IV Estimation of Panel Data Tobit Models with Normal Errors.[*]

Bo E. Honoré[†]

First complete draft: April 1996. This version: May 1998.

Abstract

Amemiya (1973) proposed a "consistent initial estimator" for the parameters in a censored regression model with normal errors. This paper demonstrates that a similar approach can be used to construct moment conditions for fixed–effects versions of the model considered by Amemiya. This result suggests estimators for models that have not previously been considered.

1. A Moment for The Truncated Bivariate Normal Distribution.

The approach taken in this paper is motivated by Amemiya's "initial consistent estimator" (Amemiya 1973). That estimator uses the observation that if

$$y_i = \max\{0, x_i'\beta + \varepsilon_i\} \tag{1.1}$$

with $\varepsilon_i \sim N(0, \sigma^2)$ (independent of x_i), then

$$E\left[y_i^2 | y_i > 0, x_i\right] = E\left[y_i | y_i > 0, x_i\right] x_i'\beta + \sigma^2$$

which implies that

$$y_i^2 = y_i x_i'\beta + \sigma^2 + \xi_i, \qquad E\left[\xi_i | y_i > 0, x_i\right] = 0. \tag{1.2}$$

One can therefore estimate β and σ^2 by applying instrumental variables to (1.2), using only observations for which $y_i > 0$, and using functions of x_i as instruments.

[*]Financial support from the Alfred P. Sloan Foundation and from the National Science Foundation are gratefully acknowledged, as are comments from numerous seminar participants. Kenneth Y. Chay, Luojia Hu and Ekaterini Kyriazidou gave useful comments on an earlier draft.

[†]Mailing address: Department of Economics, Princeton University, Princeton, NJ 08544-1021. Phone (609) 258-4014. Fax: (609) 258-5561. Email: honore@Princeton.edu.

The contribution of this paper is to demonstrate that Amemiya's result can be generalized to panel data versions of (1.1) with individual–specific effects,

$$y_{it} = \max\left\{0, x_{it}'\beta + \alpha_i + \varepsilon_{it}\right\}, \tag{1.3}$$

where $\{\varepsilon_{it}\}_{t=1}^{T}$ is normally distributed and independent of $\left(\{x_{it}\}_{t=1}^{T}, \alpha_i\right)$. Throughout, we will assume random sampling across individuals, i.

The model in (1.3) was used by Heckman and MaCurdy's (1980) to study female labor supply. In their model, α_i is a function of, among other things, the Lagrange multiplier associated with an individual's budget constraint, which in turn "is a messy, but scalar, function of wage rates, interest rates,..." This implies that it is not reasonable to try to model α_i as a function of the explanatory variables, and they therefore treat $\{\alpha_i\}_{i=1}^{n}$ as a set of parameters to be estimated. This estimation procedure is justified asymptotically as $T \to \infty$. Honoré (1992) proposed estimators[1] for β that are justified asymptotically as $N \to \infty$ (with T fixed) under the exchangeability assumption that for any two time periods t and s, $(\varepsilon_{it}, \varepsilon_{is})$ is distributed like $(\varepsilon_{is}, \varepsilon_{it})$ conditional on $(x_{it}', x_{is}', \alpha_i)$. Here Amemiya's approach will be used to construct moment conditions (and hence, implictly, instrumental variables estimators) that can be applied to versions of the fixed effects censored regression model that assume normal errors, but are more general than the model considered in Honoré (1992) in other dimensions.

Before proceeding to develop versions of (1.2) that can be applied to panel data, we note that it is possible to extend (1.2) to higher moments. A simple integration–by–parts argument can be used to show that if $U \sim N(\mu, \sigma^2)$ then for any $k \geq 1$

$$E\left[U^{k+1}|U>0\right] = \mu E\left[U^{k}|U>0\right] + \sigma^2 k E\left[U^{k-1}|U>0\right]$$

Applying this to the censored regression model $y_i = \max\left\{0, x_i'\beta + \varepsilon_i\right\}$, with $\varepsilon_i \sim N(0, \sigma^2)$, and ε_i independent of x_i, yields the moment conditions

$$E\left[y_i^{k+1} - y_i^k x_i'\beta - \sigma^2 k y_i^{k-1}|y_i > 0, x_i\right] = 0$$

so β and σ^2 can be estimated by applying instrumental variables estimation to the equation

$$y_i^{k+1} = \left(y_i^k x_i'\right)\beta + \left(k y_i^{k-1}\right)\sigma^2 + \xi_i, \tag{1.4}$$

[1] A recent paper by Honoré and Kyriazidou (1998) shows that it is possible to modify Honoré's (1992) estimator in such a way that the exchangeability assumption can be replaced by a stationarity assumption.

using only the positive y_i's and using functions of x_i as instruments. For $k = 1$ this is equation (1.2) used by Amemiya (1973).

For the panel data censored regression model, (1.3), the equation underlying Amemiya's estimator becomes

$$E\left[y_{it}^2 - y_{it}\left(x_{it}'\beta + \alpha_i\right) - \sigma_t^2 | x_{it}, \alpha_i, y_{it} > 0\right] = 0 \tag{1.5}$$

where σ_t^2 denotes the variance of ε_{it}. If ε_{it} and ε_{is} are *independent* of each other, and of $(x_{it}, x_{is}, \alpha_i)$, then (1.5) also holds if we also condition on y_{is}. This, in turn, implies that

$$E\left[y_{it}^2 y_{is} - y_{it} y_{is}\left(x_{it}'\beta + \alpha_i\right) - y_{is}\sigma_t^2 | x_{it}, x_{is}, \alpha_i, y_{is}, y_{it} > 0\right] = 0$$

and therefore by the law of iterated expectations

$$E[y_{it}^2 y_{is} - y_{it} y_{is}\left(x_{it}'\beta + \alpha_i\right) - y_{is}\sigma_t^2 | x_{it}, x_{is}, \alpha_i, y_{is} > 0, y_{it} > 0] = 0.$$

Reversing t and s, taking differences, and integrating out α_i, yields

$$\begin{aligned} &E[(y_{it}^2 y_{is} - y_{is}^2 y_{it}) - y_{it} y_{is}(x_{it}' - x_{is}')\beta \\ &- \left(y_{is}\sigma_t^2 - y_{it}\sigma_s^2\right) | x_{it}, x_{is}, y_{is} > 0, y_{it} > 0] = 0. \end{aligned} \tag{1.6}$$

or equivalently

$$(y_{it}^2 y_{is} - y_{is}^2 y_{it}) = y_{it} y_{is}(x_{it}' - x_{is}')\beta + y_{is}\sigma_t^2 - y_{it}\sigma_s^2 + \xi_{ist}, \tag{1.7}$$

with $E[\xi_{ist} | x_{it}, x_{is}, y_{is} >, y_{it} > 0]$. This implies that β, σ_t^2 and σ_s^2 can be estimated by applying instrumental variables to (1.7), using functions of the x's as instruments. The statistical properties of such an estimator then follow by standard instrumental variables arguments.

The derivation leading to (1.6) assumes that the errors are independently distributed over time. The following proposition (which is proved in the appendix) will allow us to derive moment conditions which exploit higher moments as in (1.4) and which can be applied when the errors have a joint normal distribution with arbitrary dependence.

Proposition 1.1. *If*

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$$

then for $k \geq 1$ and $m \geq 1$,

$$E\left[U_i^{k+1} U_j^m - U_i^k U_j^{m+1} \middle| U_i > 0, U_j > 0\right]$$

$$
\begin{aligned}
&= \left(\mu_i - \mu_j\right) E\left[U_i^k U_j^m \middle| U_i > 0, U_j > 0\right] \\
&\quad + \left(\sigma_i^2 - \sigma_{ij}\right) kE\left[U_i^{k-1} U_j^m \middle| U_i > 0, U_j > 0\right] \\
&\quad - \left(\sigma_j^2 - \sigma_{ij}\right) mE\left[U_i^k U_j^{m-1} \middle| U_i > 0, U_j > 0\right]
\end{aligned}
$$

2. Moment Conditions for Truncated and Censored Regression Models.

Proposition 1.1 can be used to construct moment conditions for panel data censored or truncated regression models under a variety of assumptions, provided that one is willing to assume that the errors are normally distributed conditional on the regressors of the model and conditional on the fixed effects.

2.1. Non-stationarity.

It is convenient to rewrite the censored regression model[2] in equation(1.3) in terms of an underlying latent variable, y_{it}^*,

$$
\begin{aligned}
y_{it}^* &= x_{it}'\beta + \alpha_i + \varepsilon_{it}, \qquad t = 1, \ldots, T, \qquad i = 1, \ldots, N \\
y_{it} &= \max\left\{0, y_{it}^*\right\}
\end{aligned}
\tag{2.1}
$$

In this subsection we will assume that $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ is normally distributed and independent of $(\alpha_i, x_{i1}, \ldots, x_{iT})$ with $\text{var}[\varepsilon_{it}] = \sigma_t^2$ and $\text{cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_{t,s}$. The generalization here, relatively to Honoré (1992), is that it is not necessary to assume that $\{\varepsilon_{it}\}$ is stationary. In some contexts, this can be important. For example, Chay (1995) considered panel data censored regression models for wages[3]. Since it is fairly well–documented that the distribution of earnings has changed over time in the U.S., it does not seem reasonable to assume that $\{\varepsilon_{it}\}$ is stationary (and Chay consequently took a random effects approach, rather than a fixed effects approach).

Applying Proposition 1.1 to the conditional distribution of (y_{it}^*, y_{is}^*) given $(x_{it}', x_{is}', \alpha_i)$ yields

$$
\begin{aligned}
y_{it}^{*\,k+1}\, y_{is}^{*\,m} - y_{is}^{*\,m+1}\, y_{it}^{*\,k} &= y_{it}^{*\,k}\, y_{is}^{*\,m}\left(x_{it}' - x_{is}'\right)\beta + k\, y_{it}^{*\,k-1}\, y_{is}^{*\,m}\left(\sigma_t^2 - \sigma_{s,t}\right) \\
&\quad - m\, y_{is}^{*\,m-1}\, y_{it}^{*\,k}\left(\sigma_s^2 - \sigma_{s,t}\right) + \xi_{ist}
\end{aligned}
\tag{2.2}
$$

[2] With truncation (y_{it}, x_{it}), are drawn from (2.1) conditional of $y_{it}^* > 0$. The methods described here can be applied to truncated regression models as well as to censored regression models.

[3] In his application, the censoring was induced because he used U.S. social security earnings as his dependent variable. This implies that the observations are censored from above at the taxable maximum.

where

$$E\left[\xi_{ist}|x'_{it},x'_{is},y^*_{it}>0,y^*_{is}>0\right]=0.$$

This implies that β, $(\sigma_s^2-\sigma_{t,s})$ and $(\sigma_t^2-\sigma_{t,s})$ can be estimated by using only pairs of observations for which $y_{it}>0$ and $y_{is}>0$ and then applying instrumental variable estimation to (2.2) using functions of (x'_{it},x'_{is}) as instruments.

2.2. Factor loading.

It is sometimes desirable to allow the individual specific effect to have different coefficients in different time periods. Such a model has been discussed in a number of contexts. For example, Heckman (1981) investigates a random effect discrete choice model, Holtz–Eakin, Newey and Rosen (1988) a linear autoregressive model, and Chay (1995) a random effects censored regression model, where (in all cases) the individual specific effect is multiplied by a time–specific "factor–loading". Proposition 1.1 can be used to construct an estimator of a fixed effect censored regression model of this kind.

The model is

$$y^*_{it}=x'_{it}\beta+\rho_t\alpha_i+\varepsilon_{it},\qquad t=1,\ldots,T,\qquad i=1,\ldots,N \tag{2.3}$$

where the researcher observes x_{it} and $y_{it}=\max\{0,y^*_{it}\}$.

Since α_i is not observed, it is not possible to identify the scale of the factor–loadings, $\rho_1,\rho_2,\ldots\rho_T$. However, for any two time periods, t and s, we have

$$\begin{aligned} y^{*2}_{it}y^*_{is}-y^{*2}_{is}y^*_{it}\rho_t/\rho_s &= y^*_{it}y^*_{is}\left(x_{it}-x_{is}\rho_t/\rho_s\right)'\beta \\ &\quad -y^*_{it}\left(\sigma_s^2\rho_t/\rho_s-\sigma_{t,s}\right)+y^*_{is}\left(\sigma_t^2-\sigma_{t,s}\rho_t/\rho_s\right)+\xi_{ist} \end{aligned} \tag{2.4}$$

where, again,

$$E\left[\xi_{ist}|x'_{it},x'_{is},y^*_{it}>0,y^*_{is}>0\right]=0.$$

As was the case in the previous subsection, it is also possible to use information contained in higher moments.

2.3. Fixed effects in variance.

The homoskadasticity assumption across individuals is strong. The other extreme is to assume that the errors are heteroskedastic across observations, but *i.i.d.* over time, and with different variances for different individuals. In that case (2.2) becomes

$$y_{it}^{*2}y_{is}^{*}-y_{is}^{*2}y_{it}^{*}=y_{it}^{*}y_{is}^{*}\left(x_{it}-x_{is}\right)'\beta-\left(y_{it}^{*}-y_{is}^{*}\right)\sigma_{i}^{2}+\xi_{ist} \tag{2.5}$$

with

$$E\left[\xi_{ist}|x_{it}',x_{is}',y_{it}^{*}>0,y_{is}^{*}>0\right]=0.$$

Considering also a third time period, τ, we get

$$\begin{aligned}
& \left(y_{it}^{*2}y_{is}^{*}-y_{is}^{*2}y_{it}^{*}\right)+\left(y_{is}^{*2}y_{i\tau}^{*}-y_{i\tau}^{*2}y_{is}^{*}\right)+\left(y_{i\tau}^{*2}y_{it}^{*}-y_{it}^{*2}y_{i\tau}^{*}\right) \\
= & \; y_{it}^{*}y_{is}^{*}\left(x_{it}-x_{is}\right)'\beta+y_{is}^{*}y_{i\tau}^{*}\left(x_{is}-x_{i\tau}\right)'\beta+y_{i\tau}^{*}y_{it}^{*}\left(x_{i\tau}-x_{it}\right)'\beta \\
& -\left(y_{it}^{*}-y_{is}^{*}\right)\sigma_{i}^{2}-\left(y_{is}^{*}-y_{i\tau}^{*}\right)\sigma_{i}^{2}-\left(y_{i\tau}^{*}-y_{it}^{*}\right)\sigma_{i}^{2} \\
& +\xi_{ist}+\xi_{i\tau s}+\xi_{it\tau} \\
= & \left(y_{it}^{*}y_{is}^{*}\left(x_{it}-x_{is}\right)'+y_{is}^{*}y_{i\tau}^{*}\left(x_{is}-x_{i\tau}\right)'+y_{i\tau}^{*}y_{it}^{*}\left(x_{i\tau}-x_{it}\right)'\right)\beta+\xi_{ist\tau}
\end{aligned} \tag{2.6}$$

with

$$E\left[\xi_{ist\tau}|x_{it}',x_{is}',x_{i\tau}',y_{it}^{*}>0,y_{is}^{*}>0,y_{i\tau}^{*}>0\right]=0.$$

It is also possible to allow the variance to have time-specific component, provided that the resulting variance has an additive structure of the form $\sigma_{it}^{2}=\sigma_{i}^{2}+\sigma_{t}^{2}$. In this case we have

$$\begin{aligned}
& \left(y_{it}^{*2}y_{is}^{*}-y_{is}^{*2}y_{it}^{*}\right)+\left(y_{is}^{*2}y_{i\tau}^{*}-y_{i\tau}^{*2}y_{is}^{*}\right)+\left(y_{i\tau}^{*2}y_{it}^{*}-y_{it}^{*2}y_{i\tau}^{*}\right) \\
= & \; y_{it}^{*}y_{is}^{*}\left(x_{it}-x_{is}\right)'\beta+y_{is}^{*}y_{i\tau}^{*}\left(x_{is}-x_{i\tau}\right)'\beta+y_{i\tau}^{*}y_{it}^{*}\left(x_{i\tau}-x_{it}\right)'\beta \\
& -y_{it}^{*}\left(\sigma_{s}^{2}+\sigma_{i}^{2}\right)+y_{is}^{*}\left(\sigma_{t}^{2}+\sigma_{i}^{2}\right)-y_{is}^{*}\left(\sigma_{\tau}^{2}+\sigma_{i}^{2}\right) \\
& +y_{i\tau}^{*}\left(\sigma_{s}^{2}+\sigma_{i}^{2}\right)-y_{i\tau}^{*}\left(\sigma_{t}^{2}+\sigma_{i}^{2}\right)+y_{it}^{*}\left(\sigma_{\tau}^{2}+\sigma_{i}^{2}\right) \\
& +\xi_{ist}+\xi_{i\tau s}+\xi_{it\tau} \\
= & \left(y_{it}^{*}y_{is}^{*}\left(x_{it}-x_{is}\right)'+y_{is}^{*}y_{i\tau}^{*}\left(x_{is}-x_{i\tau}\right)'+y_{i\tau}^{*}y_{it}^{*}\left(x_{i\tau}-x_{it}\right)'\right)\beta \\
& +\left(y_{i\tau}^{*}-y_{it}^{*}\right)\sigma_{s}^{2}+\left(y_{is}^{*}-y_{i\tau}^{*}\right)\sigma_{t}^{2}+\left(y_{it}^{*}-y_{is}^{*}\right)\sigma_{\tau}^{2} \\
& +\xi_{ist\tau}
\end{aligned} \tag{2.7}$$

with

$$E\left[\xi_{ist\tau}|x_{it}',x_{is}',x_{i\tau}',y_{it}^{*}>0,y_{is}^{*}>0,y_{i\tau}^{*}>0\right]=0.$$

As in the previous subsection, additional moment conditions can be obtained by applying Proposition 1 for other values of m and k.

2.4. Fixed Effects in the Slopes.

Rather than letting the fixed effect work through the level of the model, one might be interested in estimating β in the model

$$y_{it}^* = x_{it}'\beta + z_{it}\alpha_i + \varepsilon_{it}, \qquad t = 1, \ldots, T, \qquad i = 1, \ldots, N \tag{2.8}$$

where z_{it} is an observed explanatory variable. This model is in the spirit of random coefficients models.

In this case, we can apply Proposition 1 to the distribution of[4] $(y_{is}^* z_{it}, y_{it}^* z_{is})$, conditional on $(x_{it}, x_{is}, z_{it}, z_{is}, \alpha_i)$,

$$\begin{aligned} y_{it}^{*2} z_{is}^2 y_{is}^* z_{it} - y_{is}^{*2} z_{it}^2 y_{it}^* z_{is} &= y_{it}^* z_{is} y_{is}^* z_{it} \left(x_{it} z_{is} - x_{is} z_{it} \right)' \beta - \\ & \quad y_{it}^* z_{is} \left(z_{it}^2 \sigma_s^2 - z_{it} z_{is} \sigma_{t,s} \right) + y_{is}^* z_{it} \left(z_{is}^2 \sigma_t^2 - z_{it} z_{is} \sigma_{t,s} \right) + \xi_{ist} \end{aligned} \tag{2.9}$$

with

$$E\left[\xi_{ist} | z_{it}, z_{is}, x_{it}', x_{is}', y_{it}^* > 0, y_{is}^* > 0\right] = 0.$$

3. Concluding Remarks.

The moment conditions and associated instrumental variables procedures discussed in this paper all rely on strict exogeneity of the explanatory variables[5] and on normality of the errors. While these are strong assumptions, the methods proposed here are the first that can be applied to the "fixed–effects" Tobit models in previous sections.

Wales and Woodland (1980) found that the estimator of the cross–sectional censored regression model based on (1.2) is relatively inefficient. Since the approach here is based on the same insight as (1.2), one might worry that estimators based on the Proposition 1.1 will also be imprecise. The difference between the two cases, however, is that for the censored regression model with normal errors, there are other estimators to which the instrumental variables estimator can be compared. In the panel data settings discussed here, the instrumental variables estimator is the first, and so far only, available estimator.

Finally, we note that for all the example considered here, one can construct estimators based on any choice of k and m in Proposition 1.1. Combining these moment conditions optimally

[4] To simplify the derivations, we assume here that $z_{it} > 0$ and $z_{is} > 0$.

[5] Formally, assumptions are made on the error terms, conditional on future values of the explanatory variales. This is very restrictive, as it rules out feedback from the current dependent variables to future explanatory variables.

would certainly improve the asymptotic efficiency of the estimator. We also note that the moment conditions here are conditional, and that the errors are heteroskedastic. This implies additional methods for improving the asymptotic efficiency of the estimator. However, the contribution of this paper is to show that it is *possible* to estimated the models considered in the previous section, and not to re–derive results about optimal generalized method of moments estimation.

4. References.

Amemiya, T. (1973): "Regression Analysis when the Dependent Variable is Truncated Normal," Econometrica, 41, pp. 997–1016.

Chay, Kenneth Y. (1995): "Evaluating the Impact of the 1964 Civil Rights Act on the Economics Status of Black Men using Censored Longitudinal Earnings Data," unpublished manuscript, Department of Economics, Princeton University.

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

Heckman, J. J. and T. E. MaCurdy (1980): "A Life Cycle Model of Female Labour Supply." Review of Economic Studies, 47, pp. 47–74.

Holtz–Eakin, D., W. Newey and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data." Econometrica, 56, pp. 1371-1395.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," Econometrica, 60, pp. 533–565.

Honoré, B. E., and E. Kyriazidou (1998): "Estimation of Tobit–Type Models with Individual Specific Effects." Unpublished.

Wales, T. J., and A. D. Woodland (1980): "Sample Selectivity and the Estimation of Labor Supply Functions." International Economic Review 21: 437–468.

5. Appendix: Derivation of Proposition 1.1.

Using the usual notation, the bivariate normal density, $f(u_1, u_2)$ satisfies

$$\frac{\partial f(u_1,u_2)}{\partial u_1} = -\frac{1}{1-\rho^2}\left(\frac{1}{\sigma_1^2}u_1 - \frac{\rho}{\sigma_1\sigma_2}u_2 - \frac{\mu_1}{\sigma_1^2} + \frac{\rho\mu_2}{\sigma_1\sigma_2}\right) f(u_1,u_2)$$

Therefore

$$\begin{aligned} & \int_0^\infty k u_1^{k-1} u_2^m f(u_1,u_2)\, du_1 \\ = & \int_0^\infty u_1^k u_2^m \frac{1}{1-\rho^2}\left(\frac{1}{\sigma_1^2}u_1 - \frac{\rho}{\sigma_1\sigma_2}u_2 - \frac{\mu_1}{\sigma_1^2} + \frac{\rho\mu_2}{\sigma_1\sigma_2}\right) f(u_1,u_2)\, du_1 \\ = & \frac{1}{1-\rho^2}\int_0^\infty \left(\frac{1}{\sigma_1^2}u_1^{k+1}u_2^m - \frac{\rho}{\sigma_1\sigma_2}u_1^k u_2^{m+1} - \left(\frac{\mu_1}{\sigma_1^2} - \frac{\rho\mu_2}{\sigma_1\sigma_2}\right) u_1^k u_2^m\right) f(u_1,u_2)\, du_1 \end{aligned}$$

or

$$\begin{aligned} & k\left(1-\rho^2\right)\sigma_1^2\sigma_2 E\left[U_1^{k-1}U_2^m \middle| U_1>0, U_2>0\right] \\ = & E\left[\sigma_2 U_1^{k+1}U_2^m - \rho\sigma_1 U_1^k U_2^{m+1} - \left(\sigma_2\mu_1 - \rho\sigma_1\mu_2\right) U_1^k U_2^m \middle| U_1>0, U_2>0\right] \end{aligned}$$

multiplying by σ_2

$$\begin{aligned} & k\left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right) E\left[U_1^{k-1}U_2^m \middle| U_1>0, U_2>0\right] \\ = & E\left[\sigma_2^2 U_1^{k+1}U_2^m - \sigma_{12} U_1^k U_2^{m+1} - \left(\sigma_2^2\mu_1 - \sigma_{12}\mu_2\right) U_1^k U_2^m \middle| U_1>0, U_2>0\right] \end{aligned} \tag{5.1}$$

Applying the same integration by parts to $\int_0^\infty m u_1^k u_2^{m-1} f(u_1,u_2)\, du_2$, we get

$$\begin{aligned} & m\left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right) E\left[U_1^{k}U_2^{m-1} \middle| U_1>0, U_2>0\right] \\ = & E\left[\sigma_1^2 U_1^{k}U_2^{m+1} - \sigma_{12} U_1^{k+1} U_2^{m} - \left(\sigma_1^2\mu_2 - \sigma_{12}\mu_1\right) U_1^k U_2^m \middle| U_1>0, U_2>0\right] \end{aligned} \tag{5.2}$$

Multiplying (5.1) by $(\sigma_1^2 - \sigma_{12})$ and (5.2) by $(\sigma_2^2 - \sigma_{12})$, and then subtracting (5.2) from (5.1) and simplifying the right hand side yields

$$\begin{aligned} & k\left(\sigma_1^2 - \sigma_{12}\right)\left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right) E\left[U_1^{k-1}U_2^m \middle| U_1>0, U_2>0\right] \\ & -m\left(\sigma_2^2 - \sigma_{12}\right)\left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right) E\left[U_1^{k}U_2^{m-1} \middle| U_1>0, U_2>0\right] \\ = & \left[\left(\sigma_1^2 - \sigma_{12}\right)\sigma_2^2 + \left(\sigma_2^2 - \sigma_{12}\right)\sigma_{12}\right] E\left[U_1^{k+1}U_2^m \middle| U_1>0, U_2>0\right] \\ & -\left[\left(\sigma_1^2 - \sigma_{12}\right)\sigma_{12} + \left(\sigma_2^2 - \sigma_{12}\right)\sigma_1^2\right] E\left[U_1^{k}U_2^{m+1} \middle| U_1>0, U_2>0\right] \\ & -\left[\left(\sigma_1^2 - \sigma_{12}\right)\left(\sigma_2^2\mu_1 - \sigma_{12}\mu_2\right) - \left(\sigma_2^2 - \sigma_{12}\right)\left(\sigma_1^2\mu_2 - \sigma_{12}\mu_1\right)\right] E\left[U_1^{k}U_2^{m} \middle| U_1>0, U_2>0\right] \\ = & \left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right) E\left[U_1^{k+1}U_2^m \middle| U_1>0, U_2>0\right] \\ & -\left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right) E\left[U_1^{k}U_2^{m+1} \middle| U_1>0, U_2>0\right] \\ & -\left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right)\left(\mu_1 - \mu_2\right) E\left[U_1^{k}U_2^{m} \middle| U_1>0, U_2>0\right] \end{aligned}$$

dividing through by $(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)$ then yields the result.