The nutrition transition in Vietnam: some recent empirical insights

Michel Simioni

The Nutrition Transition in Vietnam

Some recent empirical insights

Michel Simioni

MOISA, INRA, University of Montpellier, France
and
IREEDS-VCREME, Hanoi, Vietnam

VEAM, Hanoi, June 2018

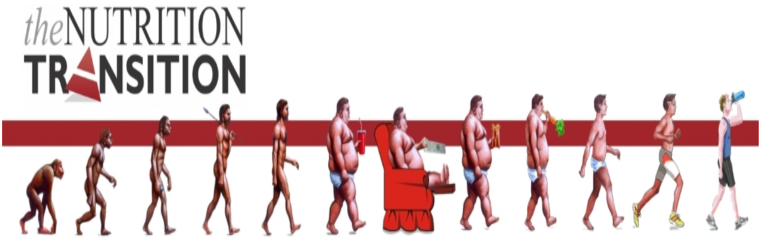
- Doi Moi economic reforms ⇒ Vietnam from one of the world's poorest country to a lower middle-income country
- Two majors features:
  1. Gross National Income: 431 in 1989 ↑ 1691 in 2016 (USD of 2010, World Bank)
  2. Poverty rate: 58% in 1993 ↓ 12% in 2011.
- But, at the same time, "**nutrition transition**".
- Your first reaction: nutrition transition = question for nutritionist and epidemiologists, not for economists.
- But, what we have observed in developed countries: nutrition transition and epidemics (obesity, non communicable diseases) ⇒ costs for the health sector and the economy in general.
- Nutrition transition → questions: economic determinants, nutritional policy efficiency. . .
- Literature at the crossroads of development economics, food economics, health economics. . .
- **Aim of my presentation**: How applied econometrics can provide quantified insights on this transition in Vietnam, using techniques at the frontier of the discipline.

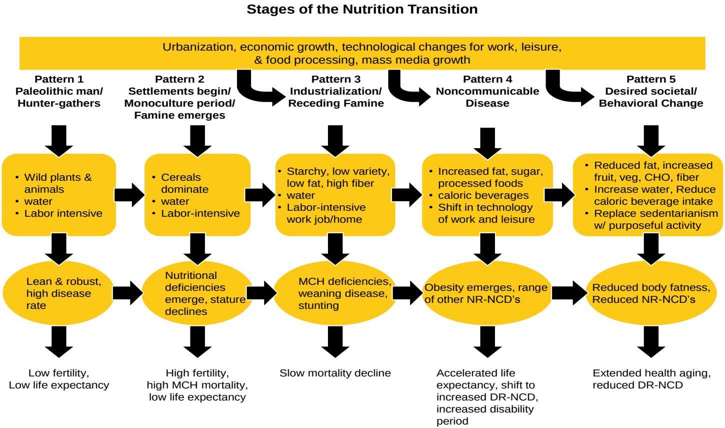
- A project I began five years ago, thanks to a joint funding program of INRA and CIRAD: **GloFoodS**.
- I am an econometrician, not a nutritionist.
- Collective work with
  - Christine Thomas Agnan, Professor of Statistics, TSE, Toulouse.
  - Huong Trinh thi, PhD student, INRA-TSE, Toulouse.
  - Joanna Morais, former PhD student, TSE, Toulouse, and now Data Scientist Consultant, BVA, Paris.

Some appetizers

Nutrition transition refers to the changes in the composition and structure of the diet, usually accompanied by changes in physical activity levels (Popkin, 1993 and 1994).

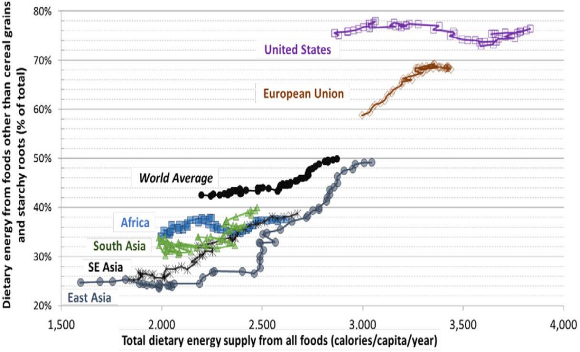
For the public at large:

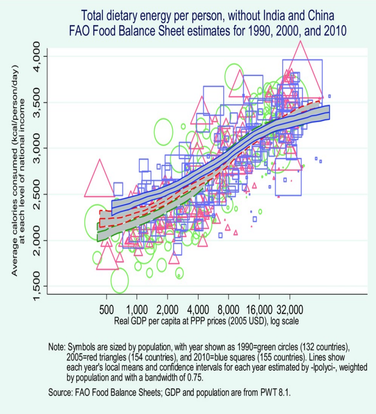


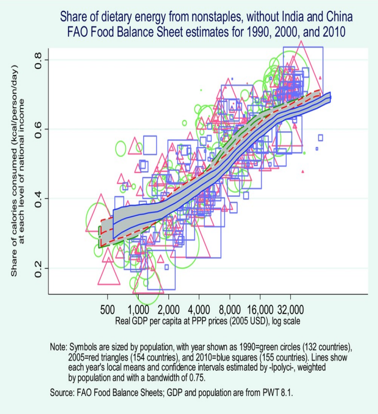


Source: Popkin 2002 revised 2006.

From more food to different foods: the dietary transition by region, 1961–2011 (Masters et al., 2016)

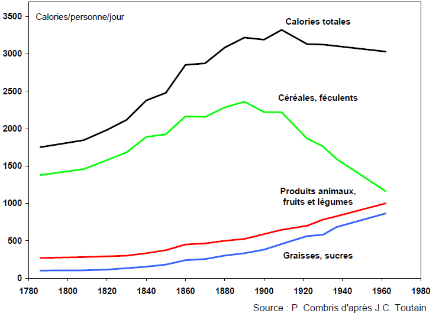




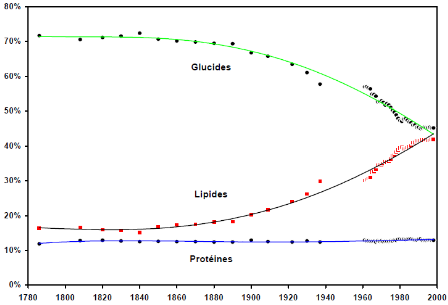


Source: Masters et al. (2016)

Evolution of the energy intake structure in France in the long term (Combris, 2006)

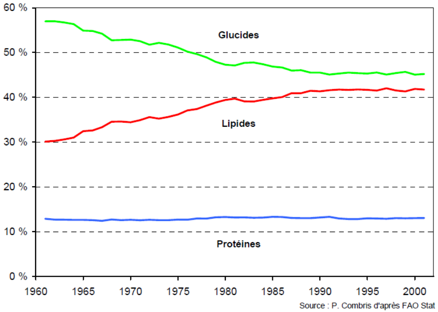


Evolution of the level of energy supply in France in the long term (Combris, 2006)

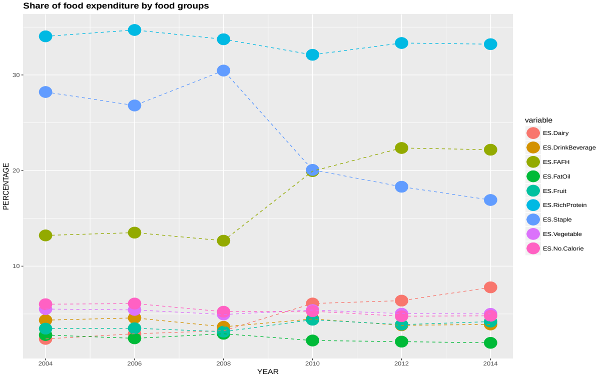


Source : P. Combris, d'après J.C. Toutain, FAO Stat

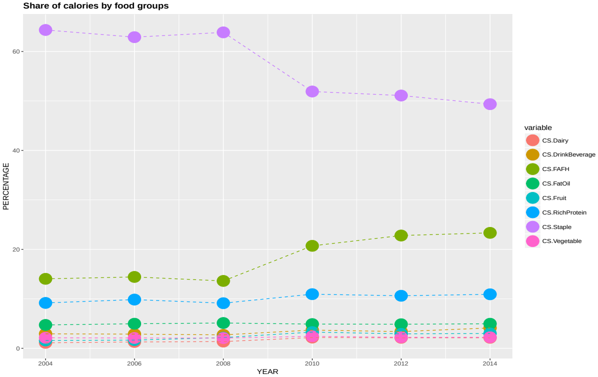
Evolution of the level of energy supply in France from 1961 to 2000 (Combris, 2006)



Trends in food expenditures from 2004 to 2014 (Trinh Thi et al., 2018)

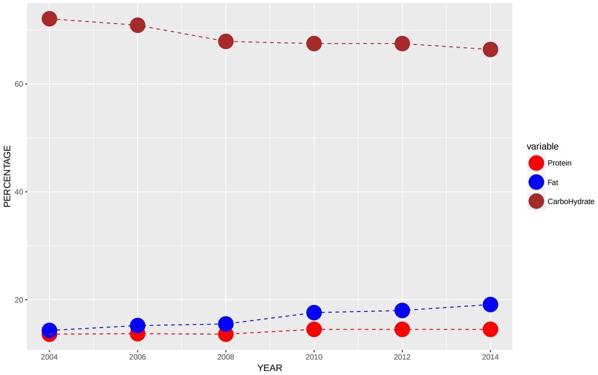


Trends in calories intakes from 2004 to 2014 (Trinh Thi et al., 2018)

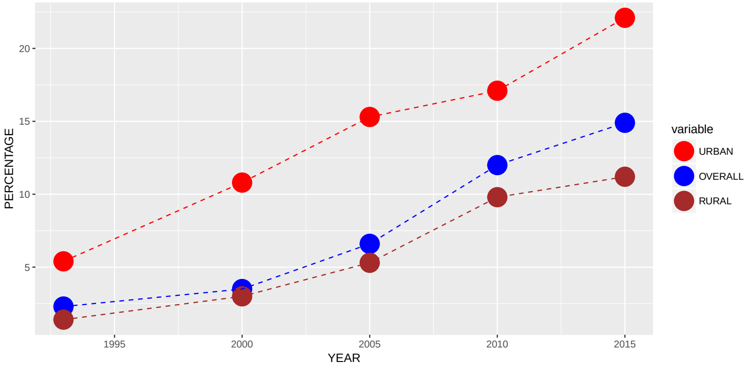


Trends in macronutrient shares from 2004 to 2014

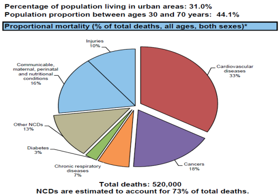
(Trinh Thi et al., 2018)

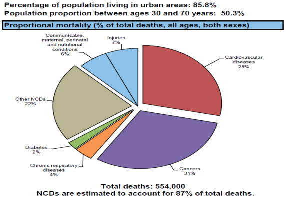


The trend of overweight and obesity in adult population from 1993 to 2015 (Nguyen and Hoang, 2018)



Non communicable diseases: Vietnam and France profiles (WHO, 2014)





Adult risk factors:	Male	Female	Total
Current tobacco smoking (%) (2011)	46 (39)	2 (32)	6.6 (36)
Total alcohol consumption, in liters of pure alcohol (2010)	12.1 (17.8)	0.2 (7.1)	6.6 (12.2)
Raised blood pressure (%) (2008)	25.7 (33.5)	20.5 (22.5)	23.1 (27.1)
Obesity (%) (2008)	1.2 (19.1)	2.1 (17.4)	1.7 (18.2)

The prevalence of selected NCD risk factors for adults aged 25-64 years (Nguyen and Hoang, 2018)

	2010			2015		
	Both sexes	Males	Females	Both sexes	Males	Females
Physiological and metabolic risk factors						
Overweight (BMI ≥ 25kg/m^2)	12.0	12.5	11.4	17.5	16.9	18.1
High blood pressure[1]	19.2	23.1	15.5	20.3	24.7	16.1
Impaired fasting blood glucose[2]	3.6	3.9	3.1	1.6	2.0	1.2
Diabetes mellitus[3]	2.7	2.8	2.6	4.1	4.6	3.6
Elevated blood cholesterol[4]	30.1	27.8	32.3	32.4	27.9	36.1
Behavioral factors						
Low level of physical activity[5]	28.7	26.4	30.8	26.1	19.0	32.6
Drank alcohol in the past 30 days	37.0	69.6	5.6	44.8	80.3	11.2
Consume < 5 servings of fruit and/or vegetables per day	80.4	80.2	80.6	57.2	63.2	51.5
Ever smokers	29.6	59.4	1.7	.	.	.
Smoke tobacco daily	28.2	56.5	1.7	.	.	.

Notes
[1]: Systolic blood pressure ≥ 140 and/or diastolic blood pressure ≥ 90 mmHg, or currently on medication
[2]: Whole blood value ≥ 6.1 mmol/L (or 110 mg/L) but < 7.2 mmol/L (or 126 mg/L
[3] Whole blood value ≥ 7.2 mmol/L or currently on medication
[4] Total cholesterol ≥ 5.0 mmmol/L (or 190 mg/dl) or currently on medication
[5] Level of physical activity < 600 MET-min per week

INRA SCIENCE & IMPACT IREEDS SCIENCE NATURE CONNECTION VCREME Connecting talents for development

Nutritional situation of children in 2011: double burden of undernutrition and overnutrition (Le Nguyen et al., 2013)

Table 4. Prevalence (%) of undernutrition and overnutrition by age group, sex and area of residence‡

	0·5–1·9 years			2·0–4·9 years			5·0–11·9 years		
	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total
Urban									
Stunting	0·0†	10·5	5·8†	3·4†	5·2†	4·4†	9·2†	8·7†	8·9†
Underweight	2·2	9·3	6·2	3·4	5·2	4·4†	13·5†	14·1†	13·8†
Thinness	2·1	8·1†	5·3†	1·6	1·1†	1·3	10·4	8·8	9·6†
Overweight	4·3	5·0	4·7	15·2†	11·6†	13·3†	16·6†	14·8†	15·7†
Obesity	0·0	4·9	2·7	3·4†	13·0*†	8·7†	11·0†	25·1*†	18·0†
Rural									
Stunting	8·1	16·7	13·0	17·5	19·1	18·3	14·7	20·6*	17·7
Underweight	3·2	6·8	5·3	11·7	12·4	12·1	21·7	28·2	25·0
Thinness	1·6	0·0	0·7	4·6	9·0	6·8	15·2	12·2	13·7
Overweight	3·1	3·6	3·4	0·5	2·6	1·5	5·2	3·9	4·5
Obesity	1·6	0·0	0·7	0·0	0·5	0·2	1·0	3·0*	2·0

*Values were significantly different from those of girls after correction for age differences ($P<0·05$).
†Values were significantly different from those of rural areas after correction for age.
‡Stunting: height-for-age z-score ≤ -2 SD; underweight: weight-for-age z-score ≤ -2 SD; thinness: BMI-for-age z-score ≤ -2 SD; overweight: ≥ 1 SD and obese: >2 SD in children aged >5 years; overweight: >2 SD and obese: >3 SD in children aged <5 years ($P<0·05$).

- "National Nutrition Strategy for 2011-2020, with a vision toward 2030" (Ministry of Health, 2012) $\rightarrow$ Various objectives $\supset$

    1. To reduce the proportion of households with low calorie intake (below 1800 Kcal) to 5%
    2. To reach a proportion of households equal to 75% in 2020, with a balanced diet: (Protein: 14%; Lipid: 18%; Carbohydrates: 68%)

- Tools: specific food and nutrition interventions to improve the nutritional status of target groups with a focus on mothers and children, and a priority to the poor, disadvantaged and ethnic minorities areas.

- Nothing about income based policies (subsidized prices for basic foodstuffs or cash transfers)

Thirsty?

- Vu, Linh Hoang (2008). *Essays on the economics of food production and consumption in Vietnam*. University of Minnesota, PhD Thesis.

↓

Vietnam**Household Living Standard** Survey, or VHLSS
(General Statistical Office of Vietnam and World Bank)

- Every two years since 2002: 8 surveys.
- VHLSS ⊃ Household Consumption Expenditure Survey.
- HCES: Deaton (1997), Smith et al. (2014) and a recent Special Issue of *Food Policy* (Zezza *et al.*, 2017).

- Vietnamese Food Composition Table (National Institute of Nutrition, Ministry of Health, 2007):

VHLSS $\supset$ expenditures and/or quantities for 56 food items

$\downarrow$

Total energy intake $\rightarrow$ per capita calorie intake

$\downarrow$

Energy intakes coming from **macronutrients** (Protein, Fat, and Carbohydrates)

- **Remarks**:
  1. Per capita = adult equivalent $\rightarrow$ OECD rules, Aguiar and Hurst (2013)
  2. Calorie **availability** = calorie intake + **leakages** (≥ 0) (Bouis and Haddad, 1992, and Muth *et al.*, 2011)
  3. **Quality** of the diet $\rightarrow$ micronutrient intakes (General Nutrition Survey)

Main course

- Three contributions:

  1. Estimating the calorie intake - income relationship
  2. Analyzing the evolution of macronutrients consumption between 2004 and 2014
  3. Analyzing the evolution of nutritional diet between 2004 and 2014

- Three contributions → some empirical insights on nutrition transition in Vietnam

- Ms. H. Trinh Thi's PhD thesis in Applied Mathematics, University of Toulouse (defense: July 9th 2018): "Adapting recent statistical techniques to the study of nutrition in Vietnam." (Co supervisor: Christine Thomas-Agnan).

- Banerjee (2016): Policies aimed at reducing starvation and redressing nutrition deficiencies.

- Income based policies

  ↓

  Assumption that **nutrition (calorie intake) is conditioned by income**.

- Banerjee (2016): Issue of the measurement of calorie intake - income elasticity

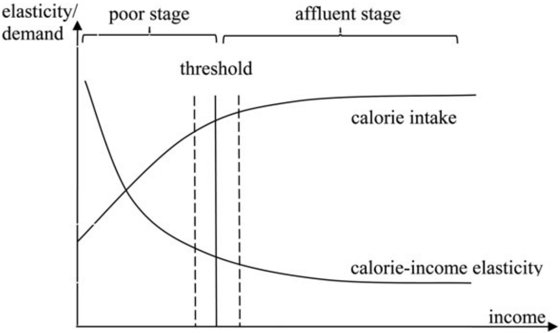
  ↓

  Meta-analyzes: Ogundari and Abdulai (2013), Santeramo and Shabnamb (2015), and Zhou and Yu (2015)

  ↓

  Controversial results (discussions tracing back to Bouis and Haddad, 1992).

The changes in calorie consumption and calorie-income elasticity with income dynamics (Zhou and Yu, 2014)



- Contribution to the literature studying calorie intake - income relationship using household-level data:

  Trinh Thi, Simioni, and Thomas-Agnan (2018): "Assessing the nonlinearity of the calorie-income relationship: an estimation strategy - With new insights on nutritional transition in Vietnam," *World Development*.

- Methodological contribution: Revisit the issue of the choice of a functional form for the calorie intake - income relationship

  1. Estimate parametric and semiparametric specifications: **Generalized Additive Models** (Wood, 2017).
  2. Choose among these alternative specifications: **Revealed Performance Test** (Racine and Parmeter, 2014).
  3. Compare changes over time in the relationship between calorie intake and income in Vietnam: **Decomposition methods** (Fortin *et al.*, 2011)
  4. Testing **exogeneity** if income (Blundell and Horowitz, 2007)

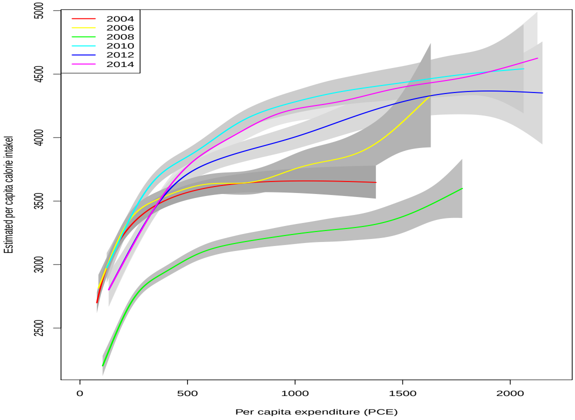
- Empirical contribution: New insights on nutritional transition in Vietnam

Methodology

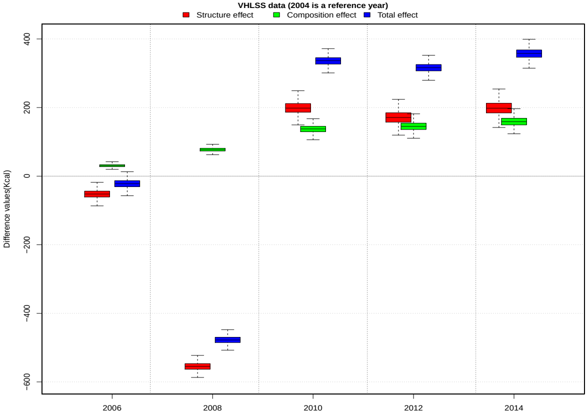
Year	Model	GAMGauId	GAMGauLog	GAMGamLog	Choice
2004	DLM	-11.64***	-10.20***	-14.70***	DLM
	GAMGauId		4.67***	-7.78***	
	GAMGauLog			-11.89***	
2006	DLM	17.14***	12.79***	9.3***	GAMGauId
	GAMGauId		-19.6***	-29.49***	
	GAMGauLog			-11.9***	
2008	DLM	62.38***	21.77***	13.67***	GAMGauId
	GAMGauId		-87.88***	-95.8***	
	GAMGauLog			-19.98***	
2010	DLM	19.26***	-10.02***	-16.74***	GAMGauId
	GAMGauId		-73.06***	-79.93***	
	GAMGauLog			-15.04***	
2012	DLM	58.25***	2.41*	-5.34***	GAMGauId
	GAMGauId		-164.72***	-149.59***	
	GAMGauLog			-16.28***	
2014	DLM	70.01***	-23.97***	-49.93***	GAMGauId
	GAMGauId		-174.34***	-163.31***	
	GAMGauLog			-31.15***	

Note: *, **, and *** mean significant at 10%, 5%, and 1%, respectively

Estimated calorie-income relationships from 2004 to 2014 (Trinh et al., 2018)

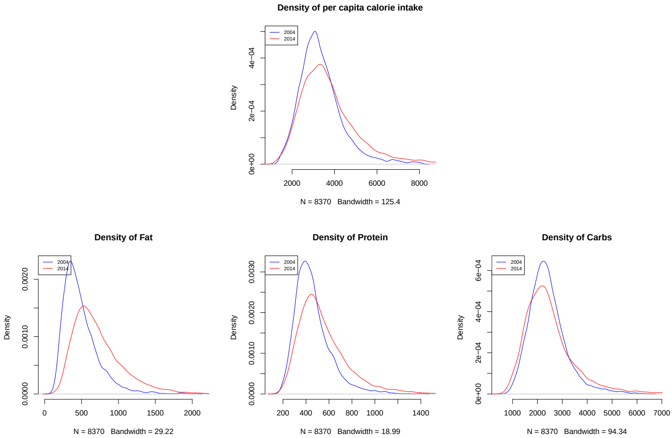


A byproduct: Decomposition of average per capita calorie intake difference (Trinh et al., 2018)

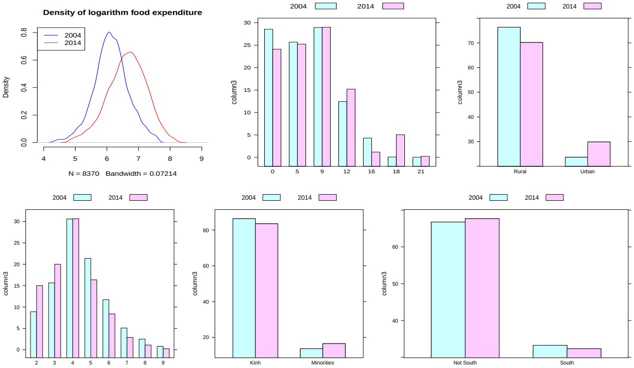


Year	Base case	Bandwidth sensitivity		
		0.80	1.25	1.50
	(1)	(2)	(3)	(4)
2004	0.1070	**0.0867**	0.1419	0.1902
2006	0.3273	0.3067	0.3701	0.4118
2008	**0.0053**	**0.0045**	**0.0061**	**0.0084**
2010	0.1911	0.1742	0.2320	0.3019
2012	0.3897	0.3505	0.4244	0.4749
2014	0.3417	0.2589	0.4803	0.6615

What we observe: Densities of per capita calorie intake, total and by macronutrient



What we observe: Population changes between 2004 and 2014



- **Question**:
  - Outcome Y: calorie intake (total or by macronutrient) with distributions F_Y^{2014} and F_Y^{2004}.
  - Object of interest:

$$\Delta_Y^{\nu} = \nu(F_Y^{2014}) - \nu(F_Y^{2004})$$

    where $\nu(.)$ measures a feature (mean, median, quantile. . .) of the considered distribution.
  - How the observed difference Δ_Y^{ν} is related to differences between F_X^{2004}, and F_X^{2014}, the joint distributions of covariates in 2004 and 2014?

- **Decomposition methods in economics** (Fortin *et al.*, 2011)

$$\downarrow$$

Trinh Thi, Simioni, and Thomas-Agnan (2018): "Decomposition of changes in the consumption of macronutrients in Vietnam between 2004 and 2014," submitted.

- Decomposition methods (1):

  1. Contributions to economic growth (Solow, 1957)
  2. Labor economics: Oaxaca (1973) and Blinder (1973)
  3. Many developments (Fortin *et al.*, 2011) $\supset$ Rothe (2015)

- Decomposition methods (2):

  1. Basic idea: Total variation = structure effect + composition effect
  2. Main tools: **Conterfactuals** or

     For example: $F_Y^{2004|2014}$*, i.e. what would have been the distribution of outcome in 2004 if the covariates had been distributed as in 2014. Then*

$$\Delta_Y^{\nu} = \left(\nu\left(F_Y^{2014}\right) - \nu\left(F_Y^{2004|2014}\right)\right) + \left(\nu\left(F_Y^{2004|2014}\right) - \nu\left(F_Y^{2004}\right)\right)$$

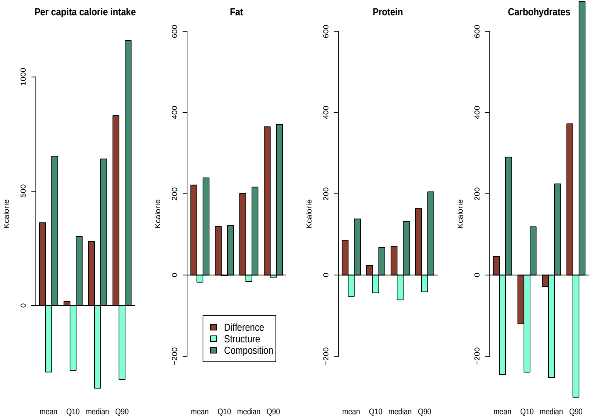
  3. Decompose decomposition effect with respect to each covariate

- Rothe (2015) $\rightarrow$ **Semiparametric approach using copulas** (Sklar, 1959):

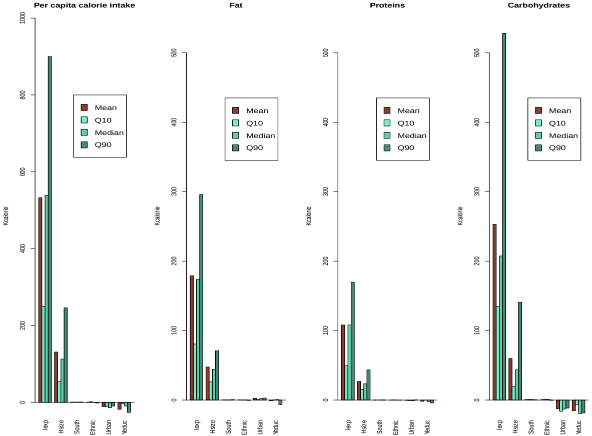
  Composition effect = Direct effects + interactions + dependence effects

Methodology

Total differences, composition and structure effects (Trinh Thi et al., 2018)



Direct contributions to the composition effects (Trinh Thi et al., 2018)



- So far:

Total energy intake $\rightarrow$ **calorie-income relationship**

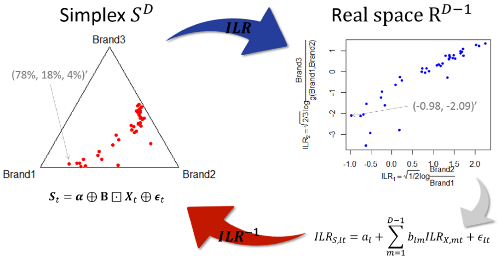
$\downarrow$

Energy intakes coming from **macronutrients** (Protein, Fat, and Carbohydrates):
Main drivers of their evolution between 2004 and 2014

- But, national nutritional strategy $\rightarrow$ "ideal" balanced diet: (Protein: 14%; Lipid: 18%; Carbohydrates: 68%)
- Focus on a new object: vector of **shares**, i.e. (S_P, S_F, S_C) with $S_P + S_F + S_C = 1$
- Statistical tools: **CO**mpositional **D**ata **A**nalysis (Pawlowsky-Glahn and Buccianti, 2011, and Pawlowsky-Glahn *et al.*, 2015)
- Trinh Thi, Morais, Thomas-Agnan and Simioni (2018). "Relations between socio-economic factors and nutritional diet in Vietnam from 2004 to 2014: new insights using compositional data analysis." *Statistical Methods in Medical Research*.

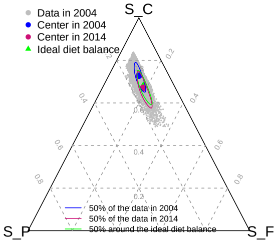
From Joanna Morais's PhD thesis defense slides (October 2017):

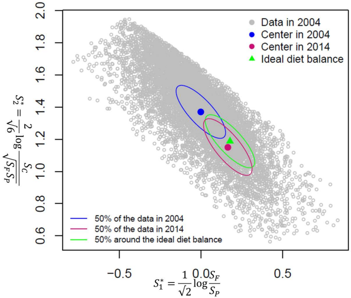
Compositional Data Analysis (CODA) method



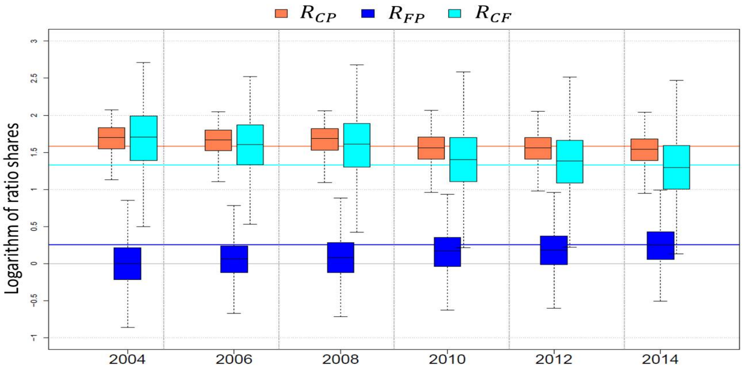
ILR^{-1}: inverse isometric log ratio transformation

Plot centers in 2004 and 2014 compared to the "ideal" diet balance (S_P=14%, S_F=18%,S_C=68%) in ternary diagram in the simplex and in ILR coordinates.





Boxplots of macronutrients log–ratio of shares by year. The line shows the value corresponding to the ideal diet for each log–ratio of share.



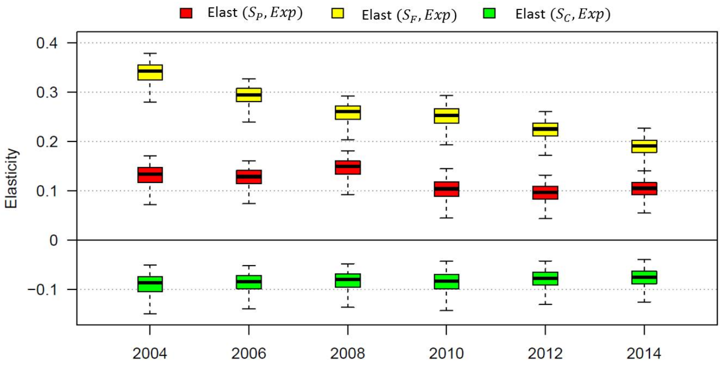
Estimated coefficients when $S_1^* = \frac{2}{\sqrt{6}} \log \frac{S_C}{\sqrt{S_F S_P}}$ (Carbohydrate against other shares) is the response

Estimator	Description	2004	2006	2008	2010	2012	2014
$Constant$		2.722***	2.561***	2.505***	2.369***	2.269***	2.136***
$log(Exp)$	food expenditure per year (US$) (in log)	-0.265***	-0.241***	-0.232***	-0.214***	-0.194***	-0.182***
$Urban$	Rural	0.064***	0.069***	0.093***	0.072***	0.045***	0.037***
$HSize$	3 people	0.178***	0.142***	0.152***	0.135***	0.119***	0.115***
	4 people	0.250***	0.212***	0.232***	0.2***	0.187***	0.16***
	5 people	0.320***	0.281***	0.301***	0.271***	0.24***	0.212***
	$\geq$ 6 people	0.423***	0.36***	0.384***	0.345***	0.305***	0.27***
$Ethnic$	Minorities	0.067***	0.05***	0.061***	0.049***	0.053***	0.069***
$Gender$	Female	-0.020***	-0.027***	-0.025***	-0.028***	-0.035***	-0.031***
$Educ$	Secondary, High school	-0.024***	-0.019***	-0.018***	-0.033***	-0.024***	-0.014*
	University	-0.071***	-0.047***	-0.063***	-0.061***	-0.063***	-0.035**
$Area$	Midlands Northern Mountains	0.001	0.011	0.03***	0.038***	0.042***	0.055***
	Northern Central Coast	0.02**	0.048***	0.033***	0.076***	0.098***	0.129***
	Central Highlands	0.011	0.05***	0.042***	0.096***	0.095***	0.128***
	South East	-0.020*	0.009	-0.007	0.025 **	0.036***	0.048***
	Mekong River Delta	0.014*	0.044***	0.057***	0.064***	0.061***	0.142***

Estimated coefficients when $S_2^* = \frac{1}{\sqrt{2}} \log \frac{S_F}{S_P}$ (Fat against Protein) is the response.

$Constant$		-0.719***	-0.524***	-0.276***	-0.455***	-0.380***	-0.139***
$log(Exp)$	food expenditure per year (US$) (in log)	0.147***	0.117***	0.079***	0.105***	0.091***	0.061***
$Urban$	Rural	-0.04***	-0.034***	-0.057***	-0.04***	-0.015 **	-0.011 *
$HSize$	3 people	-0.1***	-0.07***	-0.061***	-0.066***	-0.047***	-0.033***
	4 people	-0.137***	-0.102***	-0.105***	-0.089***	-0.074***	-0.038***
	5 people	-0.174***	-0.144***	-0.145***	-0.129***	-0.097***	-0.058***
	$\geq$ 6 people	-0.244***	-0.184***	-0.195***	-0.175***	-0.136***	-0.086***
$Ethnic$	Minorities	-0.039***	-0.026***	-0.024 **	0.017 **	0.014 *	-0.032***
$Gender$	Female	0.015 **	0.023***	0.017 **	0.021***	0.026***	0.023***
$Educ$	Secondary, High school	0.035***	0.028***	0.026***	0.042***	0.045***	0.023***
	University	0.058***	0.035***	0.042***	0.045***	0.067***	0.028 **
$Area$	Midlands Northern Mountains	0.015	0.009	0.009	-0.017 *	-0.015 .	0.002
	Northern Central Coast	-0.055***	-0.077***	-0.051***	-0.079***	-0.104***	-0.11***
	Central Highlands	-0.005	-0.042***	-0.007	-0.069***	-0.077***	-0.088***
	South East	-0.053***	-0.072***	-0.029***	-0.032***	-0.047***	-0.056***
	Mekong River Delta	-0.125***	-0.145***	-0.134***	-0.103***	-0.104***	-0.173***

Boxplot of food expenditure elasticities of macronutrient consumption shares (application of Morais et al., 2018).



I know you need some sugar now

1. Le D.T., Trinh Thi H., Thomas-Agnan C., Simioni M., Beal T. and D.S. Nguyen (2018). "Macronutrient balances and body mass index: a new insight using **compositional data analysis** with a **total** at various **quantile** orders." In progress.

2. Changes in food marketing channel (development of supermarkets + more processed food) and nutrition in Vietnam (joint work with CIAT, Hanoi)

3. Food security and climate change: "Forest as insurance mechanism against climatic shocks" (PERENA project funded by INRA-CIRAD program GloFoodS)

4. Big issue: energy intake → quality of the diet (Europe: Mediterranean diet)

5. Ex ante evaluation of the efficiency of nutritional policies (for example, taxes on sweetened food products: colas. . .)

6. . . .

It is all a matter of choice.

"What if everything is an illusion and nothing exists? In that case, I definitely overpaid for my carpet." (Woody Allen quoted by D. McFadden in 2006).

INRA SCIENCE & IMPACT IREEDS SCIENCE NATURE CONNECTION VCREME Connecting talents for development

- Aguiar, M. and Hurst, E. (2013). "Deconstructing life cycle expenditure." Journal of Political Economy **121(3)**: 437-492.
- Banerjee, A. V. (2016). "Policies for a better-fed world." Review of World Economics **152(1)**: 3-17.
- Blinder, A.S. (1973). "Wage discrimination: Reduced form and structural estimates." *Journal of Human Resources* **8**: 436-455.
- Blundell, R. and Horowitz, J. L. (2007). "A non-parametric test of exogeneity." *Review of Economic Studies* **74**: 1035-1058.
- Bouis, H. E. and Haddad, L. J. (1992). "Are estimates of calorie-income elasticities too high? A recalibration of the plausible range." *Journal of Development Economics* **39**: 333-364.
- Combris, P. (2006). "Le poids des contraintes économiques dans les choix alimentaires." *Cahiers de Nutrition et de Diététique* **41(5)**: 279-284.
- Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Baltimore and London: The John Hopkins University Press.
- Fortin, N. and Lemieux, T. and Firpo, S. (2011). "Decomposition methods in economics." In: Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, Amsterdam: North-Holland, 1-102.
- Le Nguyen B.K., Le Thi H.,Nguyen Do V.A., Tran Thuy N., Nguyen HUU C., Thanh Do T., Deurenberg P. and I. Khouw (2013). "Double burden of undernutrition and overnutrition in Vietnam in 2011: results of the SEANUTS study in 0511-year-old children." British Journal of Nutrition **110(S3)**: S45-S56.
- Masters, W.A. *et al.* (2016)."The nutrition transition and agricultural transformation: A Preston curve approach." *Agricultural Economics* **47(Supplement)**: 97-114.
- Morais, J., Thomas-Agnan C. and M. Simioni (2018). "Model selection and interpretation of explanatory variables impacts in compositional models." *Austrian Journal of Statistics*, forthcoming.
- Morais, J. (2017). *Impact of media investments on brands' market shares: a compositional data analysis approach*. PhD thesis manuscript, University of Toulouse.

- Muth, M. K. and Karns, S. A. and Nielsen, S. J. and Buzby, J. C. and Wells, H. F. (2011). "Consumer-level food loss estimates and their use in the ERS loss-adjusted food availablity data." USDA-ESR, Technical Bulletin Number 1927.
- Nguyen, T.T. and Hoang, M.V. (2018). "Non-communicable diseases, food and nutrition in Vietnam from 1975 to 2015: The burden and national response." *Asia Pacific Journal of Clinical Nutrition* **27(1)**: 19-28.
- Ogundari, K. and Abdulai, A. (2013). "Examining the heterogeneity in calorie–income elasticities: A meta-analysis." *Food Policy* 40: 119-128.
- Oaxaca, R. (1973). " Male-female wage differentials in urban labor markets." *International Economic Review* **14**: 693-709.
- Pawlowsky-Glahn, V. and Buccianti, A. (eds) (2011). *Compositional data analysis: Theory and applications*. John Wiley & Sons.
- Pawlowsky-Glahn, V. and Egozcue, J. and Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. John Wiley & Sons.
- Popkin, B.M. (1993). "Nutritional patterns and transitions." *Population and Development Review* **19(1)**: 138-157.
- Popkin, B.M. (1994). "The nutrition transition in low-income countries: an emerging crisis." *Nutrition Review* **52(9)**:285-298.
- Racine, J.S. and Parmeter, C. (2014). "Data-Driven Model Evaluation: A Test for Revealed Performance." In: Racine, J.D and Su, L. and Ullah, A. (eds), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford: Oxford University Press, 308-345.
- Rothe, C. (2015. "Decomposing the composition effect: The role of covariates in determining between-group differences in economic outcomes." *Journal of Business and Economic Statistics* **33(3)**: 323-337.

- Santeramo, F G. and Shabnamb, N. (2015). "The income-elasticity of calories, macro- and micro-nutrients: What is the literature telling us?" *Food Research International* **76**: 932-937.
- Sklar, A. (1959). "Fonctions de répartition à n dimensions et leurs marges." *Publications de l'Institut de Statistique de l'Université de Paris* **8**: 229-231.
- Solow, R. (1957). "Technical change and the aggregate production function." *Review of Economics and Statistics* **39**: 312-320.
- Trinh Thi H., Simioni M., and C. Thomas-Agnan (2018): "Assessing the nonlinearity of the calorie-income relationship: an estimation strategy - With new insights on nutritional transition in Vietnam," *World Development* 110: 192-204.
- Trinh Thi H., Morais J., Thomas-Agnan C. and M. Simioni (2018). "Relations between socio-economic factors and nutritional diet in Vietnam from 2004 to 2014: new insights using compositional data analysis." *Statistical Methods in Medical Research*, DOI:10.1177/0962280218770223.
- Trinh Thi H., Simioni M., and C. Thomas-Agnan (2018): "Decomposition of changes in the consumption of macronutrients in Vietnam between 2004 and 2014," submitted.
- Zezza, A. and Carletto, C. and Fiedler, J. L. and Gennari, P. and D. Jolliffe (eds) (2017). Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES). *Food Policy*, Special Issue, Volume 72.
- Zhou, J. and Yu, X. (2015). "Calorie elasticities with income dynamics: Evidence from the literature." *Applied Economic Perspectives and Policy* **37(4)**:575-601.

Xin cam on

- Following Abdulai and Aubert (2004), most of the empirical studies use the "classical" **double-log** (parametric) specification

$$\log(PCCI) = \alpha_0 + \alpha_1 \log(INCOME) + \alpha_2(\log(INCOME))^2 + \textstyle\sum_j \beta_j x_j + \epsilon$$

- But, recently, Tian and Yu (2015), and Nie and Sousa-Poza (2016):

$$PCCI = \alpha_0 + \underbrace{s(INCOME)}_{s(.)\text{unknown}} + \sum_j \beta_j x_j + \epsilon$$

$\rightarrow$ **Semiparametric** specifications.

- General structure of GAM (Wood, 2017):

$$g\left(\mathbb{E}(Y_i|X = x_i, Z = z_i)\right) = x_i'\beta + \sum_j f_j(z_{ij})$$

where

- Y is a response variable $\sim$ distribution belonging to the exponential family
- $g(.)$ is a known link function ($\log$ for example)
- X is a vector of covariates acting linearly, and β are the associated parameters
- Z is a vector of covariates acting nonlinearly.
- Each $f_j(.)$ is a smooth function of the corresponding covariate which acts nonlinearly

- For instance, the double-log model can be expressed as a GAM model with

$$\mathbb{E}(\log(PCCI)|INCOME, x_j) = \alpha_0 + \alpha_1 \log(INCOME) + \alpha_2(\log(INCOME))^2 + \sum_j \beta_j x_j$$

and $\log(PCCI)$ normally distributed.

- Similarly, Tian and Yu (2015), or Nie and Sousa-Poza (2016):

$$\mathbb{E}(PCCI|INCOME, x_j) = \alpha_0 + s(INCOME) + \sum_j \beta_j x_j$$

and $PCCI$ normally distributed.

- More general semiparametric model:

$$\log\left(\mathbb{E}(PCCI|INCOME, x_j)\right) = \alpha_0 + s(INCOME) + \sum_j \beta_j x_j$$

- Estimation of GAM $\rightarrow$ Regression splines (Wood, 2006)

- Thin plate regression splines (Wood, 2006):

$$s(x) = \gamma_0 + \gamma_1 x + \sum_{i=1}^{n} \delta_i (x - x_i)^3$$

$\rightarrow$ Usefulness when testing the linearity of $s(.)$ (tests in the appendix of the paper)

- **Data-driven** test proposed by Racine and Parmeter (2014).
- Simple idea when using a cross-section $n = n_1 + n_2$

$$\underbrace{1, 2, \cdots, n_1}_{\text{training sample: } S_T} \qquad \underbrace{n_1 + 1, \cdots, n}_{\text{evaluation sample: } S_E}$$

$$\downarrow \qquad\qquad \downarrow$$

$$\text{Estimation} \qquad \text{Prediction: } \widehat{\text{PCCI}}_{i,\text{training}}$$

- The loss criterion

$$\underbrace{\sum_{i \in S_E} \left(\text{PCCI}_i - \widehat{\text{PCCI}}_{i,\text{training}} \right)^2}_{\text{Estimate of the true error (Efron, 1982)}} \tag{1}$$

- Resample 10.000 times $\rightarrow$ empirical distribution of some loss criterion
- $t-$ paired test to compare the empirical distributions of the loss criterion for two competing models, whatever their types.

- **Problem: Comparison between years ?**
- Given two year t_0 and t_1, with $t_0 < t_1$, we may want to measure:

$$\Delta PCCI_{t_0 \to t_1} = \mathbb{E}_{t_1}\left(PCCI\right) - \mathbb{E}_{t_0}\left(PCCI\right)$$

- By law of iterated expectations:

$$\begin{aligned} \Delta PCCI_{t_0 \to t_1} = \quad & \mathbb{E}_{t_1}\left(\mathbb{E}\left(PCCI|INCOME, Z\right)\right) \\ & - \mathbb{E}_{t_0}\left(\mathbb{E}\left(PCCI|INCOME, Z\right)\right) \end{aligned}$$

- Note that

$$\mathbb{E}\left(PCCI|INCOME, Z\right) = m_t\left(INCOME, Z\right)$$

where $m_t\left(.\right)$ denotes the model chosen for year t by the revealed performance test.

- Then

$$\Delta PCCI_{t_0 \rightarrow t_1} = \mathbb{E}_{t_1}\left(m_{t_1}\left(INCOME, Z\right)\right) - \mathbb{E}_{t_0}\left(m_{t_0}\left(INCOME, Z\right)\right)$$

- Basic idea of decomposition method $\rightarrow$ contrefactuals.
- Here:

$$\mathbb{E}_{t_1}\left(m_{t_0}\left(INCOME, Z\right)\right)$$

- Thus

$$\Delta PCCI_{t_0 \rightarrow t_1} = \underbrace{\mathbb{E}_{t_1}\left(m_{t_1}\left(INCOME, Z\right)\right) - \mathbb{E}_{t_1}\left(m_{t_0}\left(INCOME, Z\right)\right)}_{\text{“structure” effect}} + \underbrace{\mathbb{E}_{t_1}\left(m_{t_0}\left(INCOME, Z\right)\right) - \mathbb{E}_{t_0}\left(m_{t_0}\left(INCOME, Z\right)\right)}_{\text{“Composition” effect}}$$

- Computation using empirical counterparts of the expectations, i.e. simple average values.

- Setup:
  1. Response variable: $Y \in \mathbb{R}$,
  2. Endogenous regressor: $X \in \mathbb{R}$,
  3. Unknown relationship: $Y = g(X) + \varepsilon$
  4. Endogeneity: $\mathbb{E}(\varepsilon|X = x) \neq 0$
- Instrumental variable: $W \in \mathbb{R} \Leftrightarrow$ Conditional mean restriction:

$$\mathbb{E}(Y - g(X)|W) = 0$$

- Now define the conditional mean function $G(x) = \mathbb{E}(Y|X = x)$
- Blundell and Horowitz (2007): Testing exogeneity of $X \Leftrightarrow$ testing that

$$\mathbb{E}(Y - G(X)|W) = 0$$

- Test statistics:

$$\tau_n = \int S^2(x)dx \approx \sum_{k=1}^{K} \text{Weight}_k \ \chi_k^2$$

where $S_n(x)$ is the sample analog of
$S(x) = \mathbb{E}([Y - G(X)]f_{X,W}(x, W))$

Back

- Economic outcome: Y^{2004} and Y^{2014}
- Covariates: $X^{2004} = (X_1^{2004}, X_2^{2004})$ and $X^{2014} = (X_1^{2014}, X_2^{2014})$
- Corresponding cumulative distributions functions: F_Y^{2004}, F_Y^{2014}, F_X^{2004}, and F_X^{2014}
- Total difference:

$$\Delta_Y^{\nu} = \nu(F_Y^{2014}) - \nu(F_Y^{2004})$$

  where $\nu(.)$ measures a features (mean, median, ...) of the considered distribution.

- **Aim of decomposition method**: to understand how the observed difference Δ_Y^{ν} is related to differences between the distributions F_X^{2004}, and F_X^{2014}?

- Reminder:

$$F_Y^{2004}(y) = \int F_{Y|X}^{2004}(y,x)dF_X^{2004}(x) \equiv F_Y^{2004|2004}(y)$$

- Similarly:

$$F_Y^{2014}(y) = \int F_{Y|X}^{2014}(y,x)dF_X^{2014}(x) \equiv F_Y^{2014|2014}(y)$$

- $\rightarrow$ **Counterfactual distribution**:

$$F_Y^{2004|2014}(y) = \int F_{Y|X}^{2004}(y,x)dF_X^{2014}(x)$$

- **Basic idea of the decomposition**:

$$\begin{aligned}\Delta_Y^{\nu} &= \underbrace{\left(\nu\left(F_Y^{2014}\right) - \nu\left(F_Y^{2004|2014}\right)\right)}_{\text{structure effect}} + \underbrace{\left(\nu\left(F_Y^{2004|2014}\right) - \nu\left(F_Y^{2004}\right)\right)}_{\text{composition effect}} \\ &= \Delta_S^{\nu} + \Delta_X^{\nu}\end{aligned}$$

- **Problem**: How to break down the composition effect for the different covariates?
- $\exists$ solutions (Oaxaca and Blinder, for instance) $\rightarrow$ More general framework.
- **General framework**: direct contributions, interaction effects, and "dependence effect".
- **Dependence effect**: effect of between-year difference in the dependence pattern among the covariates.
- Rothe (2015) $\rightarrow$ **Copulas**, i.e. (Sklar's Theorem, 1959):

$$F_X^t(x) = C^t(F_{X_1}^t(x_1), F_{X_2}^t(x_2)) \quad \text{for} \quad t \in \{2004, 2014\}$$

- Example: Gaussian copula, or

$$C_\Sigma(u) = \Phi_\Sigma^d \left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d) \right)$$

- Usefulness: Prevent the problem of the curse of dimension when estimating $F_X^t(x)$.

- New notations:

$$F_Y^{t|s,\mathbf{k}} = \int F_{Y|X}^t(y,x)dF_X^{s,\mathbf{k}}(x)$$

with

$$F_X^{s,\mathbf{k}}(x) = C^s(F_{X_1}^{k_1}(x_1), F_{X_2}^{k_2}(x_2))$$

and $\mathbf{k} = (k_1, k_2)$ where k_1 (resp. k_2) is equal to either 2004 or 2014.

- So

$$\begin{aligned}
\Delta_X^\nu &= \nu\left(F_Y^{2004|2014}\right) - \nu\left(F_Y^{2004}\right) \\
&= \nu\left(F_Y^{2004|2014,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right) \\
&= \underbrace{\left(\nu\left(F_Y^{2004|2014,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{1}}\right)\right)}_{\text{Dependence effect}} + \underbrace{\left(\nu\left(F_Y^{2004|2004,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right)\right)}_{\text{Differences in marginal dist. of covariates}} \\
&= \Delta_D^\nu + \beta^\nu(\mathbf{1})
\end{aligned}$$

- And so on $\rightarrow$

$$\Delta_X^\nu = \beta^\nu(\mathbf{e^1}) + \beta^\nu(\mathbf{e^2}) + \Delta_M^\nu(\mathbf{1}) + \Delta_D^\nu,$$

- Nonparametric estimation of univariate CDF:

$$\widehat{F}^t_{X_j}(x_j) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{I}(X^t_{ji} \leq x_j)$$

- Estimation of conditional CDFs using the distributional approach of Foresi and Peracchi (1995):

$$F^t_{Y|X}(y, x) \equiv \Phi(x'\delta^t(y))$$

- Choice of Gaussian copula and estimation of dependence parameters using minimum distance estimator (Weiß, 2015, and Jiryaie et al., 2016):

$$\widehat{\theta}^t = \arg\min_{\theta} \sum_{i=1}^{n_t} \left(\widehat{F}^t_X(X^t_{1i}, \ldots, X^t_{di}) - C_\theta(\widehat{F}^t_{X_1}(X^t_{1i}), \ldots, \widehat{F}^t_{X_d}(X^t_{di})) \right)$$

- Asymptotic normality of estimated effects (Rothe, 2015) $\rightarrow$ nonparametric bootstrap to get standard errors.

It's okay?

Back