Simultaneous Equations

Dataset Background:

The dataset is taken from the 'The Review of Economics and Statistics' paper by KMenta based on the Dynamics of Household Budget Allocation to Food Expenditures between the years 1922 to 1941. There are 2 equations that are used to represent the data, making it a case of simultaneous equations.

Dataset Glimpse:

Y	Q	P	D01	F
1922	98.49	100.323	87.4	98
1923	99.187	104.264	97.6	99.1
1924	102.163	103.435	96.7	99.1
1925	101.504	104.506	98.2	98.1
1926	104.24	98.001	99.8	110.8
1927	103.243	99.456	100.5	108.2
1928	103.993	101.066	103.2	105.6
1929	99.9	104.763	107.8	109.8
1930	100.35	96.446	96.6	108.7
1931	102.82	91.228	88.9	100.6
1932	95.435	93.085	75.1	81
1933	92.424	98.801	76.9	68.6
1934	94.535	102.908	84.6	70.9
1935	98.757	98.756	90.6	81.4
1936	105.797	95.119	103.1	102.3
1937	100.225	98.451	105.1	105
1938	103.522	86.498	96.4	110.5
1939	99.929	104.016	104.4	92.5
1940	105.223	105.769	110.7	89.3
1941	106.232	113.49	127.1	93

Total Number of Rows: 20.

Total Number of Columns: 5.

Column Details:

- Y – Year of observation.
- Q – Food Consumption per Head.
- P – Ratio of Food Prices to General Prices.
- D01 – Disposable Income in Constant Prices.
- F – Ratio of Preceding Year's Prices.

Using SPSS Software EViews, we have analysed the data:

Descriptive Statistics:

	Q	P	D01	F
Mean	100.8982	100.0191	97.53500	96.62500
Median	100.9270	99.88950	97.90000	99.10000
Maximum	106.2320	113.4900	127.1000	110.8000
Minimum	92.42400	86.49800	75.10000	68.60000
Std. Dev.	3.756498	5.926086	11.83048	12.70880
Skewness	-0.614316	-0.162638	0.198615	-0.899172
Kurtosis	2.726139	3.512654	3.659242	2.826790
Jarque-Bera	1.320447	0.307182	0.493659	2.720036
Probability	0.516736	0.857623	0.781274	0.256656
Sum	2017.964	2000.381	1950.700	1932.500
Sum Sq. Dev.	268.1143	667.2515	2659.246	3068.758
Observations	20	20	20	20

Inferences:

- The variable Q is slightly left skewed, ranging between 92.42 to 106.23
- The variable P is slightly left skewed, ranging between 86.49 to 113.49.
- The variable D01 is slightly right skewed, ranging between 74.1 to 127.1.
- The variable F is slightly left skewed, ranging between 68.6 to 110.8.
- There is no missing data.

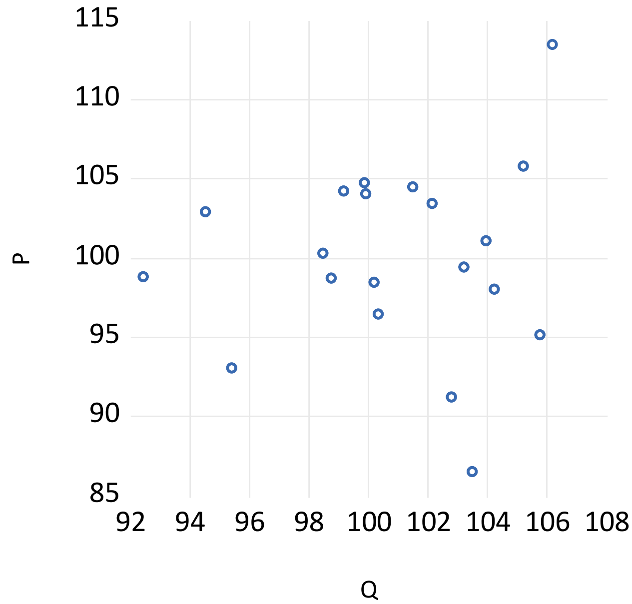
Correlation Analysis:

	Q	P	D01	F
Q	1.000000	0.098096	0.771184	0.680567
P	0.098096	1.000000	0.566549	-0.184428
D01	0.771184	0.566549	1.000000	0.492350
F	0.680567	-0.184428	0.492350	1.000000

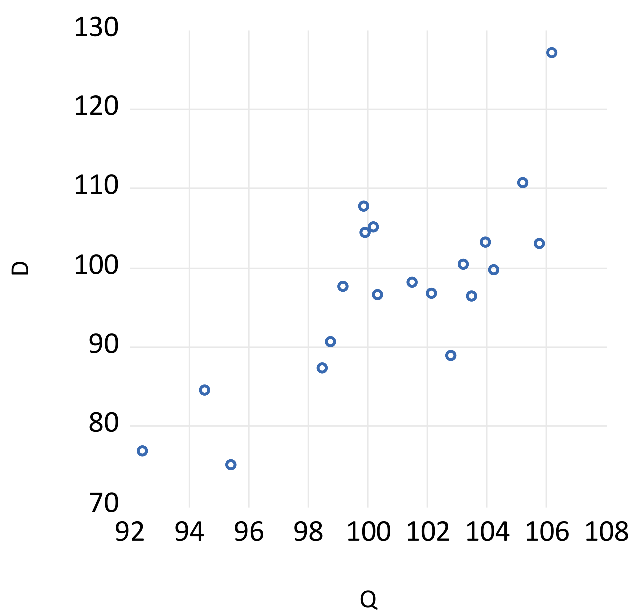
Inferences:

- The variables Q and P have possibly no linear correlation, having correlation coefficient 0.98.
- The variables Q and D01 have sufficient high degree of positive linear correlation, having correlation coefficient 0.77.
- The variables Q and F have moderate degree of positive linear correlation, having correlation coefficient 0.68.
- The variables P and D01 have only the possibility of positive linear correlation, having correlation coefficient 0.56.
- The variables P and F have possibly no linear correlation, having correlation coefficient -0.18.
- The variables D01 and F have only the possibility of positive linear correlation, having correlation coefficient 0.49.

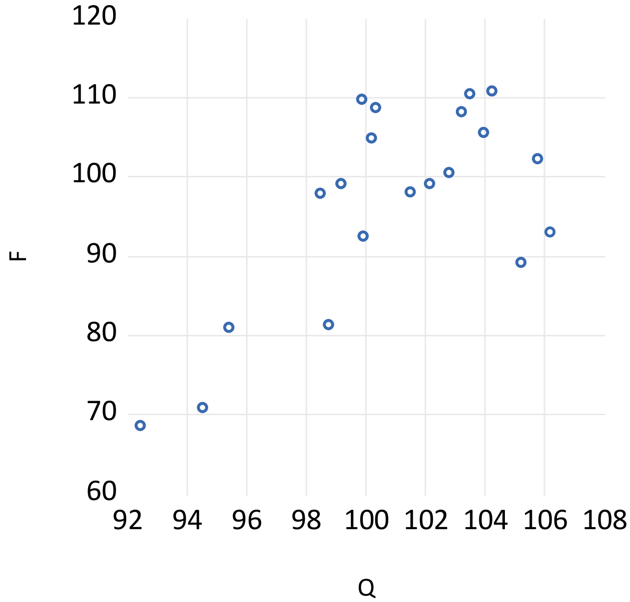
Scatter Plots:



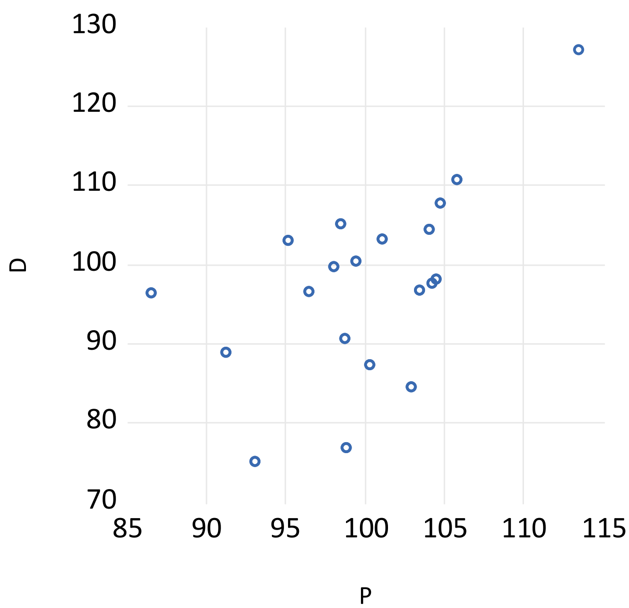
Inference: the variables Q and P have possibly no linear correlation.



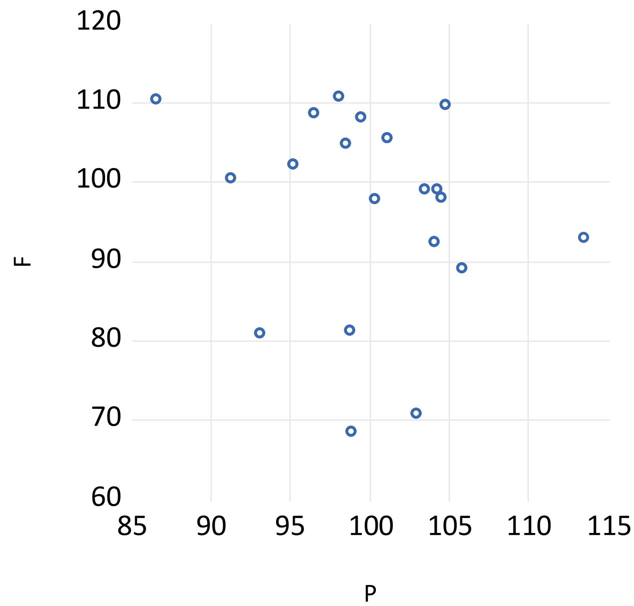
Inference: the variables Q and D01 have sufficient high degree of positive linear correlation.



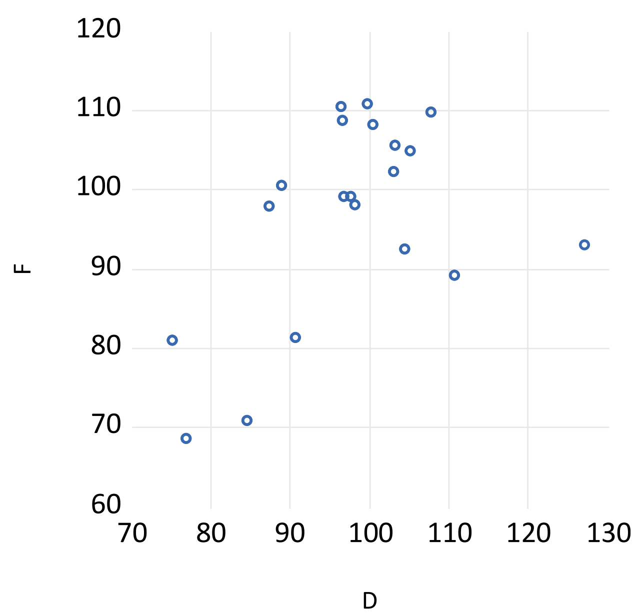
Inference: the variables Q and F have moderate degree of positive linear correlation.



Inference: the variables P and D01 have only the possibility of positive linear correlation.



Inference: the variables P and F have possibly no linear correlation.



Inference: the variables D01 and F have only the possibility of positive linear correlation.

Simultaneous Equations Analysis:

From the data, we can obtain two equations:

$$Q = \alpha 1 + (\beta 1)(P) + (£1)(D) + \mu 1$$

Here, there are 2 endogenous variables (k) – Q and P, 1 exogenous variable – D, and 2 missing variables (g) – F and Y. Since k > g -1, the equation is over-identified. The OLS Regression is as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	99.89542	7.519362	13.28509	0.0000
P	-0.316299	0.090677	-3.488177	0.0028
D01	0.334636	0.045422	7.367285	0.0000

R-squared	0.763789	Mean dependent var	100.8982
Adjusted R-squared	0.735999	S.D. dependent var	3.756498
S.E. of regression	1.930127	Akaike info criterion	4.290530
Sum squared resid	63.33165	Schwarz criterion	4.439890
Log likelihood	-39.90530	Hannan-Quinn criter.	4.319687
F-statistic	27.48472	Durbin-Watson stat	1.744203
Prob(F-statistic)	0.000005		

$$P = \alpha 2 + (\beta 2)(Q) + (£2)(F) + (\Theta 2)(Y) + \mu 2$$

Here, there are 2 endogenous variable (k) – Q and P, 2 exogenous variables – F and Y, and 1 missing variable (g) – D. Since k > g -1, the equation is over-identified. The OLS Regression is as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	920.6539	156.8278	5.870475	0.0000
Q	1.939332	0.298326	6.500711	0.0000
F	-0.426977	0.096604	-4.419862	0.0006
Y	-0.505072	0.087941	-5.743287	0.0001
D	-1.533123	0.507758	-3.019400	0.0092
SIGMASQ	15.59170	7.233805	2.155394	0.0490

R-squared	0.532659	Mean dependent var	100.0191
Adjusted R-squared	0.365752	S.D. dependent var	5.926086
S.E. of regression	4.719518	Akaike info criterion	6.501176
Sum squared resid	311.8339	Schwarz criterion	6.799895
Log likelihood	-59.01176	Hannan-Quinn criter.	6.559489
F-statistic	3.191343	Durbin-Watson stat	1.515432
Prob(F-statistic)	0.039506		

Analyzing and identifying the best instrumental equation using TSLS:

$$Q = \alpha 1 + (\beta 1)(P) + (£1)(D) + \mu 1$$

Using instrumental variables:

P(-1) D01(-1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	57.40432	42.86352	1.339235	0.1992
P	0.133682	0.477082	0.280208	0.7829
D01	0.308481	0.104953	2.939240	0.0096

R-squared	0.305233	Mean dependent var	101.0252
Adjusted R-squared	0.218388	S.D. dependent var	3.815060
S.E. of regression	3.372850	Sum squared resid	182.0179
F-statistic	6.964694	Durbin-Watson stat	2.071682
Prob(F-statistic)	0.006671	Second-Stage SSR	103.5219
J-statistic	0.000000	Instrument rank	3

Using instrumental variables:

P(-1) P(-2) D01(-1) D01(-2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	68.59077	37.24375	1.841672	0.0854
P	0.033193	0.424520	0.078189	0.9387
D01	0.297928	0.092341	3.226401	0.0056

R-squared	0.515884	Mean dependent var	101.1273
Adjusted R-squared	0.451335	S.D. dependent var	3.898850
S.E. of regression	2.887952	Sum squared resid	125.1040
F-statistic	9.045397	Durbin-Watson stat	2.198340
Prob(F-statistic)	0.002648	Second-Stage SSR	107.5355
J-statistic	1.543699	Instrument rank	5
Prob(J-statistic)	0.462158		

As we can see from the above equations, the better instrumental equation is formed using P(-1) P(-2) D01(-1) D01(-2) since the variables have lower p-values and the model has a higher value of R^2.

$P = \alpha 2 + (\beta 2)(Q) + (£2)(F) + (\Theta 2)(Y) + \mu 2$

Using instrumental variables:

Q(-1) F(-1) Y(-1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1468.913	1506.102	0.975308	0.3449
Q	4.332612	3.627975	1.194223	0.2509
F	-0.861587	0.752357	-1.145184	0.2701
Y01	-0.892042	0.896218	-0.995340	0.3354

R-squared	-1.924872	Mean dependent var	100.0031
Adjusted R-squared	-2.509846	S.D. dependent var	6.088031
S.E. of regression	11.40567	Sum squared resid	1951.341
F-statistic	0.481022	Durbin-Watson stat	2.358377
Prob(F-statistic)	0.700352	Second-Stage SSR	479.4269
J-statistic	0.000000	Instrument rank	4

Using instrumental variables:

Q(-1) Q(-2) F(-1) F(-2) Y(-1) Y(-2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1193.219	1379.064	0.865238	0.4015
Q	4.313484	3.354391	1.285922	0.2193
F	-0.840441	0.611109	-1.375273	0.1907
Y01	-0.749618	0.807956	-0.927795	0.3692

R-squared	-1.945540	Mean dependent var	99.76633
Adjusted R-squared	-2.576727	S.D. dependent var	6.173901
S.E. of regression	11.67623	Sum squared resid	1908.680
F-statistic	0.642560	Durbin-Watson stat	2.327641
Prob(F-statistic)	0.600256	Second-Stage SSR	385.1809
J-statistic	0.158408	Instrument rank	6
Prob(J-statistic)	0.923852		

As we can see from the above equations, the better instrumental equation is formed using Q(-1) Q(-2) F(-1) F(-2) Y(-1) Y(-2) since the variables have lower p-values and the model has a higher value of R^2.