Seemingly Unrelated Regressions with Identical Regressors: a Note

Debopam Bhattacharya*
Department of Economics,
Dartmouth College

February 8, 2004.

Abstract

I show that in a seemingly unrelated, multivariate normal model with identical regressors, where some equations are linear and others are in the limited dependent variable (LDV) form, joint MLE can lead to efficiency gains for parameters of the LDV equations but not for parameters of the linear equations.

JEL code: C0,C3. Keywords: SUR, Bivariate Normal, Likelihood.

1 Introduction

It is well-known that in a system of linear seemingly unrelated regression equations with identical regressors, equation by equation OLS yield efficient estimates of the coefficient vectors (See Greene, 1989, pages 488-89, for a text-book treatment). This note extends that result to identical regressors in a seemingly unrelated, (latent) multivariate normal model, where some equations are linear and others are in the limited dependent variable form (here we investigate only the binary dependent variables). An example would be where the dependent variable in first equation denotes whether an infant survived for up to 1 year after birth and the second equation denotes weight at birth and the common regressors are mother's health and educational status and family income. Note that this is different from usual selection-type models where one variable is observed only for individuals whose second variable has crossed a threshold. One may well have more than two equations in the model; as an example, consider multi-children households. The dependent variables of interest are the school enrollment status of different siblings- one for the oldest, one for the second oldest etc. and another equation where the dependent variable is the family income. Although, family income is likely a major determinant of school-enrollment, it is clearly an endogenous variable and absent plausible instruments, one may still estimate the reduced-form equations (where the dependent variables are

*Address for correspondence: Department of Economics, 327, Rockefeller Hall, Dartmouth College, Hanover, NH 03755. Phone: 603-359-5994, e-mail: debopam.bhattacharya@dartmouth.edu

the enrollment status of the different siblings plus income) with all exogenous variables included as regressors. Given that unobserved components for the different equations are likely to be correlated, a reasonable specification for this set of reduced-form equations is the SUR and a question of interest is whether using all the equations in the system improves the efficiency of the estimates.

The main result of this note is that in such a system with structural errors being multivariate normal, equation-by-equation OLS still yields efficient estimates of the coefficient vector for the linear equations. However, a (multivariate) probit for the limited dependent variable (LDV) equations alone, ignoring the linear equations in the system, will not necessarily yield asymptotically equivalent estimates of the LDV equation parameters which are obtained by maximizing the full-information joint likelihood (FIML). The result continues to hold if the regressors in the linear equation are a linear combination (in particular, a subset) of the regressors in the non-linear equations.

In section 2, we show the results with one pair of equations, to get the idea across in a simplified manner and state (without proof) the general case. In section 3, we provide an empirical illustration.

2 Theory

2.1 One pair of equations

For simplicity, we consider a two equation normal SUR model where the first equation is a probit and the second is normal linear regression. We also simplify the notation by normalizing both variances to 1 (our main results do not depend on this normalization). Specifically, then, consider the model

$$\begin{aligned} y_{1i}^* &= X_i\beta + \varepsilon_i \\ y_{2i} &= X_i\gamma + \upsilon_i \end{aligned} \tag{1}$$

with $(\varepsilon_i, \upsilon_i)$ are i.i.d. across i, independent of X_i and

$$\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

$\beta, \gamma \in \mathbb{R}^k$, $\rho \in (-1,1)$, the sequence $\{x_n\}$ belongs to $\mathbb{R}^k$ and are i.i.d. with finite second moments.

The observed version of the first equation is given by

$$y_1 = \left\{ \begin{array}{c} 0 \text{ if } y_{1i}^* \leq 0 \\ 1 \text{ if } y_{1i}^* > 0 \end{array} \right\}.$$

Denoting by ϕ and Φ the standard normal density and cdf, the log-likelihood for a simple random sample is given by:

$$\begin{aligned} l &= \sum_1^N \ln \phi \left(y_{2i} - X_i\gamma\right) + \sum_1^N y_{1i} \ln \Phi \left(\frac{X_i\beta}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} \left(y_{2i} - X_i\gamma\right)\right) \\ &\quad + \sum_1^N \left(1 - y_{1i}\right) \ln \left(1 - \Phi \left(\frac{X_i\beta}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} \left(y_{2i} - X_i\gamma\right)\right)\right) \end{aligned}$$

where we have first conditioned on y_2. Reparametrizing

$$\theta = \frac{\rho}{\sqrt{1-\rho^2}}$$

and rewriting

$$\begin{aligned} \varepsilon_{2i} &= y_{2i} - X_i\gamma \\ z_i &= \frac{X_i\beta}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} \left(y_{2i} - X_i\gamma\right) \\ &= X_i\beta\sqrt{1+\theta^2} - \theta\left(y_{2i} - X_i\gamma\right) \end{aligned}$$

likelihood is given by

$$l = \sum_1^N \ln \phi \left(\varepsilon_{2i}\right) + \sum_1^N y_{1i} \ln \Phi\left(z_i\right) + \sum_1^N \left(1 - y_{1i}\right) \ln \left(1 - \Phi\left(z_i\right)\right)$$

First order conditions are given by

$$0 = \frac{\partial l}{\partial \beta} = \sqrt{1+\theta^2} \sum_1^N X_i \phi\left(z_i\right) \left\{ \frac{y_{1i} - \Phi\left(z_i\right)}{\Phi\left(z_i\right)\left(1 - \Phi\left(z_i\right)\right)} \right\}, \tag{2}$$

$$\begin{aligned} 0 &= \frac{\partial l}{\partial \gamma} = -\sum_1^N x_i \varepsilon_{2i} + \sum_1^N \theta X_i \phi\left(z_i\right) \left\{ \frac{y_{1i} - \Phi\left(z_i\right)}{\Phi\left(z_i\right)\left(1 - \Phi\left(z_i\right)\right)} \right\} \\ &= -\sum_1^N x_i \varepsilon_{2i} + \frac{\theta}{\sqrt{1+\theta^2}} \frac{\partial l}{\partial \beta}, \end{aligned} \tag{3}$$

and

$$0 = \frac{\partial l}{\partial \theta} = \frac{\theta}{\left(1+\theta^2\right)} \left(\frac{\partial l}{\partial \beta}\right)' \beta - \sum_1^N \varepsilon_{2i} \phi\left(z_i\right) \left\{ \frac{y_{1i} - \Phi\left(z_i\right)}{\Phi\left(z_i\right)\left(1 - \Phi\left(z_i\right)\right)} \right\} \tag{4}$$

Note from (2), that the second term in (3) is 0 so that (3) reduces to

$$\sum_1^N x_i \varepsilon_{2i} = 0 \tag{5}$$

which implies the OLS solution for γ (since the likelihood function is globally concave in the parameters[1], there is a unique solution to the first order conditions and that solution must satisfy (5)). Note that this result continues to hold if $X_2 \subset l(X_1)$ where X_1 are the regressors in the 1st equation and X_2 are those in the second and $l(X_1)$ denotes the linear space spanned by the columns of X_1.

We now derive the asymptotic covariance matrix for the MLE's and show that the asymptotic variance of $\hat{\beta}$ in FIML does not, in general, equal the asymptotic variance of the MLE of the simple probit estimator of β obtained by ignoring the linear equation. This shows, therefore, that not only are the estimates for β in the FIML and the simple probit numerically different in finite samples, they are also asymptotically different.

2.2 Asymptotic variances

The asymptotic covariance matrix for $\hat{\alpha} = \left(\hat{\gamma}, \hat{\beta}, \hat{\theta}\right)$ is given by $\Sigma = \left\{ E\left\{ \left(\frac{\partial l}{\partial \alpha}\right)\left(\frac{\partial l}{\partial \alpha}\right)' \right\}_{\alpha=\alpha_0} \right\}^{-1}$. In what follows, all expectations are computed at the true parameter values and we drop the subscript $\alpha = \alpha_0$. Now,

$$E\left\{\left(\frac{\partial l}{\partial \alpha}\right)\left(\frac{\partial l}{\partial \alpha}\right)'\right\}$$

$$= \begin{pmatrix} E(x_i x_i') + \frac{\theta^2}{1+\theta^2} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & . & . \\ \frac{\theta^2}{\sqrt{1+\theta^2}} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & . \\ \begin{array}{c} \frac{\theta^2}{(1+\theta^2)^{3/2}} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) \\ +E\left(\frac{\varepsilon_{2i}^2 x_i \phi(z_i)}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} & \frac{\theta}{1+\theta^2} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta & \begin{array}{c} \frac{\theta^2}{(1+\theta^2)^2} \beta' E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta \\ +E\left(\frac{\varepsilon_{2i}^2 \phi^2(z_i)}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} \end{pmatrix}.$$

Standard results on inverses of partitioned matrices (e.g. Rao, 1989, page 33) implies that the second diagonal block in Σ^{-1} (which equals the asymptotic variance of $\hat{\beta}$ when the system is jointly estimated) equals the inverse of

$$E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) - \begin{pmatrix} \frac{\theta^2}{\sqrt{1+\theta^2}} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & \frac{\theta}{1+\theta^2} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta' \end{pmatrix}$$

$$\times \begin{pmatrix} E(x_i x_i') + \frac{\theta^2}{1+\theta^2} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & \begin{array}{c} \frac{\theta^2}{(1+\theta^2)^{3/2}} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta \\ +E\left(\frac{\varepsilon_{2i}^2 x_i \phi(z_i)}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} \\ \begin{array}{c} \frac{\theta^2}{(1+\theta^2)^{3/2}} \beta' E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) \\ +E\left(\frac{\varepsilon_{2i}^2 \phi(z_i) x_i'}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} & \begin{array}{c} \frac{\theta^2}{(1+\theta^2)^2} \beta' E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta \\ +E\left(\frac{\varepsilon_{2i}^2 \phi^2(z_i)}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \frac{\theta^2}{\sqrt{1+\theta^2}} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & \frac{\theta}{1+\theta^2} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta \end{pmatrix}'.$$

[1] This follows from the fact that the likelihood for a bivariate normal distribution is concave in the parameters and our likelihood is simply the integral of the bivariate normal likelihood, integrated over the range of one of the variables.

Letting

$$\begin{pmatrix} P^{11} & P^{12} \\ P^{21} & P^{22} \end{pmatrix} = \begin{pmatrix} E\left(x_i x_i'\right) + \frac{\theta^2}{1+\theta^2} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) & \begin{array}{c} \frac{\theta^2}{\left(1+\theta^2\right)^{3/2}} E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta \\ +E\left(\frac{\varepsilon_{2i}^2 x_i \phi(z_i)}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} \\ \begin{array}{c} \frac{\theta^2}{\left(1+\theta^2\right)^{3/2}} \beta' E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right) \\ +E\left(\frac{\varepsilon_{2i}^2 \phi(z_i) x_i'}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} & \begin{array}{c} \frac{\theta^2}{\left(1+\theta^2\right)^{2}} \beta' E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)\beta \\ +E\left(\frac{\varepsilon_{2i}^2 \phi^2(z_i)}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{array} \end{pmatrix}^{-1} \text{ and}$$

$$A_2 = E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right),$$

the asymptotic variance of $\hat{\beta}$ in the FIML estimation is the inverse of

$$V = A_2 - \frac{\theta^4}{1+\theta^2} A_2 P^{11} A_2 - \frac{\theta^2}{\left(1+\theta^2\right)^{3/2}} A_2 \left(P^{12} + P^{21}\right) A_2 - \frac{\theta^2}{\left(1+\theta^2\right)^2} A_2 \beta\beta' A_2 P^{22}.$$[2]

The asymptotic variance of the probit estimator that ignores the linear equation is the inverse of

$$A_2^0 = E\left(\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \beta'}\right)_{\beta=\beta_0, \theta=0} = E\left(\frac{x_i x_i' \left\{\phi\left(x_i'\beta_0\right)\right\}^2}{\Phi\left(x_i'\beta_0\right)\left(1 - \Phi\left(x_i'\beta_0\right)\right)}\right)$$

While these expressions do not permit a direct comparison of the matrices V and A_2^0, using the expressions for P^{12}, P^{21}, P^{22} one may verify that in general $V \neq A_2^0$ (no term in the expression for V involves the quantity $E\left\{\frac{x_i x_i' \left\{\phi\left(x_i'\beta_0\right)\right\}^2}{\Phi\left(x_i'\beta_0\right)\left(1-\Phi\left(x_i'\beta_0\right)\right)}\right\}$) although the definiteness of $\left(A_2^0 - V\right)$ is ambiguous from these expressions[3]. What we have shown here is that the asymptotic variance of the FIML estimator for β is different from that of the marginal likelihood maximizer. Also note that in the special case that $\rho = 0$ and therefore $\theta = 0$, the expression for V implies that $A_2^0 = V$.

[2] In the special case where $k = 1$ i.e. x is a scalar, the asymptotic variance of the FIML estiamte of β is the inverse of

$$M \frac{E\left(x_i^2\right) E\left(\frac{\phi^2(z_i)\varepsilon_2^2}{\Phi(z_i)(1-\Phi(z_i))}\right) - \left(E\left(\frac{x_i \phi(z_i)\varepsilon_2^2}{\Phi(z_i)(1-\Phi(z_i))}\right)\right)^2}{\Delta}$$

where

$$\begin{aligned} M &= E\left(\frac{\phi^2(z_i) x_i^2}{\Phi(z_i)(1-\Phi(z_i))}\right), \\ \Delta &= E\left(\frac{\phi^2(z_i)\varepsilon_2^2}{\Phi(z_i)(1-\Phi(z_i))}\right) E x_i^2 - \left(E\left(\frac{x_i \phi(z_i)\varepsilon_2^2}{\Phi(z_i)(1-\Phi(z_i))}\right)\right)^2 \\ &\quad + M \frac{\theta^2}{1+\theta^2} E\left(\frac{\phi^2(z_i)\varepsilon_2^2}{\Phi(z_i)(1-\Phi(z_i))}\right) + M \frac{\beta^2\theta^2}{\left(1+\theta^2\right)^2} E x_i^2 \\ &\quad - 2M \frac{\beta\theta^2}{\left(1+\theta^2\right)^{3/2}} E\left(\frac{x_i \phi(z_i)\varepsilon_2^2}{\Phi(z_i)(1-\Phi(z_i))}\right) \end{aligned}$$

[3] An analogy is the tobit model where one can estimate the slopes using the probit part of the likelihood alone and those estimates are less efficient than the one that maximizes the full Tobit likelihood (c.f. Amemiya (1985), page 366). The situation is different here in that the FIML also estimates additional parameters ρ and γ which affect the asymptotic variance of β and makes the comparison ambiguous.

2.3 The general case

We state the general result without proof. The proof follows exactly the same steps as the one for a pair of equations but is only notationally far unwieldy. Consider a set-up where we have k linear and m nonlinear equations with identical regressors (which are assumed to possess finite second moments), and correlated structural errors that follow a joint normal distribution. Specifically:

$$\begin{aligned} Y_j &= X\gamma_j + u_j,\ j = 1, ...k \\ Y_j^* &= X\beta_{j-k} + v_{j-k},\ j = k+1...k+m \\ Y_j &= 1\left(Y_j^* > 0\right), j = k+1...k+m \end{aligned}$$

where $\begin{pmatrix} u_{k\times 1} \\ v_{m\times 1} \end{pmatrix}$ are independent of X, $\begin{pmatrix} u \\ v \end{pmatrix} \sim N_{k+m}(0, V)$ where V is conformably partitioned as

$$V = \begin{pmatrix} \Sigma_{uu} & \Sigma_{vu} \\ \Sigma_{uv} & \Sigma_{vv} \end{pmatrix}.$$

One observes realizations of $(Y_j)_{j=1,...k+m}$ and X in the data.

First note that

$$(v|u) \sim N_k\left(\Sigma_{uv}\Sigma_{uu}^{-1}u, \Sigma_{vv} - \Sigma_{uv}\Sigma_{uu}^{-1}\Sigma_{vu}\right).$$

Let

$$\begin{aligned} Y_k &= (y_{11}, y_{12}, ...y_{1n}; y_{21}...y_{2n}; ...y_{k,1}...y_{k,n})', \\ \gamma &= (\gamma_1, ...\gamma_k)', \beta = (\beta_1, ...\beta_m)'. \end{aligned}$$

For scale identification, of β, a scale normalization of $diag(\Sigma_{vv}) = I_m$ is essential.

Then, ignoring constants in the likelihood that do not depend on (β, γ), likelihood for an i.i.d. sample is then given by

$$\begin{aligned} & L(y_1...y_k, y_{k+1...k+m}|X, \theta) \\ = \; & L(y_1...y_k|X, \theta) \times L(y_{k+1...}, y_{k+m}|y_1...y_k, X, \theta) \\ = \; & \exp\left(-\frac{1}{2}(Y_k - (I_k \otimes X)\gamma)'\left(\Sigma_{uu}^{-1} \otimes I_k\right)(Y_k - (I_k \otimes X)\gamma)\right) \times L_2(\beta, \gamma). \end{aligned}$$

Therefore the log-likelihood is given by

$$-\frac{1}{2}(Y_k - (I_k \otimes X)\gamma)'\left(\Sigma_{uu}^{-1} \otimes I_m\right)(Y_k - (I_k \otimes X)\gamma) + l_2(\beta, \gamma).$$

Then, first order conditions are given by

$$0 = \frac{\partial l}{\partial \beta} = \frac{\partial l_2}{\partial \beta} \tag{6}$$

and

$$0 = \frac{\partial l}{\partial \gamma} = (I_k \otimes X)' \left(\Sigma_{uu}^{-1} \otimes I_m\right) (Y_k - (I_k \otimes X)\,\gamma) + \frac{\partial l_2}{\partial \gamma}.$$

One may verify that here, $\frac{\partial l_2}{\partial \beta} = 0$ implies $\frac{\partial l_2}{\partial \gamma} = 0$ so that one gets identical estimates for γ by maximizing the full likelihood as one would get by k separate OLS estimation of the first k equations. Also, the asymptotic variance for FIML estimates of β differ from those that ignore the linear equations.

3 Empirical illustration

In this section we illustrate our results using a dataset of infant births. The data come from the national maternal and infant health survey (NMIHS), conducted in the US in 1988. We estimate the following model SUR model

$$\begin{aligned} BO_{1i}^* &= X_i\beta + \varepsilon_i \\ BO_i &= 1\,(BO_{1i}^* > 0) \\ BW_i &= X_i\gamma + \upsilon_i \end{aligned}$$

with $(\varepsilon_i, \upsilon_i)$ independent of X_i and

$$\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}\right).$$

Here, i indexes the birth, BO denotes whether the infant survived up to one year after birth (BO^* is the latent health stock), BW denotes birthweight in kilograms and the common regressors are mother's years of education, household income (in \$000) and its square, average number of cigarettes smoked by mother during pregnancy, a dummy for whether child was female and a dummy for whether mother was overweight (Body Mass Index greater than 30). The number of observations equals 2529. These estimations were performed using the statistical package Stata, where the full information maximum likelihood was performed using stata's maximum likelihood routine where the likelihood function was manually entered. In table 1, I report the results of the joint estimation, in table 2, we report the OLS coefficients on the second (birthweight) equation and in table 3, the coefficients from a probit of the first (survival) equation. From table 2 and panel 1 of table 1, one can see that for the second (i.e. birthweight) equation we get identical coefficients and standard errors. From table 3 and panel 2 of table 1, we get non-identical but close estimates for the coefficients in the survival equation but the standard errors are smaller when the system is jointly estimated.

References

[1] Amemiya, T. (1985): Advanced Econometrics, Harvard University Press, Cambridge, Massachusetts.

[2] Rao, C.R. (1989): “Linear Statistical Inference and its Applications.” Wiley eastern Limited.

[3] Greene, William (1989): ”Econometric Analysis” 2nd edition, Macmillan Book company.

[4] Zellner, A. (1962): “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias", Journal of the American Statistical Association, 57, 500-509.

Table 1: Results from Joint Estimation

Panel 1 (birthweight)			
momed	.0152	.0064	2.37
income	.0021	.0048	0.43
income_square	-.0002	.0001	-1.44
female_child	-.1162	.0282	-4.12
smoke	-.0112	.0026	-4.26
mom_overweight	.0207	.0599	0.35
intercept	3.4432	.0899	38.30
Panel 2 (survival)			
momed	.0173	.0081	2.08
income	.0124	.0050	2.39
income_square	-.0002	.0001	-2.20
female_child	.0474	.0340	1.39
smoke	-.0064	.0020	-2.13
mom_overweight	-.1619	.0628	-2.56
intercept	2.5124	.1145	21.95
rho	0.4126	.0113	36.37
sigma	0.5827	.0182	32.01

Notes: Panel 1 shows the results for infant survival up to one year after birth, panel 2 shows the results for the birthweight equation when estimation is done jointly by maximizing the full likelihood. rho and sigma are as defined in section 3 of the text.

Table 2: OLS estimates for birthweight equation

birthweight	*Coef.*	*Std. Err.*	*t*
momed	.0152	.0064	2.37
income	.0021	.0048	0.43
income_square	-.0002	.0001	-1.44
female_child	-.1162	.0282	-4.12
smoke	-.0112	.0026	-4.26
mom_overweight	.0207	.0599	0.35
intercept	3.4432	.0899	38.30

Table 3: Probit estimates for survival equation

Survival	*Coef.*	*Std. Err.*	*t*
momed	.0156	.0087	1.80
income	.0119	.0056	2.12
income_square	-.0003	.0001	-2.11
female_child	.0709	.0347	2.05
smoke	-.0058	.0028	-2.03
mom_overwight	-.1919	.0641	-2.99
intercept	1.8677	.1174	15.91