5/ Sampling theory approach to estimation of the model:

Within the framework of sampling theory, there are two familiar estimation techniques, namely least squares (LS) estimation and maximum likelihood (ML) estimation. The labour supply model has two major specific characteristics, namely censored data, and two problems related to a simultaneous equation system, where the error terms are correlated: an endogenous wage regressor and the estimator efficency. These characteristics require special consideration when applying the above conventional estimation techniques. The purpose of this section is to show how these problems are solved with appropriate estimation techniques, which then are organized into formal estimation procedures. Finally, their advantages and disadvantages are discussed.

5.1/ Least squares estimation:

Following the above outline, techniques are developed to deal with the problems of censored data and a simultaneous equation system. These techiques are then organized into formal estimation procedures, and computer programs to implement the procedures are introduced. Finally, the results are presented and interpreted.

5.1.1/ Dealing with censored data and problems of a simultaneous equation system:

5.1.1.1/ Dealing with censored data problems:

First of all, the problems of a multivariate model are ignored and only the problems related to censored data are considered, i.e. the wage equation and the hours worked equation are to be considered separately.

5.1.1.1.1/ Censored data in the wage equation:

If wage equation (4.10) is estimated with LS, discarding the zero-values of LWW_i, the error term is truncated by the condition for $LWW_i>0$, namely $LWW_i>LWW_i^*$. The error term will have a truncated normal distribution, where the expectation of the truncated error term is different from zero and depends on the values of the included regressors. Hence, the estimates will suffer from sample selectivity and become biased. To deal with this problem, we develop an appropriate wage equation, where LS can be applied, when the zero-values of LWW_i are discarded.

The condition $LWW_i > LWW_i^*$, can be reformulated as $LWW_i - LWW_i^* > 0$, or $K_i > 0$.

where $$K_i = LWW_i - LWW_i^* \tag{5.1}$$

Recall that equation (4.10) has shorter form (4.2):

$$LWW_i = Z_i\Gamma + e_{wi} \qquad K_i > 0 \tag{5.2}$$

where $$Z_i\Gamma = a_0 + a_1WA_i + a_2WA2_i + a_3WE_i + a_4WPED_i + a_5UN_i + a_6CIT_i \tag{5.3}$$

We define a new error term:

$$e_{wi}^* = (LWW_i | K_i > 0) - E(LWW_i | K_i > 0) \tag{5.4}$$

Because equation (5.4) is only relevant when $LWW_i > LWW_i^*$, the new error term is truncated with a random truncation point. Some properties of the expectation and variance of this error term are derived (Heckman 1979):

$$E(e_{wi}^*) = E(LWW_i | K_i > 0) - E(LWW_i | K_i > 0) = 0 \tag{5.5}$$

$$\begin{aligned} var(e_{wi}^*) &= var[(LWW_i | K_i > 0) - E(LWW_i | K_i > 0)] \\ &= var(LWW_i | K_i > 0) \\ &= var[Z_i\Gamma + e_{wi} | K_i > 0] \\ &= var(e_{wi} | K_i > 0) \\ &= \sigma_{wi}^{*2} \end{aligned} \tag{5.6}$$

$var(e_{wi}) = \sigma_w^2$ is constant, but the condition $K_i > 0$ depends on observation i, so we can not expect $var(e_{wi} | K_i > 0) = \sigma_{wi}^{*2}$ to be a constant; it will depend on observation i. This can cause heteroskedasticity (Berndt 1991, p. 623). On the other hand, observations are assumed independent, so we can expect $\text{cov}(e^*_{wi}, e^*_{wj}) = 0$ for $i \neq j$. That means the new error term e^*_{wi} has zero-mean, is uncorrelated but its variance is not constant. In this last respect it is not well behaved. The properties of e^*_{wi} will have impacts on estimation techniques to be discussed later.

Rearranging equation (5.4) results in:

$$LWW_i | K_i > 0 = E(LWW_i | K_i > 0) + e^*_{wi} \tag{5.7}$$

Taking the expectation of both sides of equation (5.2), conditional on $K_i > 0$ we have:

$$E(LWW_i | K_i > 0) = Z_i \Gamma + E(e_{wi} | K_i > 0) \tag{5.8}$$

Equation (5.8) is then substituted into equation (5.7):

$$LWW_i | K_i > 0 = Z_i \Gamma + E(e_{wi} | K_i > 0) + e^*_{wi} \tag{5.9}$$

The next step is to derive $E(e_{wi} | K_i > 0)$. Equations (4.2) and (4.3) are substituted into equation (5.1):

$$\begin{aligned} K_i &= (Z_i \Gamma + e_{wi}) - (X_i^{**} \Theta + e_{ri}) \\ &= Z_i \Gamma - X_i^{**} \Theta + e_{wi} - e_{ri} \\ &= J_i + e_{di} \end{aligned} \tag{5.10}$$

where

$$J_i = Z_i \Gamma - X_i^{**} \Theta \tag{5.11}$$

$$e_{di} = e_{wi} - e_{ri} \tag{5.12}$$

We assume that $e_{wi} \sim N(0, \sigma_{ww})$, $e_{ri} \sim N(0, \sigma_{rr})$, and $\text{cov}(e_{wi}, e_{ri}) = \sigma_{wr}$. So $E(e_{di}) = E(e_{wi}) - E(e_{ri}) = 0$, furthermore $\text{var}(e_{di}) = \text{var}(e_{wi}) + \text{var}(e_{ri}) - 2\text{cov}(e_{wi}, e_{ri})$, which is independent of the observations: var(e_{di})= $\sigma_{ww}+\sigma_{rr}-2\sigma_{wr}=\sigma_d^2$.The distribution of e_{di} is then: $e_{di} \sim N(0,\sigma_d^2)$. Furthermore, we have:

$$E(e_{wi}|K_i>0)=E(e_{wi}|J_i+e_{di}>0)$$
$$=E(e_{wi}|e_{di}>-J_i)$$
$$=E(e_{wi}|\frac{e_{di}}{\sigma_d}>-\frac{J_i}{\sigma_d}) \tag{5.13}$$

Because $\frac{e_{di}}{\sigma_d} \sim N(0,1)$,

$$E(e_{wi}|\frac{e_{di}}{\sigma_d}>-\frac{J_i}{\sigma_d})=\frac{\sigma_{wd}}{\sigma_d}\frac{f(-J_i/\sigma_d)}{[1-F(-J_i/\sigma_d)]}=\frac{\sigma_{wd}}{\sigma_d}\frac{f(J_i/\sigma_d)}{F(J_i/\sigma_d)} \tag{5.14}$$

Where f and F are the density and distribution functions of the standard normal distribution respectively. σ_{wd} is the covariance between e_{wi} and e_{di}, which is independent of the observations, as can be proved as follows:

$$\text{cov}(e_{wi},e_{di})=E(e_{wi}*e_{di})=E[e_{wi}*(e_{wi}-e_{ri})]=E(e_{wi}^2-e_{wi}e_{ri})=\sigma_{ww}-\sigma_{wr}$$

The ratio $\frac{f(J_i/\sigma_d)}{F(J_i/\sigma_d)}=\lambda_i$ is called the inverse Mill's ratio. Its substitution into equation (5.14) results in:

$$E(e_{wi}|K_i>0)=\frac{\sigma_{wd}}{\sigma_d}\lambda_i \tag{5.15}$$

Substitution of equation (5.15) into equation (5.9) yields the appropriate wage equation if zero values of LWW_i are discarded (Killingsworth 1983d, p.159) :

$$LWW_i=Z_i\Gamma+\frac{\sigma_{wd}}{\sigma_d}\lambda_i+e_{wi}^* \qquad \text{if } K_i>0 \tag{5.16}$$

Recall that the error term of model (5.16) has zero-mean. This property ensures that LS-estimates of the coefficients will be consistent and makes equation (5.16) the appropriate wage equation, providing we assume that the inverse Mill's ratios λ_i are available. The generation of these ratios will be discussed later (see section 5.1.1.3). Furthermore, the error term is not autocorrelated, but its variance is not constant, causing heteroskedasticity.

Compare equation (5.16) with equation (5.2); we can easily see that equation (5.2) omits the relevant variable λ_i and this leads to biased and inconsistent estimates. In the case of discarding zero values of LWW_i, the bias is called sample selectivity bias.

For the case of retaining zero values of LWW_i in estimating equation (5.2), similar reasoning also indicates that some relevant variable is omitted and the resulting estimates are biased and inconsistent.

5.1.1.1.2/ Censored data in the hours worked equation:

In contrast to the wage equation, the hours worked equation contains wage as an endogenous regressor. This fact has implications when dealing with censored data. To simplify analysis, first we ignore the issue of correlation between LWW_i and e_{hi} and see LWW_i as an exogenous variable in the derivation of an appropriate hours worked equation for using LS with the zero values of $WHRS_i$ discarded. Then this assumption is lifted.

Assuming LWW_i is an exogenous variable, we apply the same approach as in the preceding section.

Consider the hours worked equation in structural form (4.4):

$$WHRS_i = c_1 LWW_i + X_i^{**}C + e_{hi} \qquad (5.17)$$

We define a new error term:

$$e_{hi}^* = (WHRS_i | K_i > 0) - E(WHRS_i | K_i > 0) \qquad (5.18)$$

The new error term is truncated with a random truncation point. Some properties of the expectation and variance of this error term are derived (Heckman 1979):

$$E(e_{hi}^*) = E(WHRS_i | K_i > 0) - E(WHRS_i | K_i > 0) = 0 \qquad (5.19)$$

$$
\begin{aligned}
\text{var}(e_{hi}^{*}) &= \text{var}[(\text{WHRS}_i | K_i > 0) - E(\text{WHRS}_i | K_i > 0)] \\
&= \text{var}(\text{WHRS}_i | K_i > 0) \\
&= \text{var}[c_0 + c_1 \text{LWW}_i + X_i^{**} C + e_{hi} | K_i > 0] \\
&= \text{var}(e_{hi} | K_i > 0) \\
&= \sigma_{hi}^{*2}
\end{aligned}
\tag{5.20}
$$

$\text{var}(e_{hi})=\sigma_h^2$ is constant, but the condition $K_i>0$ depends on observation i, so we expect $\text{var}(e_{hi}^{*}) = \sigma_{hi}^{*2}$ to be dependent on observation i too. On the other hand , the observations are assumed to be independent, so we expect $\text{cov}(e_{hi}^{*}, e_{hj}^{*}) = 0$ for $i \neq j$. That means the new error term e_{hi}^{*} has zero-mean, is uncorrelated, but its variance is not constant, and because of this last property it is not well behaved.

Rearranging equation (5.18) results in:

$$
\text{WHRS}_i | K_i > 0 = E(\text{WHRS}_i | K_i > 0) + e_{hi}^{*} \tag{5.21}
$$

Taking the expectation of both sides of equation (5.17), conditional on $K_i > 0$ we have:

$$
E(\text{WHRS}_i | K_i > 0) = c_0 + c_1 \text{LWW}_i + X_i^{**} C + E(e_{hi} | K_i > 0) \tag{5.22}
$$

Note again that equation (5.17) is treated as if the wage equation did not exist, so we ignore the issue of correlation between LWW_i and e_{hi} and see LWW_i as an exogenous variable.

Equation (5.22) is then substituted into equation (5.21):

$$
\text{WHRS}_i | K_i > 0 = c_0 + c_1 W_i + X_i^{*} C + E(e_{hi} | K_i > 0) + e_{hi}^{*} \tag{5.23}
$$

The next step is to derive $E(e_{hi} | K_i > 0)$.

$$E(e_{hi}|K_i > 0) = E(e_{hi}|J_i + e_{di} > 0)$$
$$= E(e_{hi}|e_{di} > -J_i)$$
$$= E(e_{hi}|\frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) \tag{5.24}$$

Because $\frac{e_{di}}{\sigma_d} \sim N(0,1)$,

$$E(e_{hi}|\frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) = \frac{\sigma_{hd}}{\sigma_d} \frac{f(-J_i / \sigma_d)}{[1 - F(-J_i / \sigma_d)]} = \frac{\sigma_{hd}}{\sigma_d} \frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} \tag{5.25}$$

where f and F are the density and distribution functions of the standard normal distribution respectively and σ_{hd} =cov(e_{hi},e_{di}) is independent of the observation, as can be proved in the same way as in (5.1.1.1.1).

The inverse Mill's ratio $\frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} = \lambda_i$ appears again here . Its substitution into equation (5.25) results in:

$$E(e_{hi}|K_i > 0) = \frac{\sigma_{hd}}{\sigma_d} \lambda_i \tag{5.26}$$

Substitution of equation (5.26) into equation (5.23) yields the appropriate hours worked equation if zero values of $WHRS_i$ are discarded:

$$WHRS_i | K_i > 0 = c_1 LWW_i + X_i^{**} C + \frac{\sigma_{hd}}{\sigma_d} \lambda_i + e_{hi}^* \tag{5.27}$$

Recall again that the error term of equation (5.27) has zero-mean; this property ensures that LS-estimates of the coefficients will be consistent and makes equation (5.27) the appropriate hours worked equation.

Comparing equation (5.27) with equation (5.17), we can easily see that equation (5.17) omits the relevant variable λ_i and this leads to biased and inconsistent estimates. In the case of discarding zero values of $WHRS_i$, the bias is called sample selectivity bias.

Equation (5.27) is appropriate on the condition that LWW_i is treated as exogenous. If it is treated as endogenous and becomes correlated with e_{hi}, the equation is no longer appropriate. This is a weakness that does not appear in the appropriate wage equation, where there is no endogenous variable involved. To avoid this problem, the reduced form hours worked equation should be employed. Consider reduced form equation (4.5):

$$WHRS_i = c_1 Z_i \Gamma + X_i^{**} C + e_{gi} \qquad \text{if } K_i > 0 \tag{5.28}$$

where $$e_{gi} = e_{hi} + c_1 e_{wi} \tag{5.29}$$

With the same reasoning as for the structural form of the hours worked equation, we have the following corrected reduced form:

$$WHRS_i | K_i > 0 = c_1 Z_i \Gamma + X_i^{**} C + E(e_{gi} | K_i > 0) + e_{hi}^{*} \tag{5.30}$$

Also with the same reasoning, the new error term has zero-mean and is uncorrelated. But its variance depends on observation i and is not constant. This last property makes the error term not well behaved.

The next step is to derive $E(e_{gi}|K_i>0)$.

$$\begin{aligned}
E(e_{gi} | K_i > 0) &= E(e_{gi} | J_i + e_{di} > 0) \\
&= E(e_{gi} | e_{di} > -J_i) \\
&= E(e_{gi} | \frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) \\
&= E(e_{hi} + c_1 e_{wi} | \frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) \\
&= E(e_{hi} | \frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) + c_1 E(e_{wi} | \frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d})
\end{aligned} \tag{5.31}$$

Because $\frac{e_{di}}{\sigma_d} \sim N(0,1)$,

$$E(e_{hi} | \frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) = \frac{\sigma_{hd}}{\sigma_d} \frac{f(-J_i / \sigma_d)}{[1 - F(-J_i / \sigma_d)]} = \frac{\sigma_{hd}}{\sigma_d} \frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} \tag{5.32}$$

$$E(e_{wi} | \frac{e_{di}}{\sigma_d} > -\frac{J_i}{\sigma_d}) = \frac{\sigma_{wd}}{\sigma_d} \frac{f(-J_i / \sigma_d)}{[1 - F(-J_i / \sigma_d)]} = \frac{\sigma_{wd}}{\sigma_d} \frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} \tag{5.33}$$

Substitution of equations (5.32), (5.33) and then the inverse Mill's ratio $\frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} = \lambda_i$ into equation (5.31) yields

$$E(e_{gi} | K_i > 0) = \frac{\sigma_{hd} + c_1 \sigma_{wd}}{\sigma_d} \lambda_i \tag{5.34}$$

Substitution of equation (5.34) into equation (5.30) yields the appropriate hours worked equation if zero values of $WHRS_i$ are discarded (Killingsworth 1983, p. 160):

$$WHRS_i | K_i > 0 = c_1 Z_i \Gamma + X_i^{**} C + \frac{\sigma_{hd} + c_1 \sigma_{wd}}{\sigma_d} \lambda_i + e_{hi}^{*} \tag{5.35}$$

Recall again that the error term of equation (5.35) has zero-mean; this property ensures that LS-estimates of the coefficients will be consistent and makes equation (5.39) appropriate. Furthermore, the error term is not autocorrelated, but its variance is not constant, causing heteroskedasticity.

Comparing equation (5.35) with equation (5.28), we can easily see that equation (5.28) omits the relevant variable λ_i and this leads to biased and inconsistent estimates.

For the case of retaining zero values of $WHRS_i$ in estimating equation (5.17) or (5.28), similar reasoning also indicates that some relevant variable is omitted and the resulting estimates are biased and inconsistent.

5.1.1.2/ Dealing with problems from a bivariate model:

The system consists of three equations, so the model might be trivariate. But only observations on the dependent variables of two equations, namely the wage and the hours worked equations, are available, so the model is in fact bivariate in terms of LS estimation. We have established in section (2.2.2) that the two equations are correlated, so there are two problems related to this bivarite labour supply model, namely the endogenous wage regressor in the hours worked equation and the estimator efficiency.

5.1.1.2.1/ Dealing with endogenous wage regressor:

Recall that when dealing with problems related to censored data, we ignored the fact that the wage equation and the hours worked equation coexist in an equation system. The first problem in this regard is the simultaneous equation bias, because LWW_i as a regressor in the hours worked equation is correlated with the error term e_{hi}, as pointed out in section (2.2.2). But this problem has been removed by choosing reduced form equation (5.35) as the appropriate hours worked equation, which takes into account both the problem of censored data and the problem of an endogenous wage regressor. So the equation system now consists of equations (5.16) and (5.35) with zero-values of the dependent variables discarded (for this moment the reservation wage equation is put aside). The second problem related to a simultaneous equation system is the problem of identification. To illustrate it, we employ the rearrangement of equation (5.35) as we did for equation (4.9), and have the new equation system:

$$LWW_i=a_0+a_1WA_i+a_2WA2_i+a_3WE_i+a_4WPED_i +a_5UN_i+a_6CIT_i+\frac{\sigma_{wd}}{\sigma_d}\lambda_i+e^*_{wi} \quad \text{if} \quad LWW_i>LWW^*_i \tag{5.36}$$

$$WHRS_i=d_0+d_1WA_i+d_2WA2_i+d_3WE_i+d_4WPED_i +d_5UN_i+d_6CIT_i+d_7PRIN_i+d_8KL6_i +d_9K618_{i1}+\frac{\sigma_{hd}+c_1\sigma_{wd}}{\sigma_d}\lambda_i+e^*_{hi} \quad \text{if} \quad LWW_i>LWW^*_i \tag{5.37}$$

$$\text{where } d_0=c_0+c_1a_0,\ d_1=c_1a_1+c_2,\ d_2=c_1a_2+c_3,\ d_3=c_1a_3+c_4,\ d_4=c_1a_4,\ d_5=c_1a_5+c_5,\ d_6=c_1a_6+c_6,\ d_7=c_7,\ d_8=c_8,\ d_9=c_9 \tag{5.38}$$

To simplify the estimation techniques involved, second generation studies generally ignore the fact that the error terms of the two equations are heteroskedastic and proceed as if they were homoskedastic.

If LS is applied to estimate equations (5.36) and (5.37), the problem of identification will emerge. The structural parameters c_i of hours worked equation (5.35) are dependent on the reduced form parameters d_i through algebraic equation system (5.38).

LS estimates of the reduced form parameters a_i (i=0,1...7) and d_i (i=0,1.....10) are unique (note that the a_i are at the same time structural parameters of the wage equation, because the structural form and the reduced from of the wage equation are the same), based on these estimates, structural parameter estimates for c_i (i=0,1...9) are derived. If it is possible to obtain the structural parameters estimates and the algebraic relationship between the structural and reduced form parameters is unique, then there is only one way solving for the structural parameters from reduced form parameters, i.e. the hours worked equation is just identified. But the derivation of the structural parameters from a_i and d_i can also be either impossible, i.e unidentified hours worked equation, or there is more than one algebraic relationship between the structural parameters and the reduced form parameters, i.e over identified hours worked equation. Based on the above explanation, we consider the identification status of our model.

From equation system (5.38), c_7, c_8 and c_9 are immediately determined uniquely. The remaining 7 unknowns (c_0,c_1.....c_6) can also be solved for uniquely from the remaining 7 linear equations. That means the model is just identified. As was mentioned in section (4.3), the regressor WPED is not included in the structural hours worked equation to ensure just identification of the equation. Now, general conditions for identification will be discussed.

First, we assume that the structural hours worked equation contains all the regressors of the wage equation. Then the number of coefficients d_i of the reduced form equation remains unchanged, i.e. the number of linear equations of system (5.38) is unchanged, while the number of unknowns has been increased by one and hence becomes more than the number of equations by one. The equation system (5.38) cannot produce unique structural parameters c_i, and the hours worked equation becomes unidentified. Now, assume that more than one regressor of the wage equation are dropped from the hours worked equation. Again, the number of linear equations of system (5.38) is unchanged, but the number of unknown has been decreased by more than one and becomes less than the number of equations by one or more. Equation system (5.38) can produce different sets of c_i, depending on different combinations of the equations of the system, and the hours worked equation becomes over identified. E.g., if two regressors WPED and WE are dropped, then system (5.38) contains two equations: d_3=c_1a_3 and d_4=c_1a_4, that means c_1 can be either d_3/a_3 or d_4/a_4. In conclusion, we can say that if no regressor from the wage equation is dropped, the model is unidentified, if one regressor is dropped the model is just identified, and if more than one regressor are dropped, the model is over identified.

If the model is just identified, indirect least squares (ILS) techniques can be applied to equations (5.36) and (5.37) by solving for c_i from equation system (5.38) as already done above.

If the model is over identified, modified two stage least squares (2SLS) techniques are necessary and the two equations, involved in estimation, are now (5.16) and (5.35). 2SLS here is not in the conventional sense, because the initial structural hours worked equation is not contained in the model. Instead there is the selection bias corrected reduced form hours worked equation (5.35), where $Z_i\Gamma$ is present in both reduced form equations (5.16) and (5.35). The estimate of Γ,denoted as $\hat{\Gamma}$, can be obtained from the selection bias corrected wage equation and then be substituted into the selection bias corrected reduced form hours worked equation to calculate the predicted value of LWW_i from the original wage equation (5.2), namely $PLWW_i=Z_i\hat{\Gamma}$, i.e., we can not use a prediction for LWW_i from the selection bias corrected wage equation (5.16) as a substitution in the conventional way. If the model is just identified ILS and 2SLS produce the same results. It is worth noting that if the model is unidentified, 2SLS also does not work, like ILS, because $\hat{\Gamma}$ is a linear combination of all the regressors of the selection bias corrected wage equation and so a linear combination of some of the regressors of the selection bias corrected hours worked equation. Then, use of $\hat{\Gamma}$ in the selection bias corrected hours worked equation would create exact multicolinearity and LS would not work.

Now we turn to the identification of variances and covariances σ_d, σ_{wd} and σ_{hd} of the error terms of the selection bias corrected reduced form equations.

We define $a_7 = \dfrac{\sigma_{wd}}{\sigma_d}$ and $c_{10} = d_{10} = \dfrac{\sigma_{hd} + c_1\sigma_{wd}}{\sigma_d}$. Here there are two equations for two unknowns, because $\hat{\sigma}_d$ can be obtained in relation with the reservation wage equation (discussed in section 5.1.2.2 below). So the remaining two parameters σ_{wd} and σ_{hd} can be identified in this model.

5.1.1.2.2/ The problem of estimator efficiency:

So far we have discussed the first problem related to a bivariate model, namely the endogenous wage regressor, that leads to simultaneous equation bias and identification problems. As a result, ILS and 2SLS techniques were suggested. The next problem is that of estimator efficiency. Until now, second generation studies ignore this problem.

ILS and 2SLS techniques both estimate each equation of the system separately, ignoring the fact that there may be efficiency gains by considering the contemporaneous correlation between the error terms in different equations. 3SLS (Judge et al., p. 636-670) and SUR, the Zellner estimation techniques for multivariate regression (Judge et al., p. 443-494), take into account the problem of correlated disturbances. In the case of over identification, 3SLS is applied to equations (5.16) and (5.35). As mentioned earlier, we assume the error terms of the equations are homoskedastic to make the procedure simple. The first two stages are identical to those for modified 2SLS, and are applied to each equation of the system. For the third stage, the 2SLS residuals are used to estimate the disturbance variances and covariances, and these estimates are used to obtain estimated generalised least squares (EGLS) estimates for the whole system. The third stage can be repeated until some convergence criterion is reached. Because 2SLS is modified, so is 3SLS. In the case of just identification, SUR can be applied to equations (5.36) and (5.37). SUR is similar to 3SLS, but based on ILS. First, LS is applied to the two equations separately. Second, residuals are used to estimate the disturbance variances and covariances, and these estimates are used to obtain estimated generalised least squares (EGLS) estimates for the whole system. The second stage can be repeated until some convergence criterion is reached. Finally, the structural parameter estimates can be solved for from the reduced from parameter estimates, using equation (5.38). If the model is correctly specified, 3SLS and SUR estimators will be more efficient than 2SLS and ILS estimators, because 3SLS and SUR use information about the complete system. However, these techniques are more sensitive to misspecification, because misspecification can have widespread effects. If the model is just identified, modified 3SLS and SUR produce the same results, so do ILS and 2SLS.

The final problem is the problem related to heteroskedastic error terms, as pointed out in section (5.1.1.1). In this case, LS estimates of coefficients are consistent, but the estimates of variances of LS coefficient estimates are biased if using the conventional formula from most computer packages:

$$\hat{V}(b) = \hat{\sigma}^2 (X' X)^{-1} \tag{5.39}$$

The formula for unbiased estimates of variances of OLS coefficient estimates is:

$$\hat{V}(b) = \hat{\sigma}^2 (X' X)^{-1} (X' \Psi X)(X' X)^{-1} \tag{5.40}$$

In (5.39) and (5.40) we have used X and Ψ to denote a regressor matrix and an error covariance matrix.

Nowadays, some packages support unbiased estimates of variances of LS coefficient estimates. But the story is very different with ILS, 2SLS, 3SLS and SUR. Even in the case of homoskedastic error terms, the exact distribution of coefficient estimates and their variances are rather complex, even in simple situations. Under some conditions, coefficient and variance estimates are consistent and have asymptotic normal distributions. Expressions for the variances of these distributions can also be derived.

5.1.1.3/ Generation of the inverse Mill's ratios:

So far we have assumed that the inverse Mill's ratios are readily available for using LS in estimation of the appropriate wage and hours worked equations, when the zero values of LWW_i and $WHRS_i$ are discarded. Now we show how they can be generated. We define $LFP_i=1$ if an individual works and $LFP_i=0$ if an individual does not work (see section 3), and together with equation (5.10) we can infer that:

$$LFP_i=1 \qquad \text{if} \qquad J_i / \sigma_d > -e_{di}^* \tag{5.41}$$

$$LFP_i=0 \qquad \text{if} \qquad J_i / \sigma_d \leq -e_{di}^* \tag{5.42}$$

where $e_{di}^* = \dfrac{e_{di}}{\sigma_d}$, and $e_{di}^* \sim N(0,1)$

$$\begin{aligned} J_i / \sigma_d &= (Z_i\Gamma - X_i^{**}\Theta) / \sigma_d \\ &= [(a_0+a_1WA_i+a_2WA2_i+a_3WE_i+a_4WPED_i+a_5UN_i+a_6CIT_i) \\ &\quad - (b_0+b_1WA_i+b_2WA2_i+b_3WE_i+b_4UN_i+b_5CIT_i+b_6PRIN_i \\ &\quad\quad +b_7KL6_i+b_8K618_i)] / \sigma_d \\ &= p_0 + p_1WA_i + p_2WA2_i + p_3WE_i + p_4WPED \\ &\quad + p_5UN_i + p_6CIT_i + p_7PRIN_i + p_8KL6_i + p_9K618_i \end{aligned}$$

where:

$$\begin{aligned} &p_0 = \frac{a_0 - b_0}{\sigma_d}, p_1 = \frac{a_1 - b_1}{\sigma_d}, p_2 = \frac{a_2 - b_2}{\sigma_d}, p_3 = \frac{a_3 - b_3}{\sigma_d}, p_4 = \frac{a_4}{\sigma_d} \\ &p_5 = \frac{a_5 - b_4}{\sigma_d}, p_6 = \frac{a_6 - b_5}{\sigma_d}, p_7 = \frac{b_6}{\sigma_d}, p_8 = \frac{b_7}{\sigma_d}, p_9 = \frac{b_8}{\sigma_d}, \end{aligned} \tag{5.43}$$

The equations (5.41) and (5.42) form a conventional PROBIT model, so J_i / σ_d can be estimated using the conventional PROBIT estimation procedure. Recall that the inverse Mill's ratios are defined as $\frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} = \lambda_i$, where f and F are the density and distribution functions of the standard normal distribution respectively. Thus, the estimates of the inverse Mill's ratios can be obtained by substitution of the estimates of J_i / σ_d into the formula for the inverse Mill's ratios.

5.1.2/ Formal estimation procedure:

In the sections above, we have dealt with the problems caused by censored data and the bivariate model and suggested suitable equations and estimation techniques. Now, these techniques are organised into a formal procedure. First, we show how the wage and the hours worked equations are estimated, then how the reservation wage equation can be estimated indirectly, based on the estimates of the wage equation. The estimation of the reservation wage equation is a contribution of this dissertation to second generation studies; it has not been discussed before.

5.1.2.1/ Estimation of the wage and hours worked equations:

5.1.2.1.1/ The three stage method:

a/ Stage 1:

It was established in sections (5.1.1.1.1) and (5.1.1.1.2) that the inverse Mill's ratio should be added as a new regressor to form the selection bias corrected reduced form equations (for both wage and hours worked equations). Recall that the ratio is defined as:

$$\frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} = \lambda_i \qquad (5.44)$$

Where f and F are the density and distribution functions of the standard normal distribution respectively.

In order to have λ_i, J_i / σ_d should be estimated. As mentioned in section (5.1.1.3), PROBIT estimation will do this.

b/ Stage 2:

Based on the estimate of J_i / σ_d, inverse Mill's ratios $\frac{f(J_i / \sigma_d)}{F(J_i / \sigma_d)} = \lambda_i$ are calculated. LS is then applied to estimate the selection bias corrected wage equation (5.36). One result is $\hat{\Gamma}$, which is used to calculate the prediction of LWW_i:

$$PLWW_i = Z_i \hat{\Gamma} \tag{5.45}$$

c/ Stage 3:

$PLWW_i$ is then substituted into the selection bias corrected hours worked equation (5.35) and LS is applied to estimate this equation, where the inverse Mill's ratio is added as a regressor. Stage 2 and stage 3 constitute the modified 2SLS estimation.

If a model is just identified as in our case, ILS can be applied to equations (5.36) and (5.37) to produce the same result. Here, we can skip calculation of $PLWW_i$ and proceed directly to stage 3 using LS to estimate equation (5.37) and derive structural parameters c_i for hours worked equation (4.4) using equation system (5.38). So stage 3 is in fact independent of stage 2.

5.1.2.1.2/ The four stage method:

This method, which is a contribution of this dissertation, can be seen as an extension of the three stage method because it takes into account the correlation of the error terms of each equation to improve efficiency. It is recommended, if the risk of model misspècification is limited. In the case of over identification, modified 3SLS is applied, where the first three stages are those of the three stage method. In the fourth stage, the three stage residuals are used to estimate the disturbance variances and covariances, and these estimates are used to obtain estimated generalised least squares (EGLS) estimates for the whole system. The fourth stage can be repeated until some convergence criterion is reached. In the case of just identification, SUR is applied, where the first two stages are those of the three stage method, but $PLWW_i$ is not calculated. In the third stage, LS is used to estimate equation (5.37). In the fourth stage, residuals of equations (5.36) and (5.37) are used to estimate the disturbance variances and covariances, and these estimates are used to obtain estimated generalised least squares (EGLS) estimates for the whole system. The process can be repeated until some convergence criterion is reached, then equation system (5.38) is used to derive structural parameters c_i for the hours worked equation. This method could be further extended by recognizing the explicit form of heteroskedasticity implied through truncation of the original error terms.

5.1.2.2/ Estimation of the reservation wage equation:

There are no observations of the dependent variable LWW_i^*, so estimation of the reservation wage equation (4.19) should be based on estimation of the wage equation. As mentioned in section (5.1.1.3), conventional PROBIT estimation of equations (5.41) and (5.42) will provide estimates $\hat{p}_k$ (k=0,1,2,.....,9) of the algebraic system (5.43). Estimates $\hat{a}_k$ (k=0,1,2,...6) are provided by estimation of the wage equation, applying either the three stage or the four stage method. Using (5.43), estimates $\hat{b}_k$ (k=0,1,2....8) for the reservation wage equation can be derived. There are 10 equations for 10 unknowns, so a unique set of $\hat{b}_k$ (k=0,1,2....8) and $\hat{\sigma}_d$ can be obtained. If $WPED_i$ were not dropped in the reservation wage equation, there would be 11 unknown with 10 equations, and $\hat{b}_k$ (k=0,1,2....8) and $\hat{\sigma}_d$ could not be derived, so the reservation wage equation would be unidentified. On the other hand, if more than one regressor of the wage equation was dropped, there would be 10 equations for less than 10 unknowns and $\hat{b}_k$ (k=0,1,2....8) and $\hat{\sigma}_d$ can not be uniquely determined and the reservation wage equation would be over identified.

5.1.3/ Shazam-program, estimation results and interpretation:

5.1.3.1/ Program for estimation of the model: (see appendix 2 for details)

The program applies both the three stage method and the four stage method. First, data from the MROZ file is read into corresponding variables, then secondary variables are calculated, namely PRIN, WA2 and WPED. In the first stage, the PROBIT model (5.41),(5.42) is estimated using the PROBIT command. The option IMR=IR is used to generate the inverse Mill's ratios, which are stored in the variable IR. In the second stage, the sample is reduced to contain only working people and the logarithm of WW_i is taken. The OLS command is used to estimate the selection bias corrected wage equation, where the inverse Mill's ratio is an additional regressor. The option HETCOV is used to produce unbiased variances of the LS coefficient estimates (White 1991, p. 153-167). The coefficient estimates are stored in the vector A, which is then used to calculate the predicted value of LWW_i, namely $PLWW_i$. A zero vector is included in the regressor matrix X to exclude the inverse Mill's ratio from the prediction. In the third stage, both 2SLS techniques and ILS techniques are applied in estimation of the selection bias corrected hours worked equation. It is worth noting that the 2SLS command is not used, because the second stage is modified 2SLS. The fourth stage is based on either modified 3SLS or SUR, where the form of the hours worked equation in each case is different. Due to the fact both equations are just identified, the result is the same for both. The SYSTEM command is used to estimate both equations simultaneously with the option DN used to make the sample size the divisor for computing the covariance matrix. This option recognizes the asymptotic properties of the estimate. Option DN is also used in the three stage method to make the two methods comparable.

5.1.3.2/ Presentation of the estimation results:

5.1.3.2.1/ For the three stage method:

Table 1: Estimates of the PROBIT model

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	p_0	-1.2651	1.5496	-0.81673	no
WA	p_1	0.050983	0.071303	0.71502	no
WA2	p_2	-0.000987	0.000817	-1.2079	no
WE	p_3	0.16098	0.026451	6.0862	yes
WPED	p_4	-0.003358	0.009371	-0.35839	no
UN	p_5	-0.011071	0.015909	-0.69593	no
CIT	p_6	0.02253	0.10750	0.20958	no
PRIN	p_7	-0.0000215	0.0000047	-4.5422	yes
KL6	p_8	-0.85956	0.11675	-7.3624	yes
K618	p_9	-0.048524	0.041561	-1.1675	no

Table 2: Estimates of the selection bias corrected wage equation (OLS)

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	a_0	-1.1475	1.061	-1.082	no
WA	a_1	0.04563	0.0472	0.9666	no
WA2	a_2	-0.0004685	0.0005583	-0.8391	no
WE	a_3	0.11928	0.01655	7.208	yes
WPED	a_4	-0.011766	0.0065	-1.81	no
UN	a_5	-0.00385	0.009525	-0.4042	no
CIT	a_6	0.070374	0.06487	1.085	no
λ	a_7	-0.047546	0.1543	-0.3082	no

Based on the estimates of TOBIT model (5.41), (5.42) and the wage equation, the estimates of the reservation wage equation can be obtained using equation system (5.43). But the variances of the estimates can not be derived, because the parameters of the reservation wage equation are a function of the parameters of the TOBIT model and the wage equation as well as σ_d and there are no estimates of the covariances between all these parameters.

Table 3: Estimates of the reservation wage equation

Variables	Coefficient	Estimates
Constant	b_0	3.285
WA	b_1	0.133
WA2	b_2	0.00289
WE	b_3	-0.44477
UN	b_4	0.0349
CIT	b_5	0.008568
PRIN	b_6	-0.0000753
KL6	b_7	-3.01179
K618	b_8	-0.17002

Table 4: Estimates of the selection bias corrected hours worked equation by modified 2SLS:

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	c_0	869.94	1873	0.4646	no
LWW	c_1	257.66	603.9	0.4267	no
WA	c_2	51.013	66.69	0.7649	no
WA2	c_3	-0.7477	0.8454	-0.8844	no
WE	c_4	-21.033	86.17	-0.2441	no
UN	c_5	-17.751	13.33	-1.332	no
CIT	c_6	-12.242	86.63	-0.1413	no
PRIN	c_7	-0.0074729	0.0111	-0.671	no
KL6	c_8	-432.63	433.8	-0.9972	no
K618	c_9	-126.87	38.33	-3.31	yes
λ	c_{10}	230.13	848.8	0.2711	no

Table 5: Estimates of the selection bias corrected hours worked equation by ILS

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	d_0	574.28	1825	0.3146	no
WA	d_1	62.769	62.53	1.004	no
WA2	d_2	-0.8684	0.8323	-1.043	no
WE	d_3	9.7006	76.85	0.1262	no
WPED	d_4	-3.0316	7.105	-0.4267	no
UN	d_5	-18.743	13.37	-1.402	no
CIT	d_6	5.89	80.58	0.0731	no
PRIN	d_7	-0.0074729	0.0111	-0.671	no
KL6	d_8	-432.63	439.5	-0.9853	no
K618	d_9	-126.87	38.33	-3.31	yes
λ	d_{10}	230.13	848.8	0.2711	no

If we use the estimates of d_i and a_i to derive the structural parameters c_i with equation system (5.38), the same set of estimates c_i will be obtained as the set of estimates of c_i from the modified 2SLS.

5.1.3.2.2/ For the four stage method:

The PROBIT model remains the same as that of the three stage method. The estimates of the selection bias corrected wage equation and the selection bias corrected hours worked equation are the same for both methods, namely modified 2SLS and ILS. But they are different to those from the three stage method. However, the estimates of the coefficients of the wage equation remain unchanged, only the standard errors change. So the estimates of the reservation wage equation do not change and will not be presented again.

Table 6: Estimates of the selection bias corrected wage equation (3SLS and SUR)

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	a_0	-1.1475	1.0513	-1.091	no
WA	a_1	0.04563	0.0472	0.9666	no
WA2	a_2	-0.0004685	0.0005505	-0.851	no
WE	a_3	0.11928	0.01867	6.389	yes
WPED	a_4	-0.011766	0.0062	-1.898	no
UN	a_5	-0.00385	0.01109	-0.3471	no
CIT	a_6	0.070374	0.07042	0.9993	no
λ	a_7	-0.047546	0.146	-0.3257	no

Table 7: Estimates of the selection bias corrected hours worked equation (3SLS)

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	c_0	829.94	1872	0.4428	no
LWW	c_1	264.15	603.9	0.4374	no
WA	c_2	51.921	66.69	0.7786	no
WA2	c_3	-0.7667	0.8453	-0.907	no
WE	c_4	-19.206	86.16	-0.2229	no
UN	c_5	-17.9	13.33	-1.343	no
CIT	c_6	-12.657	86.63	-0.1461	no
PRIN	c_7	-0.00778	0.0111	-0.6988	no
KL6	c_8	-449.44	433.7	-1.036	no
K618	c_9	-128.9	38.32	-3.363	yes
λ	c_{10}	261.2	848.6	0.3078	no

Table 8: Estimates of the selection bias corrected hours worked equation (SUR)

Variables	Coefficient	Estimated coefficient	Standard error	T - Ratio	Significant (5%)
Constant	d_0	526.02	1825	0.2882	no
WA	d_1	63.973	62.53	1.023	no
WA2	d_2	-0.89045	0.8322	-1.070	no
WE	d_3	12.302	76.84	0.1601	no
WPED	d_4	-3.108	7.105	-0.4374	no
UN	d_5	-18.917	13.37	-1.415	no
CIT	d_6	5.9322	80.57	0.07362	no
PRIN	d_7	-0.00778	0.0111	-0.6988	no
KL6	d_8	-449.44	433.7	-1.036	no
K618	d_9	-128.9	38.32	-3.363	yes
λ	d_{10}	261.2	848.6	0.3078	no

If we use the estimates of d_i and a_i to derive the structural parameters c_i with equation system (5.38), the same set of estimates c_i will be obtained as the set of estimates of c_i from 3SLS.

5.1.3.3/ Interpretation of estimation results:

a/ The wage equation:

$$LWW_i=a_0+a_1WA_i+a_2WA2_i+a_3WE_i+a_4WPED_i+a_5UN_i+a_6CIT_i+e_{wi} \qquad (5.46)$$

As far as signs, significance and size of the coefficient estimates are concerned, the estimates are consistent over all procedures and are shown in Table 2. Consider the signs first, $\hat{a}_1$ is positive and $\hat{a}_2$ is negative as expected. These ensure that the relationship between wage and age have a concave parabolic shape peaking some where in midlife WA=$\frac{\hat{a}_1}{-2\hat{a}_2}$=47 (setting first order derivative of LWW with respect to WA equal to zero). Wage is expected to be positively correlated with education of the woman and her parents; $\hat{a}_3$ is positive as expected. But $\hat{a}_4$ is negative not as expected. Unemployment rate UN has a negative effect on wage, because the higher UN is, the more competetive the labour market is. Thus $\hat{a}_5$ is negative as expected. Recall that $a_7=\frac{\sigma_{wd}}{\sigma_d}$, σ_d is positive, so the sign of σ_{wd} will determine the sign of a_7. But, because the sign of σ_{wd} can not be established in advance, neither can that of a_7. Overall we can say that the signs of the wage equation estimates are as expected. But there is some problem with their significance. Except for a_3, all the estimates are insignificant. The cause might lie in the heteroskedastic nature of the equation, where OLS estimation is not efficient, leading to high standard errors and low significance.

b/ The reservation wage equation:

$$\begin{aligned} LWW_i^* &= b_0+b_1WA_i+b_2WA2_i+b_3WE_i+b_4UN_i+b_5CIT_i \\ &\quad +b_6PRIN_i+b_7KL6_i+b_8K618_i+e_{ri} \end{aligned} \qquad (5.47)$$

The estimates are shown in table 3. The signs of $\hat{b}_1$ and $\hat{b}_2$ are expected to be like those of $\hat{a}_1$ and $\hat{a}_2$ with the same reasoning. Thus $\hat{b}_2>0$ is not as expected. Also, $\hat{b}_3<0$ is not as expected, because the reservation wage should be positively correlated with education. Further, unemployment rate should be negatively correlated with the reservation wage, but $\hat{b}_4 > 0$. Property income is expected to be positively related to the reservation wage, but $\hat{b}_6 < 0$, not as expected. The number of children are expected to be positively correlated with reservation wage, but $\hat{b}_7 < 0$ and $\hat{b}_8 < 0$ not as expected. In general, the signs of the estimates are not as expected. Because the estimates of the TOBIT model and the wage equation have low significance, we expect those of the reservation wage equation to also be insignificant. It does appear that the reservation wage specification is a poor one or there is just not enough information to get a decent set of estimates.

c/ The hours worked equation:

$$\begin{aligned} WHRS_i = {} & c_0+c_1LWW_i+c_2WA_i+c_3WA2_i+c_4WE_i+c_5UN_i+c_6CIT_i \\ & +c_7PRIN_i+c_8KL6_i+c_9K618_i+e_{hi} \end{aligned} \tag{5.48}$$

In contrast to the wage equation, the three stage and four stage estimates of the hours worked equation are slightly different. The estimates from the three stage method are shown in table 4 and those from the four stage method in table 7. As far as signs are concerned, there is no difference. These facts indicate that the problem of correlation between the two equations is not severe. So we can examine just one set of estimates, particularly that of the three stage method.

First we consider the signs of the estimates. Hours supplied should be positively correlated with wage, so $\hat{c}_1$ is positive as expected. The signs of $\hat{c}_2$ and $\hat{c}_3$ are expected to be like those of $\hat{a}_1$ and $\hat{a}_2$, with the same reasoning; they turn out to be as expected. The age, where the peak of labour supply should be reached, namely $WA=\frac{\hat{c}_2}{-2\hat{c}_3}=36$, alsoseems reasonable. Unemployment rate UN should be inversely correlated with labour supply, and $\hat{c}_5$ is negative as expected. Property income PRIN as an alternative source of income enables women to be independent from work, and $\hat{c}_7$ is negative as expected. Children require a woman to spend a lot of time to take care of them, so the number of children KL6 and K618 should be negatively correlated with hours supplied $\hat{c}_8$ and $\hat{c}_9$ have these signs as expected. The effect of small children is bigger than that of older children, so $\hat{c}_8 > \hat{c}_9$ in absolute terms. The effect of education on labour supply is not simple. Higher education enables a woman to get a job, but it also enables her to chose a more sophisticated one with fewer working hours. The same situation applies for the size of city. So we can not predict the signs of $\hat{c}_4$ and $\hat{c}_6$.

Recall that $c_{10} = d_{10} = \frac{\sigma_{hd} + c_1\sigma_{wd}}{\sigma_d}$ and we face the same issue for the sign of c_{10} as with a_7. It also can not be determined in advance. In conclusion, the signs of the estimates satisfy economic theory requirements. But there is also some problem with their significance. Except for c_9, all the remaining estimates are insignificant. The cause might lie in the heteroskedastic nature of the equation, where LS estimation is not efficient, leading to high standard errors and low significance.

Following Mroz (1987), the average WHRS is 1500, that of PRIN is 1000 and that of WW is 4.5. The average of KL6 is 0.23772 and of K618 is 1.3533. Then we can derive the following effects:

The uncompensated wage effect on labour supply is:

$$\frac{\partial WHRS_i}{\partial WW_i} = \frac{c_1}{WW_i} \approx \frac{\hat{c}_1}{\overline{\overline{WW}}} = 57.26 \qquad (5.49)$$

The uncompensated wage elasticity with respect to labour supply is:

$$\frac{\partial \ln(WHRS_i)}{\partial \ln(WW_i)} = \frac{c_1}{WHRS_i} \approx \frac{\hat{c}_1}{\overline{\overline{WHRS}}} = 0.1718 \qquad (5.50)$$

With the same derivation the uncompensated property income effect on labour supply is 7.4 hours per \$1000 property income and the uncompensated property income elasticity with respect to labour supply is 0.00519.

The uncompensated effect of KL6 on WHRS is 432.63, while the uncompensated elasticity is 0.069. The uncompensated effect of K618 on WHRS is 126.87, while the uncompensated elasticity is 0.115.

5.2/ Maximum likelihood estimation:

Maximum likelihood (ML) estimators are popular because they have good asymptotic properties: they are consistent, asymptotically efficient and have limiting normal distributions. Critical to the concept of ML is that different statistical distributions generate different samples, and any particular sample is more likely to have been generated from some distributions than from others. The ML estimate of an unknown parameter is defined as that value of the parameter which maximizes the probability of randomly drawing the sample of observations which we have actually observed. To do that, first of all a likelihood (LH) function should be established, which is equal to the joint density function evaluated at the sample observations in the case of continuous random observations or equal to the joint probability function evaluated at the sample observations in the case of discrete random observations. When the sample involves both continuous and discrete observations, the LH function is formed by the product of the continuous part and the discrete part. If we assume the observations are independent, the joint density function is the product of the individual observation density functions and the joint probability function is the product of the individual observation probability functions. To facilitate mathematical maximization, the logarithm of the LH function is preferred. However, sometimes differentiation of a log LH function yields a set of first order conditions which are highly nonlinear functions of the parameters of interest and are accordingly extremely difficult to solve. In such cases, ML estimates are found numerically, involving a systematic search over the parameter space to identify the optimal parameter values, starting from a set of values and changing them in a systematic way, such as iteration, until a maximum is reached. Unfortunately, the procedure is not foolproof. Sometimes it does not converge. Sometimes a local rather than a global maximums is reached, so that several sets of starting values should be tried to compare the obtained maximums. Finally, it is worth noting that ML is a very general technique which can be used whenever the distribution of the error term is known and certain regularity conditions, governing the shape of the LH function, hold.

5.2.1/ The Likelihood Function:

There are no observations on the reservation wage, so it is not included directly in the model. The model is bivariate in LH estimation:

$$LWW_i=a_0+a_1WA_i+a_2WA2_i+a_3WE_i+a_4WPED_i+a_5UN_i+a_6CIT_i+e_{wi} \qquad \text{if} \qquad LWW_i>LWW_i^* \qquad (5.51)$$

$$LWW_i=0 \qquad \text{if} \qquad LWW_i \leq LWW_i^* \qquad (5.52)$$

$$WHRS_i=c_0+c_1LWW_i+c_2WA_i+c_3WA2_i+c_4WE_i+c_5UN_i+c_6CIT_i$$
$$+c_7PRIN_i+c_8KL6_i+c_9K618_i+e_{hi} \qquad \text{if} \qquad LWW_i>LWW_i^* \qquad (5.53)$$

$$WHRS_i=0 \qquad \text{if} \qquad LWW_i \leq LWW_i^* \qquad (5.54)$$

To form the joint density function for the continuous part, we consider equations (5.51) and (5.53) for one observation only, namely observation (i). If the error terms e_{wi} and e_{hi} are assumed to have a normal distribution with zero mean, the general formula for joint density function (Judge et al 1988, p. 636-670) is:

$$f(y_i) = (2\Pi)^{-1}|\Sigma|^{-1/2}|J|\exp[-\frac{1}{2}(y_i - R_i\delta)'\,\Sigma^{-1}(y_i - R_i\delta)] \qquad (5.55)$$

$y_i=(LWW_i, WHRS_i)'$ is the vector of endogenous variables.

$$R_i = \begin{pmatrix} R_{1i} & \\ & R_{2i} \end{pmatrix}$$

$R_{1i}=(1\ \ WA_i\ \ WA2_i\ \ WE_i\ \ WPED_i\ \ UN_i\ \ CIT_i)$

$R_{2i}=(1\ \ WA_i\ \ WA2_i\ \ WE_i\ \ WPED_i\ \ UN_i\ \ CIT_i\ \ PRIN_i\ \ KL6_i\ \ K618_i)$

$$\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$$

$\delta_1=(a_0\ \ a_1\ \ a_2\ \ a_3\ \ a_4\ \ a_5\ \ a_6)'$

$\delta_2=(c_0\ \ c_1\ \ c_2\ \ c_3\ \ c_4\ \ c_5\ \ c_6\ \ c_7\ \ c_8\ \ c_9)'$

$$J = \begin{pmatrix} 1 & -c_1 \\ 0 & 1 \end{pmatrix} \quad \rightarrow \quad |J| = 1 \text{ is the Jacobian}$$

To ensure that the estimates of variances are nonnegative, we need the following Cholesky decomposition of the variance-covariance matrix of the error terms:

$$\Sigma = \begin{pmatrix} \sigma_w^2 & \sigma_{wh} \\ \sigma_{wh} & \sigma_h^2 \end{pmatrix} = \begin{pmatrix} h_{11} & 0 \\ h_{12} & h_{22} \end{pmatrix} \begin{pmatrix} h_{11} & h_{12} \\ 0 & h_{22} \end{pmatrix}$$

$$\sigma_w^2 = h_{11}^2$$

$$\sigma_{wh} = h_{11} h_{12}$$

$$\sigma_h^2 = h_{12}^2 + h_{22}^2$$

So the inverse covariance matrix is:

$$\Sigma^{-1} = \frac{1}{(h_{11} h_{22})^2} \begin{pmatrix} h_{12}^2 + h_{22}^2 & -h_{11} h_{12} \\ -h_{11} h_{12} & h_{11}^2 \end{pmatrix}$$

Substitution of the above expressions into equation (5.55), we have the joint density function for observation (i):

$$f(y_i) = \frac{1}{2\Pi h_{11} h_{22}} \exp\{-\frac{1}{2(h_{11} h_{22})^2} [(LWW_i - R_{1i}\delta_1)^2 (h_{12}^2 + h_{22}^2) + (WHRS_i - R_{2i}\delta_2)^2 h_{11}^2$$

$$-2h_{11} h_{12} (LWW_i - R_{1i}\delta_1)(WHRS_i - R_{2i}\delta_2)]\} \qquad (5.56)$$

The LLH function is then:

$$LLH_i^c = -\log(2\Pi h_{11} h_{22}) - \frac{1}{2(h_{11} h_{22})^2} [(LWW_i - R_{1i}\delta_1)^2 (h_{12}^2 + h_{22}^2) + (WHRS_i - R_{2i}\delta_2)^2 h_{11}^2$$

$$-2h_{11} h_{12} (LWW_i - R_{1i}\delta_1)(WHRS_i - R_{2i}\delta_2)] \qquad (5.57)$$

Now, it is assumed that observation (i) is discrete. In this case, what concerns us is that the woman is not working (the case of a working woman is covered by the continuous part) and the probability function is to be found. Recall equation (5.42):

$$LFP_i = 0 \qquad \text{iff} \qquad J_i / \sigma_d \leq -e_{di}^* \qquad (5.58)$$

where $e_{di}^{*} \sim N(0,1)$

So
$$\begin{aligned} P(LFP_i{=}0) &= P(J_i / \sigma_d \le -e_{di}^{*}) \\ &= P(e_{di}^{*} \le -J_i / \sigma_d) \\ &= F(-J_i / \sigma_d) \end{aligned} \tag{5.59}$$

where F is the distribution function of the standard normal distribution.

The LLH function is then:

$$LLH_i^d = \log F(-J_i / \sigma_d) \tag{5.60}$$

An observation can be either continuous or discrete and we can combine the density function and the probability function of one single observation to form the general likelihood function for observation (i) by employing the dummy variable LFP_i.

$$LLH_i = LLH_i^c * LFP_i + LLH_i^d * (1-LFP_i) \tag{5.61}$$

The LLH function for the whole sample size of 753 is then:

$$LLH = \sum_{i=1}^{753} LLH_i \tag{5.62}$$

5.2.2/ SHAZAM program to maximize the LLH function:

First of all, new variables are generated as in the program for LS estimation, such as PRIN, WA2, WPED and LWW. Then the Shazam command NL is applied to the whole sample size of 753.

```
nl 1/ncoef=31 iter=200 logden
eq (-log(2*$pi*h11*h22)-0.5*(1/((h11+h22)**2))* &
(( lww-a0-a1*wa-a2*wa2-a3*we-a4*wped-a5*un-a6*cit)**2*(h12**2+h22**2)+ &
(whrs-c0-c1*lww-c2*wa-c3*wa2-c4*we-c5*un-c6*cit-c7*prin-c8*kl6-c9*k618)**2* &
h11**2-2*h11*h12*( lww-a0-a1*wa-a2*wa2-a3*we-a4*wped-a5*un-a6*cit)* &
```

```
(whrs-c0-c1*lww-c2*wa-c3*wa2-c4*we-c5*un-c6*cit-c7*prin-c8*kl6-c9*k618)))*  &
lfp+  &
(log(ncdf((b0+b1*wa+b2*wa2+b3*we+b4*wped+b5*un+b6*cit+b7*prin+b8*kl6+  &
b9*k618-a0-a1*wa-a2*wa2-a3*we-a4*wped-a5*un-a6*cit)/h33)))*(1-lfp)
coef h11 2 h22 3 h12 4 h33 1                    &
a0 0 a1 1 a2 2 a3 0 a4 0 a5 0 a6 0              &
c0 0 c1 1 c2 2 c3 3 c4 4 c5 5 c6 0 c7 0 c8 0 c9 0  &
b0 0 b1 0 b2 0 b3 0 b4 0 b5 0 b6 0 b7 0 b8 0 b9 0
end
```

The number of parameters is 31 and the LOGDEN option tells Shazam that the equation given in the EQ command is the log-density function for a single observation rather than a regression equation. The complete LLH function is then computed by summing the log-density functions. Note that the probability that a woman is not working is computed by command NCDF (Normal cumulative Distribution Function) and h33 stands for σ_d.

5.2.3/ Comments on LH estimation of the labour supply model and comparison with LS estimation:

LH techniques have some advantages over LS techniques in estimation of the labour supply model. LH techniques deal conveniently with the two problems of the labour supply model, namely censored data and the simultaneous equation system. The LH function can accommodate both continuous and discrete observations for working and nonworking women respectively, so there is no problem from censored data. As a result, there is no need to find selection bias corrected equations and there is no problem of heteroskedasticity. With regard to the problems of a simultaneous equation system, LH techniques provide a good way to estimate all three equations in one go, taking into account correlations between the equations. On the other hand, there are some disadvantages with LH estimation. Although the reservation wage equation parameters are estimated together with the parameters of the other equations in one go, the reservation wage equation itself is not directly included in the system of equations, so that the joint density function is bivariate and not trivariate. Further, the number of parameters to be estimated is large relatively to the number of observations available, resulting in low degrees of freedom. But the major disadvantage is that LH estimation is not foolproof. After many trials the LLH function did not converge. Due to time limit, in this dissertation LH estimation was not pursued further.

6/ Bayesian estimation of the labour supply model:

So far we have discussed LS estimation of the labour supply model. Now we develop a new approach, namely the Bayesian one. First, the general concept of the Bayesian approach is explained, then its numerical realization, the Gibbs sampling technique, is introduced. Finally we show how the Bayesian approach can be applied to estimate the labour supply model and deal with its major problems of censored data and a simultaneous equation system.

6.1/ The Bayesian concept:

The main characteristic of the Bayesian approach is that uncertainty about the value of an unknown parameter can be expressed in terms of a probability distribution, which is subjective, representing an individual's degree of belief (Judge et al. 1988, p.275-325). Information about the parameter is presented in the form of subjective probability statements; such an approach is thought of as intuitively reasonable. Parameters are treated as random variables in the sense that a parameter is associated with a subjective probabiltity density function that describes one's state of knowledge about the parameter. Knowledge that exists before observing the sample is called a prior distribution. Knowledge derived from both prior and sample information is called a posterior distribution. The procedure combining a prior distribution with sample information to form a posterior distribution is based on Bayes' theorem. The prior probability density function (p.d.f.) can be informative or noninformative. An informative prior p.d.f. may come from economic theory, past studies or both, while a noninformative prior p.d.f. assumes that all possible values of the parameter are equally likely, hence taking the form of a uniform density function. As mentioned above, Bayes' theorem is the foundation of the Bayesian approach and should be explained in more detail. Consider the following model:

$$\begin{aligned} y &= x\beta + e \\ e &\sim N(0, \Sigma) \end{aligned} \qquad (6.1)$$

where β and Σ are to be estimated.

The prior p.d.f. is assumed to be $g(\beta, \Sigma)$. Sample information about β and Σ is expressed in the form of the likelihood function for β and Σ, conditional on sample y. By definition, it is the joint p.d.f. of y, but y is considered as fixed while β and Σ are considered as variables. The likelihood function is denoted as $l(\beta, \Sigma|y)$. Bayes' theorem combines the prior p.d.f. and the likelihood function into the posterior p.d.f.

$$g(\beta, \Sigma|y) \alpha\, l(\beta, \Sigma|y)\ g(\beta, \Sigma) \qquad (6.2)$$

where α denotes proportionality.

For convenience, a prior informative p.d.f. often takes the form of a natural conjungate prior, which combines nicely with the likelihood function to produce a posterior p.d.f. with the same functional form, i.e. the posterior p.d.f. of one sample will be the natural conjungate for a future sample, and at the same time it represents a large variety of prior opinions.

The posterior p.d.f. is the souce of all inference about an unknown parameter, such as interval and point estimates and hypothesis testing. For interval estimation, the marginal p.d.f. of the parameter is derived from the joint posterior p.d.f., then a probability statement about the parameter is made. Point estimation involves the specification of a loss function. The point estimate is the value minimizing expected loss, which is also derived from the marginal p.d.f of the parameter. In the case of a quadratic loss function, the point estimate is the expected value of the parameter. In the follwing discussion we assume a quadratic loss function. Finally, we examine hypothesis testing in a Bayesian framework, assuming H_0 is tested against H_1. The test will involve two steps. First, the posterior probabilities $P(H_0|y)$ and $P(H_1|y)$ are calculated and their ratio, the posterior odds ratio, is formed:

$$K_{0,1} = \frac{P(H_0|y)}{P(H_1|y)} \qquad (6.3)$$

The ratio can favour one hypothesis over the other, but we neither accept nor reject the null hypothesis. If a decision is required, the second step should be carried out by introducing a loss function. Based on the posterior probabilities $P(H_0|y)$ and $P(H_1|y)$, expected loss for each kind of decision is derived. The hypothesis minimizing expected loss will be accepted and the other rejected.

A specific feature of the Bayesian approach is the possibility of including prior information in the form of inequality restrictions. Very often, economic theory suggests that the possible values of a parameter lie within a certain range. If prior information does not preclude the impossible values, the posterior p.d.f. will attach positive mass to them and interval and point estimates may not be sensible. If inequality prior restrictions are included, then the prior p.d.f is truncated, and so is the posterior p.d.f.. This can make the derivation of the mean and variance of the parameter concerned more complicated. One way to solve this problem is to employ Monte Carlo numerical integration (Griffiths 1987).

The conventional techniques of the Bayesian approach are directed towards derivation of the posterior p.d.f., the marginal posterior p.d.f.s of the parameters, and finally.the relevant moments of the parameters, using analytical or numerical integration. But in many circumstances, these derivations are very difficult and even impossible. An alternative way to estimate the moments of interest is to use simulation techniques such as Gibbs sampling. To employ Gibbs sampling, we need not know the form of the marginal posterior p.d.f's., thus avoiding complicated integration. In other situations, even if the posterior p.d.f. is known, it can still be very difficult to generate observations from it, then Gibbs sampling can also be useful.

6.2/ The concept of Gibbs sampling:

Monte Carlo numerical integration can be used to generate posterior moments of any posterior marginal p.d.f. of a parameter (or a subset of parameters), assuming observations on the posterior marginal p.d.f are available (Geweke 1986, 1994). The posterior marginal p.d.f. of a parameter is in fact conditional on the sample observations and, as mentioned above, is sometimes impossible to derive. Even if it can be derived, it is often difficult to generate observations from the posterior marginal p.d.f.. But in many cases, if we concentrate on only a subset of the parameters, while assuming the rest of them to be known, it is possible to get the posterior p.d.f of the subset of interest, conditional on the rest and, of course, also on the sample information. This kind of posterior p.d.f. facilitates the generation of parameter observations, which can not be used directly as output from Monte Carlo numerical integration to generate marginal posterior moments of interest, because these parameter observations are conditional not only on the sample observations, but also on the rest of the parameters. So we need to find a way to neutralise the effect of the rest of the parameters on these parameter observations. Gibbs-sampling will solve this problem (Casella et al. 1992). For the convenience of explanation, in the following we use the term "conditional" only for the dependence on the rest of the parameters and consider the dependence on the sample observations as understood. The concept of Gibbs sampling is best exposed with a simple example.

The vector of the parameters to be estimated is a = (a_1 a_2 a_3) and we need to generate parameter observations from the joint posterior p.d.f. f(a | y), where y is the vector of the sample observations. Suppose we are able to derive all the conditional posterior p.d.f's. of the parameters:

$$f_1(a_1|a_2,a_3,y) \tag{6.5}$$

$$f_2(a_2|a_1,a_3,y) \tag{6.6}$$

$$f_3(a_3|a_1,a_2,y) \tag{6.7}$$

Based on these p.d.f's. and given sample observations y, we assume further that we can generate the conditional parameter observations, which are denoted by a_i^t, where i=1,2,3, indicating the parameter and t=0,1,2,..........,n indicating the number of an observation. t=0 indicates the starting values of the observations. Based on the starting values for a_2 and a_3, we can use the p.d.f. (6.5) to generate the starting value for a_1. We then use p.d.f. (6.6) together with a_1^0, a_3^0 to generate a_2^1. The next step is using p.d.f (6.7) and a_1^0, a_2^1 to generate a_3^1. So by the end of this cycle we have generated a_1^0, a_2^0, a_3^0, a_2^1 and a_3^1. The cycle is then repeated n times to generate three sets of parameter observations, namely a_1^0, a_1^1, a_1^2................, a_1^n; a_2^0, a_2^1, a_2^2................, a_2^n; and a_3^0, a_3^1, a_3^2................, a_3^n. We now look at one set in more detail, namely a_1^t to see that Gibbs-sampling neutralises the effect of the rest the parameters. a_1^0 depends on a_2^0 and a_3^0, while a_1^1 depends on a_2^1 and a_3^1. So, in general we can say that a_1^t depends on a_2^t and a_3^t. In other words as the number of an observation of a_1 changes, the number of the observations of the rest of the parameters also change accordingly, i.e. they are not fixed, but varying, so that their effect can become neutral. Statistical theory suggests that after a burn-in period, Gibbs sampling will generate the unconditional posterior parameter observations $a_1|y$, $a_2|y$ and $a_3|y$. There is no accurate rule for determination of the length of the burn-in period. What we can do is to check if the generated parameter observations have converged to some stable p.d.f. and this can be done by calculating moments such as means, and variances and observing minimum and maximum values of the parameter observations over some periods.

If inequality restrictions are imposed on parameters, Gibbs sampling can still be applied by throwing away the generated observations, which violate the restrictions.

Assuming the burn-in period consists of the first k observations, then the converging parameter observations are the remaining n-k and can be used to estimate the moments of interest, such as means and variances. Under the assumption of quadratic loss functions, the means provide point estimates of the parameters.

Sometimes, an economic model involves latent variables with unobservable values. In the Gibbs sampling context, these unobservable values are treated as unknown parameters and are called augmented data (Chib 1992).

6.3/ Bayesian estimation of the labour supply model, using Gibbs sampling:

The Bayesian approach provides convenient ways to deal with the two major problems of the labour supply model, namely censored data and the simultaneous equation system. The problem of censored data can be solved by considering the unobservable data as augmented data. Furthermore, the correlations between the equations are taken into account by using joint p.d.f's. In the Bayesian approach, all three equations are considered at the same time, whereas the LS approach considers only the correlation between the wage and the hours worked equation.

To apply the Bayesian approach, we first derive the necessary posterior p.d.f's., which are the basis for Gibbs sampling. Finally, a practical approach to estimating the labour supply model is introduced to deal with some computational problems.

6.3.1/ Conditional posterior p.d.f's. of the parameters:

In the following, augmented data are treated like unknown parameters, so we only explicitly mention augmented data, when it is necessary. Furthermore, we always assume the prior p.d.f. to be a uniform p.d.f., in some instances with inequality restrictions.

Recall that the labour supply model consists of three equations, namely the wage equation, the reservation wage equation and the hours worked equation.

$$LWW_i = Z_i \Gamma + e_{wi} \qquad (6.8)$$

$$LWW_i^* = X_i^{**}\Theta + e_{ri} \qquad (6.9)$$

$$WHRS_i = c_1 Z_i \Gamma + X_i^{**} C \; + e_{gi} \qquad (6.10)$$

where $C = (c_0 \; c_2 \; c_3 \; c_4 \; c_5 \; c_6 \; c_7 \; c_8 \; c_9 \; c_{10})$ and CN=(c_1 C)

The error vector is $E = \begin{pmatrix} E_w \\ E_r \\ E_g \end{pmatrix}$ (6.11)

$E_w^{'} = (e_{w1} \; e_{w2} \ldots\ldots e_{w753})$, $E_r^{'} = (e_{r1} \; e_{r2} \ldots\ldots e_{r753})$, $E_g^{'} = (e_{g1} \; e_{g2} \ldots\ldots e_{g753})$

with $e_{wi} \sim N(0, \sigma_{ww})$, $e_{ri} \sim N(0, \sigma_{rr})$, $e_{gi} \sim N(0, \sigma_{gg})$.

It is further assumed that the error terms within each equation are independent and identically distributed but the error terms are contemporaneously correlated across the equations, with cov(e_{wi},e_{ri})=σ_{wr}, cov(e_{wi},e_{gi})=σ_{wg} and cov(e_{ri},e_{gi})=σ_{rg}. We denote E_i=(e_{wi} e_{ri} e_{gi})', then we have:

$$COV(E_i) = E(E_i E_i^{'}) = \Sigma = \begin{pmatrix} \sigma_{ww} & \sigma_{wr} & \sigma_{wg} \\ \sigma_{wr} & \sigma_{rr} & \sigma_{rg} \\ \sigma_{wg} & \sigma_{rg} & \sigma_{gg} \end{pmatrix} \qquad (6.12)$$

So the error covariance matrix of the model is:

$$COV(E)=E(EE')=\Sigma \otimes I_{753} \qquad (6.13)$$

6.3.1.1/ The conditional posterior p.d.f. of the augmented data:

LWW and WHRS are available for workers only and LWW* are not available at all. LWW and WHRS for nonworkers are denoted by NLWW and NWHRS respectively, they are, together with LWW*, the above mentioned augmented data that need to be generated. LWW for workers is denoted by WLWW. Further, LWW* for workers is denoted by WLWW* and LWW* for nonworkers is denoted by NLWW*. If we assume that Σ, Γ, Θ and CN are given, then the conditional posterior p.d.f. for augmented data can be derived. Because the data are assumed to be only contemporaneously correlated across the three equations, i.e. they are correlated only if observation number (i) is the same, we need to consider only one set of observations with observation number (i). Equation system (6.8), (6.9) and (6.10) can be written in the matrix form for one observation (i) for a nonworking woman:

$$\begin{pmatrix} NLWW_i \\ NLWW_i^* \\ NWHRS_i \end{pmatrix} = \begin{pmatrix} Z_i & 0 & 0 \\ 0 & X_i^{**} & 0 \\ c_1 Z_i & 0 & X_i^{**} \end{pmatrix} \begin{pmatrix} \Gamma \\ \Theta \\ C \end{pmatrix} + \begin{pmatrix} e_{wi} \\ e_{ri} \\ e_{gi} \end{pmatrix}$$

or

$$Y_i = X_i \alpha + E_i \tag{6.14}$$

So the conditional posterior p.d.f. for Y_i is:

$$Y_i \sim TN(X_i \alpha, \Sigma) \tag{6.15}$$

where TN denotes "truncated normal distribution". The distribution is truncated because the condition for nonworking is $NLWW_i \leq NLWW_i^*$. We consider the p.d.f. of all three components as a joint p.d.f. to take into account the correlation across equations when generating data, thus making use of all available information.

If (i) stands for a working woman, then Σ, Γ, Θ, CN and LWW_i and $WHRS_i$ are given. That means the error terms e_{wi} and e_{gi} in equations (6.8) and (6.10) are observed, they are no longer random. We need to consider only equation (6.9) to derive the conditional posterior p.d.f. of $WLWW_i^*$:

$$WLWW_i^* \sim TN(X_i^{**}\Theta, \sigma_{rr}) \tag{6.16}$$

The distribution is truncated because the condition for working is $WLWW_i > WLWW_i^*$

6.3.1.2/ The conditional posterior p.d.f. of Γ, Θ :

If we assume that the augmented data, CN and Σ are given, we can derive the conditional posterior p.d.f. of Γ, Θ. The model is rewritten as follows:

$$LWW_i = Z_i \Gamma + e_{wi} \tag{6.17}$$

$$LWW_i^* = X_i^{**} \Theta + e_{ri} \tag{6.18}$$

$$WHRS_i - X_i^{**} C = (c_1 Z_i) \Gamma + e_{gi} \tag{6.19}$$

In matrix form, the model is:

$$\begin{pmatrix} LWW \\ LWW^* \\ WHRS - X^{**}C \end{pmatrix} = \begin{pmatrix} Z & 0 \\ 0 & X^{**} \\ c_1 Z & 0 \end{pmatrix} \begin{pmatrix} \Gamma \\ \Theta \end{pmatrix} + E \tag{6.20}$$

or

$$Y = X\gamma + E \tag{6.21}$$

So the conditional posterior p.d.f. of Γ, Θ is (Judge et al. 1988, p. 443-494):

$$\gamma \sim TN(\hat{\gamma}, (X'(\Sigma^{-1} \otimes I_{753})X)^{-1}) \tag{6.22}$$

where $\hat{\gamma}$ is the GLS-estimate of γ. Because some inequality restrictions can be imposed on γ, the p.d.f. is truncated.

6.3.1.3/ The conditional posterior p.d.f. of CN:

If we assume that the augmented data, γ and Σ are given, the conditional posterior p.d.f. of CN can be derived. The model is rewritten as follows:

$$LWW_i - Z_i \Gamma = e_{wi} \tag{6.23}$$

$$LWW_i^* - X_i^{**}\Theta = e_{ri} \tag{6.24}$$

$$WHRS_i = (Z_i\Gamma)c_1 + X_i^{**}C + e_{gi} \tag{6.25}$$

In matrix form, the model is:

$$\begin{pmatrix} LWW - Z\Gamma \\ LWW^* - X^{**}\Theta \\ WHRS \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ Z\Gamma & X^{**} \end{pmatrix} \begin{pmatrix} c_1 \\ C \end{pmatrix} + E \tag{6.26}$$

Or

$$Y2 = X2 * CN + E \tag{6.27}$$

So the conditional posterior p.d.f. of CN is:

$$CN \sim TN(\hat{CN}, (X2'(\Sigma^{-1} \otimes I_{753})X2)^{-1}) \tag{6.28}$$

where $\hat{CN}$ is the GLS-estimate of CN (Judge et al. 1988, p. 443-494). Because some inequality restrictions can be imposed on CN, the p.d.f. is truncated.

6.3.1.4 The conditional posterior p.d.f. of Σ :

If we assume that the augmented data, CN and γ are given, the conditional posterior p.d.f. of Σ can be derived, which is known as the inverse Wishart distribution (Press 1982):

$$\Sigma \sim IW(A, d.f., m) \tag{6.29}$$

Where m=3 is the number of equations, d.f.=n-(k+m)+1 is the degree of freedom, n=2259 is the number of observations, k=26 is the number of parameters to be estimated, so d.f.=2231. A is the matrix of cross sums of squares of residuals of the equations.

$$A=\begin{pmatrix} E_w^{'}E_w & E_w^{'}E_r & E_w^{'}E_g \\ E_r^{'}E_w & E_r^{'}E_r & E_r^{'}E_g \\ E_g^{'}E_w & E_g^{'}E_r & E_g^{'}E_g \end{pmatrix} \tag{6.30}$$

where $E_w=(e_{w1}\ e_{w2}\ ....e_{w753})'$, $E_r=(e_{r1}\ e_{r2}\ ....e_{r753})'$, and $E_g=(e_{g1}\ e_{g2}\ ....e_{g753})'$.

6.3.2/ The Gibbs sampling procedure:

The Gibbs sampling procedure is the repetition of the same cycle of observation generation. The cycle consists of four steps: firstly, generation of the augmented data, given Σ, Γ, Θ and CN, secondly generation of Γ, Θ, given the augmented data, CN and Σ, thirdly generation of CN, given the augmented data, γ and Σ, and finally generation of Σ, given the augmented data, CN and γ. It is worth reiterating, that in all the four steps, the sample observations, namely LWW and WHRS for workers, are given. In order to start the Gibbs sampling procedure, the starting values of the given parameters should be available, namely Σ, Γ, Θ and CN. One possible set of values is the values from the least squares estimation in part (5).

In the previous section, the conditional posterior p.d.f's. of the parameters have been derived. These p.d.f's. form the basis for Gibbs sampling to generate the posterior parameter observations. However, they can not be used directly in the generation, as far as the software in use, namely SHAZAM, is concerned. In SHAZAM, only observations from the standard normal p.d.f's. can be generated directly. For nonstandard normal p.d.f's., we make use of the theorem on linear transformations of a vector of normal random variables to generate observations for nonstandard normal p.d.f's. Assume Z~N(0,I) and E~N(μ, Σ), where Σ is positive definite, Cholesky decomposition matrix S can be generated, where SS'$=\Sigma$, then E can be generated as:

$$E=\mu+S*Z=\mu+NE \tag{6.31}$$

where NE=S*Z takes into account only the effects of the nonscalar identity covariance matrix Σ.

6.3.2.1/ Generation of the augmented data, given Σ, Γ, Θ and CN:

Our approach in this section is the same as in (6.3.1.1), i.e. we focus on one observation, namely (i) only. Since the observations are independent, they can be generated separately, while taking into account the corresponding inequality condition.

a/ Generation of WLWW_i^* : The conditional posterior p.d.f. for WLWW_i^* is given by equation (6.16). Using formula (6.31), Cholesky matrix S=$\sqrt{\sigma_{rr}}$ is generated, NE=S*Z is calculated. NE is added to $X_i^{**}\Theta$ to get WLWW_i^* . If $\mathrm{WLWW}_i^* < \mathrm{WLWW}_i$, the generated value is accepted, otherwise the generation process is repeated, until the inequality restriction is satisfied. The whole process is repeated 428 times, covering all workers.

b/ Generation of NLWW_i, NLWW_i^* and NWHRS_i : the conditional posterior p.d.f. of Y_i is given by 6.15). Using formula (6.31), Cholesky matrix S is generated, then the the observation $Z_i=(Z_{1i}\ Z_{2i}\ Z_{3i})'$ is generated. $NE_i=S*Z_i$ is calculated. Then NE_i is added to $X_i\gamma$ (see 6.14) to get NLWW_i, NLWW_i^* and NWHRS_i. If $\mathrm{NLWW}_i < \mathrm{NLWW}_i^*$, the three generated values are accepted, otherwise the generation process is repeated until the inequality restriction is satisfied. The whole process is repeated 325 times, covering all nonworkers.

6.3.2.2/ Generation of Γ, Θ, given the augmented data, CN and Σ :

The conditional posterior p.d.f. for Γ, Θ is given by (6.22). Using formula (6.31), Cholesky matrix S is generated, where SS'=$(X'(\Sigma^{-1} \otimes \mathrm{I}_{753})X)^{-1}$. The joint observation $Z=(Z_1\ Z_2.............Z_7\ Z_8...................Z_{16})'$ is generated. NE=S*Z and $\gamma = \hat{\gamma} + NE$ are generated. The 16 generated values are used to check the inequality restrictions. The process is repeated until all inequality restrictions are satisfied.

6.3.2.3/ Generation of CN, given the augmented data, γ and Σ :

The conditional posterior p.d.f. for CN is given by (6.28). Using formula (6.31), Cholesky matrix S is generated, where SS'=$(X2'(\Sigma^{-1} \otimes \mathrm{I}_{753})X2)^{-1}$. The joint observation $Z=(Z_1\ Z_2..........Z_{10})'$ are generated. NE=S*Z and $\mathrm{CN} = \hat{\mathrm{CN}} + \mathrm{NE}$ are generated. The 10 generated values are used to check the inequality restrictions. The process is repeated until all inequality restrictions are satisfied.

6.3.2.4/ Generation of Σ, given the augmented data, CN and γ :

The conditional posterior p.d.f. for Σ is given by (6.29). Matrix A is generated by calculating the residuals for each equation, then Cholesky decomposition matrix C of A^{-1}, where $CC' = A^{-1}$, is calculated. The joint observation $Z_i=(Z_{1i}\ Z_{2i}\ Z_{3i})'$ is generated for i=1,...,753 to form $Z=(Z_1\ Z_2\ ......Z_{753})$. Then W=C*Z is generated. Finally we calculate $\Sigma = (WW')^{-1}$.

6.3.3/ A practical approach to estimating the labour supply model:

The Gibbs sampling procedure, described in section (6.3.2), is theoretically correct, but it would require a large amount of computer time to be realized. The problem lies in the large number of augmented data to be generated, relative to the available sample observations, and the related large number of inequality restrictions for the augmented data to satisfy, namely 428 inequality restrictions for the generation of WLWW*, and 325 inequality restrictions for the generation of the augmented data for nonworkers. Furthermore, the generation of parameter observations involves a number of inequality restrictions, which should be satisfied simultaneously. Due to the limit of time available for the dissertation, a practical approach to estimating the labour supply model is employed to cope with the time constraint problem. In (6.3.2), the generation of LWW* for the whole sample size serves not only the estimation of the reservation wage equation, but also the generation of missing observations for the other two equations, namely LWW and WHRS for nonworkers, by functioning as a limit for the inequality restrictions. If we do not estimate the reservation wage directly using Gibbs sampling, we can skip the generation of LWW*. As the limit for the inequality restriction for generating LWW and WHRS for nonworkers, we can choose the maximum value of LWW=3.3 for workers as the lower limit and discard the values of LWW for nonworkers, which are above this limit, with the rationale that people with such high wages would prefer to work, so do not belong to the set of nonworkers. Once the wage and hours worked equations have been estimated, the reservation wage equation can be estimated indirectly using the estimates of the probit model in section (5), as it was done in the least squares approach (see section 5.1.2.2).

6.3.3.1/ The practical Gibbs sampling procedure:

a/ Generation of the starting values for the Gibbs sampling first cycle: These are the values for Γ, CN and Σ, where $\Sigma = \begin{pmatrix} \sigma_{ww} & \sigma_{wg} \\ \sigma_{wg} & \sigma_{gg} \end{pmatrix}$, because we now concentrate on two equations: the wage and hours worked equations. The values for Γ and CN are the least squares estimates from part (5). Substitution of these values in the two equations enables us to calculate the error variances, based on observations for workers only:

$$\hat{\sigma}_{ww} = \hat{E}_w^{'}\hat{E}_w / 428 \qquad (6.32)$$

$$\hat{\sigma}_{gg} = \hat{E}_g^{'}\hat{E}_g / 428 \qquad (6.33)$$

The generation of the starting value for σ_{wg} is more involved. Recall that the hours worked equation can be either in structural or reduced form and the relationship between error term e_{gi} of the reduced form and error term e_{hi} of the structural form is:

$$e_{gi}=c_1e_{wi}+e_{hi} \qquad (6.34)$$

So
$$\sigma_{gg} = c_1^2\sigma_{ww} + \sigma_{hh} + 2c_1\sigma_{wh} \qquad (6.34)$$

or
$$\sigma_{hw} = (\sigma_{gg} - \sigma_{hh} - c_1^2\sigma_{ww}) / 2c_1 \qquad (6.35)$$

On the other hand, also based on (6.34):

$$\sigma_{wg} = c_1\sigma_{ww} + \sigma_{hw} \qquad (6.36)$$

Substitution of (6.35) into (6.36) results in:

$$\sigma_{wg} = c_1\sigma_{ww} + (\sigma_{gg} - \sigma_{hh} - c_1^2\sigma_{ww}) / 2c_1 \qquad (6.37)$$

From the hours worked equation in the structural form we can calculate:

$$\hat{\sigma}_{hh} = \hat{E}_h^{'}\hat{E}_h / 428 \qquad (6.38)$$

Substitution of (6.32), (6.33) and (6.38) into (6.37) results in $\hat{\sigma}_{wg}$.

b/ The Gibbs sampling cycle:

- Generation of the augmented data, given Γ, Σ and CN: The augmented data are now $NLWW_i$ and $NWHRS_i$, we proceed as in (6.3.2.1), with the exception that $Z_i=(Z_{1i}\ Z_{2i})'$ and $\Sigma = \begin{pmatrix} \sigma_{ww} & \sigma_{wg} \\ \sigma_{wg} & \sigma_{gg} \end{pmatrix}$. If $NLWW_i<3.3$, then $NLWW_i$ and $NWHRS_i$ are accepted, otherwise the generation process is repeated, until the inequality restriction is satisfied. The whole process is repeated 325 times to cover all nonworkers. However, SHAZAM is unstable, if the program involves three levels of do-loops with relatively large ranges. So the above generation procedure should be modified to involve only two levels of do-loops. $NLWW_i$ and $NWHRS_i$ are generated for all 325 nonworkers. If $NLWW_i<3.3$ for all nonworkes, then $NLWW_i$ and $NWHRS_i$ are accepted. Otherwise, the generation process is repeated, until all the inequality restrictions are satisfied.

- Generation of Γ, given the augmented data, CN and Σ: We proceed as in (6.3.2.2), with the exception that only Γ is generated and the reservation wage equation is not involved. From economic theory, the following inequality restrictions are imposed:

+ $a_1>0$ and $a_2<0$ to ensure concave wage equation, peaking somewhere in midlife,

+ $a_3>0$ to ensure education is positively correlated with wage,

+ $a_5<0$ to ensure that unemployment rate is negatively correlated with wage.

- Generation of CN, given the augmented data, Γ, and Σ: We proceed as in (6.3.2.3), with the exception that the reservation wage equation is not involved. From economic theory, the following inequality restriction are imposed:

+ $c_1>0$ to ensure wage is positively correlated with labour supply,

+ $c_2>0$ and $c_3<0$ to ensure labour supply equation is concave, peaking somewhere in midlife,

+ $c_7<0$ to ensure that property income is negatively correlated with labour supply.

In order to reduce computation time, only the most essential inequality restrictions on parameters are imposed.

- Generation of Σ, given the augmented data, Γ and CN: We proceed like in (6.3.2.4), with the exception that the reservation wage equation is not involved.

6.3.3.2/ The SHAZAM program to realise the pratical Gibbs sampling procedure:

The program can be seen in appendix 3.

Despite the fact that the procedure had already been much simplified, SHAZAM could still not cope with the large amount of memory required. So the program was divided into two smaller programs and some tricks were employed. The first program generates the starting values and the first 500 observations, storing the last generated observation in two files as starting values for the second program, one file for Γ and (c_1 C), the other for Σ. The second program consists of two PROCEDURES, each procedure generates 500 observations. After the first procedure is executed, all the unnecessary variables are deleted and compressed to recover memory. After the second procedure is executed, the last generated observation is stored in two files as in the first program. The second program is then repeated as long as required.

6.3.3.3/ Computing results and interpretation:

In total, 4500 observations were generated and for each set of 500 observations, the STAT command was used to check for convergence. After discarding the first set of 500 observations, applications of the STAT command to each set 500 of observations revealed that the means, variances, minimum and maximum values of the observations are very constant with little deviation. So we can conclude that the first set of 500 observations formed the burn-in period and the remaining 4000 observations converged to the observations from the marginal posterior p.d.f's. of the parameters. We then used the CAT command of UNIX to combine the remaining 4000 observations into one file and used the STAT command to generate the moments of interest for the parameters, such as means and variances.

6.3.3.3.1/ The wage equation:

Table 9: Estimates of the wage equation

Variables	Coefficient	Estimates	Standard deviation
Constant	a_0	-1.3231	0.72289
WA	a_1	0.052899	0.032956
WA2	a_2	-0.00055542	0.00038168
WE	a_3	0.12041	0.01603
WPED	a_4	-0.0096028	0.0062
UN	a_5	-0.010674	0.0075913
CIT	a_6	0.069734	0.070485

All the sign conditions, set up in form of the inequality restrictions, are satisfied, namely $\hat{a}_1 > 0, \hat{a}_2 < 0, \hat{a}_3 > 0, \hat{a}_5 < 0$. The wage reaches maximum in midlife WA=$\frac{\hat{a}_1}{-2\hat{a}_2}$=47.

In the Bayesian approach, it does not really make sense to talk about significance of the estimates, particularly when inequality restrictions are imposed.

6.3.3.3.2/ The hours worked equation:

Table 10: Estimates of the hours worked equation

Variables	Coefficient	Estimates	Standard deviation
Constant	c_0	1093.8	998.52
LWW	c_1	716.73	940.66
WA	c_2	57.209	40.388
WA2	c_3	-0.80009	0.46881
WE	c_4	-91.154	108.15
UN	c_5	-10.708	16.145
CIT	c_6	-36.814	122.25
PRIN	c_7	-0.00556	0.00319
KL6	c_8	-308.99	101.84
K618	c_9	-123.81	30.714

All the sign conditions, set up in form of the inequality restrictions, are satisfied, namely $\hat{c}_1 > 0, \hat{c}_2 > 0, \hat{c}_3 < 0, \hat{c}_7 < 0$ The hours supplied reaches maximum in midlife $WA=\frac{\hat{c}_2}{-2\hat{c}_3}=36$. Following Mroz (1987), the average WHRS is 1500, that of PRIN is 1000 and that of WW is 4.5. The average of KL6 is 0.23772 and of K618 is 1.3533. Then, we can derive the following effects:

The uncompensated wage effect on labour supply is:

$$\frac{\partial WHRS_i}{\partial WW_i} = \frac{c_1}{WW_i} \approx \frac{\hat{c}_1}{\overline{WW}} = 159.27 \text{ hours}$$

The uncompensated wage elasticity with respect to labour supply is:

$$\frac{\partial \ln(WHRS_i)}{\partial \ln(WW_i)} = \frac{c_1}{WHRS_i} \approx \frac{\hat{c}_1}{\overline{WHRS}} = 0.478$$

Using the same derivation, the uncompensated property income effect on labour supply is 5.56 hours per \$1000 property income and the uncompensated property income elasticity with respect to labour supply is 0.003707

The signs of $\hat{c}_8, \hat{c}_9$ are negative, indicating negative effects of children numbers on female labour supply, where the smaller the child is, the greater the effect is. The uncompensated effect of KL6 on WHRS is 308.99 hours, while the uncompensated elasticity is 0.049. The uncompensated effect of K618 on WHRS is 123.81 hours, while the uncompensated elasticity is 0.1117.

6.3.3.3.3/ The covariance matrix Σ :

Table 11: Estimates of Σ

σ	Estimates	Standard deviation
σ_{ww}	0.45773	0.029
σ_{wg}	-11.184	24.96
σ_{gg}	571910	37555

The estimation of the covariance matrix is an advantage compaired to LS-approach, where it is impossible.

The estimates of the reservation wage equation can be derived from those of the wage and hours worked equation, using the estimates of the PROBIT model in part (5). But the estimates of the wage and hours worked equations using the Bayesian approach are very similar to those with LS-approach, so there is not much information to gain from derivation of the estimates of the reservation wage equation.