Department
for Environment
Food & Rural Affairs

Calculating farm household income from banded data in the Farm Business Survey

Jo Hutchinson, Farm Business Survey,
Defra, England

Overview

- Introduction
- Methodology for estimating mid-point of banded data
- Evaluation of methods
- Results
- Recommendations

- Defra's Farm Household Income release methodology
- Definition of Farm Household Income

Introduction

- In 2021/22 Defra reintroduced questions on farm household income into the FBS to a sub-sample of farms (940 farms, 59% of full survey sample).

- Last collected in 2014/15

- Questions asked to farmer:
  1. Share of drawings
  2. Banded off-farm income for farmer and spouse

- 3 measures available for analysis:
  1. Off-farm income
  2. Household Share of Drawings
  3. Household Share of Farm Business Income

			Structure of Household								
			CODE for Farmer & Spouse Columns 10 & 11		NUMBER of Persons Columns 12, 13 and 14					Sources of	
					Other Adults in Household			Drawings as Proportion of Total	Other	Other	Household
Household Code	Dwelling Code	Status of Household Code	Code for Farmer	Code for Spouse (even if not doing farm work)	Below State Pension Age	Of State Pension Age	Children	Drawings at FAS 24, G 73 (rounded to nearest 5%)	Household Income Code	Household Income Code	Reliance on off-farm income
1	2	3	10	11	12	13	14	7	8	15	16

Introduction

Household Off-Farm Income

- In the FBS, Household Off-Farm Income is collected at a gross level as banded data.

- Previous publications used the midpoint of each band, which isn't necessarily the true midpoint.

- Carried out literacy review and found methods that better estimate the mid point with different methods developed for the open top band.

- To estimate the true value from banded data for each farm, two methods were tested and then used:

- Mean Constrained Integration over Brackets (MCIB)

- Robust - Pareto Midpoint Estimator (RPME).

Other Household Income Codes (column 8)
for negative income, the corresponding negative c

Zero		
£1	to below	£1,000
£1,000	to below	£2,500
£2,500	to below	£5,000
£5,000	to below	£7,500
£7,500	to below	£10,000
£10,000	to below	£15,000
£15,000	to below	£20,000
£20,000	to below	£25,000
£25,000	to below	£30,000
£30,000	to below	£40,000
£40,000	to below	£50,000
£50,000	to below	£75,000
£75,000	to below	£100,000
£100,000	to below	£150,000
£150,000	to below	£200,000
£200,000	or more	

Methodology

Mean Constrained Integration over Brackets (MCIB)

- MCIB was developed by Paul Jargowsky and Christopher Wheeler (2018) and has been used particularly in cases of high income inequality

- The method was used to estimate the density functions of the closed-end brackets by assuming that the density of households within each bracket could be described by a linear function.

- By estimating these density functions, the MCIB methodology provided a more accurate representation of the income distribution.

- The density of each bracket represented its contribution to the total probability function.

- The mean of each band was calculated using integration.

- This mean was then applied as the Household Off-Farm Income figure for the corresponding bands.

Methodology

Mean Constrained Integration over Brackets (MCIB)

- The MCIB estimation process involves three steps.

***Step 1:* Estimate the density functions of the closed-end brackets.**

- The MCIB method estimates the density functions of the closed-end brackets by assuming that the density of households within each bracket can be described by a linear function.

$$f_b(y) = \frac{m_b y + c_b}{N},$$

where m_b is the slope and c_b is the constant of the line that describes the relative frequency of households in bracket b.

- The slopes and intercepts for the brackets are calculated using the number of households in each bracket and the width of the bracket.
- The slopes are calculated as the average of the slopes from neighbouring brackets.
- The constants are calculated to force the line of slope m through the relative frequency point.

Methodology

Mean Constrained Integration over Brackets (MCIB)
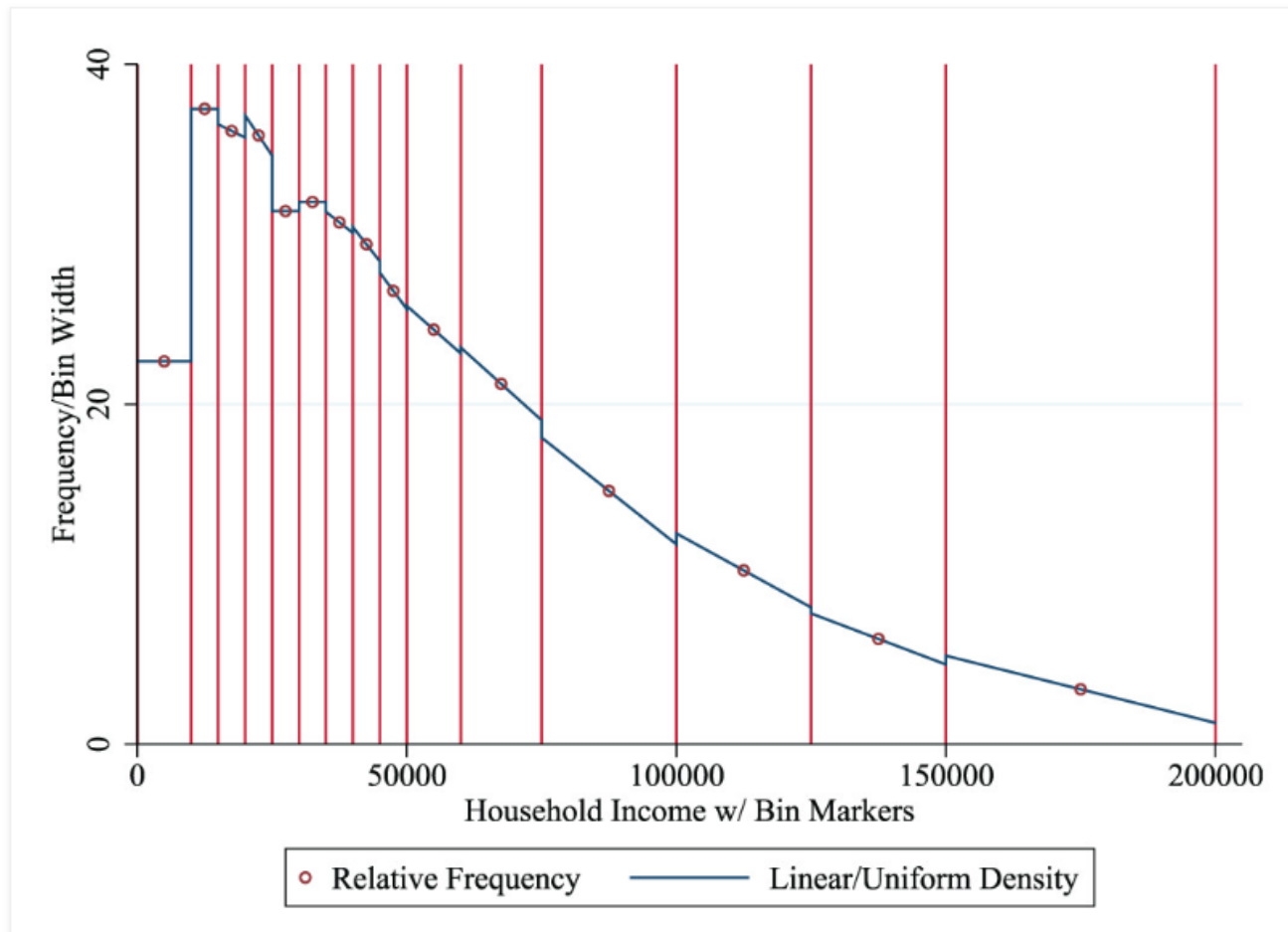
Methodology

Mean Constrained Integration over Brackets (MCIB)

Step 2: Estimate the mean for the open-ended bracket.

- The grand mean, G, is related to the bracket means through the following identity

$$G = \frac{1}{N} \sum_{b=1}^{B} n_b \mu_b.$$

- Removing the top bracket from the summation and rearranging we get

$$\mu_B = \left(NG - \sum_{b=1}^{B-1} n_b \mu_b \right) / n_B$$

- The mean of the top bracket is calculated by taking the total aggregate income, subtracting the income in all the brackets below the top bracket, and dividing by the number of households in the top bracket. By using the rearranged formula the top bracket mean is constrained by the grand mean providing an upper limit.

Methodology

Mean Constrained Integration over Brackets (MCIB)

Step 3: Estimate the variance (and other parameters).

The formula for the variance of the income distribution on the basis of individual households is additively separable by income brackets:

$$\sigma^2 = \frac{1}{N} \sum_{b=1}^{B} \sum_{i=1}^{n_b} (y_i - G)^2 \approx \sum_{b=1}^{B} \int_{L_b}^{U_b} (y_i - G)^2 f_b(y) \; dy.$$

In other words, MCIB estimates the contributions to the variance separately in each bracket and sums them up to obtain the estimate of the variance
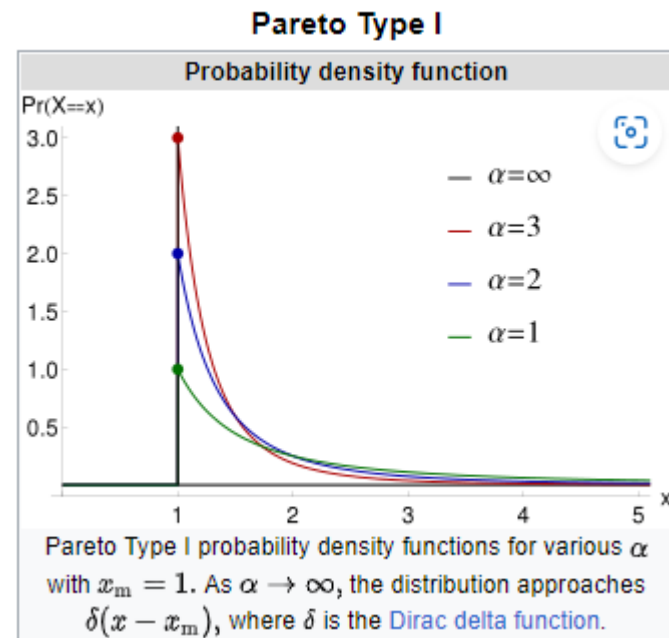
$$
\begin{aligned}
\int_{L_b}^{U_b} (y - G)^2 f_b(y) \; dy &= \int_{L_b}^{U_b} (y^2 - 2Gy - G^2) \left(\frac{m_b y + c_b}{N} \right) dy \\
&= \frac{1}{N} \int_{L_b}^{U_b} (m_b y^3 + c_b y^2 - 2G m_b y^2 - 2G c_b y + G^2 c_b) \; dy \\
&= \frac{1}{N} \int_{L_b}^{U_b} (m_b y^3 + (c_b - 2G m_b) y^2 + (G^2 m_b - 2G c_b) y + G^2 c_b) \; dy \\
&= \frac{1}{N} \left(\frac{m_b}{4} y^4 + \frac{(c_b - 2G m_b)}{3} y^3 + \frac{(G^2 m_b - 2G c_b)}{2} y^2 + G^2 c_b y \right) \Bigg|_{L_b}^{U_b} .
\end{aligned}
$$

Methodology

Two-point Pareto Estimator

- If the highest income bracket follows a pareto distribution, its average income can be estimated using the shape parameter (α).
    - This is done by assuming that the entire population in the top bracket has the same mean income as the average income of the top bracket.

- May not be reliable in small samples, the accuracy of the estimated top bracket mean depends on the estimate of α.

Pareto Type I

Probability density function

Pareto Type I probability density functions for various α with $x_m = 1$. As $\alpha \to \infty$, the distribution approaches $\delta(x - x_m)$, where δ is the Dirac delta function.

Methodology

Robust Pareto Midpoint Estimator (RPME)

- RPME was developed by Paul von Hippel et al (2014).

- It is designed for estimating income in the open top bracket, a common issue in banded data, and to address issues with the two-point estimator.

- RPME uses a Pareto distribution assumption to model the upper tail of the income distribution and estimates the parameters of the Pareto distribution using robust methods to account for potential outliers or skewness in the data.

    - Robust methods include
        - Median
        - Geometric mean
        - Harmonic mean

Evaluation of methods

- For all our analysis we used Farm Business Income (FBI) as true values are known to test the methods. Used data from 2004 – 2021 (18 years)

- Estimated the mean in each of the 14 off farm income bands

- We used several statistical measures to evaluate the performance of each prediction method

  - **Bias**: measures the average difference between the predicted values and the true values. A lower bias indicates that the predictions are closer to the true values on average.

  - **RMSE** (Root Mean Squared Error): measures the average magnitude of the errors in the predictions. It considers both the bias and the variance of the predictions. A lower RMSE indicates that the predictions are closer to the true values, on average.

Evaluation

- **RMSPE** (Root Mean Squared Percentage Error): measures the average percentage difference between the predicted values and the true values. It is like RMSE, but it considers the relative size of the errors. A lower RMSPE indicates that the predictions are more accurate, on average.

- **MAE** (Mean Absolute Error): measures the average absolute difference between the predicted values and the true values. It is a simpler measure than RMSE, but it doesn't penalize large errors as much. A lower MAE indicates that the predictions are closer to the true values, on average.

- **MAPE** (Mean Absolute Percentage Error): measures the average percentage difference between the predicted values and the true values. It is like MAE, but it considers the relative size of the errors. A lower MAPE indicates that the predictions are more accurate, on average.

- **MAD** (Mean Absolute Deviation): measures the mean absolute difference between the predicted values and the true values. A lower MAD indicates that the predictions are closer to the true values, on average.
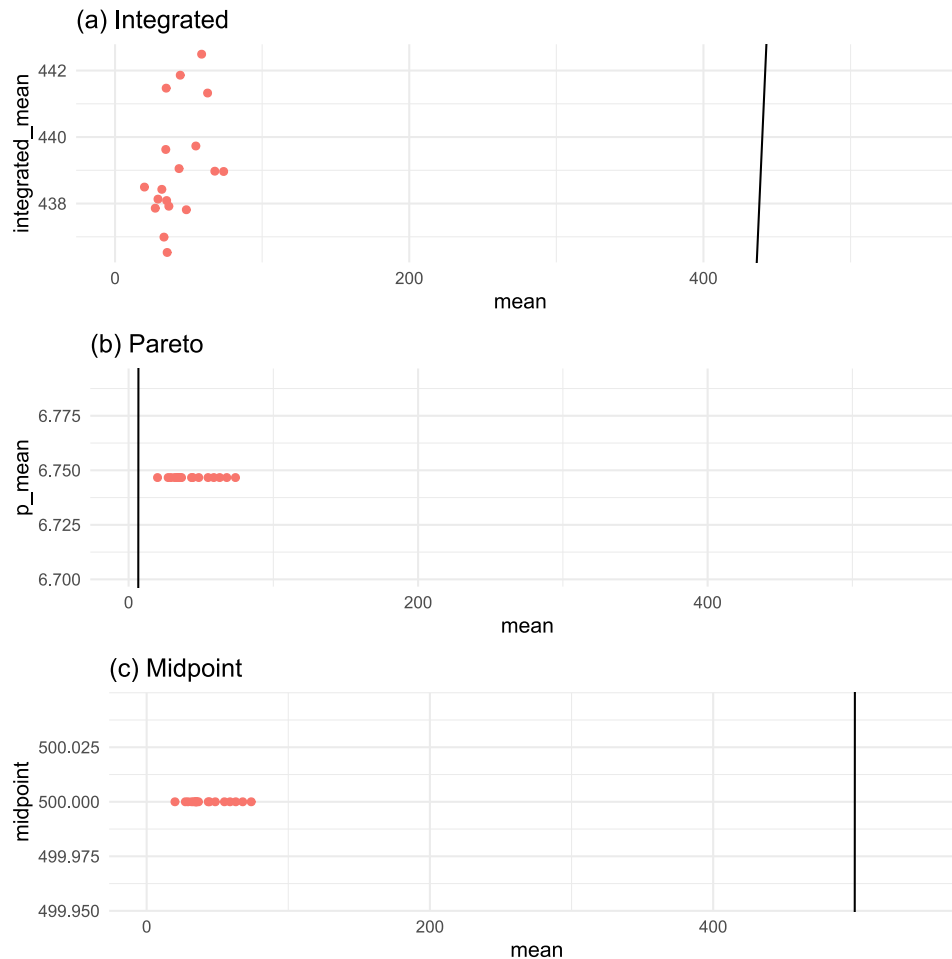
Results

Evaluation of estimators – for 18 years of FBI data

		Bias	RMSE	RMSPE	MAE	MAPE	MAD
1st band	Midpoint	457	457	1,283.69	457	1,205.06	0
	Integration	396	396	1,114.73	396	1,045.50	1
	Pareto (α = 1)	-36	39	82.61	36	82.39	0
	Pareto (α = 2)	-41	43	94.83	41	94.81	0
Below Median	Midpoint	80	217	468.74	167	162.33	6,968
	Integration	51	206	407.06	166	141.54	6,968
	Pareto (α = 1)	-250	299	30.83	252	15.37	6,890
	Pareto (α = 2)	-292	340	35.67	292	18.32	6,913
Median	Midpoint	25	222	0.75	159	0.58	5,617
	Integration	-9	217	0.77	163	0.61	5,639
	Pareto (α = 1)	-414	531	2.02	447	1.79	5,611
	Pareto (α = 2)	-414	531	2.02	447	1.79	5,611
Above median	Midpoint	632	925	1.49	689	1.20	19,861
	Integration	32	459	0.82	345	0.65	19,437
	Pareto (α = 1)	-692	929	1.78	734	1.48	19,097
	Pareto (α = 2)	-780	986	1.82	800	1.56	18,995

Results - Evaluation of estimators
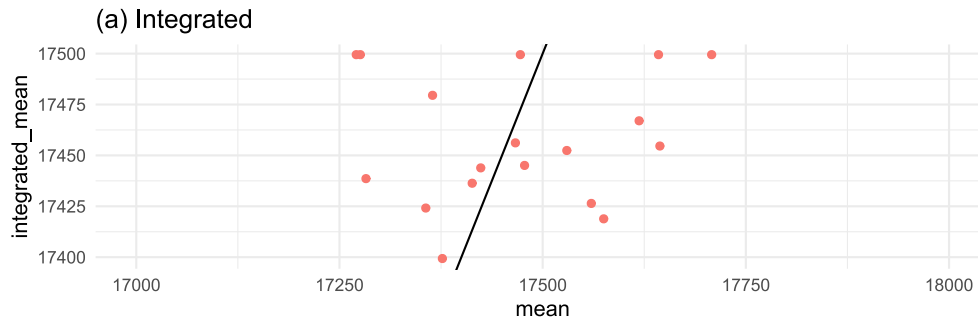
First FBI income band (£1 - £999)

(a) Integrated
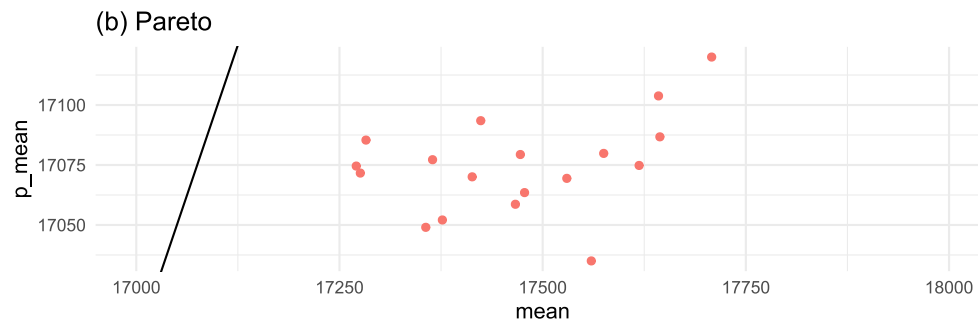
(b) Pareto

(c) Midpoint

Pareto estimate demonstrating significantly greater proximity to the true mean than the integrated and midpoint estimates
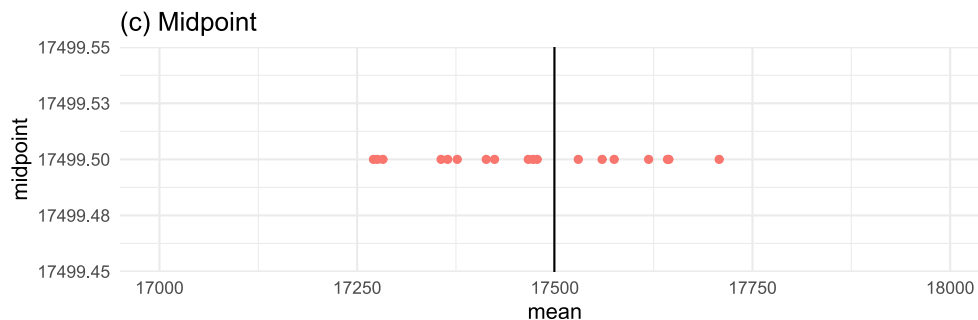
Results - Evaluation of estimators

FBI below median income band (£15,000 - £19,999)

(a) Integrated

(b) Pareto

(c) Midpoint

The Pareto estimates consistently fall below the true mean.
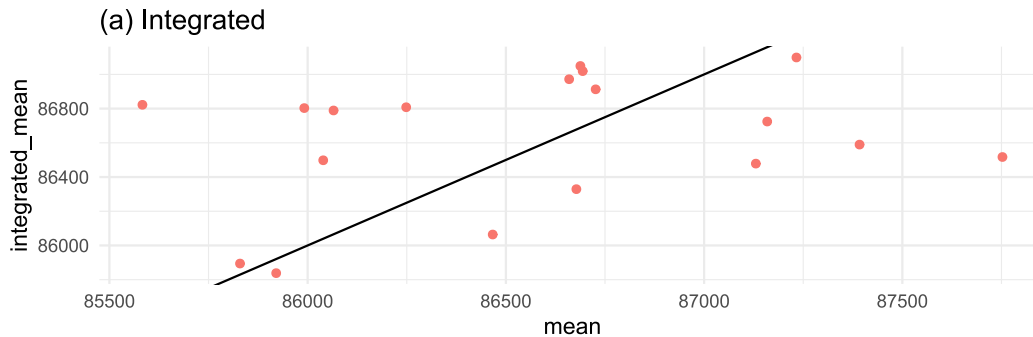
The midpoint and integration methods tend to overestimate and underestimate, respectively, with similar distribution of values
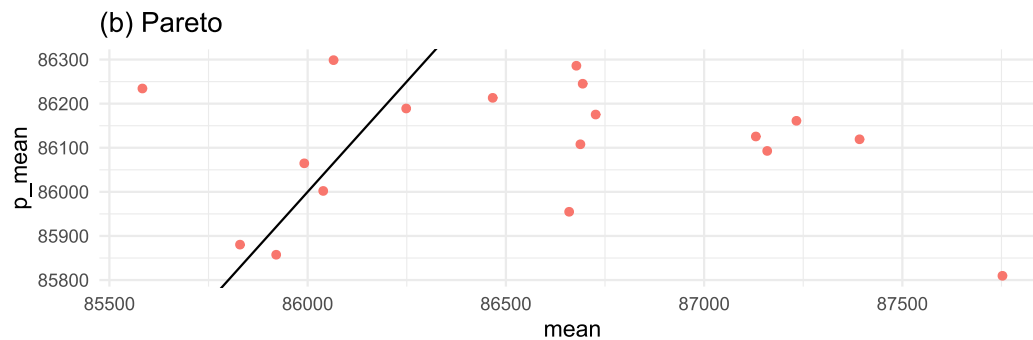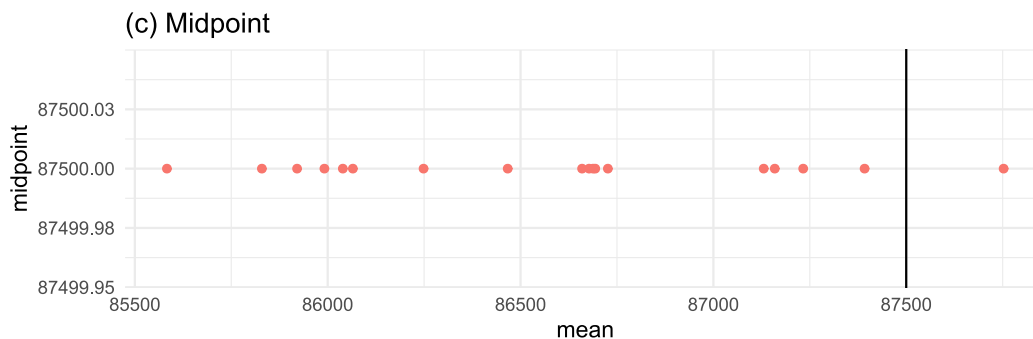
Results - Evaluation of estimators

FBI above median income band (£75,000 - £100,000)

(a) Integrated

(b) Pareto

(c) Midpoint

The midpoint method tends to overestimate the mean in all but one estimate.

The Pareto estimates are generally underestimated, whereas the integrated estimates are more evenly distributed and better clustered

Recommendations

- For closed bands use the integration (MCIB) method for all bands.
  - By explicitly accounting for the variation of values within each bracket and the changing relative frequency of households, the MCIB method provides accurate estimates of inequality statistics even with limited data.

  - MCIB is very good at estimating income, and has been proven to work with large datasets (US census data) and small datasets (FBS)

- For the top open bracket use Robust Pareto Midpoint Estimator (as we don't have a grand mean)

- The estimates for each band were calculated for each survey year and farm type, and then these were assigned to the corresponding farms as the Household Off-Farm Income value.

- The median income could then be calculated conventionally.

Farm Household Income 2021/22 release
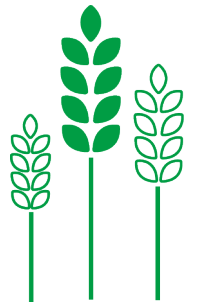
Methodology for other Household Income measures

Household Share of Drawings

- In the Farm Business Survey, total private drawings is reported directly, and the percentage of drawings taken by the principal farmer and their spouse or common law partner is also reported.

- The Household Share of Drawings is calculated by multiplying total private drawings by this percentage.

Household Share of Farm Business Income

- The Household Share of Farm Business Income is also calculated using the percentage of drawings taken by the principal farmer and their spouse or common law partner, rather than that farmer's legal share of the business as defined in partnership agreements (which frequently reflects tax considerations).

- The total Farm Business Income is multiplied by the percentage of drawings taken.

Methodology

Equivalisation

- All three measures (Household Share of Drawings, Household Share of Farm Business Income and Household Off-Farm Income) were equivalised; the figures were adjusted to the level of a single adult

- This allows us to fairly compare households of differing compositions.

- The OECD-modified equivalence scale assigns a value of 1 to the household head, 0.5 to each additional person aged 14 and over and 0.3 to each child aged under 14.

- The FBS does not collect the age of children, therefore, we assigned a value of 0.5 to each additional adult and 0.3 to each child.

- For example, the total equivalence value for a household with a married couple and one child is calculated as follows:

    1 (first adult) plus 0.5 (second adult) plus 0.3 (child) is 1.8

- The income is then divided by this value to give the equivalised income for a single adult.

Methodology

Equivalalisation

Indicative medians of farming household income for differing household compositions in England, 2021/22

Measure	1 adult[†]	1 adult 1 child	2 adults	2 adults 1 child	2 adults 2 children
Household Share of FBI	£22,200	£28,800	£33,300	£39,900	£46,600
Household Share of Drawings	£12,800	£16,700	£19,300	£23,100	£27,000
Household Off-Farm Income	£3,300	£4,300	£5,000	£6,000	£7,000

[†]Published figures

Methodology

Adjusting for inflation

- After equivalisation, the measures were adjusted for inflation using the Consumer Prices Index, including owner-occupiers' housing costs (CPIH).

- Each of the annual CPIH values was divided by the value for 2021/22 to create a new, rebased, index.

- The equivalised measures were multiplied by this new index to adjust them for inflation.

- Once the equivalised, inflation-adjusted measures were calculated for every farm in each of the survey years, weighted medians for each year, farm type and farm size were calculated.

Methodology

Effects of methodology changes between 2015 and 2023

- The 2023 publication shows median figures, as opposed to mean (as was used in the 2015 publication), because the income data are highly skewed.

- The 2023 published figures have also been equivalised, which reduces the median value to make the household comparable to a single person household.

- Off-farm income values have been estimated using advanced mathematical methods, rather than simply taking the midpoint of the bands.

The effect of different methodologies on the 2014/15 survey data

| Measure	Mean	Median	Equivalised median
Household Share of FBI	£31,900	£19,000	£11,000
Household Off-Farm Income	£13,400	£8,700	£5,800

FBI or Drawings?

Definitions

- Farm Business Income is the financial return to all unpaid labour on their capital invested in the farm business, including land and buildings; in essence, Net Profit.

- The Household Share of FBI is the theoretical maximum disposable income for the household.

- Private drawings is the gross income taken from the farm business for investments in capital outside of the farm business and private expenditure.

- Household Share of Drawings is the actual amount taken by the household.

FBI or Drawings?

Trends
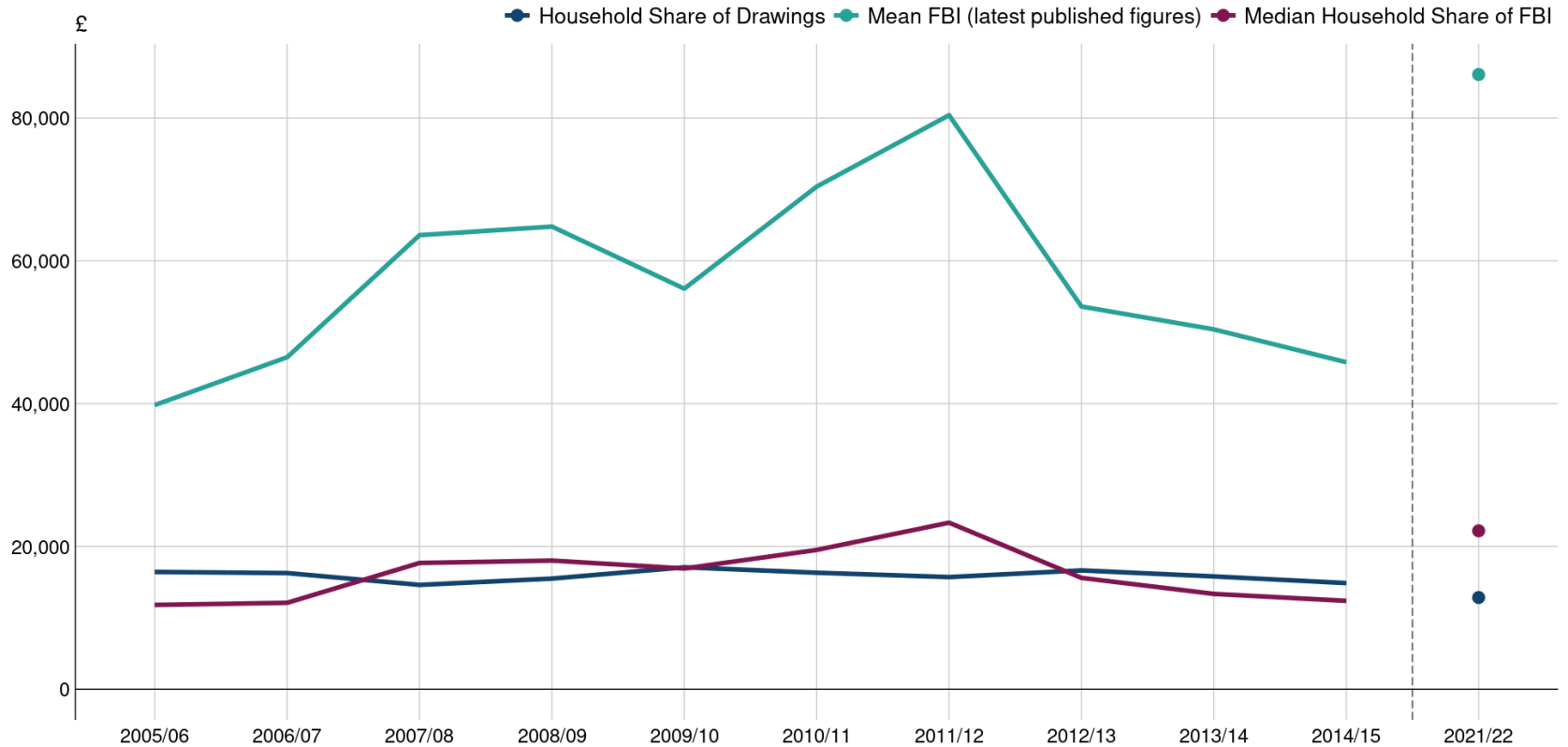
FBI or Drawings?

Debate

- For most farms in England, the farmer and their family live on the farm and the farmhouse is often used as an office.

- Therefore, some household expenses, including things that an average employed person spends their income on, can legitimately be claimed as farm expenditure.

- Although private shares of expenditure, for example, energy costs, insurance, motor costs and property costs, are all calculated in the FBS, **drawings** may slightly underestimate the total disposable income available to the farm household.

- **FBI** can be negative for some farms, so it will definitely underestimate disposable income for these farms.

- If **FBI** is positive, the household may choose not to draw the full amount and instead save some of it in the farm business or use some for making additional investments in the farm – so this may overestimate disposable income.

FBI or Drawings?

Other sources

- The Office for National Statistics presents measures of disposable household income for all UK households; for self-employed individuals, they use the **drawings** from the business (not the net profit)

- Previously, the OECD presented farm household income using FBI, but this measure has since been removed [link to paper]

- The USDA currently publish farm household income, which uses FBI.

- We would like to understand what other countries do to help us feed into our development work

Thank you

Any questions?

AND

Any thoughts on Farm Household Definition?

Joanna.Hutchinson@defra.gov.uk