Omega X(X'X)^{-1} \quad (3)$$

The MLE or GLS estimator is:

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

with covariance matrix:

$$Cov(\tilde{\beta}) = (X'\Omega^{-1}X)^{-1}$$

(4)

In order to make a simple comparisons among (2), (3) , and (4), we assume that there is no constant in the model, all the X's have mean zero, there are exactly N individuals in each community, and all the correlation in the correlated individual level variable is due to the community level variable (i.e. Cov($X^{IC}_{i'c}, X^{IC}_{ic}$)=$\tau^2$) . After some tedious algebra (see Angeles and Mroz, 2001), one can show that:

$$Cov(\hat{\beta})_{OLS} = \frac{1}{N}\begin{bmatrix} \frac{1}{(1-\tau^2)} & -\frac{\tau}{(1-\tau^2)} & 0 \\ -\frac{\tau}{(1-\tau^2)} & \frac{1}{(1-\tau^2)} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Cov(\hat{\beta})_{OLS-C} = \frac{1}{N}\begin{bmatrix} \frac{1+\rho(N-1)(1-\tau^2)}{(1-\tau^2)} & -\frac{\tau}{(1-\tau^2)} & 0 \\ -\frac{\tau}{(1-\tau^2)} & \frac{1}{(1-\tau^2)} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$Cov(\tilde{\beta}) = \frac{1}{N}\begin{bmatrix} \frac{[1+\rho(N-1)][1+\rho((N-2)-\tau^2(N-1))]}{[1+\rho(N-2)](1-\tau^2)} & \frac{-\tau(1-\rho)[1+\rho(N-1)]}{[1+\rho(N-2)](1-\tau^2)} & 0 \\ \frac{-\tau(1-\rho)[1+\rho(N-1)]}{[1+\rho(N-2)](1-\tau^2)} & \frac{(1-\rho)[1+\rho(N-1)]}{[1+\rho(N-2)](1-\tau^2)} & 0 \\ 0 & 0 & \frac{(1-\rho)[1+\rho(N-1)]}{1+\rho(N-2)} \end{bmatrix}$$

The variances of the estimators for the coefficients of the community level variables are given by the (1,1) elements of theses matrices, the variances of the correlated individual level variable coefficients are given by the (2,2) elements, and the variances of the independent individual level variable coefficients is given by the (3,3) elements. A comparison of the three covariance matrices yields several interesting results:

1. The naive OLS estimated variance for the effect of the community level variable understates the true variance for $\rho \neq 0$, and the degree of understatement increases as ρ or N (of individuals per community) increases. [Compare the (1,1) elements of $Cov(\hat{\beta})_{OLS}$ *and* $Cov(\hat{\beta})_{OLS-C}$].

2. The OLS estimator for the coefficient of the community level variable's effect is almost always inefficient relative to the MLE whenever $\rho \neq 0$. [A comparison of the (1,1) elements of $Cov(\hat{\beta})_{OLS-C}$ *and* $Cov(\tilde{\beta})$].

3. The interesting exception to the preceding result is when the community level covariates are uncorrelated with all of the individual level covariates (i.e., $\tau=0$). In this instance, there is no efficiency loss from using OLS instead of MLE.for the estimation of the impact of the community level covariate. A pertinent example when there would be a zero correlation is for the case where particular treatments (e.g.,

facilities or programs) are assigned randomly across communities. [Compare the (1,1) elements of

$Cov(\hat{\beta})_{OLS-C}$ *and* $Cov(\tilde{\beta})$ evaluated at $\tau^2 = 0$].

4. The standard errors reported by naive OLS for the impacts of both the individual level covariates are correct even when there is a multilevel structure. [Comparison of the (2,2) and (3,3) elements across $Cov(\hat{\beta})_{OLS}$ *and* $Cov(\hat{\beta})_{OLS-C}$]. This result, however, is an artifact of our simple example. If there were correlations of the individual level covariates across individuals within the same community that were not completely captured by the observed community level covariates, then the naive OLS standard errors would be too small.

5. There can be significant improvements in the precision of the estimators of the impacts of individual level covariates by using MLE instead of OLS. [Compare the (2,2) and (3,3) elements between

$Cov(\hat{\beta})_{OLS-C}$ *and* $Cov(\tilde{\beta})$.]

B. Graphical Illustrations

The OLS estimator with corrected standard errors is much simpler computationally than the MLE. Additionally, in more complex situations it might be able to produce unbiased estimators better than the maximum likelihood estimators because it does not depend upon auxiliary and often arbitrary assumptions such as homoscedastic normally distributed error components. Consequently, it is useful to see how the efficiency gain from MLE varies by ρ and N (the number of individuals per community).

We use the analytic formulae for the two covariance matrices to generates this comparison. Figures 1A and 1B graph how the standard deviations of the OLS estimators of the impacts of the three variables vary by the number of observations per community and the level of correlation between the community variable and the individual-level variable. Figure 1A examines the case when the intraclass correlation, ρ, is 0.25, and Figure 1B examines the case for ρ=0.75. Each of the four graphs within these figures refers to a different value of the correlation between the community-level explanatory variable and

the correlated individual-level explanatory variable (0.00, 0.25, 0.50, and 0.95). The horizontal axis measures the number of observations per community.

The vertical axis measures the standard deviation of the OLS estimator as a fraction of the standard deviation of the efficient maximum likelihood estimator.[1] This provides a measure of how much efficiency loss one can expect by using the less precise OLS estimator instead of the maximum likelihood estimator. The relative efficiencies for the estimators of the impacts of the two individual-level variables (indicated by the diamonds and plus signs) are identical analytically, and they do not depend on the degree of correlation between the individual-level variables and the community-level variable.

At the moderate level of the intraclass correlation in Figure 1A, $\rho=0.25$, there would be little efficiency gain from using the maximum likelihood estimator instead of the OLS estimator. When there is no correlation among the community-level variables and the individual-level variables, there are zero efficiency gains from using maximum likelihood in the estimation of the impact of the community level variable. Only when the correlation of the community-level variable and the individual-level variable is quite high (τ above 0.50) is there any discernable efficiency gain for the estimator of the impact of the community-level variable from using maximum likelihood estimation. The efficiency gain for the estimation of the impact of the community-level variable initially increases as one adds more observations per community, but then it falls. But with $\rho=0.25$, even when the correlation of the regressors is as high as 0.95 the standard error improves by using maximum likelihood by less than 10 percent.

Figure 1B examines the case where there is a high level of intraclass correlation, $\rho=0.75$. As above, there is little efficiency gain from using maximum likelihood to estimate the impact of the community-level variable, unless the correlation of the community-level variable and the individual-level variable (τ) is quite high. But even when there can be substantial efficiency gains in estimating the

[1]We evaluate the analytic formulae for the variances of the two estimators of each regression coefficient, calculate their ratio and take the square root. This provides a ratio of the standard deviations of the estimators. A value of 1.10, for example, would mean that the OLS estimator would have a standard error of estimate 10 percent higher than the maximum likelihood estimator of the coefficient for the same DGP; heuristically, t-statistics would tend to be about 10% smaller for OLS estimator than they would be for maximum likelihood estimator.

community-level variable impact by maximum likelihood, the gains diminish rapidly with increases in the number of observations per community.

The interaction among the number of observations per community, the intraclass correlation, and the correlation of the community-level regressor with the individual-level regressor appears to be the key determinant of efficiency gains from maximum likelihood estimation when estimating the impact of the community-level covariate. In Figures 2A and 2B we examine this relationship in finer detail along the dimension of the correlation of the community-level variable and the individual-level variable. As in Figure 1, the top panel in Figure 2 is for ρ=0.25, and the lower panel is for ρ=0.75. The graphs in each figure are for different numbers of individuals per community (NIPC=2, 5, 25, and 50). The horizontal axis measures the level of correlation between the community level and individual level explanatory variables(τ). In Figure 2 we only examine the relative efficiency for the estimators of the impact of the community-level variable.

For the estimation of the impact of the community level variable, there appears to be almost no efficiency gain from using maximum likelihood estimators instead of OLS estimators for values of the regressor correlation being less than 0.50. For the moderate level of the intraclass correlation, 0.25, there never are efficiency gains over 15 percent for all values of the regressor correlation at 0.99 or lower. When the intraclass correlation is high, ρ=0.75, there can be some substantial gains in efficiency, with the larger gains happening when there are several individuals per community. These gains are quite small unless the regressor correlation is well over 0.50. While not displayed in Figure 2, there are substantial efficiency gains from using MLE instead of OLS for measuring the impacts of the two individual level covariates.

III. Simulation Results

In Section II we showed that the efficiency gain from using the more complex MLE estimator instead of the OLS estimator was not large for a wide range of true parameter values. We also showed that the naive OLS standard error estimates could seriously understate the true coefficient standard errors

which means that statistical hypotheses about the true impact of the explanatory variables would be incorrect. Equation (3) presents the correct formula for the covariance matrix for the OLS estimator given the homoskedastic error components structure specified above. Robust standard error estimators developed by Eicker (1963, 1967), Huber (1967) and White (1980), however, provide for more general corrections of the standard errors. They allow for arbitrary forms of error correlation within communities and for error variances that are arbitrary functions of the explanatory variables. In addition this more robust correction yields consistent standard errors even if the true regression coefficients are random, if the error structure has more than two levels, or if the errors are heteroscedastic. Finally, these standard error estimators can be implemented using the cluster option in STATA, a widely available commercial statistical software package.

Unfortunately, the robust standard errors are only asymptotically valid (i.e. as the number of communities increases to infinity) and so it is important to see how well they work in sample sizes typically encountered in standard demographic data sets. To do this, we use Monte Carlo simulation methods to assess their accuracy in comparison to the naive OLS standard errors and the standard errors for the MLE. We focus on statistical inference on the community level variable since it has not been studied in the past and since it is typically the policy variable in demographic research. Results for individual level variables are discussed in Angeles and Mroz (2001). Results for a two level error structure are presented first followed by a discussion for a three level error structure.

A. Two Level Error Simulation Results

Equation (1) specifies the data generating process (DGP) for the Monte Carlo simulations. We set the true values of the constant term to zero and the other coefficients to one. Even though the true constant is zero, we estimate a constant in all models. The rest of the specification of the DGP is as follows:

1. We set the squared correlation coefficients between X_c^C, the community level variable

and X_{ic}^{IC} , the correlated individual level variable, to .5 (τ^2 in the notation of the previous section).

2. We examine 21 values for ρ (0, 0.05, 0.10, 0.15,..., 0.95,1.00).

3. We focus on sample sizes with approximately 20,000 individual-level observations with 800 communities each containing between 1 and 50 individual observations, with a mean of 25 persons per community.[2] Angeles and Mroz (2001) also present results for 400 communities each containing exactly 50 individual-level observations which are quite similar to the results presented here.

4. We choose an $R^2 = 0.10$. This restriction has almost no substantive impact on our comparisons of estimators.

All of our calculations were done using STATA. For each specification of the data generating process we draw 1000 independent samples, each with 800 communities. Each community contains, on average, 25 individuals (standard deviation 9.5). We simulate community- and individual-level explanatory variables, community-level disturbances, and individual-level disturbances according to fixed, specific rules. For each of these 1000 replications of the DGP, we estimate the model specified in equation (1) by OLS and by a maximum likelihood procedure that allows for the hierarchical error structure. For the OLS estimates we calculate estimates of the standard errors of the point estimates by using standard, naive OLS formulae and by using the robust (Eicker-Huber-White) formulae that adjust for possible heteroscedasticity and clustering within communities (i.e., $\rho \neq 0$). For the maximum likelihood procedure, we use STATA's report of the square root of the diagonal elements of the inverse of Hessian matrix as the standard error estimator.

We assess the accuracy of alternative estimators by determining whether hypothesis tests that use the standard error estimator yield accurate probabilities under the null hypothesis. In particular, a standard

[2]We used this range of individuals to represent roughly the distribution of the number of adult women per community in the Demographic and Health surveys. For each community in each replication of each data generating process we selected the number of individuals per community by taking draw from a truncated normal distribution with mean 25.5 and standard deviation 10 with the truncation points set at 1 and 50. We then took the integer portion of this truncated normal random variable as the choice of the number of individual level observations per community. This yields a mean number of individuals per community of 25 and a standard deviation of 9.5. Using this procedure, 91% of the time the number of individuals per community lies in the range [9,41].

error estimator would be considered accurate if hypothesis tests that use this estimator reject a true null hypothesis with a frequency given by the specified size of a test. If one tests at the 5% level, for example, then one should reject correct null hypotheses 5% of the time. Otherwise the standard error estimator does not allow one to carry out meaningful tests.

A standard, simple hypothesis test is of the form: $H_0: \beta = \beta_0$ *vs* $H_1: \beta \neq \beta_0$. One typically undertakes a hypothesis test of this form by using a two-tailed test under the assumption that the estimator of β follows an approximate Student t or normal distribution. To carry out such a test, one sets a size of the test, α, the probability of rejecting the null hypothesis when the null hypothesis is actually true. Our evaluation of the accuracy of the estimator for the standard error of the estimate is an assessment of how closely the frequency of rejecting a true null hypothesis in the Monte Carlo experiments matches the specified size α. If we specify α = 0.05, then we would want to have the null hypothesis that $\beta=\beta_0$ to be rejected five percent of the time when β actually does equal the value β_0.

In the Monte Carlo experiments we know exactly the true parameter value (all β's equal to 1), so we can examine how frequently a true null hypothesis is rejected when we use various standard error estimators for the different point estimation procedures. We examine the size of the tests for three configurations of the test. If the ρ=0, all testing configurations should provide close to identical results.

Two configurations use the ordinary least squares point estimators for the parameter estimates. The first of these uses the standard error of the estimate as reported by the simple ordinary least squares procedure to evaluate the hypothesis test. This procedure corresponds to using a standard OLS procedure and using the "default" estimators of the standard errors that assume uncorrelated, homoscedastic disturbances. This will usually provide biased tests when the intraclass correlation is non-zero for the DGPs examined here.

The second testing configuration uses a robust standard error estimator that accounts for the fact that there could be arbitrary correlations of disturbances within communities along with the simple OLS parameter point estimator (Eicker-Huber-White). The third testing configuration we examine uses the

maximum likelihood estimator. We use the point estimates and the standard error estimates from the maximum likelihood procedure to carry out hypothesis tests. For each of these three testing configurations, we examine whether the standard error estimators used with the point estimators provide tests of the correct size.

Figure 3 provides evidence on the probability that each of the testing procedures incorrectly rejects a true null hypothesis about the impact of the community level variable (i.e., we test H_0:β_C =1 vs. H_1:$\beta_C \neq 1$.) The vertical axis measures the fraction of times out of the 1000 replications that the hypothesis test rejects a true null hypothesis for a particular testing procedure. The horizontal axis measures the level of the intraclass correlation coefficient. The left-hand graph presents tests where the desired size of the test is five percent (0.05), and the right hand graph contains tests where the desired size of the test is ten percent (0.10). The vertical scales vary within Figure 3.

When the intraclass correlation coefficient (ρ) equals 0.00, all three of the testing procedures yield approximately the correct size of 0.05. As the intraclass correlation rises, the procedure using the ordinary least squares point estimate with the simple OLS standard error estimate (labeled olstest, with circles) has an empirical probability of false rejection that greatly exceeds the specified 5% size. For all intraclass correlations above 0.10, the empirical size of this testing procedure exceeds 20% when one specifies a probability of false rejection of only 5%. With fewer communities and the same number of total individual-level observations, the empirical size from this approach can be much greater than 50% for a specified size of 5%.

From the same graph in Figure 3, tests that use the same OLS point estimate for the coefficient on the community-level variable but with the standard error adjusted to correct for arbitrary forms of correlation within communities (labeled osthtest, with triangles) yield approximately the correct size for all values of the intraclass correlation coefficient. Similarly the tests based upon the maximum likelihood point estimates and the corresponding maximum likelihood estimates of the standard errors of the estimates appear to have approximately the correct size. The right-hand graph provides similar evidence

for the case where the requested size is increased to 10%. Little of substance changes with this increase in requested size.

In summary for the community-level variable, relying on the OLS point estimates and simple, default standard error estimates results in hypothesis tests that too frequently reject true null hypotheses. The simple OLS standard error estimates are, in a sense, too small. This propensity to reject null hypotheses much too frequently can be fixed with either of the other approaches. One can use the same OLS point estimates in conjunction with robust standard errors estimators to accommodate possible residual correlations within communities. Or, one can use maximum likelihood point and standard error estimators that correctly specifies the form of the within community disturbance correlation.

Given that there are estimators and testing procedures that appear to have the correct size for the forms of multilevel models that we have examined, we can now examine the ability of these procedures to reject null hypotheses that are incorrect. Holding the size of the test constant, one would prefer to have estimators and procedures that reject false null hypotheses more frequently. The ability of a test taking the form $H_0: \beta = \beta_0 \; vs \; H_1: \beta \neq \beta_0$ to reject the null hypothesis when the alternative is in fact correct depends crucially on the true value of the parameter. If the true value is quite close to the value specified under the null, then the probability of rejection is quite close to the specified size of the test (e.g., only five or ten percent), while if the true parameter value differs dramatically the probability of rejection should be close to 1.0. A graph displaying the probability of rejecting $H_0: \beta = \beta_p$ versus $H_1: \beta \neq \beta_p$ for a possible set of values β_p when the true parameter $\beta = \beta_0$ is one way of displaying the power of the test. We assess the power of each test empirically by using the estimates and standard errors from the Monte Carlo experiments. For each null hypothesis examined, we present the fraction of times (out of 1000 replications) that the testing procedure rejects each specified null hypothesis.

Figures 4 contain graphs of the power functions corresponding to tests for involving the coefficient of the community level variable of the form $H_0: \beta = \beta_p \; vs \; H_1: \beta \neq \beta_p$. The definition of the power function displayed here is the probability of rejecting the null hypothesis $\beta = \beta_p$ (against the

alternative $\beta \neq \beta_p$) as a function of the value of β_p when the true parameter value is 1.0.[3] We set the size of the tests to 0.05.

Only two testing approaches had the correct size: OLS point estimates with standard errors adjusted for possible clustering of disturbances within communities and maximum likelihood point estimates and standard errors. These are displayed in Figure 4 as "olshtest" and "mletest." When the intraclass correlation is 0.10 and the true value of $\beta_c=1$, one would reject the null hypothesis that $\beta_c=0.75$ (or $\beta_c=1.25$)[4] about 76% of the time with OLS and 79% of the time with the maximum likelihood procedure. As the intraclass correlation rises to 0.25, the power to reject H_0: $\beta_c=0.75$ (or $\beta_c=1.25$) when the true value is 1 falls to 50% for the OLS-based procedure and 55% for the maximum likelihood procedure; when the intraclass correlation is a high 0.75, the power for the same test is only 25% for the OLS based test and 27% for the maximum likelihood based test. In all cases examined here, the largest discrepancy in size between the two testing procedures is only eight percentage points. In over half of the tests displayed in Figure 4, the probability of rejection using the maximum likelihood estimator is less than one percentage point larger than the probability of rejection from using the OLS point estimate with a robust standard error estimator.

Overall, one would conclude that there is little difference between the performance of tests based on the OLS point estimates of the coefficient on the community-level variable (and with the standard error adjustment for arbitrary within community correlation) and the tests based on the maximum likelihood estimates. This should not be surprising; the comparisons of the asymptotic standard deviations of these two estimators discussed above indicates that there would only be sizable differences if the intraclass correlation were exceptionally high with only a few observations per community.

---

[3]This definition differs from the usual definition of a power function. For the more standard definition of the power function, one tests an identical hypothesis (e.g. H_0:$\alpha=2$ vs.H_a:$\alpha \neq 2$) and graphs the probability of rejection as a function of a varying true value of the parameter. Here, we graph the probability of rejecting a varying null hypothesis, when the true parameter value 1.0, as a function of hypothesized values specified in the null and alternative hypotheses.

[4]The alternative hypothesis for all power and size tests discussed in this study is the complement of the null hypothesis under examination.

B. Three Level Error Simulation Results

The analysis we reported on above indicated that one could obtain correct inferences from ordinary least squares model estimates that ignored the multilevel error structure provided one adjusted the standard errors to ex post account for the within community error correlation. The approach we used to adjust the standard errors of the estimates allowed for arbitrary forms of heteroscedasticity and error correlation within communities, but we only examined the performance of the standard error estimators when there were two levels in the analysis. It could be the case, for example, that individuals live in families which reside in communities, and there may be determinants of the individuals' behaviors that depend on unobserved family characteristics as well as unobserved individual and community characteristics. Another possibility is that we could have longitudinal data on individuals within communities. In either case, we could have an error term with more than two levels. Here we consider the performance of standard error estimators when the error term has up to three levels.

Extending the descriptive notation used above, the three levels in this model are the individual level, the family level, and the community level. The DGPs we consider have the individual-level outcome being influenced by one explanatory variable from each level. Let X_c^C be the community-level explanatory variable (c=1,2,...,C) , X_{fc}^F be the family level variable (f=1,2,...,F_c), and X_{ifc}^{IC} level variable (i=1,2,...,N_{fc}). We allow these explanatory variables to be correlated within communities and families, and we set Cor[X_c^C,X_{fc}^F]=Cor[X_{fc}^F , X_{ifc}^{IC}]=0.5 and Cor[X_c^C , X_{ifc}^{IC}]=0.667. We permit there to be unobservable determinants of the individual-level outcomes associated with each of these three levels.

The linear regression model we examine takes the following form:

$$Y_{ifc} = \beta_0 + \beta_C X_c^C + \beta_{FC} X_{fc}^F + \beta_{IC} X_{ic}^{IC} + \mu_c + \lambda_{fc} + \varepsilon_{ifc} \quad (5)$$

μ_c gives rise to the within level three error correlation (community), λ_{fc} gives rise to level-two error correlation (family), and ϵ_{ifc} is the level-one error term (individual). The μ_c are independent across different communities (level-three observations) with variance σ^2_μ, the λ_{fc} are independent across families (level-two observations) with variance σ^2_λ, and the $\epsilon_{i\,fc}$ are independent across all individuals with variance σ^2_ϵ. If σ^2 is defined to be the sum of the three variances, we can define $\rho_c = \sigma^2_\mu/\sigma^2$ and $\rho_F = \sigma^2_\lambda/\sigma^2$ as the proportions of the overall error variance due to level 3 (community) and level 2 (family) unobserved factors.

In the simulations, the community and family error components are distributed as independent N(0,1) random variables. We set ρ_C and ρ_F, to achieve different correlation patterns for the error terms and R^2 values (R^2 = .1 in the reported results). We set the three regression coefficients equal to 1.0 and we specify four level-one units (individuals) within each of 25 level-two units (families) for each of 200 level-three units (communities), for a total of 20,000 individual-level observations.

Our primary concern here is how one can carry out unbiased tests on the coefficient of the community level variable in these three-level models. Figures 5, 6, and 7 contain pertinent information about the size performance of various estimators of standard errors for different configurations of the multilevel error correlations. The left-hand side graphs examine various ways to estimate standard errors for the OLS point estimators. We consider three standard error estimators for these OLS estimates. The first is the naive standard errors as reported by standard OLS procedures assuming completely uncorrelated disturbances (labeled olstest). The second is a robust standard error estimator assuming that only observations within the second level are correlated (labeled olshfam). These standard error estimators would be appropriate, for example, if there could be non-zero error correlation among individuals within the same family ($\rho_F \neq 0$) but no correlation of disturbances across families living within the same community (ρ_C=0). The third standard error estimator is similar to the second, except that it

allows for possible error correlation at the third level among level-two units (e.g., error correlation among families and individuals living within the same community, labeled olshcom).

The right-hand side graphs are based on maximum likelihood point and standard error estimators that naively assume a two-level error hierarchy.[5] The first assumes that all level-one observations are equally correlated within the level-three units (labeled mlecomm). This would be the case, for example, if community-level unobserved factors could influence an individual's outcomes ($\rho_C \neq 0$), but there are no unobserved family-level factors influencing the individual-level outcome ($\rho_F = 0$). The second set of maximum likelihood point and standard error estimators again assumes that there is only error correlation among level-one units within the same level-two unit (e.g., only disturbances for individuals within the same family are correlated, i.e., $\rho_C = 0$ and $\rho_F \neq 0$, labeled mlefam).

The graphs display the empirical Type I error (size) for null hypotheses of the form $H_0: \beta = \beta_0 \; vs \; H_1: \beta \neq \beta_0$, where β_0 is the true value of the parameter in the DGP (i.e., 1.00 for all parameters examined), as a function of the intraclass correlation coefficient among individuals at level one within each level-two unit.[6] Each of these tests take place at a five percent level, and we carry out each test for each of the 1000 Monte Carlo replications. As in the analysis of standard error estimators in the simpler models, a point on the graph represents the fraction of times the true null hypothesis is rejected using that particular point and standard error estimator at the specified level of the intraclass correlation. An accurate standard error estimator for a particular point estimator would exhibit a straight, horizontal line at 0.05 for all values of the intraclass error correlation. Note that the vertical scales vary across graphs within these figures.[7]

---

[5]What we attempt to evaluate here are simple-to-use estimators that are available in many multi-purpose statistical packages. Consequently we do not examine correctly specified maximum likelihood estimators that recognize the possible three level error structure. Such models should provide accurate estimates and unbiased hypothesis tests.

[6]If individuals are members of families located within communities, then the intraclass correlation we consider is the correlation of the disturbances among individuals within the same family.

[7]Figures 5, 6, and 7 only examine tests at size 0.05. We obtained quite similar results for size 0.10.

Figure 5 considers the case where there is only error correlation among level-one units within the same level-two unit (e.g., only error correlation among individuals within the same family). In particular, $\rho_C = 0$, while $\rho_F \neq 0$. We see that the naive standard error estimator for the OLS point estimator performs quite poorly (olstest) . The empirical size exceeds twice the specified size even for some intraclass correlations below 0.50, with the empirical size rising to about 0.30 at the highest levels of intraclass correlation. Both of the robust standard error estimators yield tests of the correct size. It is important to recognize that for these robust standard error estimators to perform correctly, one only needs to specify the highest level at which there could be error correlations.[8] Hence, the estimator allowing there to be correlations among all individuals within the same community (olshcom) provides unbiased hypothesis tests, even though there is no community-level (level 3) error correlation. Its assumption of clustering up to as high as the community level (level 3) incorporates as a special case clustering only within families (level 2). For this standard error estimator to perform well, there need not be the same form of error correlation for all observations within the level specified as being the highest level with correlation.

The maximum likelihood estimator does not generalize this way. The right-hand graph in Figure 5 indicates that the maximum likelihood estimator that models the within level two (family) correlation does provide unbiased tests; this estimation procedure coincides with this specification of the DGP (only level-two correlation). The maximum likelihood estimator assuming only the higher level, community-level error correlations, however, does appear increasingly biased as the level of the intraclass correlation rises; this estimation method does not contain Figure 5's DGP as a special case. But for ρ less than 0.50, this bias appears quite small. Even at the highest levels of ρ the incorrectly specified maximum

[8]There could be a cost of specifying the "clustering" level higher than is necessary. It is important for there to be enough independent higher level observations for this estimator to work well. In fact, the estimator will not provide a positive definite covariance matrix unless there are at least as many independent higher level units as parameters being estimated. Typically, one would like to have many more than this number of observations in order to obtain accurate estimators of the standard errors of the parameter estimates. If there is no community level error correlation but one specifies that there could be error correlation within communities, this will yield valid standard error estimators as long as the number of communities is large. But, if there are only a few communities the estimators might not work well.

likelihood estimator provides tests that reject at most about eight percent of the time when the requested size is five percent.[9]

Figure 6 provides the same information as Figure 5, but the DGP used for Figure 6 has all of the error correlation taking place at the community (third) level. Here $\rho_C > 0$, while $\rho_F = 0$. After controlling for the level-three correlation (e.g., community), there is no additional correlation among level-one units (e.g., individuals) within the same level-two unit (e.g., families). The performance of the testing procedures in this instance are somewhat different that those discussed for Figure 5. The naive OLS standard error estimator continues to provide biased tests. The "robust" standard error estimator with only family level (level two) error correlation and the maximum likelihood procedure that allows for error correlation only at the family level (level two) now provide quite biased tests; this is because these procedures do not recognize the level-three error correlation. The two-level maximum likelihood model that allows there to be error correlation at the community level (level three), not surprisingly, provides unbiased tests for the impact of the community-level covariate because it is correctly specified. Tests using the OLS point estimates along with the robust standard error estimators allowing for up to level-three error correlation also provide unbiased tests for the impact of the community-level covariate.

Figure 7 presents the perhaps more realistic case when there are error correlations at level two (within the family) and at level three (within the community). The OLS standard error estimator with possible within community error (level three) correlations and the maximum likelihood approach that allows for error correlation at the community level (level three) provide unbiased tests. It is somewhat surprising that this maximum likelihood estimator performs correctly here. It performed somewhat poorly when there was only family level (level two) error correlation as in Figure 5, while here there is family error correlation as well as community error correlation. For the community-level variable effect, any approach that fails to recognize that there can be correlated errors at the community level provides biased hypothesis tests.

---

[9]We obtained approximately the same probabilities of false rejections for all approaches and for all data generating processes when we examined R^2 values of 0.90 instead of the 0.10 examined in these figures.

The main conclusion about the performance of standard error estimators when there are three-level models is that it is most important to control for the error correlation at the highest possible level at which it might exist. For the OLS estimates with robust standard errors (Eicker-Huber-White), it does not matter whether one "over-controls" and allows for possible error correlations at a higher level than is actually the case. For these Eicker-Huber-White standard error estimators, as long as the highest level of actual error correlation is nested within the level specified in the estimation, hypothesis tests will be unbiased. For the two-level maximum likelihood estimator, it is more important to specify exactly the level at which the error correlation takes place. But, if one is going to use two-level maximum likelihood estimators in the presence of three level error components, these results suggest it would be best to assume that all error correlation takes place at the highest level (e.g., community level).

IV. Continuous Dependent Variable Application: The Determinants of Child's Weight

To illustrate the methods presented above, we estimate a reduced form model of the determinants of child weight using data from the Cebu Longitudinal Health and Nutrition Survey (CLHNS). This data set provides an excellent illustration of the methods since there is data at three levels: community, individual and time-varying individual. The level one units are bi-monthly observations on an infant's weight from birth to age 2 (up to a maximum of 13 per infant). A level two unit is the child (or, equivalently, its mother or its family), and the level three units are the Barangays (communities) where the family resides. There are 33 Barangays in this study, with 3,327 infants, and a total of 34,293 bi-monthly recordings of the child's weight. Table 1 contains summary statistics, and Guilkey et al (1989) provides a detailed description of the data. This data set is 50% larger than the Monte Carlo data set and has more observations per community (level 3) unit. As a result, a priori, we would expect the naive OLS standard errors to be badly biased

Table 2 reports six sets of estimation results for the effect of community, family and individual level covariates on an infant's weight in kilograms for up to 13 points in time from birth to age two:

1. OLS with naive standard errors.

2. OLS with robust (Eicker-Huber-White) standard errors with correction at the person level.

3. OLS with robust (Eicker-Huber-White) standard errors with correction at the Barangay level.

4. Two-level MLE at the person level.

5. Two-level MLE at the Barangay level.

6. Three-level GLS with Barangay and person levels.

All estimation methods provide consistent point estimates of the parameters, but in general only methods 3 and 6 will provide correct standard errors. We used the MLwiN program (Goldstein, et. al., 1998) to do iterated feasible GLS for the three-level model since three-level MLE cannot be estimated using STATA. Iterated feasible GLS is asymptotically equivalent to MLE under the assumption of error term normality. As a check on the possible differences between MLwiN and STATA, we re-estimated columns 4 and 5 using MlwiN (MLE with only two levels) and obtained results identical to the reported STATA results.

The top half of Table 2 presents results for three Barangay level variables: the price of formula, the price of corn, and whether or not the Barangay is urban; the columns refer, in order, to the estimation procedures just outlined. If we look at the first three columns of the table for these three variables, we see that the point estimates of the coefficients are identical as they should be. However, we see that the naive OLS standard errors are badly biased and the OLS standard errors with the Eicker-Huber-White correction at the individual level (column 2) understate the standard errors relative using the correction at the Barangay (community) level – the outmost level.

Consider the impact of the price of formula on child's weight. The positive sign is somewhat unexpected for a price variable. However, we see that the p values for a two tailed test of significance decline from 0.04, to 0.19 and 0.32 as one moves from OLS with naive standard errors to robust standard errors with corrections first at the individual level and then at the Barangay level. A similar pattern, although not as pronounced is seen with the price of corn, the p values are 0.02, 0.05. and 0.07. These results are in line with the simulation results reported above: any correction for error correlation that ignores the outmost level of correlation can lead to seriously biased inferences and statistical tests.

The last three columns of Table 2 present MLE results. The corrections to the standard errors we observed in the first three columns suggest that there may be important correlations at both level 2 and level 3 for these data. This implies that columns 4 and 5 could be incorrectly specified, as column 4 ignores the Barangay level error correlation and column 5 ignores the level two (child) error correlation. The bottom of the table reports the estimated error component variances and we see that both the Barangay level variance and the fixed individual level variance are highly significantly different from zero in the three level model (p values of 0.03 and 0.00 respectively). The fraction of error variance due to child level unobservables, from column 6, is 0.66, while the fraction of variance due to Barangay is only 0.01. The analytical results presented in Section II suggest that such a combination of a large intraclasscorrelation with a small number of observations per level two unit (at most 13 observation per child) is a case where there could be substantial efficiency gains from using MLE instead of OLS with standard error corrections. The results presented in columns 4 and 6, where we typically find standard error estimates substantially smaller than the those presented in column 3 (OLS with correct standard errors), are supportive of this possibility. The results in columns 4 and 6 are quite similar in this case because the Barangay level error variance, while quite significant, is not very large.

A somewhat interesting result is that the point estimate of the coefficient for the price of infant formula is positive for OLS and negative for all MLE estimations. However, only the naive OLS coefficient is significantly different from zero (p=0.04), with 95% confidence intervals for any of the three MLE coefficients including positive as well as negative values. The estimated impacts of the price of corn are always small and negative. Confidence intervals using the three level MLE model, however, are only half as large as those from the OLS model with corrected standard errors. This might be due to the fact that these community level price variables are not constant across level one and level two units; they vary by calendar month. Note that the price of corn is only statistically significant for the naive OLS model that fails to account for correlated errors. All other specifications would fail to reject the null hypothesis that the effect is zero at the 5% level (2-tailed tests).

Table 2 reveals similar impacts of the various estimation procedures on the estimates for the the impact of the child's gender and the mother's age at birth, education , and height on the child's weight. Note that these maternal variables, unlike the community level variables, do not have any time series variation that is not captured by the child age variables. Naive OLS standard errors suggest that the mother's age is a significant determinant of the child's weight; its significance disappears after controlling appropriately for the multilevel error structure. Similarly, the naive OLS standard errors of the gender, education, and height effects understate the corrected standard errors by factors of two to five. The three level MLE model does provide somewhat more accurate estimators of these effects than the OLS model, with the largest efficiency gains being for the level-one indicators of the child's age.

This example clearly reflects the statistical and Monte Carlo evidence presented in the previous two sections. It clearly demonstrates the need to correct the OLS standard errors using the outmost level of clustering if one wants to make correct inferences. Maximum likelihood estimation for these data provides a substantial efficiency gain over OLS estimation primarily because of the very large individual level ρ.

V. Extension to the Probit Model with an Application to Use of Family Planning in the Philippines

Probit and logit models are typically used for models with dichotomous dependent variables, and extensions to multilevel models have been developed for both methods. Unlike the case of the continuous outcome models, it is not possible to make simple, theoretical (analytic) statements about the relative performance of estimators. Instead, in this section we point out some important interpretation issues that arise when one considers multilevel models with discrete outcomes and present an example illustrating the efficiency gains from using models that control for multilevel error structures.

Probit models are slightly more convenient for multilevel models because they depend on an underlying normal error distribution. Sums of normal random variables from different level units will remain in the class of a normal distributions, and models controlling for possible correlations across

different levels can fit into a common and internally consistent statistical model. The probit extension involves modifying equation (1) as follows:

$$Y_{ic}^{*} = \beta_0 + \beta_C X_c^{C} + \beta_{ic} X_{ic}^{IC} + \beta_i X_{ic}^{I} + \mu_c + \varepsilon_{ic} \qquad (6)$$

where all terms are defined as above except that Y_{ic}^{*} is a latent variable. The observed variable, Y_{ic}, is an indicator variable that takes on the value "1" when the latent variable is positive and "0" otherwise. As in the continuous dependent variable case, we assume the two error components are independent. What makes this a probit model instead of a logit or other binary outcome model is the assumption that both μ_c and ϵ_{ic} follow normal distributions. We define $\sigma^2 = \sigma_\varepsilon^2 + \sigma_\mu^2$ and $\rho = \sigma_\mu^2/\sigma^2$ where ρ is the fraction of the total error variance due to the community level component of the error term.

The difference from the continuous outcome case is that we only observe the sign of the dependent variable Y_{ic}^{*}; we must therefore impose a normalization. The normalization used in all simple probit procedures is $\sigma^2 = 1$ (Heckman(1981)). In more complex models some computer packages instead impose $\sigma_\varepsilon^2 = 1$. Because of this need to make an arbitrary normalization, only ratios of coefficients (i.e., relative effects) and significance levels can be identified. Robinson (1982) has shown that simple probit applied to (6) will consistently estimate the model's coefficients. Just as in the continuous case, however, the coefficient standard errors from the simple estimation will be incorrect. Robust (Eicker-Huber-White) standard errors will be asymptotically valid as long as one corrects at the highest level.

In order to discuss the MLE for this two level probit model, and to see the effect of the normalization on estimated coefficients, we re-write equation (6) with an arbitrary normalization as follows:

$$\frac{Y_{ic}^{*}}{\sigma_\eta} = \frac{\beta_0}{\sigma_\eta} + \frac{\beta_C}{\sigma_\eta} X_c^{C} + \frac{\beta_{ic}}{\sigma_\eta} X_{ic}^{IC} + \frac{\beta_i}{\sigma_\eta} X_{ic}^{I} + \left(\frac{\rho}{1-\rho}\right)^{1/2} \frac{\sigma_\epsilon}{\sigma_\mu} \frac{1}{\sigma_\eta} \mu_c + \frac{\varepsilon_{ic}}{\sigma_\eta}$$

(7)

The only restriction on σ_η is that it is positive. Since the μ_c are not observed, the MLE integrates out with respect to the μ_c's using either a simulation method or numerical quadrature (Bultler and Moffit, 1982). Some maximum likelihood procedures impose $\sigma_\eta = \sqrt{\sigma_\varepsilon^2 + \sigma_\mu^2}$. The overall error variance is 1 and these procedures estimate as coefficients $\beta_{..} / \sqrt{\sigma_\varepsilon^2 + \sigma_u^2}$. Since they impose the same normalization as a simple probit procedure ($\sigma^2 = 1$), one can directly compare the coefficient estimates of these MLE to simple probit. Other maximum likelihood procedures, such as STATA's xtprobit for example, impose the normalization that $\sigma_\eta = \sigma_\varepsilon$ so that the estimated coefficients equal $\beta_{..} / \sigma_\varepsilon$. This means that direct comparisons of the point estimates of simple probit and the MLE do not make sense unless ρ=0. However, significance levels are not affected and it is possible to compare ratios of coefficients across estimation procedures.

A Monte Carlo study (Guilkey and Murphy, 1993) obtained results that were similar to the results of Angeles and Mroz (2001) for the continuous dependent variable case: probit with Eicker-Huber-White standard errors performed quite well while the naive probit model produced standard errors that were badly biased. In addition, there was only a small efficiency gain from using the much more complicated MLE estimator in this case. Note, however, that their Monte Carlo study was designed to examine longitudinal data models with many individuals and a small number of observations per individual (either 2, 5, or 10 observations); it was not designed to examine the community level multilevel model with more than 10 observations per community.

We can also extend the three level continuous dependent variable model to probit:

$$Y_{ifc}^* = \beta_0 + \beta_C X_c^C + \beta_{FC} X_{fc}^F + \beta_{IC} X_{ic}^{IC} + \mu_c + \lambda_{fc} + \varepsilon_{ifc}$$

(8)

where all terms are as defined above. As in the two-level model, numerical methods can be used to integrate out with respect to μ_c and λ_{jc} . Again, it is important to recognize that one needs to impose some arbitrary normalization. The most common are to normalize by $\sqrt{\sigma_\varepsilon^2 + \sigma_\mu^2 + \sigma_\lambda^2}$ or by σ_ε . Ratios of coefficients and significance levels do not depend on the chosen normalization. Just as in the two level model, Eicker-Huber-White standard errors are asymptotically correct as long as one corrects at the outmost level – community in this example. However, there is no Monte Carlo evidence on the finite sample performance of these standard errors or the MLE.

Our example for the probit model uses data with three levels from the Philippines where the outcome is interest is whether a reproductive aged woman uses family planning.. The data is cross sectional with province (level 3), municipality (level 2), and individual (level 1) observations. We use province and municipality level health care expenditure data from 1998 matched to individual level women from the 1998 Demographic and Health Survey (DHS). The higher level data come from a Commission on Audit survey; Schwartz et al (2000) provides a more detailed description of these data. Schwartz, Guilkey and Racelis (2001) provide more information about the DHS data and the merged data set. There are 7,492 level one observations residing in 484 communities contained in 75 provinces. Descriptive statistics for all variables are presented on the right hand side of Table 1. In the multilevel estimation models we impose the restriction that the total error variance is 1.00, so we can directly compare coefficients to those from simple probit models.

The original reason for gathering the expenditure data was to measure the effect of devolution on health outcomes. In discussing the results, we focus on the effects of public health expenditures at the province and municipality level on contraceptive use that are reported at the top of Table 3. A comparison of the standard errors of the three simple probit estimators shows that the standard errors almost triple for both variables when we correct for possible error correlation at the municipality level and further increase when we correct at the province level. Province level public expenditures are

insignificantly different from zero at any reasonable level of significance. The point estimate of the impact of the public municipality health expenditures is much larger, with a two standard deviation increase in municipal health expenditures implying, on average, a five percentage point increase the in fraction of women using contraception. This effect, however, is not significantly different from zero at the 5% level once one recognizes that there could be error correlations at the municipal (p=0.10) or province level (p=0.13).

We now turn to the MLE results reported in the last three columns of Table 3. When one only estimates two level models, the intraclass correlations are .45 and .21 for municipality and province level two components. When the three level model is estimated, the municipality level ρ is .40 and the province level ρ is.06. The p-values for the ρ's for the three level model are 0.00 and 0.05 respectively. Corrections for error correlations at both levels are necessary, so correct inferences can only be made using the standard errors reported in columns 3 and 6.

For the individual level coefficients in columns 3 and 6, the changes in point estimates and standard errors follow the same patterns as seen in Table2 for the continuous outcome. The point estimates change by relatively small amounts from using multilevel models, and one can obtain more efficient estimators by using multilevel models instead of simple models. A much different situation arises for the higher level explanatory variables. The estimated impact of municipal health expenditures, for example, falls by over 60% from the probit estimates to the multilevel estimates with controls for province and municipal level correlations. With the multilevel model that has error correlation controls only at the province level, the estimated effect actually becomes negative. Such changes should not be statistically significant if the underlying model is correct.

Even though neither estimate of the impact of municipal level expenditures in columns 3 and 6 is significantly different from zero, the point estimate does change significantly between the two columns. To see this, note that maximum likelihood estimation of the three level model is an efficient estimator. The simple probit estimator should also be consistent in this situation, so one can carry out a simple Hausman (1978) test. The change in the point estimates is 0.1643 between columns 3 and 6. The estimate

of the standard error of this difference, calculated from the differences in the estimated variances, is 0.056. This change has a p-value of 0.003, indicating a model specification problem. This significant change might be due to the fact that government health expenditures are not allocated independent of perceived need, resulting in the endogeneity of the health expenditures. Schwartz, Guilkey, and Racelis (2001) tested the endogeneity of health expenditures using both the 1993 and 1998 data, and they found strong evidence that health expenditures at the municipality level are endogenous to contraceptive use. When an important point estimates does change significantly when one uses a standard multilevel model, it is an indication that the researcher should spend more time examining the underlying assumptions of the model instead of trying to obtain a more refined multilevel model.

V. Conclusion

This paper presents both analytical and simulation evidence on the usefulness of the OLS estimator in multilevel models with up to three levels of data in comparison to the MLE. We focus on the correct measurement of the impacts of community level variables which are often the variables of primary policy interest. Even though the OLS point estimators appear to perform quite well relative to the maximum likelihood estimators in most applied situations, the standard error estimators provided by standard Ordinary Least Squares formulae are incorrect in the presence of multilevel error correlations. For two-level models, we find that the robust asymptotic approximations to the standard errors of the OLS model due to Eicker, Huber, and White provide approximately unbiased tests for all parameter estimators when one uses formulae that allow error correlations at the higher level. The maximum likelihood standard error estimators perform flawlessly for these two-level models.

When we examine three-level models, the robust standard error estimators allowing for error correlations within the highest level continue to perform quite well, while the maximum likelihood estimators that assume only two error levels often perform poorly. This failure of the maximum likelihood estimators is due to the fact that they are incorrectly specified for the three-level models we examine. It is important to note that one will usually obtain biased tests with these “maximum likelihood” estimators

even though they control for correlations at the highest level. This could be an important factor to consider when using maximum likelihood estimators if there could be a missing "middle" level in a researcher's empirical model. Ordinary least squares estimators with the robust, Eicker-Huber-White standard error estimators do not have this limitation.

If one is primarily interested in estimating the impacts of a community-level variable on individual-level outcomes, as is frequently the case in the evaluation of health and family planning programs in developing countries, then the results of this paper provide some important guidelines. First, unless intra-class correlations are large, there are typically small efficiency gains for the estimates of the impacts of community-level factors on the individual-level behavior from using maximum likelihood procedures instead of simple ordinary least squares estimation. Second, it is crucial to adjust the estimated standard errors of the ordinary least squares estimators to reflect the fact that there can be correlated error terms at higher levels; the robust standard error estimators appear to provide adequate adjustments. Third, even if there are complex multilevel error correlations in the data, the robust standard error adjustments always provide unbiased tests, as long as one allows for error correlation at the highest level; simple two-level maximum likelihood models do not provide unbiased tests when lower level error correlations are present.

In addition to the theoretical results, we also present empirical illustrations of the methods for the continuous dependent variable case and present extensions of the probit model to three level error structures along with an empirical example. The empirical examples dovetail nicely with the theoretical results; they demonstrate how the standard error estimates change when one allows for error correlations at higher levels. Our final example highlights what might be a most important concern: all estimation approaches could yield potentially misleading results if the model is not correctly specified. Since it is straightforward to obtain robust standard error estimators for simple estimation approaches, a researcher's time might be spent better by evaluating key maintained assumptions in a model rather than by trying to incorporate multilevel error structures into the estimation of point estimates.

References

Angeles, G., Dietrich J., Guilkey D., Mancini D., Mroz T., Tsui A. and F. Zhang. 2001. "A Meta-Analysis of the Impact of Family Planning Programs on Fertility Preferences, Contraceptive Method Choice and Fertility." MEASURE Evaluation Working Paper Series No. 30.

Angeles, G. and T. Mroz. 2001. "A Guide to Using Multilevel Models for the Evaluation of Program Impacts," MEASURE Evaluation Working Paper Series No. 33.

Bollen, K., Guilkey D. and T. Mroz. 1995. "Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia," Demography: 32, 1.

Bryk A. and S. Raudenbush. 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park: Sage

Butler, J.S. and R. Moffitt. 1982. "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," Econometrica 50: 761-64.

Eicker, F., 1963. "Asymptotic Normality and Consistency of Least Squares Estimators for Families of Linear Regressions," Annals of Mathematical Statistics, 24, 447-456.

Eicker, F., 1967. "Limit Theorems for Regressions with Unequal and Dependent Errors," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics, Vol. 1, Berkeley: University of California Press, pp. 59-82.

Gertler, P. and J. Molyneaux. 1994. "How Economic Development and Family Planning Programs Combined to reduce Indonesian Fertility," Demography: 31, 1.

Goldstein, H., 1995. Multilevel Statistical Models. London: E. Arnold; New York: Halsted Press.

Guilkey, D. and S. Cochrane. 1995. "The Effects of Fertility Intentions and Access to Services on Contraceptive Use in Tunisia," Economic Development and Cultural Change 43.

Guilkey, D. and S. Jayne. 1997. "Zimbabwe: Determinants of Contraceptive Use at the Leading Edge of Fertility Transitions in Sub-Saharan Africa," Population Studies 51.

Guilkey, D. and J. Murphy. 1993. "Estimation and Testing in the Random Effects Probit Model," Journal of Econometrics 59.

Guilkey D., Popkin B,, Akin, J, and E. Wong. 1989. " Prenatal Care and Pregnancy Outcomes in the Philippines," Journal of Development Economics 30:241-272.

Hausman, J.A., 1978. "Specification Tests in Econometrics," Econometrica 46: 1251-71.

Heckman, J.J. 1981. "Statistical Models for Discrete Panel Data," In: C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.

Huber, P. J., 1967. "The Behavior of Maximum Likelihood Estimates under Non-Standard Conditions," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics, Vol. 1, Berkeley: University of California Press, pp. 221-233.

Kreft, I. and J. de Leeuw. 1998. Introducing Multilevel Modeling. Thousand Oaks, California: Sage.
Matyas, L. 1992. "Error Component Models," in The Econometrics of Panel Data (Matyas and Sevestre, eds), Kluwer Academic Publishers: Boston.

Robinson, P., 1982. "On the Asymptotic Properties of Models Containing Limited Dependent Variables," Econometrica 50:27-41.

Rous, J., 2001. "Is Breast-feeding a Substitute for Contraception in Family Planning," Demography 38, 4.

Schwartz, B., Guilkey, D., and R. Racelis. 2001. "Decentralization, Allocative Efficiency, and Health Service Outcomes in the Philippines," MEASURE *Evaluation* Working Paper Number 02-44.

Stewart, J., Popkin B., Guilkey. D., Akin J., Adair L. and W. Flieger. 1991. "Influences on the Extent of Breast-Feeding: A Prospective Study in the Philippines," Demography 28, 2.

Thomas, D. and J. Maluccio, 1995. "Contraceptive Choice, Fertility, and Public Policy in Zimbabwe." The World Bank, Washington, D.C. Living Standards Measurement Study Working Paper No. 109.

Tsui A., 1985. "Community Effects on Contraceptive Use," in: Casterline, B.J. (ed) The Collection and Analysis of Community Data, Voorburg, The Netherlands, International Statistical Institute, World Fertility Survey.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica 48: 817-830.

Figure 1

Standard Deviations of Ordinary Least Squares Estimators as a Fraction of the Standard Deviations of the Maximum Likelihood Estimators as a Function Number of Observations per Community

Figure 1A: Community Level, Correlated Individual Level, and Independent Individual Level Coefficient Estimators for Intraclass Correlation 0.25 and Four Regressor Correlations

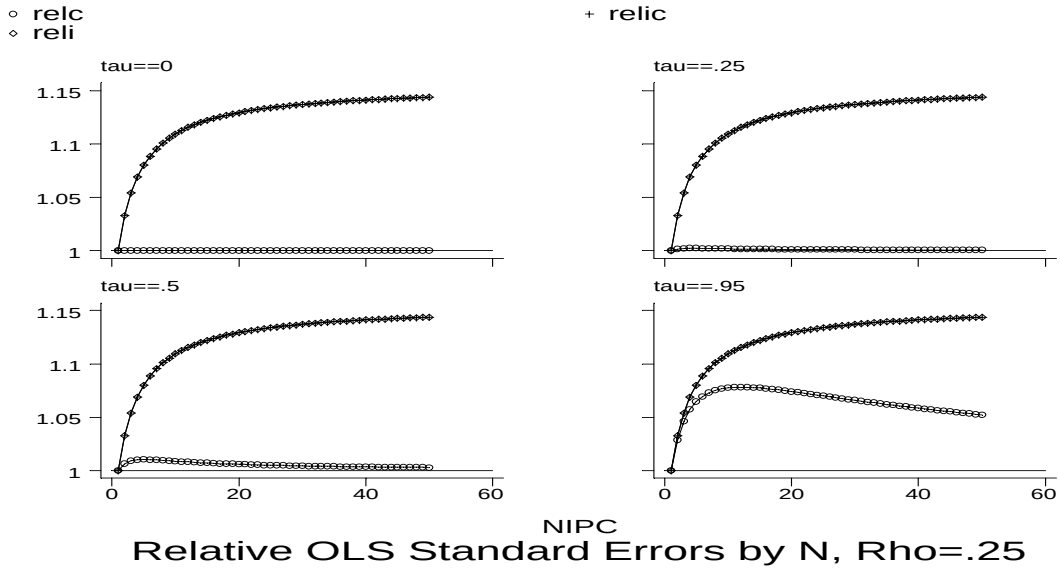


Figure1B: Community Level, Correlated Individual Level, and Independent Individual Level Coefficient Estimators for Intraclass Correlation 0.75 and Four Regressor Correlations

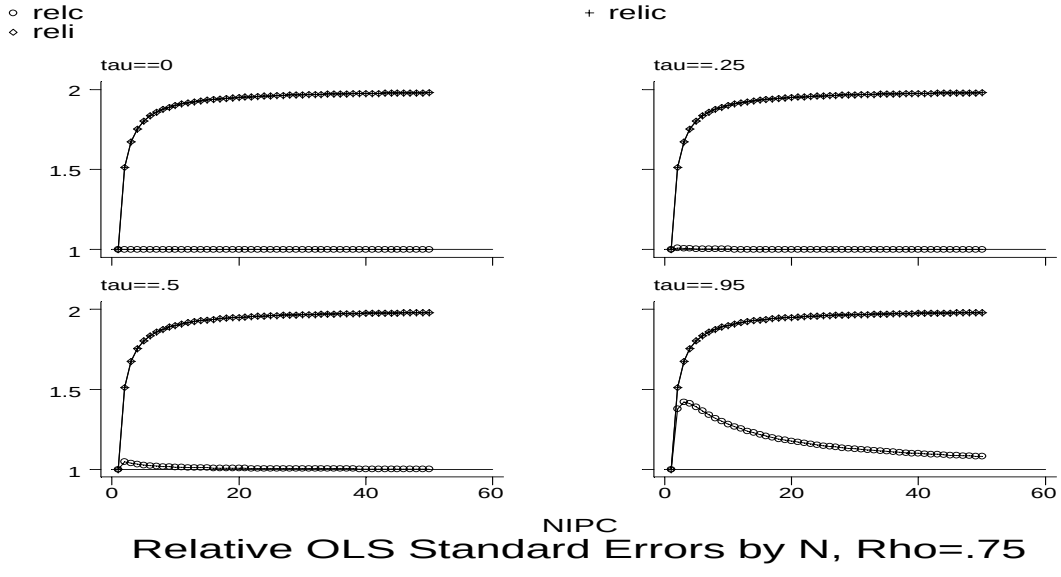


Figure 2

Standard Deviation of Ordinary Least Squares Estimator as a Fraction of the Standard Deviation of the Maximum Likelihood Estimator of the Impact of the Community Level Variable, as a Function of the Correlation of the Community and Individual Level Regressors (τ)

Figure 2A: Community Level Coefficient Estimators with an Intraclass Correlation of 0.25 and Four Specifications of the Number of Observations per Community

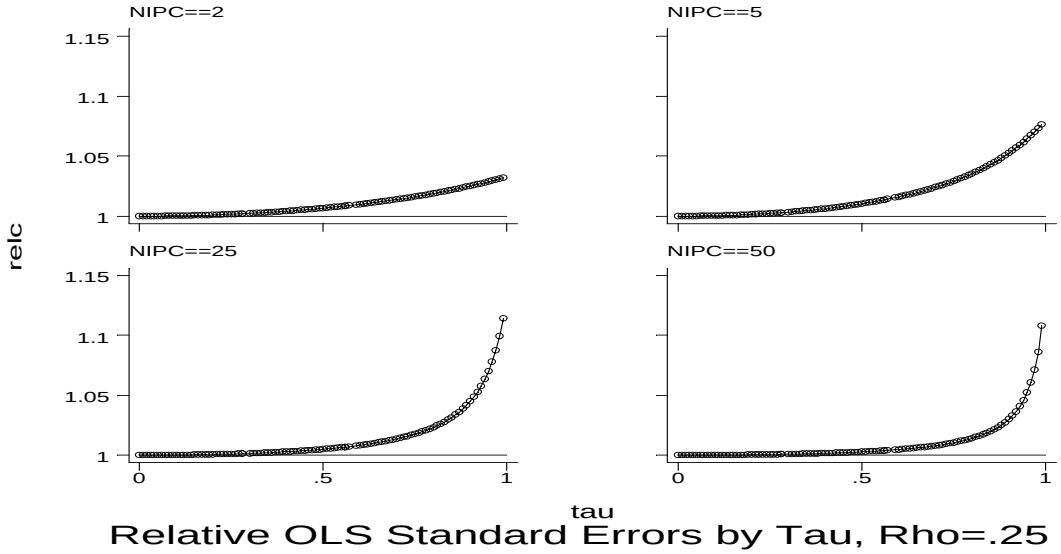


Figure 2B: Community Level Coefficient Estimators with an Intraclass Correlation of 0.75 and Four Specifications of the Number of Observations per Community

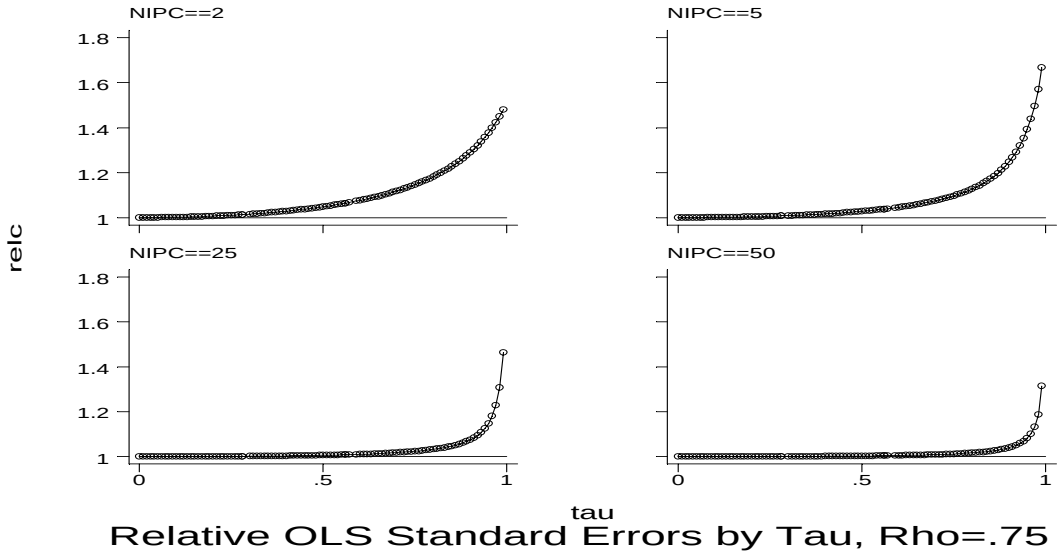


Figure 3

Performance of Standard Error Estimators: Probability of Rejecting a True Null Hypothesis

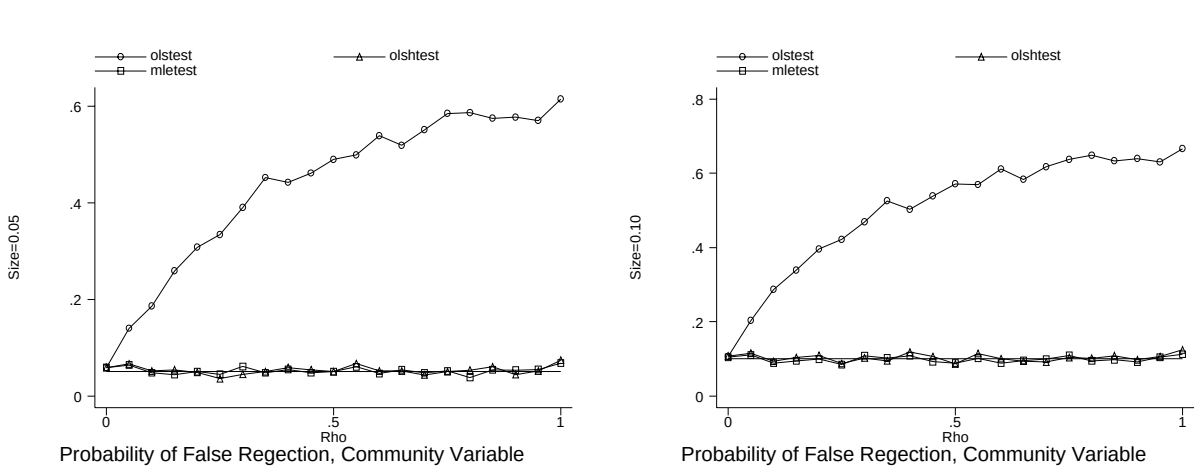


Figure 4

Power to Reject Null Hypotheses as a Function of the Intraclass Error Correlation
Ordinary Least Squares Estimators with Eicker-Huber-White Standard Errors and
Two-Level Maximum likelihood Estimators

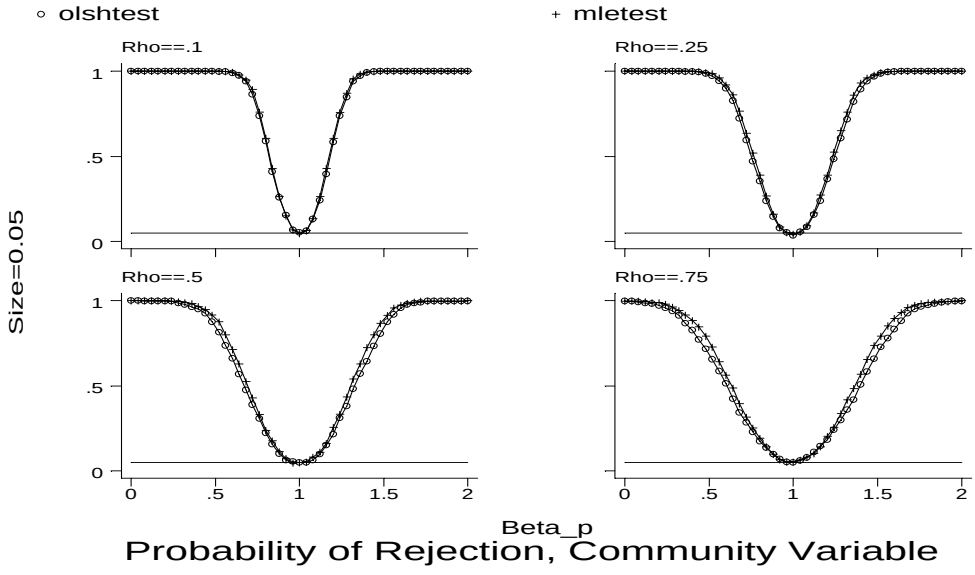
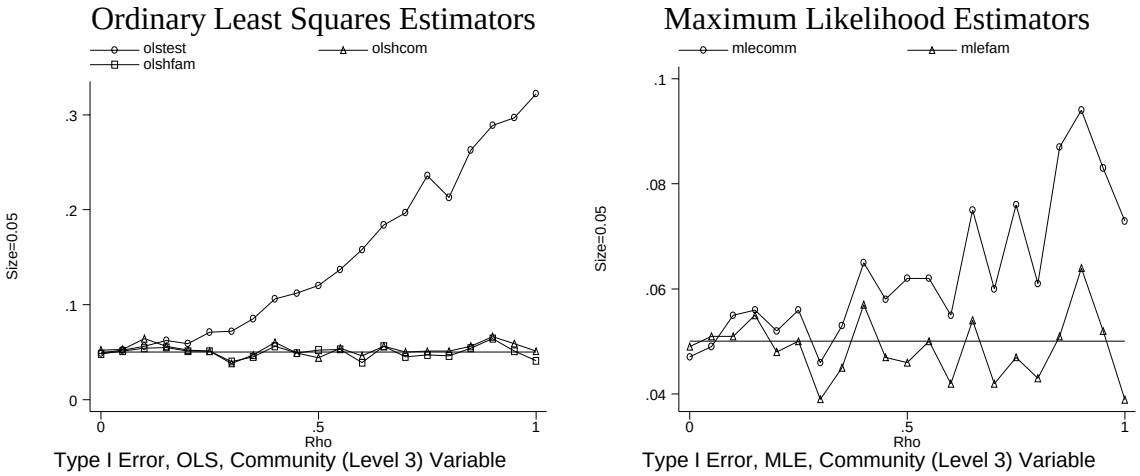


Figure 5

Performance (Size) of Standard Error Estimators for Three Level Models
Only Level Two Error Correlation

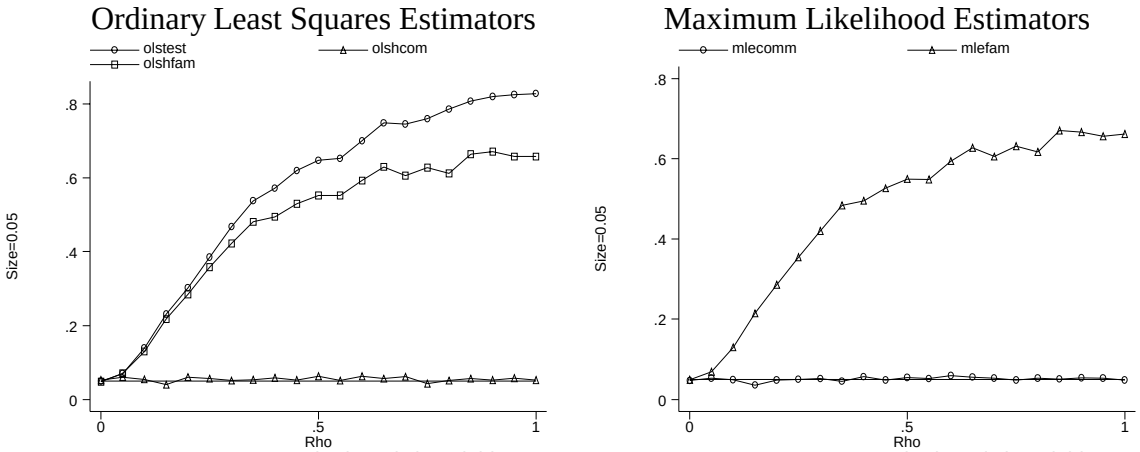


Type I Error, OLS, Community (Level 3) Variable

Type I Error, MLE, Community (Level 3) Variable

Figure 6

Performance (Size) of Standard Error Estimators for Three Level Models
Only Level Three Error Correlation



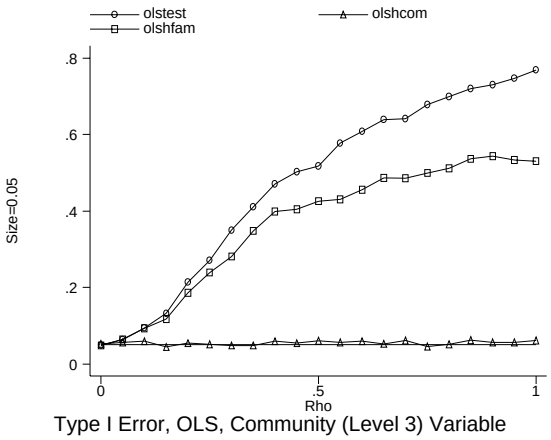
Type I Error, OLS, Community (Level 3) Variable

Type I Error, MLE, Community (Level 3) Variable

Figure 7

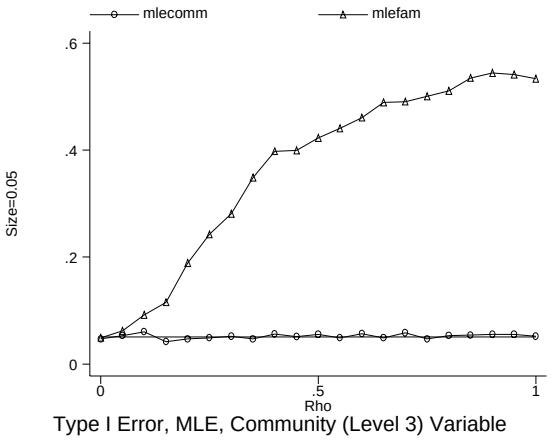
Performance (Size) of Standard Error Estimators for Three Level Models
Level Two and Level Three Error Correlations

Ordinary Least Squares Estimators



Type I Error, OLS, Community (Level 3) Variable

Maximum Likelihood Estimators



Type I Error, MLE, Community (Level 3) Variable

Table 1. Descriptive Statistics			
Cebu: Child's Weight		**Philippines: Use of Family Planning**	
Variable	Mean (SD)	Variable	Mean (SD)
Community Variables			
Price of Formula (n=429)	4.4553 (1.4513)	Public Province Health Exp. (n=75)	.4353 (.4716)
Price of Corn (n=429)	5.9831 (1.5006)	Public Mun. Health Exp. (n=484)	.5094 (.3288)
Urban Barangay (n=33)	.5151 (.5000)		
Individual Variables			
Child's Weight (n=34,297)	7.3889 (2.1513)	Use of FP (n=7,492)	.6884 (.4631)
Mother's Age (n=3073)	26.0680 (5.9915)	Age 15 to 19 (n=7,492)	.0119 (.1083)
Mother's Education (n=3073)	7.1132 (3.3115)	Age 20 to 24 (n=7,492)	.1481 (.3552)
Mother's Height (n=3073)	150.5611 (5.0042)	Age 25 to 29 (n=7,492)	.2714 (.4448)
Child is a Boy (n=3073)	.5307 (.4991)	Age 30 to 34 (n=7,492)	.2474 (.4316)
		Age 35 to 39 (n=7,492)	.1817 (.3856)
		Age 40 to 44 (n=7,492)	.0874 (.2825)
		Education 0 to 5 Yrs (n=7,492)	.1517 (.3588)
		Education 6 Yrs (n=7,492)	.1969 (.3977)
		Education 7 to 10 (n=7,492)	.3812 (.4857)
		Partner's Ed. 0 to 5 Yrs (n=7,492)	.1931 (.3948)
		Partner's Ed. 6 Yrs (n=7,492)	.1771 (.3818)
		Partner's Ed. 7 to 10 Yrs (n=7,492)	.3509 (.4773)
		Respondent Lives in Urban Area (n=7,492)	.3990 (.4897)
		Asset Index (n=7,492)	2.1307 (2.3751)
		Religion is Catholic (n=7,492)	.7623 (.4257)

Table 2. The Determinants of Child's Weight in the Philippines (Standard Errors in Parentheses)						
Variable	OLS	OLS Individual Correction	OLS Community Correction	MLE Individual Only	MLE Community Only	MLE Both
Community Variables						
Price of Formula	.0108 (.0052)	.0108 (.0083)	.0108 (.0109)	-.0031 (.0033)	-.0031 (.0057)	-.0035 (.0033)
Price of Corn	-.0138 (.0060)	-.0138 (.0070)	-.0138 (.0077)	-.0010 (.0035)	-.0084 (.0061)	-.0009 (.0035)
Urban	-.0301 (.0122)	-.0301 (.0351)	-.0301 (.0635)	-.0371 (.0342)	-.0799 (.0459)	-.06444 (.0486)
Individual Variables						
Mother's Age	-.0028 (.0008)	-.0028 (.0025)	-.0028 (.0023)	-.0029 (.0023)	-.0030 (.0008)	-.0032 (.0023)
Mother's Years of Education	.0570 (.0016)	.0570 (.0047)	.0570 (.0074)	.0531 (.0045)	.0591 (.0016)	.0544 (.0046)
Mother's Height	.0406 (.0010)	.0406 (.0029)	.0406 (.0031)	.0396 (.0029)	.0399 (.0010)	.0391 (.0028)
Child is a Boy	.5127 (.0099)	.5127 (.0292)	.5127 (.0252)	.4782 (.0280)	.5062 (.0099)	.4754 (.0279)
Age 2 months	1.8941 (.0239)	1.8941 (.0111)	1.8941 (.0157)	1.8944 (.0134)	1.8983 (.0237)	1.8943 (.0134)
Age 4 months	3.1415 (.0243)	3.1415 (.0154)	3.1415 (.0213)	3.1393 (.0137)	3.1506 (.0242)	3.1394 (.0137)
Age 6 months	3.8753 (.0250)	3.8753 (.0184)	3.8753 (.0261)	3.8661 (.0141)	3.8860 (.0248)	3.8662 (.0141)
Age 8 months	4.3267 (.0259)	4.3267 (.0236)	4.3267 (.0285)	4.3120 (.0146)	4.3388 (.0257)	4.3122 (.0146)
Age 10 months	4.6715 (.0270)	4.6715 (.0236)	4.6715 (.0327)	4.6472 (.0153)	4.6841 (.0269)	4.6474 (.0153)
Age 12 months	4.9814 (.0283)	4.9814 (.0261)	4.9814 (.0371)	4.9511 (.0161)	4.9952 (.0282)	4.9513 (.0161)
Age 14 months	5.2641 (.0294)	5.2641 (.0282)	5.2641 (.0438)	5.2284 (.0167)	5.2783 (.0293)	5.2287 (.0167)
Age 16 months	5.5459 (.0301)	5.5459 (.0293)	5.5459 (.0478)	5.5109 (.0171)	5.5631 (.0300)	5.5112 (.0171)
Age 18 months	5.8302 (.0301)	5.8302 (.0301)	5.8302 (.0483)	5.7965 (.0172)	5.8487 (.0300)	5.7969 (.0172)
Age 20 months	6.1396 (.0296)	6.1396 (.0304)	6.1396 (.0509)	6.1085 (.0169)	6.1608 (.0296)	6.1090 (.0169)
Age 22 months	6.4593 (.0281)	6.4593 (.0286)	6.4593 (.0468)	6.4327 (.0160)	6.4776 (.0280)	6.4331 (.0160)
Age 24 months	6.8048 (.0268)	6.8048 (.0269)	6.8048 (.0443)	6.7748 (.0152)	6.8204 (.0267)	6.7750 (.0152)
Constant	-3.7223 (.15414)	-3.7223 (.4454)	-3.7223 (.4221)	-3.4903 (.4269)	-3.5361 (.1575)	-3.3931 (.4266)

Table 2. Continued						
Variances						
σ^2 columns 1,2,3 σ^2_ε columns 4,5,6	.8378 (.00004)	.8378 (.00004)	.8378 (.00004)	.2573 (.0021)	.8238 (.0063)	.2573 (.0020)
σ^2_λ				.5666 (.0153)		.5585 (.0151)
σ^2_μ					.0156 (.0042)	.0083 (.0039)

Table 3. Determinants of Contraceptive Use in the Philippines (Probit with standard errors in parentheses)

Variable	Probit	Probit Municipality Correction	Probit Province Correction	MLE Municipality Only	MLE Province Only	MLE Both
Province and Municpality						
Public Province Health Expenditures	.0006 (.0438)	.0006 (.1324)	.0006 (.1325)	-.0735 (.0948)	.0502 (.0965)	-.0556 (.1515)
Public Municipality Health Expenditures	.2622 (.0559)	.2622 (.1614)	.2622 (.1733)	.1119 (.1607)	-.0381 (.0771)	.0979 (.1639)
Individual						
Age 15 to 19	.9492 (.1533)	.9492 (.2900)	.9492 (.2906)	.9231 (.1791)	.7943 (.1528)	.9326 (.1797)
Age 20 to 24	1.3329 (.0823)	1.3329 (.1572)	1.3329 (.1612)	1.3860 (.0989)	1.3890 (.0803)	1.3834 (.0993)
Age 25 to 29	1.3994 (.0773)	1.3994 (.1439)	1.3994 (.1398)	1.6209 (.0924)	1.4565 (.0772)	1.6157 (.0930)
Age 30 to 34	1.2681 (.0773)	1.2681 (.1403)	1.2681 (.1259)	1.5868 (.0929)	1.3240 (.0744)	1.5775 (.0936)
Age 35 to 39	1.1715 (.0795)	1.1715 (.1544)	1.1715 (.1613)	1.4074 (.0960)	1.2444 (.0785)	1.4055 (.0956)
Age 40 to 44	.7975 (.0867)	.7975 (.1898)	.7975 (.1918)	.8808 (.1062)	.8880 (.0863)	.8774 (.1061)
Education 0 to 5 years	-.4444 (.0657)	-.4444 (.1644)	-.4444 (.1508)	-.5286 (.0879)	-.4339 (.0716)	-.5329 (.0882)
Education 6 years	-.1519 (.0577)	-.1519 (.1444)	-.1519 (.1574)	-.3116 (.0748)	-.2313 (.0624)	-.3209 (.0756)
Education 7 to 10 years	.1269 (.0475)	.1269 (.1095)	.1269 (.1242)	.1609 (.0595)	.0599 (.0509)	.1495 (.0599)
Partner's Education 0 to 5 Years	.1206 (.0608)	.1206 (.1683)	.1206 (.1611)	.0185 (.0777)	.0528 (.0657)	.0212 (.0785)
Partner's Education 6 Years	.3344 (.0596)	.3344 (.1548)	3344 (.1529)	.2844 (.0762)	.3318 (.0644)	.2813 (.0770)
Partner's Education 7 to 10 Years	.2607 (.0462)	.2607 (.1174)	.2607 (.1007)	.2010 (.0577)	.2855 (.0500)	.2102 (.0584)
Urban	-.0484 (.0352)	-.0484 (.0894)	-.0484 (.0945)	-.1484 (.0627)	-.1293 (.0398)	-.1611 (.0678)
Asset Index	.0082 (.0081)	.0082 (.0251)	.0082 (.0242)	-.0077 (.0109)	.0128 (.0093)	-.0066 (.0107)
Religion is Catholic	.0675 (.0377)	.0675 (.0929)	.0675 (.1140)	.0213 (.0534)	-.0360 (.0443)	.0096 (.0542)
Constant	-.9751 (.0964)	-.9751 (.2163)	-.9751 (.2649)	-.7323 (.1507)	-.8755 (.1078)	-.7477 (.1606)
ρ_λ				.4542 (.0228)		.3960 (.0315)

ρ_μ					.2084 (.0267)	.0560 (.0288)