NEW FUNCTIONAL ESTIMATOR IN QUADRATIC ERRORS-IN-VARIABLES MODEL

ELENA USOLTSEVA AND ALEXANDER KUKUSH

Abstract. A quadratic structural errors-in-variables model is considered. Functional estimators are studied that are generated by estimating functions conditionally unbiased given the latent variable. Those estimators are constructed without the knowledge of the latent variable distribution. A problem is studied how to construct an estimator from the class, which has the smallest, in certain sense, asymptotic covariance matrix.

1. Introduction

We study a structural regression model

$$y = \beta^T \rho(x) + \varepsilon, \quad \rho(x) = (1, x, \cdots, x^m)^T. \tag{1}$$

Here $m \geq 1$ is fixed, $\beta = (\beta_0, \beta_1, \cdots, \beta_m)^T \in \mathbb{R}^{(m+1)\times 1}$, the regressor x is a random variable, and the observation error ε is a centered random variable. The x is unobserved, instead a surrogate data

$$w = x + u \tag{2}$$

is observed, whereas x, u, and ε are independent. The model is normal; that is

$$x \sim N(\mu, \sigma_x^2), \quad \varepsilon \sim N(0, \sigma_\varepsilon^2), \quad u \sim N(0, \sigma_u^2). \tag{3}$$

All the variances are positive and σ_u^2 is the only known parameter.

In the model (1) to (3), a version of the Quasi-Likelihood estimator is the optimal estimator for β; see [2]. The construction of this estimator is based on the normality of x. Therefore, it is reasonable to consider less efficient but robust estimator, e. g., the Corrected Score (CS) estimator (see [4] for the definition of the CS estimator; another name for this estimator is Adjusted Least Squares). This estimator is robust in the sense that it is consistent for any distribution of x (the only restriction is that certain moment of x should be finite).

In the present paper, **E**, **Var**, and **Cov** denote expectation, variance (of a random variable), and covariance matrix, respectively.

Now, we introduce the class S_L of linear-in-y estimating functions of the form

$$S_L = S_L(w, y, \beta) = p(w)y - Q(w)\beta \tag{4}$$

such that for all $\beta \in \mathbb{R}^{(m+1)\times 1}$,

$$\mathbf{E}_\beta(S_L|x) = 0. \tag{5}$$

The functions $p(\cdot)$ and $Q(\cdot)$ are C^2-smooth functions valued in $\mathbb{R}^{(m+1)\times 1}$ and $\mathbb{R}^{(m+1)\times(m+1)}$, respectively. We assume that components of those functions belong to the Schwarz space S' of slowly growing distributions; therefore, the deconvolution problems considered below deal with the functions from S' and have unique solutions.

Received by the editors 15/03/2010.

2000 *Mathematics Subject Classification.* 62J05, 62F12.

Key words and phrases. Asymptotic covariance matrix, efficient estimator, functional estimator, quadratic errors-in-variables model.

Both authors are supported by the Swedish Institute grant SI-01424/2007.

Consider independent copies (x_i, w_i, y_i) of the model (1) to (3). We observe the couples (w_i, y_i), $i = 1, \dots, n$.

For an estimating function $s_L \in S_L$, the estimator $\widehat{\beta}_L$ is defined as a measurable solution to the equation

$$\sum_{i=1}^{n} s_L(w_i, y_i, \beta) = 0, \quad \beta \in \mathbb{R}^{(m+1)\times 1}. \tag{6}$$

In fact,

$$\widehat{\beta}_L = \left(\frac{1}{n} \sum_{i=1}^{n} Q(w_i) \right)^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} p(w_i) y_i. \tag{7}$$

We assume that within the class S_L, the matrix $\mathbf{E}Q(w)$ is nonsingular. By the strong low of large numbers $T_n := \frac{1}{n}\sum_{i=1}^{n} Q(w_i) \longrightarrow \mathbf{E}Q(w)$ as $n \to \infty$, a.s. Here the limit is nonsingular, and then the matrix T_n is nonsingular, for all $n \ge n_0(w)$, a.s. Thus, the estimator (7) is well-defined for all $n \ge n_0(w)$, a.s. To be precise we set $\widehat{\beta}_L(w) = 0$ if the matrix $T_n(w)$ is singular.

We mention that S_L contains the estimating function of the CS estimator. It is straightforward that the estimator $\widehat{\beta}_L$ is strictly consistent. Then according to the theory of estimating equations it is asymptotically normal, i.e., $\sqrt{n}(\widehat{\beta}_L - \beta) \to N(0, \Sigma_L)$ in distribution. The matrix Σ_L is called the asymptotic covariance matrix (ACM) of the estimator and can be computed by the sandwich formula, see [1],

$$\Sigma_L = A_L^{-1} B_L A_L^{-T}, \quad A_L = -\mathbf{E}\frac{\partial S_L}{\partial \beta}, \quad B_L = \mathbf{E} S_L S_L^T. \tag{8}$$

Hereafter $A_L^{-T} := \left(A_L^{-1}\right)^T$.

In [6] an attempt was made to prove the optimality of the CS estimator within the class S_L. This was done only for the case of small non-intercept coefficients $\beta_1, \cdots, \beta_m$. Moreover in that paper it was mentioned that there exists β such that the CS estimator is not optimal within the class S_L.

In the present paper we are looking for the estimator within this class, which is more efficient, to some extent, compared with the CS estimator. We consider the case $m = 2$ only, which corresponds to the quadratic model.

Let s_1, $s_2 \in S_L$ and Σ_1, Σ_2 be the ACMs of the corresponding estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$. We call $\widehat{\beta}_1$ strictly more efficient than $\widehat{\beta}_2$ if $\Sigma_1 < \Sigma_2$. Hereafter the inequality between symmetric matrices of the same size is understood in Loewner order, i.e., $\Sigma_1 < \Sigma_2$ and $\Sigma_1 \le \Sigma_2$ means that $\Sigma_2 - \Sigma_1$ is positive definite or positive semidefinite, respectively.

The paper is organized as follows. Section 2 computes the ACM of the estimator $\widehat{\beta}_L$ and presents our main result, Section 3 concludes, and proofs are given in Appendix.

2. Asymptotic covariance matrix and main result

For any $s_L \in S_L$, we compute the A_L and B_L given in (8). We have

$$A_L = -\mathbf{E}\frac{\partial S_L}{\partial \beta^T} = \mathbf{E}(Q(w)) = \mathbf{E}[\mathbf{E}(Q|x)] = \mathbf{E}[\mathbf{E}(p|x)\rho^T] =$$

$$= \mathbf{E}[\mathbf{E}(p\rho^T|x)] = \mathbf{E}(p\rho^T) = \mathbf{E}[\mathbf{E}(p\rho^T|w)] = \mathbf{E}[p\mathbf{E}(\rho^T|w)] = \mathbf{E}(p\rho_w^T),$$

where we set

$$\rho_w^T = \mathbf{E}(\rho^T(x)|w);$$

$$B_L = \mathbf{E}S_L S_L^T = \mathbf{E}(py - Q\beta)(py - Q\beta)^T = \mathbf{E}([(y-M)p + (Mp - Q\beta)]\times$$

$$\times[(y-M)p + (Mp - Q\beta)]^T) = \mathbf{E}vpp^T + \mathbf{E}(Mp - Q\beta)(Mp - Q\beta)^T =$$

$$= \mathbf{E}vpp^T + \mathbf{Cov}(Mp - Q\beta).$$

Here we denote

$$M = \mathbf{E}(y|w) = \beta^T \mathbf{E}\rho(x), \tag{9}$$

$$v = \mathbf{Var}(y|w) = \mathbf{Var}(\beta^T \rho(x)|w) + \sigma_\varepsilon^2. \tag{10}$$

From (8) we have finally

$$\Sigma_L = \left(\mathbf{E}(p\rho_w^T)\right)^{-1} \left(\mathbf{E}vpp^T + \mathbf{Cov}(Mp - Q\beta)\right) \left(\mathbf{E}(p\rho_w^T)\right)^{-T}.$$

We consider the model (1) to (3) with $m = 2$. The CS estimator is generated by $S_C \in S_L$,

$$S_C = p_C(w) - Q_C(w)\beta.$$

The vector function $p_C(w)$ and matrix valued function $Q_C(w)$ are polynomial in w which satisfy the deconvolution equations

$$\mathbf{E}\left(p_C(w)|x\right) = \rho(x),$$

$$\mathbf{E}\left(Q_C(w)|x\right) = \rho(x)\rho(x)^T.$$

Now, we introduce the re-corrected estimating function

$$S_{rc} = p_{rc}(w)y - Q_{rc}(w)\beta,$$

where $p_{rc}(w)$ and Q_{rc} are (polynomial) solutions to the deconvolution problems

$$\mathbf{E}(p_{rc}(w)|x) = \begin{pmatrix} 1 \\ x \\ x^2 + \delta x^3 \end{pmatrix},$$

$$\mathbf{E}(Q_{rc}(w)|x) = \begin{pmatrix} 1 \\ x \\ x^2 + \delta x^3 \end{pmatrix} \rho(x)^T.$$

Here δ is a real parameter; $|\delta|$ will be small enough.

We want to compare the corresponding ACMs $\Sigma_{rc}(\delta)$ and Σ_c of the estimator generated by the S_{rc} and the CS estimator. Because $\Sigma_c = \Sigma_{rc}(0)$, in fact we compare $\Sigma_{rc}(\delta)$ and $\Sigma_{rc}(0)$.

Theorem 2.1. *It holds* $(\det \Sigma_{rc})'(0) \neq 0$, *for almost all parameters* $(\beta^T, \mu, \sigma_x^2, \sigma_\varepsilon^2)^T$ *w.r.t. Lebesgue measure on* $\mathbb{R}^6$.

The proof is given in Appendix.

Remark 2.1. Suppose that the true values of the parameters are "typical" in the sense that $\mathrm{d} := (\det\Sigma_{rc})'(0) \neq 0$. For the observations (w_i, y_i), $i = 1, \ldots, n$, based on the Quasi-Likelihood estimator for β and empirical mean and empirical variance of w, it is easy to construct a strongly consistent estimator $\widehat{\theta}_n$ of the parameter vector $\theta := (\beta^T, \mu, \sigma_x^2, \sigma_\varepsilon^2)^T$; see [3]. The function $\mathrm{d} = \mathrm{d}(\theta)$ is a rational function; therefore, $\widehat{d} := \mathrm{d}(\widehat{\theta}_n)$ is a strongly consistent estimator for $\mathrm{d}(\theta)$. Then we set $\widehat{\delta}_n = -\,\delta_0 \cdot \mathrm{sign}\, \mathrm{d}(\widehat{\theta}_n)$, with fixed $\delta_0 > 0$, and consider the estimating function $\widetilde{S}_{rc} = S_{rc}(\delta)|_{\delta=\widehat{\delta}_n}$. The corresponding estimator $\widetilde{\beta}_{rc}$ coincides with $\widehat{\beta}_{rc}(\delta)|_{\delta=-\,\delta_0 \cdot \mathrm{sign}\ d}$, for all $n \geq n_0(\omega)$, a.s. Therefore, the ACMs $\widetilde{\Sigma}_{rc}$ and $\Sigma_{rc}(-\,\delta_0 \cdot \mathrm{sign}\ d)$ are equal. Then for small enough δ_0, $\det \widetilde{\Sigma}_{rc} < \det \Sigma_c$; as a result the volume of the asymptotic confidence ellipsoid for β will be smaller for the estimator $\widetilde{\beta}_{rc}$ than for the $\widehat{\beta}_c$, for large n.

3. Conclusion

We considered the normal quadratic (i.e., with $m = 2$) measurement error model (1) to (3). The CS estimator of β is robust in the sense that it is (strictly) consistent without the assumption of the normality of x. We showed the way how to construct another robust estimator which is more efficient than the CS estimator in the sense that the new estimator yields smaller volume of the asymptotic confidence ellipsoid for β.

References

[1] R. J. Carrol, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models. A Modern Perspective*, 2nd ed, Chapman and Hall/CRC, London 2006.
[2] A. Kukush, A. Malenko, and H. Schneeweiss, Comparing the efficiency of estimates in concrete errors-in-variables models under unknown nuisance parameters, *Theory of Stochastic Processes* **13(29)** (2007), no. 4, 69–81.
[3] A. Kukush, A. Malenko and H. Schneeweiss, Optimality of the quasi score estimator in mean-variance model with applications to measurement error models, *Journal of Statistical Planning and Inference* **139** (2009), 69–81.
[4] S. Shklyar, H. Schneeweiss, and A. Kukush, Quasi Score is more efficient than CS in a polynomial measurement error model, *Metrika* **65** (2007), 275–295.
[5] L. A. Stefanski, Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models, *Communications of Statististics–Theory and Methods* **18** (1989), 4335–4358.
[6] E. Usoltseva and A. Kukush, Optimality of CS estimator in nonlinear errors-in-variables model when non-intercept terms are small, *Journal of Numerical and Applied Mathematics* **1(96)** (2008), 139–148.

4. Appendix

4.1. **Auxiliary computations.** For the estimating function (4) from the class S_L, the condition (5) implies that a.s.

$$\mathbf{E}_\beta\left(p(w)y|x\right)=\mathbf{E}\left(Q(w)|x\right)\beta,$$

$$\mathbf{E}\left(p(w)|x\right)\rho^T(x)\beta=\mathbf{E}\left(Q(w)|x\right)\beta.$$

Because this holds for each $\beta\in\mathbb{R}^{m+1}$, we obtain

$$p_x(x)\rho^T(x)=\mathbf{E}\left(Q(w)|x\right), \tag{11}$$

where we denote

$$p_x(x)=\mathbf{E}\left(p(w)|x\right). \tag{12}$$

Further, we want to expand the function $Q(t)$, $t\in\mathbb{R}$, for small σ_u^2. The next Lemma is a consequence of the expansions from [5].

Lemma 4.1. *Let $u\sim N(0,\sigma_u^2)$, and g, h be smooth enough functions such that*

$$\boldsymbol{E}g(t+u)=h(t),\quad t\in\mathbb{R}.$$

Then for all $t\in\mathbb{R}$, it holds

$$g(t)=h(t)-\frac{1}{2}h''(t)\sigma_u^2+R,\quad as\quad \sigma_u^2\to 0,$$

where $\boldsymbol{E}R=O(\sigma_u^4)$.

Now, all remainder terms R_i below satisfy the condition $\mathbf{E}R_i=O(\sigma_u^4)$, as $\sigma_u^2\to 0$.
We apply Lemma 4.1 to the relation (12) and obtain for all $t\in\mathbb{R}$, that

$$p(t)=p_x(t)-\frac{1}{2}p_x''(t)\sigma_u^2+R_1.$$

Next, applying Lemma 4.1 to the relation (11) we obtain

$$Q(t)=p_x(t)\rho^T(t)-\frac{1}{2}(p_x(t)\rho^T(t))''\sigma_u^2+R_2=$$

$$=p_x(t)\rho^T(t)-\frac{1}{2}(p_x''(t)\rho^T(t)+p_x'(t)\rho'^T(t)+p_x(t)\rho''^T(t))\sigma_u^2+R_2.$$

We consider $\rho_w=\mathbf{E}(\rho(x)|w)=\begin{pmatrix}1\\ \mu_1\\ \mu_1^2+\tau^2\end{pmatrix}$, where

$$\mu_1=\mu_x+\frac{\sigma_x^2}{\sigma_x^2+\sigma_x^2}(w-\mu),\quad \tau^2=\frac{\sigma_u^2\sigma_x^2}{\sigma_u^2+\sigma_x^2}.$$

Thus,

$$\rho_w = \begin{pmatrix} 1 \\ \mu_x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_x^2}(w-\mu) \\ \left(\mu_x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_x^2}(w-\mu)\right)^2 + \frac{\sigma_u^2\sigma_x^2}{\sigma_u^2+\sigma_x^2} \end{pmatrix}.$$

We have

$$\Sigma_L = \left(\mathbf{E}p\rho_w^T\right)^{-1}\left(\mathbf{E}vpp^T + \mathbf{Cov}(Mp - Q\beta)\right)\left(\mathbf{E}p\rho_w^T\right)^{-T},$$

where

$$M = \mathbf{E}(y|w) = \beta^T\rho_w = \beta_0 + \beta_1\mu_1 + \beta_2(\mu_1^2 + \tau^2);$$

$$v = \sigma_\varepsilon^2 + \mathbf{Var}(\beta_1 x + \beta_2 x^2|w) = \sigma_\varepsilon^2 + \tau^2(\beta_1^2 + 4\mu_1\beta_1\beta_2 + \beta_2^2(4\mu_1^2 + 2\tau^2)).$$

Then

$$\Sigma_L = \left(\mathbf{E}p\rho_w^T\right)^{-1}\left(\mathbf{E}\left(\sigma_\varepsilon^2 + \tau^2(\beta_1^2 + 4\mu_1\beta_1\beta_2 + \beta_2^2(4\mu_1^2 + 2\tau^2))\right)pp^T + \right.$$

$$\left. + \mathbf{Cov}(Mp - Q\beta)\right)\left(\mathbf{E}p\rho_w^T\right)^{-T},$$

and we insert the approximations

$$\tau^2 = \frac{\sigma_u^2\sigma_x^2}{\sigma_u^2+\sigma_x^2} \approx \sigma_u^2\left(1 - \frac{\sigma_u^2}{\sigma_x^2}\right) = \sigma_u^2 - \frac{\sigma_u^4}{\sigma_x^2},$$

$$\mu_1 = \mu_x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_x^2}(w-\mu) \approx w - \frac{\sigma_u^2}{\sigma_x^2}(w-\mu) + \frac{\sigma_u^4}{\sigma_x^4}(w-\mu).$$

Hereafter instead of equality $A = B + O(\sigma_u^6)$, as $\sigma_u^2 \to 0$ we write $A \approx B$.

Using these expansions we write the conditional variance in the form

$$v \approx \sigma_\varepsilon^2 + \sigma_u^2(\beta_1^2 + 4\beta_1\beta_2 w + 4\beta_2^2 w^2) + \frac{\sigma_u^4}{\sigma_x^2}(2\beta_2^2\sigma_x^2 - \beta_1^2 - 4\beta_1\beta_2 w - 4\beta_2^2 w^2). \tag{13}$$

So, the next form of ACM for S_L holds true

$$\Sigma_L \approx \left(\mathbf{E}p\rho_w^T\right)^{-1}\mathbf{E}\left(\sigma_\varepsilon^2 + \sigma_u^2(\beta_1^2 + 4\beta_1\beta_2 w + 4\beta_2^2 w^2) + \right.$$

$$\left. + \frac{\sigma_u^4}{\sigma_x^2}(2\beta_2^2\sigma_x^2 - \beta_1^2 - 4\beta_1\beta_2 w - 4\beta_2^2 w^2)\right)pp^T + \mathbf{Cov}(Mp - Q\beta)\left(\mathbf{E}p\rho_w^T\right)^{-T}. \tag{14}$$

We will compute Σ_{rc} and $\Sigma_{als} = \Sigma_{rc}(0)$. We have

$$p_{als}(w) = \begin{pmatrix} 1 \\ w \\ w^2 - \sigma_u^2 \end{pmatrix}, \; p = p_{rc}(w) = \begin{pmatrix} 1 \\ w \\ w^2 - \sigma_u^2 + \delta(w^3 - 3w\sigma_u^2) \end{pmatrix}.$$

Next,

$$\mathbf{E}p\rho_w^T = \mathbf{E}\begin{pmatrix} 1 & x & x^2 \\ x & x^2 & x^3 \\ x^2 + \delta x^3 & x^3 + \delta x^4 & x^4 + \delta x^5 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & \mu & \mu^2 + \sigma_x^2 \\ \mu & \mu^2 + \sigma_x^2 & \mu^3 + 3\mu\sigma_x^2 \\ \mu^2 + & \mu^3 + 3\mu\sigma_x^2 + & \mu^4 + 6\mu^2\sigma_x^2 + 3\sigma_x^4 + \\ +\delta\left(\mu^3 + 3\mu\sigma_x^2\right) & +\delta\left(\mu^4 + 6\mu^2\sigma_x^2 + 3\sigma_x^4\right) & +\delta\left(\mu^5 + 10\mu^3\sigma_x^2 + 15\sigma_x^4\right) \end{pmatrix}.$$

For fixed $\delta \in \mathbb{R}$, $\det \mathbf{E}p\rho_w^T \neq 0$ for almost all parameters $(\mu, \sigma_x^2)^T$ w.r.t. Lebesgue measure on $\mathbb{R}^2$,

$$det(\mathbf{E}p\rho_w^T) = 2\sigma_x^6 + \delta\mu\sigma_x^4(6\sigma_x^2 - 10\mu^2). \tag{15}$$

4.2. **Proof of Theorem 2.1.** We have proved that the determinant of matrix $\mathbf{E}p\rho_w^T\,|_{\delta=0}$ doesn't equal to zero. So we want to show that

$$(det\Sigma_{rc})'(0) = \left(det\left(\mathbf{E}p\rho_w^T\right)\right)|_{\delta=0}{}^{-3}\cdot\Big(\left(det(\mathbf{E}vpp^T+\mathbf{Cov}(Mp-Q\beta))\right)'_\delta\left(det\mathbf{E}p\rho_w^T\right)-$$

$$-2\left(det\mathbf{E}p\rho_w^T\right)'_\delta\left(det(\mathbf{E}vpp^T+\mathbf{Cov}(Mp-Q\beta))\right)\Big)|_{\delta=0}\neq 0$$

for almost all parameters.

We will prove this selecting from the matrix $\mathbf{E}vpp^T+\mathbf{Cov}(Mp-Q\beta)$ elements, that have summand of zero order in β_2 and μ. Here we use analytic functions theory. In fact, the unique decomposition for zero function has not summands with zero order of components.

From matrix $\mathbf{E}vpp^T+\mathbf{Cov}(Mp-Q\beta)$ we select the matrix A that consists of summands of the entries that do not have multipliers β_2 and μ, that is of zero order in β_2 and μ:

$$\begin{pmatrix} \sigma_\varepsilon^2+\beta_1^2\sigma_u^2-\frac{\beta_1^2\sigma_u^4}{\sigma_x^2} & 0 & \sigma_\varepsilon^2\sigma_x^2+\beta_1^2\sigma_u^2\sigma_x^2-\beta_1^2\sigma_u^4 \\ 0 & (\sigma_\varepsilon^2+\beta_1^2\sigma_u^2)(\sigma_u^2+\sigma_x^2) & 3\delta\sigma_x^2(\sigma_\varepsilon^2+\beta_1^2\sigma_u^2)(\sigma_u^2+\sigma_x^2) \\ \sigma_\varepsilon^2\sigma_x^2+\beta_1^2\sigma_u^2\sigma_x^2-\beta_1^2\sigma_u^4 & 3\delta\sigma_x^2(\sigma_\varepsilon^2+\beta_1^2\sigma_u^2)(\sigma_u^2+\sigma_x^2) & \begin{matrix}2\sigma_\varepsilon^2\sigma_u^4+4\sigma_\varepsilon^2\sigma_u^2\sigma_x^2+3\sigma_\varepsilon^2\sigma_x^4+\\+4\beta_1^2\sigma_u^2\sigma_x^2(\sigma_u^2+\sigma_x^2)\end{matrix} \end{pmatrix}$$

The derivative of determinant $(det\Sigma_{rc})'(0)$ is an analytic function of all parameters $(\beta^T,\mu,\sigma_x^2,\sigma_\varepsilon^2)^T$. Then

$$(det\Sigma_{rc})'(0)\approx\frac{\det(\mathbf{E}p\rho_w^T)\cdot(\det A)'-2(\det(\mathbf{E}p\rho_w^T))'\cdot\det A}{\left(det\left(\mathbf{E}p\rho_w^T\right)\right)^3}\,|_{\delta=0}\,.$$

Denominator of the latter fraction does not equal to zero, for almost all parameters $(\beta^T,\mu,\sigma_x^2,\sigma_\varepsilon^2)^T$ w.r.t. Lebesgue measure on $\mathbb{R}^6$; it equals $4\mu\sigma_x^4(5\mu^2-3\sigma_x^2)\cdot\det A\,|_{\delta=0}$, because $(\det A)'\,|_{\delta=0}$. Here

$$\det A\,|_{\delta=0}=(\sigma_\varepsilon^2+\beta_1^2\sigma_u^2)(\sigma_u^2+\sigma_x^2)\left(\left(\sigma_\varepsilon^2+\beta_1^2\sigma_u^2-\frac{\beta_1^2\sigma_u^4}{\sigma_x^2}\right)\left(2\sigma_\varepsilon^2\sigma_u^4+4\sigma_\varepsilon^2\sigma_u^2\sigma_x^2+\right.\right.$$

$$\left.\left.+3\sigma_\varepsilon^2\sigma_x^4+4\beta_1^2\sigma_u^2\sigma_x^2(\sigma_u^2+\sigma_x^2)\right)-\left(\sigma_\varepsilon^2\sigma_x^2+\beta_1^2\sigma_u^2\sigma_x^2-\beta_1^2\sigma_u^4\right)^2\right).$$

It does not equal to zero, for almost all parameters $(\beta^T,\mu,\sigma_x^2,\sigma_\varepsilon^2)^T$ w.r.t. Lebesgue measure on $\mathbb{R}^6$. Thus, the fraction does not equal to zero as well.