# The nutrition transition in Vietnam: some recent empirical insights 

Michel Simioni

## To cite this version:

Michel Simioni. The nutrition transition in Vietnam: some recent empirical insights. 11. Vietnam Economist Annual Meeting (VEAM 2018), Jun 2018, Hanoi, Vietnam. 65 p. [Powerpoint]. hal02787856

## HAL Id: hal-02787856 <br> https://hal.inrae.fr/hal-02787856

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Nutrition Transition in Vietnam <br> 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 $\Rightarrow$ Vietnam from one of the world's poorest country to a lower middle-income country
- Two majors features:

1. Gross National Income: 431 in $1989 \uparrow 1691$ in 2016 (USD of 2010 , World Bank)
2. Poverty rate: $58 \%$ in $1993 \downarrow 12 \%$ 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) $\Rightarrow$ costs for the health sector and the economy in general.
- Nutrition transition $\rightarrow$ 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.

Introduction

## 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:

## theNUTRITION TR_NSITION



## Stages of the Nutrition Transition



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



Note: Symbols are sized by population, with year shown as $1990=$ green circles ( 132 countries), $2005=$ red triangles ( 154 countries), and $2010=b$ bue squares ( 155 countries). Lines show each year's local means and confidence intervals for each year estimated by -polyci-, weighted by population and with a bandwidth of 0.75 .
Source: FAO Food Balance Sheets; GDP and population are from PWT 8.1.

Share of dietary energy from nonstaples, without India and China FAO Food Balance Sheet estimates for 1990, 2000, and 2010


Note: Symbols are sized by population, with year shown as $1990=$ green circles ( 132 countries), $2005=$ red triangles ( 154 countries), and $2010=$ blue squares ( 155 countries). Lines show each year's local means and confidence intervals estimated by -lpolyci-, weighted by population and with a bandwidth of 0.75 .
Source: FAO Food Balance Sheets; GDP and population are from PWT 8.1.

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)



## 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 ( $\mathrm{BMI} \geq 25 \mathrm{~kg} / \mathrm{m}^{2}$ ) | 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 ${ }^{4}$ Behavioral factors | 30.1 | 27.8 | 32.3 | 32.4 | 27.9 | 36.1 |
| 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 $\geq 140$ and/or diastolic blood pressure $\geq 90 \mathrm{mmHg}$, or currently on medication
2. Whole blood value $\geq 6.1 \mathrm{mmol} / \mathrm{L}$ (or $110 \mathrm{mg} / \mathrm{L}$ ) but $<7.2 \mathrm{mmol} / \mathrm{L}$ (or $126 \mathrm{mg} / \mathrm{L}$
${ }^{3}$ Whole blood value $\geq 7.2 \mathrm{mmol} / \mathrm{L}$ or currently on medication
${ }^{4}$ Total cholesterol $\geq 5.0 \mathrm{mmmol} / \mathrm{L}$ (or $190 \mathrm{mg} / \mathrm{dl}$ ) or currently on medication
${ }^{5}$ Level of physical activity $<600$ MET-min per week


# 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 $\ddagger$

|  | 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 | $0.0 \dagger$ | $10 \cdot 5$ | $5.8 \dagger$ | $3.4 \dagger$ | $5 \cdot 2 \dagger$ | $4.4 \dagger$ | $9.2 \dagger$ | $8.7 \dagger$ | $8.9 \dagger$ |
| Underweight | 2.2 | $9 \cdot 3$ | 6.2 | 3.4 | $5 \cdot 2$ | $4.4 \dagger$ | $13.5 \dagger$ | 14.1† | $13.8 \dagger$ |
| Thinness | $2 \cdot 1$ | $8.1 \dagger$ | 5.3† | 1.6 | $1.1 \dagger$ | 1.3 | 10.4 | 8.8 | $9.6 \dagger$ |
| Overweight | 4.3 | $5 \cdot 0$ | 4.7 | $15.2 \dagger$ | $11.6 \dagger$ | $13.3 \dagger$ | $16.6 \dagger$ | $14.8 \dagger$ | $15.7 \dagger$ |
| Obesity | 0.0 | 4.9 | 2.7 | $3.4 \dagger$ | 13.0* $\dagger$ | $8.7 \dagger$ | $11.0 \dagger$ | 25.1*† | $18.0 \dagger$ |
| Rural |  |  |  |  |  |  |  |  |  |
| Stunting | $8 \cdot 1$ | 16.7 | 13.0 | 17.5 | 19.1 | 18.3 | 14.7 | $20 \cdot 6$ * | 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 | $5 \cdot 2$ | 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 atter correction for age differences ( $P<0-05$ ).
$\dagger$ Values were significantly different from those of rural areas after correction for age.
$\ddagger$ Stunting: height-for-age $z$-score $\leq-2 \mathrm{sD}$; underweight: weight-for-age $z$-score $\leq-2 \mathrm{sD}$; thinness: BMI-for-age $z$-score $\leq-2 \mathrm{sD}$; overweight: $\geq 1 \mathrm{sD}$ and obese: $>2 \mathrm{sD}$ in children aged $>5$ years; overweight: $>2 \mathrm{sD}$ and obese: $>3 \mathrm{sD}$ 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)
- Vu, Linh Hoang (2008). Essays on the economics of food production and consumption in Vietnam. University of Minnesota, PhD Thesis.

$$
\downarrow
$$

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

- Every two years since 2002: 8 surveys.
- VHLSS $\supset$ 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

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 $(\geq 0)$ (Bouis and Haddad, 1992, and Muth et al., 2011)
3. Quality of the diet $\rightarrow$ micronutrient intakes (General Nutrition Survey)


- 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 $\rightarrow$ 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

$$
\downarrow
$$

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

$$
\downarrow
$$

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

| Year | Model | GAMGauld | GAMGauLog | GAMGamLog | Choice |
| :--- | :--- | :--- | :---: | :---: | :---: |
| 2004 | DLM | $-11.64^{* * *}$ | $-10.20^{* * *}$ | $-14.70^{* * *}$ | DLM |
|  | GAMGauld |  | $4.67^{* * *}$ | $-7.78^{* * *}$ |  |$]$

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 | $\mathbf{0 . 0 8 6 7}$ | 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:

$$
\Delta_{Y}^{\nu}=\nu\left(F_{Y}^{2014}\right)-\nu\left(F_{Y}^{2004}\right)
$$

where $\nu($.$) measures a feature (mean, median, quantile...) of the consi-$ dered distribution.

- How the observed difference $\Delta_{Y}^{\nu}$ 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)

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: $F_{Y}^{2004 \mid 2014}$, i.e. what would have been the distribution of outcome in 2004 if the covariates had been distributed as in 2014. Then

$$
\Delta_{Y}^{\nu}=\left(\nu\left(F_{Y}^{2014}\right)-\nu\left(F_{Y}^{2004 \mid 2014}\right)\right)+\left(\nu\left(F_{Y}^{2004 \mid 2014}\right)-\nu\left(F_{Y}^{2004}\right)\right)
$$

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

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

$\overline{\bar{\prime}}$

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




Proteins
Carbohydrates



- So far:

$$
\text { Total energy intake } \rightarrow \text { calorie-income relationship }
$$

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. $\left(S_{P}, S_{F}, S_{C}\right)$ with $S_{P}+S_{F}+S_{C}=1$
- Statistical tools: COmpositional Data Analysis (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

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


Plot centers in 2004 and 2014 compared to the "ideal" diet balance ( $S_{P}=\mathbf{1 4 \%}$, $S_{F}=18 \%, S_{C}=68 \%$ ) 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*** |
| $\log (E x p)$ | 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*** |

Michel Simioni (INRA)
The Nutrition Transition in Vietnam
VEAM, Hanoi, June 2018

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*** |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\log (E x p)$ | 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).



## INRA $\because$ IREEDS ${ }^{\circ}$ VCREME

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 $\rightarrow$ 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).

- 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.

Michel Simioni (INRA)
The Nutrition Transition in Vietnam
VEAM, Hanoi, June 2018

- Muth, M. K. and Karns, S. A. and Nielsen, S. J. and Buzby, J. C. and Wells, H. F. (2011). "Consumerlevel 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 metaanalysis." 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.

Michel Simioni (INRA)
The Nutrition Transition in Vietnam
VEAM, Hanoi, June 2018

- 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.

Michel Simioni (INRA)
The Nutrition Transition in Vietnam
VEAM, Hanoi, June 2018


## Xin cam on

## 

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

$$
\begin{gathered}
\log (P C C I)=\alpha_{0}+\alpha_{1} \log (I N C O M E)+\alpha_{2}(\log (I N C O M E))^{2} \\
+\sum_{j} \beta_{j} x_{j}+\epsilon
\end{gathered}
$$

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

$$
P C C I=\alpha_{0}+\underbrace{s(I N C O M E)}_{s(.) \text { unknown }}+\sum_{j} \beta_{j} x_{j}+\epsilon
$$

$\rightarrow$ Semiparametric specifications.


- General structure of GAM (Wood, 2017):

$$
g\left(\mathbb{E}\left(Y_{i} \mid X=x_{i}, Z=z_{i}\right)\right)=x_{i}^{\prime} \beta+\sum_{j} f_{j}\left(z_{i j}\right)
$$

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 $\beta$ 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

$$
\begin{aligned}
& \mathbb{E}\left(\log (P C C I) \mid I N C O M E, x_{j}\right)= \\
& \alpha_{0}+\alpha_{1} \log (I N C O M E)+\alpha_{2}(\log (I N C O M E))^{2}+\sum_{j} \beta_{j} x_{j}
\end{aligned}
$$

and $\log (P C C I)$ normally distributed.

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

$$
\mathbb{E}\left(P C C I \mid I N C O M E, x_{j}\right)=\alpha_{0}+s(I N C O M E)+\sum_{j} \beta_{j} x_{j}
$$

and PCCI normally distributed.

- More general semiparametric model:

$$
\begin{aligned}
& \log \left(\mathbb{E}\left(P C C I \mid \text { INCOME }, x_{j}\right)\right)=\alpha_{0}+s(I N C O M E)+\sum \beta_{j} x_{j}
\end{aligned}
$$

- Estimation of GAM $\rightarrow$ Regression splines (Wood, 2006)
- Thin plate regression splines (Wood, 2006):

$$
s(x)=\gamma_{0}+\gamma_{1} x+\sum_{i=1}^{n} \delta_{i}\left(x-x_{i}\right)^{3}
$$

$\rightarrow$ 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 $n=n_{1}+n_{2}$

- The loss criterion

$$
\begin{equation*}
\underbrace{\sum_{i \in S_{E}}\left(\mathrm{PCCl}_{i}-\widehat{\mathrm{PCCl}}_{i, \text { training }}\right)^{2}}_{\text {Estimate of the true error (Efron, 1982) }} \tag{1}
\end{equation*}
$$

- 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 typese
- Problem: Comparison between years ?
- Given two year $t_{0}$ and $t_{1}$, with $t_{0}<t_{1}$, we may want to measure:

$$
\Delta P C C I_{t_{0} \rightarrow t_{1}}=\mathbb{E}_{t_{1}}(P C C I)-\mathbb{E}_{t_{0}}(P C C I)
$$

- By law of iterated expectations:

$$
\begin{aligned}
\Delta P C C I_{t_{0} \rightarrow t_{1}}= & \mathbb{E}_{t_{1}}(\mathbb{E}(P C C I \mid I N C O M E, Z)) \\
& -\mathbb{E}_{t_{0}}(\mathbb{E}(P C C I \mid I N C O M E, Z))
\end{aligned}
$$

- Note that

$$
\mathbb{E}(P C C I \mid I N C O M E, Z)=m_{t}(I N C O M E, Z)
$$

where $m_{t}($.$) denotes the model chosen for year t$ by the revealed performance test.

- Then

$$
\Delta P C C I_{t_{0} \rightarrow t_{1}}=\mathbb{E}_{t_{1}}\left(m_{t_{1}}(I N C O M E, Z)\right)-\mathbb{E}_{t_{0}}\left(m_{t_{0}}(I N C O M E, Z)\right)
$$

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

$$
\mathbb{E}_{t_{1}}\left(m_{t_{0}}(I N C O M E, Z)\right)
$$

- Thus

$$
\begin{aligned}
\Delta P C C I_{t_{0} \rightarrow t_{1}}= & \underbrace{\mathbb{E}_{t_{1}}\left(m_{t_{1}}(I N C O M E, Z)\right)-\mathbb{E}_{t_{1}}\left(m_{t_{0}}(I N C O M E, Z)\right)}_{\text {"structure" effect }}+ \\
& \underbrace{\mathbb{E}_{t_{1}}\left(m_{t_{0}}(I N C O M E, Z)\right)-\mathbb{E}_{t_{0}}\left(m_{t_{0}}(I N C O M E, Z)\right)}_{\text {"Composition" effect }}
\end{aligned}
$$

- 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: $\mathbb{E}(\varepsilon \mid X=x) \neq 0$

- Instrumental variable: $W \in \mathbb{R} \Leftrightarrow$ Conditional mean restriction:

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

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

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

- Test statistics:

$$
\tau_{n}=\int S^{2}(x) d x \approx \sum_{k=1}^{K} \text { Weight }_{k} \chi_{k}^{2}
$$

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

- Economic outcome: $Y^{2004}$ and $Y^{2014}$
- Covariates: $X^{2004}=\left(X_{1}^{2004}, X_{2}^{2004}\right)$ and $X^{2014}=\left(X_{1}^{2014}, X_{2}^{2014}\right)$
- Corresponding cumulative distributions functions:
$F_{Y}^{2004}, F_{Y}^{2014}, F_{X}^{2004}$, and $F_{X}^{2014}$
- Total difference:

$$
\Delta_{Y}^{\nu}=\nu\left(F_{Y}^{2014}\right)-\nu\left(F_{Y}^{2004}\right)
$$

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

- Aim of decomposition method: to understand how the observed difference $\Delta_{Y}^{\nu}$ is related to differences between the distributions $F_{X}^{2004}$, and $F_{X}^{2014}$ ?
- Reminder:

$$
F_{Y}^{2004}(y)=\int F_{Y \mid X}^{2004}(y, x) d F_{X}^{2004}(x) \equiv F_{Y}^{2004 \mid 2004}(y)
$$

- Similarly:

$$
F_{Y}^{2014}(y)=\int F_{Y \mid X}^{2014}(y, x) d F_{X}^{2014}(x) \equiv F_{Y}^{2014 \mid 2014}(y)
$$

- $\rightarrow$ Counterfactual distribution:

$$
F_{Y}^{2004 \mid 2014}(y)=\int F_{Y \mid X}^{2004}(y, x) d F_{X}^{2014}(x)
$$

- Basic idea of the decomposition:

$$
\begin{aligned}
& \Delta_{Y}^{\nu}=\underbrace{\left(\nu\left(F_{Y}^{2014}\right)-\nu\left(F_{Y}^{2004 \mid 2014}\right)\right)}_{\text {structure effect }}+\underbrace{\left(\nu\left(F_{Y}^{2004 \mid 2014}\right)-\nu\left(F_{Y}^{2004}\right)\right)}_{\text {composition effect }} \\
& =\Delta_{S}^{\nu}+\Delta_{X}^{\nu}
\end{aligned}
$$

- 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):

$$
F_{X}^{t}(x)=C^{t}\left(F_{X_{1}}^{t}\left(x_{1}\right), F_{X_{2}}^{t}\left(x_{2}\right)\right) \quad \text { for } \quad t \in\{2004,2014\}
$$

- Example: Gaussian copula, or

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

- Usefulness: Prevent the problem of the curse of dimension when estimating $F_{X}^{t}(x)$.
- New notations:

$$
F_{Y}^{t \mid s, \mathbf{k}}=\int F_{Y \mid X}^{t}(y, x) d F_{X}^{s, \mathbf{k}}(x)
$$

with

$$
F_{X}^{s, \mathbf{k}}(x)=C^{s}\left(F_{X_{1}}^{k_{1}}\left(x_{1}\right), F_{X_{2}}^{k_{2}}\left(x_{2}\right)\right)
$$

and $\mathbf{k}=\left(k_{1}, k_{2}\right)$ where $k_{1}$ (resp. $\left.k_{2}\right)$ is equal to either 2004 or 2014.

- So

$$
\begin{aligned}
\Delta_{X}^{\nu} & =\nu\left(F_{Y}^{2004 \mid 2014}\right)-\nu\left(F_{Y}^{2004}\right) \\
& =\nu\left(F_{Y}^{2004 \mid 2014, \mathbf{1}}\right)-\nu\left(F_{Y}^{2004 \mid 2004, \mathbf{0}}\right) \\
& =\underbrace{\left(\nu\left(F_{Y}^{2004 \mid 2014, \mathbf{1}}\right)-\nu\left(F_{Y}^{2004 \mid 2004,1}\right)\right)}_{\text {Dependence effect }}+\underbrace{\left(\nu\left(F_{Y}^{2004 \mid 2004, \mathbf{1}}\right)-\nu\left(F_{Y}^{2004 \mid 2004, \mathbf{0}}\right)\right)}_{\text {Differences in marginal dist. of covariates }}
\end{aligned}
$$

$$
=\Delta_{D}^{\nu}+\beta^{\nu}(\mathbf{1})
$$

- And so on $\rightarrow$

$$
\Delta_{X}^{\nu}=\beta^{\nu}\left(\mathbf{e}^{\mathbf{1}}\right)+\beta^{\nu}\left(\mathbf{e}^{\mathbf{2}}\right)+\Delta_{M}^{\nu}(\mathbf{1})+\Delta_{D}^{\nu}
$$



- Nonparametric estimation of univariate CDF:

$$
\widehat{F}_{X_{j}}^{t}\left(x_{j}\right)=\frac{1}{n_{t}} \sum_{i=1}^{n_{t}} \mathbb{I}\left(X_{j i}^{t} \leq x_{j}\right)
$$

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

$$
F_{Y \mid X}^{t}(y, x) \equiv \Phi\left(x^{\prime} \delta^{t}(y)\right)
$$

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

$$
\widehat{\theta}^{t}=\arg \min _{\theta} \sum_{i=1}^{n_{t}}\left(\widehat{F}_{X}^{t}\left(X_{1 i}^{t}, \ldots, X_{d i}^{t}\right)-C_{\theta}\left(\widehat{F}_{X_{1}}^{t}\left(X_{1 i}^{t}\right), \ldots, \widehat{F}_{X_{d}}^{t}\left(X_{d i}^{t}\right)\right)\right)
$$

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



## It's okay?

