On synthetic median estimators in small area estimation

Tomasz Stachurski

University of Economics in Katowice, Poland
Faculty of Management
Department of Statistics, Econometrics & Mathematics

5th-7th of June 2019

Table of contents

Introduction

Aims of the study:

- Estimation of the population or domain median with higher accuracy using auxiliary information in the estimation process;
- Examination of the bias and MSE of synthetic median estimators based on simulation analyses.

The importance of median estimation

- In the case of skewed variable's distribution, median is far more appropriate measure of location than mean;
- Since 80's median is used by the European Commission (Commission of the European Communities) in comparative studies among European countries;
- Eurostat recommends to set a poverty line as a 60% of the national median equivalised disposable income;
- Currently percentage of median equivalised disposable income in EU-SILC methodology is used to estimate the number of people being at the risk of poverty.

Introduction

The importance of median estimation

- In the case of skewed variable's distribution, median is far more appropriate measure of location than mean;
- Since 80's median is used by the European Commission (Commission of the European Communities) in comparative studies among European countries;
- Eurostat recommends to set a poverty line as a 60% of the national median equivalised disposable income;
- Currently percentage of median equivalised disposable income in EU-SILC methodology is used to estimate the number of people being at the risk of poverty.

The importance of median estimation

- In the case of skewed variable's distribution, median is far more appropriate measure of location than mean;
- Since 80's median is used by the European Commission (Commission of the European Communities) in comparative studies among European countries;
- Eurostat recommends to set a poverty line as a 60% of the national median equivalised disposable income;
- Currently percentage of median equivalised disposable income in EU-SILC methodology is used to estimate the number of people being at the risk of poverty.

Introduction

The importance of median estimation

- In the case of skewed variable's distribution, median is far more appropriate measure of location than mean;
- Since 80's median is used by the European Commission (Commission of the European Communities) in comparative studies among European countries;
- Eurostat recommends to set a poverty line as a 60% of the national median equivalised disposable income;
- Currently percentage of median equivalised disposable income in EU-SILC methodology is used to estimate the number of people being at the risk of poverty.

Auxiliary information

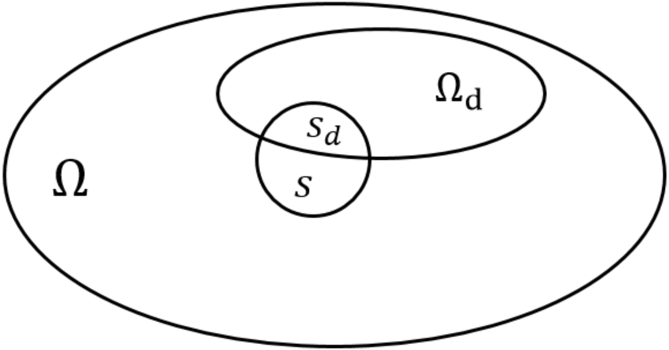
Auxiliary information could be variable that is correlated to the variable under study. The values of the auxiliary variable must be known for each sampling unit and value of the population parameter e.g. population median or total must be known as well. Examples:

- earlier census, registers;
- results from previous surveys;
- other characteristics on which it is easy to get information.

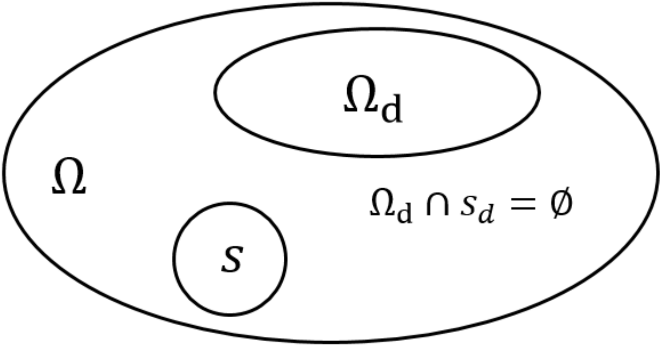
Basic notations

- Ω - the population of size N,
- Ω_d - the dth subpopulation (domain) of size N_d, where $d = 1, ..., D$,
- s - the sample of size n,
- s_d - the sample in the dth domain of size n_d,
- y_i - value of the study variable for the ith element of the population,
- x_i - value of the auxiliary variable for the ith element of the population,
- $\hat{M}_Y$ - estimator of the median of Y variable,
- M_X - known value of the median of X variable.

Small area estimation



Small area estimation



Median estimation

Särndal, Swensson and Wretman (1992), p. 200, present how to estimate population median under any sampling design using the following formula:

$$\hat{M}_Y = \begin{cases} y_l & if \;\; B_{l-1} < 0,5\hat{N} < B_l \\ 0,5(y_l + y_{l+1}) & if \;\; B_l = 0,5\hat{N} \end{cases} \tag{1}$$

where:

$B_l = \sum_{i=1}^{l} \frac{1}{\pi_i}$,

$\hat{N} = \sum_{i=1}^{n} \frac{1}{\pi_i}$ is an estimator of N.

Ratio median estimator

A ratio median estimator was firstly proposed for simple random sampling by Kuk and Mak (1989) p. 262.

A direct ratio estimator of a domain median is given by the following formula (Thompson and Godambe (2010) p. 86):

$$\hat{M}^R_{Y_d} = \frac{\hat{M}_{Y_d}}{\hat{M}_{X_d}} M_{X_d} \tag{2}$$

where $\hat{M}_{Y_d}$, $\hat{M}_{X_d}$ are obtained using (1) and M_{X_d} is the known median of auxiliary variable X in dth domain.

Remarks:

- Estimator (2) cannot be used if subpopulation sample size equals zero.

The synthetic ratio median estimator

Stachurski (2018) proposed **the synthetic ratio estimator of a domain median** given by formula:

$$\hat{M}_{Y_d}^{SR} = \frac{M_{X_d}}{\hat{M}_X}\hat{M}_Y = \frac{M_{X_d}}{M_X}\hat{M}_Y^R \tag{3}$$

where $\hat{M}_X$, $\hat{M}_Y$ are obtained using (1) and M_X, M_{X_d} are the known median of auxiliary variable X in the population and dth domain, respectively.

Remarks:

- Estimator (3) is an indirect estimator. It uses information about the study variable also from other domains (borrowing strength);
- It can be used even though the domain sample size equals 0.

Synthetic ratio median estimator - the bias

The bias of **the synthetic ratio median estimator of a domain median** is given by following formula:

$$B\left(\hat{M}^{SR}_{Y_d}\right) = \frac{M_{X_d}}{M_X} B\left(\hat{M}^{R}_{Y}\right) + M_{X_d}\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right) \tag{4}$$

The formula $M_{X_d}\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right)$ equals zero if the following condition is fulfilled:

$$\frac{M_{X_d}}{M_X} = \frac{M_{Y_d}}{M_Y}. \tag{5}$$

Synthetic ratio median estimator - MSE

The mean square error of **the synthetic ratio median estimator of a domain median** is given by the following formula:

$$MSE\left(\hat{M}_{Y_d}^{SR}\right) = \left(\frac{M_{X_d}}{M_X}\right)^2 MSE\left(\hat{M}_Y^R\right) + \\ +2\frac{\left(M_{X_d}\right)^2}{M_X} B\left(\hat{M}_Y^R\right)\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right) + \left(M_{X_d}\right)^2\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right)^2 \tag{6}$$

Synthetic ratio median estimator - MSE

MSE of **the synthetic ratio median estimator of a domain median** can be also decomposed on the sum of the variance $D^2\left(\hat{M}_Y^{SR}\right)=\left(\frac{M_{X_d}}{M_X}\right)^2 D^2\left(\hat{M}_Y^{R}\right)$ and squared bias given by (4):

$$\begin{aligned}
MSE\left(\hat{M}_{Y_d}^{SR}\right) &= \left(\frac{M_{X_d}}{M_X}\right)^2 D^2\left(\hat{M}_Y^{R}\right)+\left(\frac{M_{X_d}}{M_X}\right)^2 B\left(\hat{M}_Y^{R}\right)^2+ \\
&+2\frac{\left(M_{X_d}\right)^2}{M_X}B\left(\hat{M}_Y^{R}\right)\left(\frac{M_Y}{M_X}-\frac{M_{Y_d}}{M_{X_d}}\right)+\left(M_{X_d}\right)^2\left(\frac{M_Y}{M_X}-\frac{M_{Y_d}}{M_{X_d}}\right)= \\
&=\left(\frac{M_{X_d}}{M_X}\right)^2 D^2\left(\hat{M}_Y^{R}\right)+\left(\frac{M_{X_d}}{M_X}B\left(\hat{M}_Y^{R}\right)+M_{X_d}\left(\frac{M_Y}{M_X}-\frac{M_{Y_d}}{M_{X_d}}\right)\right)^2
\end{aligned} \tag{7}$$

Product median estimator

A product estimator of a population median is presented by Sharma, Singh (2015):

$$\hat{M}_Y^P = \frac{\hat{M}_Y \hat{M}_X}{M_X} \tag{8}$$

We propose **the product estimator of a domain median** given by:

$$\hat{M}_{Y_d}^P = \frac{\hat{M}_{Y_d} \hat{M}_{X_d}}{M_{X\,d}} \tag{9}$$

Remarks:

- A product estimator is used in the case of negative correlation between the study variable and the auxiliary variable.
- Estimator (9) is a direct estimator. It cannot be used if domain sample size equals zero.
- If $\hat{M}_{Y_d}$, $\hat{M}_{X_d}$ are obtained using (1), then (9) is applicable for any sampling design.

Synthetic product median estimator

We propose **the synthetic product estimator of a domain median** given by:

$$\hat{M}_{Y_d}^{SP} = \frac{\hat{M}_Y \hat{M}_X}{M_{X_d}} = \frac{\hat{M}_Y \hat{M}_X}{M_{X_d}} \frac{M_X}{M_X} = \frac{M_X}{M_{X_d}} \hat{M}_Y^P \tag{10}$$

Remarks:

- Estimator (10) is an indirect estimator. It uses information about the study variable also from other domains (borrowing strength);
- It can be used even though domain sample size equals 0.

Synthetic product median estimator - the bias

The bias of **the synthetic product median estimator of a domain median** is given by the following formula:

$$B\left(\hat{M}_{Y_d}^{SP}\right) = \frac{M_X}{M_{X_d}} B\left(\hat{M}_Y^{P}\right) + M_X \left(\frac{M_Y}{M_{X_d}} - \frac{M_{Y_d}}{M_X}\right) \tag{11}$$

The formula $M_X\left(\frac{M_Y}{M_{X_d}} - \frac{M_{Y_d}}{M_X}\right)$ equals zero if the following condition is fulfilled:

$$M_X M_Y = M_{X_d} M_{Y_d}. \tag{12}$$

Synthetic product median estimator - MSE

The mean square error of **the synthetic product median estimator of a domain median** is given by following formula:

$$MSE\left(\hat{M}^{SP}_{Y_d}\right) = \left(\frac{M_X}{M_{X_d}}\right)^2 MSE\left(\hat{M}^{P}_{Y}\right) + \\ +2\frac{(M_X)^2}{M_{X_d}} B\left(\hat{M}^{P}_{Y}\right)\left(\frac{M_Y}{M_{X_d}} - \frac{M_{Y_d}}{M_X}\right) + (M_X)^2\left(\frac{M_Y}{M_{X_d}} - \frac{M_{Y_d}}{M_X}\right)^2 \tag{13}$$

Regression median estimator

A direct **regression estimator of a subpopulation median** is given by:

$$\hat{M}_{Y_d}^{REG} = \hat{M}_{Y_d} + \hat{\beta}_d \left(M_{X_d} - \hat{M}_{X_d} \right) \tag{14}$$

where:

$$\hat{\beta}_d = \frac{\sum_{i \in s_d} \left(x_i - \bar{x}_{\Omega_d}^{HT} \right) \left(y_i - \bar{y}_{\Omega_d}^{HT} \right) \frac{1}{\pi_i}}{\sum_{i \in s_d} \left(x_i - \bar{x}_{\Omega_d}^{HT} \right)^2 \frac{1}{\pi_i}}, \tag{15}$$

$$\hat{\bar{y}}_{\Omega_d}^{HT} = \frac{1}{\hat{N}_d} \sum_{i \in s_d} \frac{y_i}{\pi_i}, \quad \hat{\bar{x}}_{\Omega_d}^{HT} = \frac{1}{\hat{N}_d} \sum_{i \in s_d} \frac{x_i}{\pi_i}.$$

Remarks:

- Estimator (14) cannot be used if subpopulation sample size equals zero.
- If we assume that $\Omega_d = \Omega$, $s_d = s$, then (14) is a regression estimator of a population median.

Synthetic regression median estimator

We propose **the synthetic regression estimator of a domain median** given by:

$$\hat{M}_{Y_d}^{SYN-REG} = \hat{M}_Y + \hat{\beta}\left(M_{X_d} - \hat{M}_X\right). \tag{16}$$

where:

$$\hat{\beta} = \frac{\sum_{i \in s}\left(x_i - \hat{\bar{x}}^{HT}\right)\left(y_i - \hat{\bar{y}}^{HT}\right)\frac{1}{\pi_i}}{\sum_{i \in s}\left(x_i - \hat{\bar{x}}^{HT}\right)^2\frac{1}{\pi_i}} \tag{17}$$

Estimator (16) can be also written using regression estimator of a population median, as follows:

$$\begin{aligned} \hat{M}_{Y_d}^{SYN-REG} &= \hat{M}_Y + \hat{\beta}\left(M_{X_d} - \hat{M}_X\right) + \hat{\beta}\left(M_X - M_X\right) = \\ &= \hat{M}_Y + \hat{\beta}\left(M_X - \hat{M}_X\right) + \hat{\beta}\left(M_{X_d} - M_X\right) = \\ &= \hat{M}_Y^{REG} + \hat{\beta}\left(M_{X_d} - M_X\right) \end{aligned} \tag{18}$$

Synthetic regression median estimator - the bias

The bias of **the synthetic regression median estimator of a domain median** is given by the following formula:

$$B\left(\hat{M}_{Y_d}^{SYN-REG}\right) = B\left(\hat{M}_{Y}^{REG}\right) - E\left(\hat{\beta}\right)\left(M_X - M_{X_d}\right) + \left(M_Y - M_{Y_d}\right) \tag{19}$$

Remarks:

- If $M_X = M_{X_d}$ and $M_Y = M_{Y_d}$ then $B\left(\hat{M}_{Y_d}^{SYN-REG}\right) = B\left(\hat{M}_{Y_d}^{REG}\right)$.

Synthetic regression median estimator - MSE

The mean square error of **the synthetic regression median estimator of a domain median** is given by following formula:

$$
\begin{aligned}
MSE\left(\hat{M}_{Y_d}^{SYN-REG}\right) = MSE\left(\hat{M}_{Y}^{REG}\right) + \left(M_Y - M_{Y_d}\right)^2 + \\
+\left(D^2\left(\hat{\beta}\right) + E^2\left(\hat{\beta}\right)\right)\left(M_X - M_{X_d}\right)^2 + \\
+2B\left(\hat{M}_{Y}^{REG}\right)\left(M_Y - M_{Y_d}\right) - 2E\left(\hat{\beta}\right)\left(M_Y - M_{Y_d}\right)\left(M_X - M_{X_d}\right) + \\
-2\left(M_X - M_{X_d}\right)\left(Cov\left(\hat{M}_{Y}^{REG}, \hat{\beta}\right) + E\left(\hat{\beta}\right)B\left(\hat{M}_{Y}^{REG}\right)\right)
\end{aligned}
\tag{20}
$$

Synthetic regression median estimator - MSE

MSE of **the synthetic regression median estimator of a domain median** can be also written alternatively as:

$$MSE\left(\hat{M}_{Y_d}^{SYN-REG}\right) = D^2\left(\hat{M}_Y^{REG}\right) + D^2\left(\hat{\beta}\right)\left(M_X - M_{X_d}\right) + \\ -2\left(M_X - M_{X_d}\right) Cov\left(\hat{M}_Y^{REG}, \hat{\beta}\right) + \\ + \left[B\left(\hat{M}_Y^{REG}\right) - E\left(\hat{\beta}\right)\left(M_X - M_{X_d}\right) + \left(M_Y - M_{Y_d}\right)\right]^2. \quad (21)$$

Remarks:

- First three elements of equation (21) form the variance of synthetic regression median estimator, what can be written alternatively as:

$$D^2\left(\hat{M}_{Y_d}^{SYN-REG}\right) = E\left[\left(\hat{M}_Y^{REG} - E\left(\hat{M}_Y^{REG}\right)\right) - \left(\hat{\beta} - E\left(\hat{\beta}\right)\right)\left(M_X - M_{X_d}\right)\right]^2. \quad (22)$$

- The fourth element of the equation (21) is squared bias of $\hat{M}_{Y_d}^{SYN-REG}$, which is given by formula (19).

Simulation study

Aim of the study is an analysis of the properties of synthetic estimators of a domain median.

Analysed dataset

- Considered dataset consists of 281 Swedish municipalities;
- The sample size is equal to $n = 56$ and it was drawn using Brewer's sampling scheme;
- The population was divided into 7 regions;
- The study variable is revenue from taxes in 1985 (RMT85) and the auxiliary variable is the number of municipal employees in 1984 (ME84).

Simulation study

In the simulation study, we drew samples $B = 10000$ times and calculated the relative bias in % as:

$$\theta_d^{-1} B^{-1} \sum_{i=1}^{B} \left(\hat{\theta}_d^b - \theta_d \right) \tag{23}$$

and the relative root mean square error as:

$$\theta_d^{-1} B^{-1} \sum_{i=1}^{B} \left(\hat{\theta}_d^b - \theta_d \right)^2. \tag{24}$$

Results - the synthetic ratio/product median estimator

Table 1. Results for the synthetic ratio estimator of a domain median

Domain ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Relative bias (%)	-0.44	-8.75	-1.00	0.33	-3.81	-2.73	22.65
Relative RMSE (%)	6.61	10.64	6.63	6.65	7.42	7.00	24.07
$\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right)$	0.005	-0.008	0.004	0.006	0.000	0.001	0.031

Table 2. Results for the synthetic product estimator of a domain median

Domain ID	1	2	3	4	5	6	7
Relative bias (%)	-82.26	14.09	21.24	1.13	81.38	45.27	185.77
Relative RMSE (%)	82.66	55.35	60.72	47.47	117.72	81.82	229.04
$\left(\frac{M_Y}{M_{X_d}} - \frac{M_{Y_d}}{M_X}\right)$	-0.305	-0.001	0.008	-0.018	0.067	0.034	0.121

Results - the synthetic regression median estimator

Table 3. Results for the synthetic regression estimator of a domain median

Domain ID	1	2	3	4	5	6	7
Relative bias (%)	-0.02	-9.40	-1.68	-0.19	-4.89	-3.58	20.99
Relative RMSE (%)	2.81	11.59	7.48	6.70	10.06	8.68	24.41
$M_X - M_{X_d}$	-1240	19.0	11.0	-68.0	161.0	84.0	223.5
$M_Y - M_{Y_d}$	-169.5	-3.5	4.5	-5.0	23.0	13.0	49.5

Conclusion

- The bias/MSE of the synthetic median estimators can be decomposed on a sum of the bias/MSE of a direct estimator of a population median and a function of medians of both variables in small area and in the whole population.
- If conditions of relationship between the medians of both variables are not fulfilled, then higher values of the bias or MSE are obtained.
- If the auxiliary variable is positively correlated with the study variable, then both the synthetic ratio estimator and the synthetic regression estimator of a domain median have good properties.
- For the synthetic product estimator of a domain median, in the case of positive correlation, we obtained high values of the relative bias and the relative RMSE.

Conclusion

- The bias/MSE of the synthetic median estimators can be decomposed on a sum of the bias/MSE of a direct estimator of a population median and a function of medians of both variables in small area and in the whole population.
- If conditions of relationship between the medians of both variables are not fulfilled, then higher values of the bias or MSE are obtained.
- If the auxiliary variable is positively correlated with the study variable, then both the synthetic ratio estimator and the synthetic regression estimator of a domain median have good properties.
- For the synthetic product estimator of a domain median, in the case of positive correlation, we obtained high values of the relative bias and the relative RMSE.

Conclusion

- The bias/MSE of the synthetic median estimators can be decomposed on a sum of the bias/MSE of a direct estimator of a population median and a function of medians of both variables in small area and in the whole population.
- If conditions of relationship between the medians of both variables are not fulfilled, then higher values of the bias or MSE are obtained.
- If the auxiliary variable is positively correlated with the study variable, then both the synthetic ratio estimator and the synthetic regression estimator of a domain median have good properties.
- For the synthetic product estimator of a domain median, in the case of positive correlation, we obtained high values of the relative bias and the relative RMSE.

Conclusion

- The bias/MSE of the synthetic median estimators can be decomposed on a sum of the bias/MSE of a direct estimator of a population median and a function of medians of both variables in small area and in the whole population.
- If conditions of relationship between the medians of both variables are not fulfilled, then higher values of the bias or MSE are obtained.
- If the auxiliary variable is positively correlated with the study variable, then both the synthetic ratio estimator and the synthetic regression estimator of a domain median have good properties.
- For the synthetic product estimator of a domain median, in the case of positive correlation, we obtained high values of the relative bias and the relative RMSE.

Bibliography I

Kuk Y.C.A., & Mak T.K. (1989) *Median estimation in the Presence of Auxiliary Information*, "Journal of the Royal Statistical Society. Series B (Methodological)", Vol. 51, No. 2, pp. 261-269.

Murthy M.N. (1964) *Product methods of Estimation*, Sankhya: The Indian Journal od Statistics, Series A", Vol. 26, No. 1, pp. 69-74.

Rao J.N.K., & Molina I.: *Small Area Estimation*. John Wiley and Sons, New York, 2015.

Särndal, C.E., Swensson, B. & Wretman J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Sharma P., & Singh R. (2015), *Generalized Class of Estimators for Population Median Using Auxiliary Information*, "Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics", 44(2), pp. 443 – 453.

Stachurski T. (2018) *A simulation analysis of the accuracy of median estimators for different sampling designs*, pp. 509-514, [in]: L. Váchová, V. Kratochvíl (eds.) "Proceedings of the 36th International Conference Mathematical Methods in Economics MME 2018", Praha, Czech Republic 2018, MatfyzPress, Publishing House of the Faculty of Mathematics and Physics Charles University.

Bibliography II

Thompson M.E., and Godambe V.P.: Estimating Functions and Survey Sampling. In: *Handbook of Statistics. Sample Surveys: Inference and Analysis*, Vol. 29B (Pfeffermann D., and Rao C.R., eds.), Elsevier Sci-ence, Oxford, 2009.

Wywiał J. & Żądło T. (2003) *On mean square error of synthetic ratio estimator*, „Studia Ekonomiczne", No. 29, Katowice, pp. 167-175.

Żądło T. (2004) *On mean square error of synthetic regression estimator*, „Acta Universitatis Lodziensis, Folia Oeconomica", Vol. 105, pp. 73-90.

Żądło T. (2008), *Elementy statystyki małych obszarów z programem R*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice.

Thank You for Your attention!

Appendix 1 - Synthetic ratio median estimator - the bias

$$B\left(\hat{M}^{SR}_{Y_d}\right) = E\left(\hat{M}^{SR}_{Y_d} - M_{Y_d}\right) = E\left(\hat{M}^{SR}_{Y_d}\right) - M_{Y_d} = E\left(\frac{M_{X_d}}{M_X}\hat{M}^{R}_{Y}\right) - M_{Y_d} =$$

$$= E\left(\frac{M_{X_d}}{M_X} \cdot \left(\hat{M}^{R}_{Y} + M_Y - M_Y\right)\right) - M_{Y_d} =$$

$$= E\left(\frac{M_{X_d}}{M_X}\left(\hat{M}^{R}_{Y} - M_Y\right) + \frac{M_{X_d}}{M_X} M_Y\right) - M_{Y_d} =$$

$$= \frac{M_{X_d}}{M_X} E\left(\hat{M}^{R}_{Y} - M_Y\right) + \frac{M_{X_d}}{M_X} M_Y - M_{Y_d} = \frac{M_{X_d}}{M_X} B\left(\hat{M}^{R}_{Y}\right) + \frac{M_{X_d}}{M_X} M_Y - M_{Y_d} =$$

$$= \frac{M_{X_d}}{M_X} B\left(\hat{M}^{R}_{Y}\right) + M_{X_d}\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right)$$

Appendix 2 - Synthetic ratio median estimator - MSE

$$MSE\left(\hat{M}^{SR}_{Y_d}\right) = E\left(\hat{M}^{SR}_{Y_d} - M_{Y_d}\right)^2 = E\left(\frac{M_{X_d}}{M_X}\hat{M}^{R}_{Y} - M_{Y_d}\right)^2 =$$

$$= E\left(\frac{M_{X_d}}{M_X}\left(\hat{M}^{R}_{Y} + M_Y - M_Y\right) - M_{Y_d}\right)^2 =$$

$$= E\left(\frac{M_{X_d}}{M_X}\left(\hat{M}^{R}_{Y} - M_Y\right) + \frac{M_{X_d}}{M_X}M_Y - M_{Y_d}\right)^2 =$$

$$= E\left[\left(\frac{M_{X_d}}{M_X}\right)^2\left(\hat{M}^{R}_{Y} - M_Y\right)^2 + 2\frac{M_{X_d}}{M_X}\left(\hat{M}^{R}_{Y} - M_Y\right)\left(\frac{M_{X_d}}{M_X}M_Y - M_{Y_d}\right)\right] +$$
$$+E\left(\left(\frac{M_{X_d}}{M_X}M_Y - M_{Y_d}\right)^2\right) =$$

Appendix 2 - Synthetic ratio median estimator - MSE

$$
\begin{aligned}
&= \left(\frac{M_{X_d}}{M_X}\right)^2 MSE\left(\hat{M}_Y^R\right) + 2\frac{M_{X_d}}{M_X} B\left(\hat{M}_Y^R\right)\left(\frac{M_{X_d}}{M_X} M_Y - M_{Y_d}\right) + \\
&+ \left(\frac{M_{X_d}}{M_X} M_Y - M_{Y_d}\right)^2 = \\
&= \left(\frac{M_{X_d}}{M_X}\right)^2 MSE\left(\hat{M}_Y^R\right) + 2\frac{{M_{X_d}}^2}{M_X} B\left(\hat{M}_Y^R\right)\left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right) + \\
&+ {M_{X_d}}^2 \left(\frac{M_Y}{M_X} - \frac{M_{Y_d}}{M_{X_d}}\right)^2
\end{aligned}
$$

Appendix 3 - Synthetic product median estimator - the bias

$$
\begin{aligned}
B\left(\hat{M}_{Y_d}^{SP}\right) = E\left(\hat{M}_{Y_d}^{SP} - M_{Y_d}\right) = E\left(\frac{M_X}{M_{X_d}}\hat{M}_Y^P - M_{Y_d}\right) = \\
= E\left(\frac{M_X}{M_{X_d}}\left(\hat{M}_Y^P - M_Y + M_Y\right)\right) - M_{Y_d} = \\
= \frac{M_X}{M_{X_d}}B\left(\hat{M}_Y^P\right) + \frac{M_X}{M_{X_d}}M_Y - M_{Y_d} = \\
= \frac{M_X}{M_{X_d}}B\left(\hat{M}_Y^P\right) + M_X\left(\frac{M_Y}{M_{X_d}} - \frac{M_{Y_d}}{M_X}\right)
\end{aligned}
\tag{25}
$$

Appendix 4 - Synthetic regression median estimator - the bias

$$
\begin{aligned}
B\left(\hat{M}_{Y_d}^{SYN-REG}\right) = E\left(\hat{M}_{Y_d}^{SYN-REG} - M_{Y_d}\right) = \\
= E\left(\hat{M}_{Y}^{REG} + \hat{\beta}\left(M_{X_d} - M_X\right) - M_{Y_d}\right) = \\
= E\left[\hat{M}_{Y}^{REG} + \hat{\beta}\left(M_{X_d} - M_X\right)\right] - M_{Y_d} = \\
= E\left[\hat{M}_{Y}^{REG} - M_Y + M_Y + \hat{\beta}\left(M_{X_d} - M_X\right)\right] - M_{Y_d} = \\
= E\left(\hat{M}_{Y}^{REG} - M_Y\right) + E\left(\hat{\beta}\right)\left(M_{X_d} - M_X\right) + \left(M_Y - M_{Y_d}\right) = \\
= B\left(\hat{M}_{Y}^{REG}\right) + E\left(\hat{\beta}\right)\left(M_{X_d} - M_X\right) + \left(M_Y - M_{Y_d}\right) = \\
= B\left(\hat{M}_{Y}^{REG}\right) - E\left(\hat{\beta}\right)\left(M_X - M_{X_d}\right) + \left(M_Y - M_{Y_d}\right)
\end{aligned}
\tag{26}
$$

Appendix 5 - Synthetic regression median estimator - MSE

$$
\begin{aligned}
MSE\left(\hat{M}_{Y_d}^{SYN-REG}\right) &= E\left(\hat{M}_{Y_d}^{SYN-REG} - M_{Y_d}\right)^2 = \\
&= E\left(\hat{M}_{Y}^{REG} + \hat{\beta}\left(M_{X_d} - M_X\right) - M_{Y_d}\right)^2 = \\
&= E\left(\hat{M}_{Y}^{REG} - M_Y + M_Y + \hat{\beta}\left(M_{X_d} - M_X\right) - M_{Y_d}\right)^2 = \\
&= E\left[\left(\hat{M}_{Y}^{REG} - M_Y\right) + \left(M_Y - M_{Y_d}\right) - \left(\hat{\beta} - E\left(\hat{\beta}\right) + E\left(\hat{\beta}\right)\right)\left(M_X - M_{X_d}\right)\right]^2 = \\
&= E\left[\left(\hat{M}_{Y}^{REG} - M_Y\right)^2 + \left(M_Y - M_{Y_d}\right)^2 + \hat{\beta}^2\left(M_X - M_{X_d}\right)^2\right] + \\
&\quad E\left[2\left(\hat{M}_{Y}^{REG} - M_Y\right)\left(M_Y - M_{Y_d}\right) - 2\hat{\beta}\left(M_Y - M_{Y_d}\right)\left(M_X - M_{X_d}\right)\right] + \\
&\quad -E\left[2\left(\hat{M}_{Y}^{REG} - M_Y\right)\left(M_X - M_{X_d}\right)\left(\hat{\beta} - E\left(\hat{\beta}\right) + E\left(\hat{\beta}\right)\right)\right] = \\
&= MSE\left(\hat{M}_{Y}^{REG}\right) + \left(M_Y - M_{Y_d}\right)^2 + \left(D^2\left(\hat{\beta}\right) + E^2\left(\hat{\beta}\right)\right)\left(M_X - M_{X_d}\right)^2 + \\
&\quad +2B\left(\hat{M}_{Y}^{REG}\right)\left(M_Y - M_{Y_d}\right) - 2E\left(\hat{\beta}\right)\left(M_Y - M_{Y_d}\right)\left(M_X - M_{X_d}\right) + \\
&\quad -2\left(M_X - M_{X_d}\right)\left(Cov\left(\hat{M}_{Y}^{REG}, \hat{\beta}\right) + E\left(\hat{\beta}\right)B\left(\hat{M}_{Y}^{REG}\right)\right)
\end{aligned}
\tag{27}
$$