2025 Belgian Stata Conference

Shapley value calculations: Implementation and illustrations

Philippe Van Kerm

Brussels – September 9, 2025

UNIVERSITÉ DU LUXEMBOURG

The Shapley and Owen values

A Stata package

Illustrations

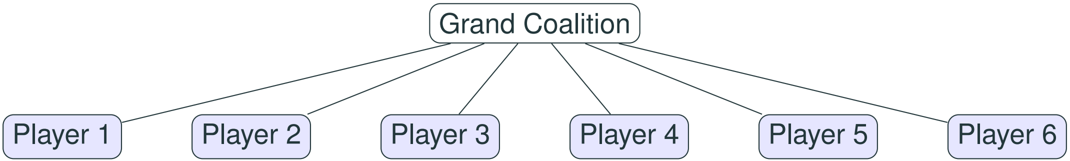
The Shapley value in a nutshell

What is the contribution of an individual element i to the aggregate "value" collectively created by a set of elements $N = \{1, 2, \ldots, N\}$?

or

How to apportion contributions fairly in order to distributed payoffs when only the collective outcome is observable?

The Shapley value in a nutshell



The Shapley value in a nutshell

Many applications under that generic formulation

- Cost sharing: payment of individual users to total cost functions (cf. the “airport problem”)
- Contribution of individual covariates to a regression model’s R^2
- Contribution of individual predictors to a machine learning predictive or classification model
- Bargaining power of political parties in coalition formation
- (Various applications to income distribution analysis (**?**) – we’ll return to these)
- ...

The Shapley value formula

- A set of elements ('players') $N = \{1, 2, \dots, n\}$. N is the "grand coalition"
- A characteristic function $v : 2^N \to \mathbb{R}$ returns the collective output of any coalition formed by elements of N
- The Shapley value (**?**) for element i is given by

$$\phi_i = \phi(i; v, N) = \underbrace{\sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!}}_{\text{weighted sum across coalitions}} \underbrace{(v(S \cup \{i\}) - v(S))}_{\text{marginal contribution to coalition } S}$$

$\Longrightarrow$ weighted average of marginal contribution to any sub-coalition from the 'grand coalition'

$\Longrightarrow$ equivalently: average of marginal contribution to any permutation of elements in N

Four key properties

- *Efficiency*: Additive decomposability

$$\sum_{i \in N} \phi_i = v(N)$$

- *Symmetry*: Equal marginal contributions imply equal ϕ_i
- *Dummy player*: Players contributing a fixed constant k (possibly 0) to any coalition get $\phi_i = k$
- *Additivity*: Contributions to additive games are additive themselves

Nested structures: The Owen value (?)

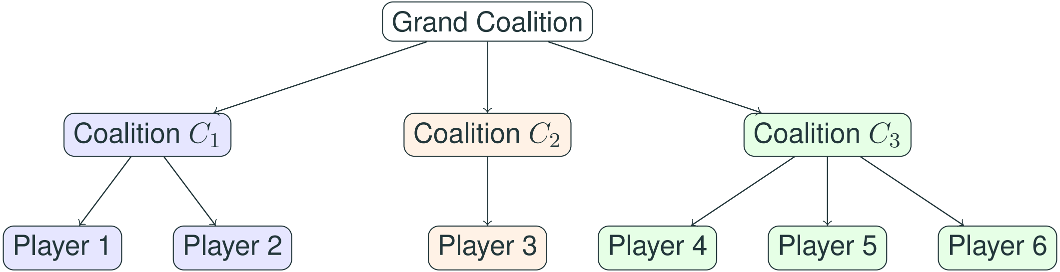
Consider now preset (sub-)coalitions within the ‘grand coalition’

$\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ is a partition of N into disjoint coalitions

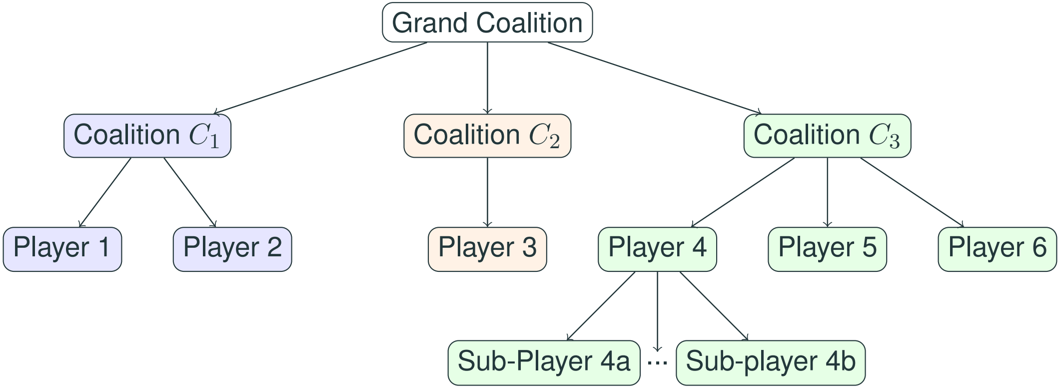
For example

- blocks of ‘similar’ covariates (e.g., demographic attributes, labour market characteristics and regional environment)
- blocks of ‘similar’ political parties (e.g., left, center, right?)

Nested structures: The Owen value (?)



Nested structures: The Owen value (?)



Nested structures: The Owen value (?)

The Owen value is the composition of Shapley values for each preset sub-coalition

$$\psi_i(v, \mathcal{C}) = \phi_i(v_{C_j}) \cdot \phi_{C_j}(v^{\mathcal{C}})$$

- Recursive additivity: the contributions of each player in a sub-coalition add up to the contribution of the sub-coalition to the grand coalition
- Evaluate the Shapley value of a player in a sub-coalition for all possible coalitions of sub-coalitions...
- ... for any level of nested structures

[Outline]

The Shapley and Owen values

A Stata package

Illustrations

`shapowen` – Simplified syntax

`shapowen` evaluates Shapley and Owen values for arbitrary Stata instructions – provided these

1. take input ‘players’ specification as a list
2. return evaluation in `r()` or `e()` scalars or matrices (or functions thereof)

Simplified syntax diagram

`shapowen` *list-of-items*

[, `scalarexpressions(`*string*`)` `matrixexpressions(`*string*`)` `substitution(`*string*`)` ...]:

cmd ... @ ...

(Syntax borrowed from the package `shapley` available on SSC (**?**).)

`shapowen` – Nested structures

Nested list of items

Nested structures are specified in *list-of-items* by grouping items within brackets, e.g,:

$$(\mathtt{a}\ \mathtt{b})\ \mathtt{c}\ (\mathtt{d}\ \mathtt{e}\ \mathtt{f})$$

or

$$\mathtt{a}\ (\mathtt{b}\ \mathtt{c}\ (\mathtt{d}\ \mathtt{e}\ \mathtt{f})\ (\mathtt{g}\ \mathtt{h})\ \mathtt{i})\ ((\mathtt{j}\ \mathtt{k}\ \mathtt{l})\ \mathtt{m})\ \mathtt{n}\ \text{``}\mathtt{o}\ \mathtt{p}\ \mathtt{q}\text{''}$$

NB: grouping by double-quotes forms an unbreakable item – here `o p q` are never evaluated separately.

`shapowen` – Implementation comments

- Shapley value calculations are relatively straightforward to implement – `shapowen` mainly does ‘bookkeeping’ on behalf of the user
- Nested structures and Owen values *are* significantly more difficult to deal with
  - » `shapowen` leverages “advanced” Mata features such as classes (objects), structures, pointers and recursivity
- Speed is potentially an issue: the number of evaluations increases exponentially with the number of ‘players’
  - » *cmd* needs to be fast or n needs to be relatively small

[Outline]

The Shapley and Owen values

A Stata package

Illustrations

Data

- The **Luxembourg Income Study**
  - Household survey data from approximately 50 countries (mostly on income and employment) from early 1980s
- Household- and individual-level representative survey data
- *ex post* harmonization by LIS staff
- Remote (or on-site) access (using (slightly restricted) Stata)
- Illustrations here on data for Belgium (for 2009 and 2019)

Example 1: Ascribing covariate contributions to a regression's R^2

```
. svy : regress lnY  c.age##c.age i.educ i.sex i.typehh2
(running regress on estimation sample)
Survey: Linear regression

Number of strata   =         1              Number of obs     =      6,715
Number of PSUs     =     6,715              Population size   = 11,448,293
                                            Design df         =      6,714
                                            F(   9,   6706)   =     203.00
                                            Prob > F          =     0.0000
                                            R-squared         =     0.2679

------------------------------------------------------------------------------------------------------------
                                            |             Linearized
                                        lnY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------------------------------------+---------------------------------------------------------------
                                        age |   .0365523   .0028338    12.90   0.000     .0309972    .0421074
                                c.age#c.age |  -.0003494   .0000255   -13.73   0.000    -.0003993   -.0002995
                                       educ |
                                 [2]medium  |   .2379809   .0209922    11.34   0.000     .1968296    .2791323
                                   [3]high  |   .4566599   .0210029    21.74   0.000     .4154875    .4978322
                                        sex |
                                 [2]female  |   -.036143   .0142383    -2.54   0.011    -.0640546   -.0082315
                                    typehh2 |
                               Couple only  |   .3099867    .016799    18.45   0.000     .2770553    .3429181
Couple with child(ren) and possibly others  |    .236953    .019542    12.13   0.000     .1986444    .2752616
Single with child(ren) and possibly others  |  -.0790149   .0229858    -3.44   0.001    -.1240743   -.0339555
                       Other configuration  |   .2094785   .0422841     4.95   0.000     .1265883    .2923688
                                      _cons |    8.95386   .0766612   116.80   0.000      8.80358     9.10414
------------------------------------------------------------------------------------------------------------
```

Example 1: Ascribing covariate contributions to a regression's R^2

```
. shapowen c.age##c.age i.educ i.sex i.typehh2  , scalar(e(r2)) : ///
>         svy : regress lnY  @
................
Instruction:  svy : regress lnY  @
Items list:   c.age##c.age i.educ i.sex i.typehh2

 ---  e(r2) ---                 Empty set value:         0                  Full set value:  .2678879
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2678879				1			
c.age##c.age	.0430312	.0419648	.0613163	.0290116	.1606314	.1566507	.2288881	.1082975
i.educ	.1329032	.1316128	.1627722	.1081956	.496115	.4912983	.6076132	.4038838
i.sex	.0069608	.0068213	.0132664	.0012132	.025984	.0254633	.0495222	.0045286
i.typehh2	.0849927	.0836965	.1113913	.0637788	.3172696	.312431	.4158132	.2380803

Example 2: Ascribing contributions to changes in a β coefficient

```
. shapowen c.age##c.age i.educ i.typehh2  , scalar(_b[2.sex]) : ///
>         svy : regress  lnY i.sex @
........
Instruction:  svy : regress  lnY i.sex @
Items list:   c.age##c.age i.educ i.typehh2

 ---  _b[2.sex] ---            Empty set value: -.1145738                 Full set value:  -.036143
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.0784308				1			
c.age##c.age	.0072613	.0073656	.0010747	.0130309	.0925822	.0939115	.0137021	.1661448
i.educ	.0060868	.0061911	.0095243	.0022323	.0776076	.078937	.1214353	.0284625
i.typehh2	.0650827	.0651869	.0651792	.0645691	.8298102	.8311395	.8310406	.8232624

Example 3: Inequality contributions of sources of income

- Let us measure inequality of household disposable income in a country by the Gini coefficient
- Disposable income is the sum of multiple sources: e.g., market incomes (labour incomes, capital incomes), public transfers (pensions, social transfers), minus taxes and social security contributions
- Question: *What is the contribution of each of these sources to overall inequality?* (see, e.g., **????**)

Example 3: Inequality contributions of sources of income

```
. sgini  Lh Lo K  PB T O [aw=hpwgt] if period==2 , sourcedecomp   // ssc install sgini

Gini coefficient for Lh, Lo, K, PB, T, O

Note: T has 5883 negative observations (used in calculations).
```

Variable	v=2
Lh	0.5515
Lo	0.6249
K	0.9316
PB	0.5599
T	-0.5297
O	0.9646

```
Decomposition by source:
  TOTAL =  Lh +  Lo +  K +  PB +  T +  O
```

Example 3: Inequality contributions of sources of income

```
...
Parameter: v=2
```

Variable	Share s	Coeff. g	Corr. r	Conc. c=g*r	Contri. s*g*r	%Contri. s*g*r/G	Elasticity s*g*r/G-s
Lh	0.5668	0.5515	0.6910	0.3811	0.2160	0.8464	0.2796
Lo	0.4177	0.6249	0.7587	0.4741	0.1980	0.7760	0.3583
K	0.0342	0.9316	0.6583	0.6133	0.0210	0.0822	0.0480
PB	0.3273	0.5599	-0.1094	-0.0613	-0.0200	-0.0786	-0.4058
T	-0.3500	-0.5297	-0.8616	0.4564	-0.1597	-0.6258	-0.2759
O	0.0040	0.9646	-0.0130	-0.0125	-0.0000	-0.0002	-0.0042
TOTAL	1.0000	0.2552	1.0000	0.2552	0.2552	1.0000	0.0000

Example 3: Inequality contributions of sources of income

```
. shapowen Lh Lo K  PB T O , scalar(r(coeff)) error(0) : ///
>         sgini @ [aw=hpwgt] , source
..Execution of --          sgini  [aw=hpwgt] , source -- failed (error 100)
Set error value 0 used
..............................................................
Instruction:  sgini @ [aw=hpwgt] , source
Items list:   Lh Lo K  PB T O
```

--- r(coeff) --- Empty set value: 0 Full set value: .2552038

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2552038				1			
Lh	-.7172844	-.6949358	.551542	-.3362231	-2.810634	-2.723062	2.161182	-1.317469
Lo	-.3582809	-.2313472	.6249195	-.1139094	-1.403901	-.9065192	2.448708	-.4463466
K	1.466194	2.491113	.9316415	.0072015	5.745189	9.761267	3.650578	.0282187
PB	-.1321958	.2120674	.559928	-.2501125	-.5180009	.8309725	2.194042	-.98005
T	.0146824	.6560417	-.5296846	-.0594907	.0575319	2.570658	-2.075536	-.2331106
O	-.0179117	-.3026071	.964634	-.0015981	-.070186	-1.185747	3.779857	-.0062619

Example 3: Inequality contributions of sources of income

```
. gen zero = 0.01
. shapowen Lh Lo K  PB T O , scalar(r(coeff)) : ///
>         sgini zero @ [aw=hpwgt]  , source
.................................................................
Instruction:  sgini zero @ [aw=hpwgt]  , source
Items list:   Lh Lo K  PB T O

 ---  r(coeff) ---             Empty set value:         0              Full set value:  .2552037
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2552037				1			
Lh	-.7172473	-.6948817	.5515417	-.3362227	-2.810489	-2.722851	2.161182	-1.317468
Lo	-.3582452	-.2312952	.6249191	-.1139092	-1.403762	-.9063157	2.448707	-.4463463
K	1.466184	2.491093	.9316334	.0072015	5.74515	9.761195	3.650548	.0282187
PB	-.1322222	.2120212	.5599275	-.2501123	-.5181043	.8307917	2.194041	-.9800497
T	.0146543	.6559929	-.5296851	-.0594907	.0574219	2.570468	-2.075538	-.2331106
O	-.0179198	-.3026018	.9645613	-.0015981	-.0702174	-1.185726	3.779574	-.0062619

Example 3: Inequality contributions of sources of income with a nested structure

```
. shapowen ((Lh Lo) K)  ((P B) T) O , scalar(r(coeff)) : ///
>         sgini zero @ [aw=hpwgt]  , source
.......................................................
Instruction:  sgini zero @ [aw=hpwgt]  , source
Items list:   ((Lh Lo) K)  ((P B) T) O

 ---  r(coeff) ---              Empty set value:        0                Full set value:  .2552037
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2552037				1			
(Lh Lo) K	8.289586	7.957234	.4881215	17.42046	32.48223	31.17993	1.912674	68.26099
Lh Lo	-1.810342	-2.278142	.4942429	-20.96295	-7.093712	-8.926759	1.93666	-82.14204
Lh	-.9820429	-1.216076	.5515417	-.3362227	-3.848074	-4.765118	2.161182	-1.317468
Lo	-.828299	-1.062066	.6249191	-.1139092	-3.245638	-4.161641	2.448707	-.4463463
K	10.09993	10.23538	.9316334	.0072015	39.57594	40.10669	3.650548	.0282187
(P B) T	-7.851159	-8.183512	-14.14202	-.2308853	-30.76428	-32.06658	-55.41464	-.9047097
P B	-3.700538	-3.866567	.5599275	-.2501123	-14.50033	-15.1509	2.194041	-.9800497
P	-1.921998	-2.005436	.8248165	-.1514365	-7.531229	-7.858175	3.231992	-.5933947
B	-1.77854	-1.861132	.6124187	-.0804831	-6.969099	-7.292729	2.399725	-.3153682
T	-4.150621	-4.316944	-.5296851	-.0594907	-16.26395	-16.91568	-2.075538	-.2331106
O	-.1832234	-.5155759	.9645613	-.0015981	-.7179494	-2.020252	3.779574	-.0062619

Example 4: Inequality contributions of sources of income: alternatives

- In the previous examples we created coalitions by “eliminating” income sources
- Alternatively one could
  - » eliminate *inequality in the source*: set it to its mean rather than to zero
  - » eliminate *association* of a source with any other income source
- `shapowen`’s `substitution(`*string*`)` option permits this: when a element is excluded from a coalition, it is *replaced* by the item in the same position in *string*

Example 4: Inequality contributions of sources of income: alternatives

Let's create some 'substitutes'

```
. foreach v of varlist Lh Lo K  PB T O {
  2.         su `v' [aw=hpwgt] , meanonly
  3.         gen mn_`v' = r(mean)
  4.         gen r_`v' = runiform()
  5.         egen rnd_`v' = clsort(`v' r_`v')  // ssc install _gclsort
  6. }
```

Example 4: Inequality contributions of sources of income: alternatives

```
. shapowen Lh Lo K  PB T O , scalar(r(coeff)) substitution(zero zero zero zero zero zero)    : ///
>         sgini @ [aw=hpwgt]  , source
.................................................................
Instruction:  sgini @ [aw=hpwgt]  , source
Items list:   Lh Lo K  PB T O

 ---  r(coeff) ---              Empty set value:         0                 Full set value:  .2552038
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2552038				1			
Lh	-.7171782	-.6947749	.5515405	-.3362227	-2.810217	-2.722431	2.161177	-1.317467
Lo	-.3581793	-.2311928	.6249173	-.1139092	-1.403503	-.9059146	2.448699	-.4463459
K	1.466175	2.491078	.9316009	.0072016	5.745116	9.761132	3.650419	.028219
PB	-.1322688	.2119414	.5599254	-.2501123	-.5182868	.830479	2.194032	-.9800491
T	.0146039	.6559077	-.5296869	-.0594906	.0572247	2.570133	-2.075545	-.2331103
O	-.0179493	-.3025758	.9642707	-.001598	-.0703331	-1.185624	3.778434	-.0062616

Example 4: Inequality contributions of sources of income: alternatives

```
. shapowen Lh Lo K  PB T O , scalar(r(coeff)) substitution(mn_Lh mn_Lo mn_K mn_PB mn_T mn_O) : ///
>         sgini @ [aw=hpwgt]  , source
.................................................................
Instruction:  sgini @ [aw=hpwgt]  , source
Items list:   Lh Lo K  PB T O

 ---  r(coeff) ---              Empty set value:         0                 Full set value:  .2552038
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2552038				1			
Lh	.1177067	.0989027	.3126045	-.0010125	.4612261	.387544	1.224921	-.0039675
Lo	.1123793	.093476	.2610494	.0402813	.4403512	.3662799	1.022906	.1578398
K	.0167581	.0134794	.0318558	.0156815	.0656654	.0528181	.1248249	.061447
PB	.0287063	.0186672	.1832561	-.08473	.1124839	.0731461	.7180774	-.3320091
T	-.0211016	-.0353484	.1853667	-.1696202	-.0826854	-.1385104	.7263477	-.6646459
O	.0007551	.0003568	.0038227	-.0005804	.0029589	.0013982	.0149791	-.0022742

Example 4: Inequality contributions of sources of income: alternatives

```
. shapowen Lh Lo K  PB T O , scalar(r(coeff)) substitution(rnd_Lh rnd_Lo rnd_K rnd_PB rnd_T rnd_O) : ///
>         sgini @ [aw=hpwgt]  , source
.................................................................
Instruction:  sgini @ [aw=hpwgt]  , source
Items list:   Lh Lo K  PB T O

 ---  r(coeff) ---             Empty set value:  .5878367                Full set value:  .2552038
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	-.3326328				1			
Lh	-.1269625	-.1250364	-.0462109	-.2153681	.3816895	.3758993	.1389247	.6474648
Lo	-.0981312	-.0978587	-.063637	-.133664	.2950135	.2941944	.1913131	.4018365
K	.001636	.0015333	.0016604	.0020847	-.0049184	-.0046095	-.0049916	-.0062671
PB	-.0313957	-.0252985	.0431575	-.1302905	.0943855	.0760553	-.1297452	.3916945
T	-.0772337	-.0715256	.0248975	-.2021507	.2321889	.2150286	-.0748497	.6077293
O	-.0005458	-.0004681	-.0002925	-.0011227	.001641	.0014072	.0008793	.0033751

Example 5: Inequality contributions of population subgroups

- Say we partition the population in population groups – by age, education, hh type, etc.
- *What is the contribution of subgroups to aggregate inequality?*
- The Shapley value can be used here too (**??**)
- (illustration of `separator()` option to handle alternative list formats)

Example 5: Inequality contributions of population subgroups

```
. shapowen 1 2 3 4 5  , scalar(r(coeff))  sep(,) error(0)  : ///
>         sgini Yalt  [aw=hpwgt] if inlist(typehh2,@)
..Execution of --          sgini Yalt  [aw=hpwgt] if inlist(typehh2,) -- failed (error 198)
Set error value 0 used
..............................
Instruction:  sgini Yalt  [aw=hpwgt] if inlist(typehh2,@)
Items list:   1 2 3 4 5

 ---  r(coeff) ---              Empty set value:         0                Full set value:  .2552038
```

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	.2552038				1			
1	.0688676	.0356221	.2683823	.0124104	.2698532	.139583	1.051639	.0486294
2	.0680083	.0357219	.2697844	.0054535	.2664862	.1399741	1.057133	.021369
3	.0282575	-.0039953	.2198652	-.0242629	.1107253	-.0156554	.8615277	-.0950727
4	.0498535	.0184645	.2320633	.0032755	.1953479	.0723521	.9093254	.0128347
5	.0402169	.011902	.2045503	-.0007813	.1575873	.0466373	.8015174	-.0030613

Example 6: Decomposing the ‘composition effect’ in inequality trends

- Changes over time in the distribution of income can be driven by changes in returns to individual or hh characteristics and by changes in the composition of the population (again, by age, education, labour market participation, hh structures, etc.) .
- The ‘composition’ factor is a combination of changes in multiple characteristics
- *What is the contribution of changes in different characteristics to inequality changes?*
- (illustration of use of small `program define` wrappers)

Example 6: Decomposing the ‘composition effect’ in inequality trends

```
. cap pr drop changini
. pr def changini , rclass
  1.          tempvar p rw
  2.          svy : logit period ‘0’
  3.          qui predict ‘p’ , rules
  4.          qui gen ‘rw’ = cond(period==1 , 1 , ‘p’/(1-‘p’))
  5.          sgini Yalt [aw=hpwgt*‘rw’] if period==0
  6.          return scalar gini = r(coeff)
  7. end
```

Example 6: Decomposing the ‘composition effect’ in inequality trends

```
. sgini Yalt [aw=hpwgt] if period==0
Gini coefficient for Yalt
```

Variable	v=2
Yalt	0.2715

```
. sgini Yalt [aw=hpwgt] if period==1
Gini coefficient for Yalt
```

Variable	v=2
Yalt	0.2552

Example 6: Decomposing the ‘composition effect’ in inequality trends

```
. changini i.typehh2 c.shearn i.educ i.sex
(running logit on estimation sample)
Survey: Logistic regression

Number of strata   =         2                Number of obs     =     10,889
Number of PSUs     =    10,889                Population size   = 20,475,368
                                              Design df         =     10,887
                                              F(   8,  10880)   =     179.16
                                              Prob > F          =     0.0000
```

period	Coef.	Linearized Std. Err.	t	P>|t|	[95% Conf.	Interval]
typehh2						
Couple only	.2168325	.08322	2.61	0.009	.0537062	.3799587
Couple with child(ren) and possibly others	-.1628252	.0799651	-2.04	0.042	-.3195713	-.006079
Single with child(ren) and possibly others	-.7765123	.1326156	-5.86	0.000	-1.036463	-.5165617
Other configuration	-.0360524	.315691	-0.11	0.909	-.6548643	.5827594
shearn	1.145561	.0851836	13.45	0.000	.9785854	1.312536
educ						
[2]medium	1.217598	.0744029	16.36	0.000	1.071754	1.363441
[3]high	1.599592	.0726049	22.03	0.000	1.457273	1.741911
sex						
[2]female	1.877692	.0671318	27.97	0.000	1.746101	2.009282
_cons	-1.619951	.078148	-20.73	0.000	-1.773135	-1.466767

Gini coefficient for Yalt

Variable	v=2
Yalt	0.2521

Example 6: Decomposing the ‘composition effect’ in inequality trends

```
. shapowen i.typehh2 c.shearn i.educ i.sex , scal(r(gini))   : ///
>         changini @
................
Instruction:  changini @
Items list:   i.typehh2 c.shearn i.educ i.sex
```

--- r(gini) --- Empty set value: .2714685 Full set value: .2521056

	Nominal contribution				Relative contribution			
	Shapley -Owen	Banzhaf -Owen	First	Last	Shapley -Owen	Banzhaf -Owen	First	Last
FULL	-.019363				1			
i.typehh2	.0008653	.0012075	.001221	-.000859	-.0446902	-.0623614	-.0630602	.0443643
c.shearn	-.0191425	-.0186721	-.0211986	-.0189679	.9886135	.9643199	1.094803	.9795988
i.educ	-.0032951	-.0028207	-.0050923	-.0033955	.1701762	.1456759	.2629924	.1753611
i.sex	.0022093	.0026497	.0024182	.0002386	-.1140995	-.1368462	-.12489	-.0123222

Example 6: Decomposing the ‘composition effect’ in inequality trends (bootstrap inference)

- Prefix commands can be combined
- Earlier examples show that `svy:` can seamlessly be used in the *cmd ... @ ...* pseudo instruction
- The `bootstrap` prefix can be used with `shapowen`

Example 6: Decomposing the ‘composition effect’ in inequality trends (bootstrap inference)

```
.         bootstrap ///
>                 a=el(r(ShapOw),2,1)  ///
>                 b=el(r(ShapOw),3,1)  ///
>                 c=el(r(ShapOw),4,1)  ///
>                 d=el(r(ShapOw),5,1)  ///
>                 , reps(499) : ///
>                 shapowen i.typehh2 c.shearn i.educ i.sex , scal(r(gini))   : ///
>                         changini @
(running shapowen on estimation sample)

Bootstrap replications (499)
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
..................................................    50
<SNIP>
.................................................
```

Example 6: Decomposing the ‘composition effect’ in inequality trends (bootstrap inference)

```
Bootstrap results                                 Number of obs     =     10,889
                                                  Replications      =        499

      command:  shapowen i.typehh2 c.shearn i.educ i.sex, scal(r(gini)) : changini @
            a:  el(r(ShapOw),2,1)
            b:  el(r(ShapOw),3,1)
            c:  el(r(ShapOw),4,1)
            d:  el(r(ShapOw),5,1)
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
a	.0008653	.0008148	1.06	0.288	-.0007317	.0024624
b	-.0191425	.002364	-8.10	0.000	-.0237759	-.0145091
c	-.0032951	.0023183	-1.42	0.155	-.0078389	.0012487
d	.0022093	.0021852	1.01	0.312	-.0020737	.0064923

Closing

- The Shapley value and Shapley value decompositions and their Owen counterparts for nested structures have many potential applications
  - » Notably in income distribution analysis as illustrated here
- `shapowen` (available on SSC very shortly) facilitates their calculation with a generic prefix-based syntax and flexible input processing

Comments, feedback and suggestions welcome!

References i

Chantreuil, F., Courtin, S., Fourrey, K. and Lebon, I. (2019), ‘A note on the decomposability of inequality measures’, *Social Choice and Welfare* **53**(2), 283–298.

Chantreuil, F. and Trannoy, A. (2013), ‘Inequality decomposition values: the trade-off between marginality and efficiency’, *Journal of Economic Inequality* **11**(1), 83–98.

Deutsch, J. and Silber, J. (2007), Decomposing income inequality by population subgroups: A generalization, *in* ‘Inequality and Poverty’, Research on Economic Inequality, Emerald Group Publishing Limited, pp. 237–253.

Giorgi, G. M. and Guandalini, A. (2018), ‘Decomposing the Bonferroni inequality index by subgroups: Shapley value and balance of inequality’, *Econometrics* **6**(2).

Kolenikov, S. (2000), ‘Shapley: Stata module to perform additive decomposition of sample statistic’.
URL: *https://EconPapers.repec.org/RePEc:boc:bocode:s411401*

Owen, G. (1977), Values of games with a priori unions, *in* R. Henn and O. Moeschlin, eds, ‘Mathematical Economics and Game Theory’, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 76–88.

References ii

Sastre, M. and Trannoy, A. (2002), ‘Shapley inequality decomposition by factor components: Some methodological issues’, *Journal of Economics* **77**(1), 51–89.
URL: *http://dx.doi.org/10.1007/BF03052500*

Shapley, L. S. (1953), A value for n-person games, *in* H. Kuhn and A. W. Tucker, eds, ‘Contributions to the Theory of Games, Vol. II’, Princeton: Princeton University Press, Princeton, NJ, pp. 307–317.

Shorrocks, A. F. (2013), ‘Decomposition procedures for distributional analysis: a unified framework based on the Shapley value’, *Journal of Economic Inequality* **11**(1), 99–126.
URL: *http://dx.doi.org/10.1007/s10888-011-9214-z*

Trannoy, A. and Sastre, M. (2002), Shapley inequality decomposition by factor components: Some methodological issues, *in* P. Moyes, C. Seidl and A. F. Shorrocks, eds, ‘Inequalities: Theory, Measurement and Applications’, number 9 *in* ‘Journal of Economics Supplement Issues’, Springer-Verlag, Wien, pp. 51–89.