

# Comparing imputation methods using expenditure surveys and unlinkable administrative data

An application to household consumption behaviour in the Netherlands<sup>1</sup>

Pim Kastelein

University of Amsterdam, Tinbergen Institute

13th May 2021

---

<sup>1</sup>ESCoE Conference on Economic Measurement 2021

# Inspiration

- Consumption heterogeneity central to modern macromodels  
(*Kaplan et al., 2018*)
- Marginal propensity to consume (MPC) out of wealth, income
- Balance sheet important for amplification of shocks  
(*liquidity, leverage*)
- Housing yield amplification → pension fund savings also!  
(*exhibit: the Netherlands*)

# Constraints in universe of Dutch households

Microdata from Statistics Netherlands (CBS), 2015:

- 30% of households with  $A < Y^{\text{monthly}}$
- 30% of homeowners with house < mortgage
- 73% of working age with mandatory pension fund participation
- Also income-rich households are constrained!

# Current policy debate

## **Due to pension and tax system design:**

- Households primarily invested in illiquid housing and pension
- Less liquidity and ability to smooth out shocks

## **Arising policy questions:**

- Should the Dutch pension fund system be more flexible?
  - ▶ Voluntary participation
  - ▶ Age-dependent contribution rates
  - ▶ Early withdrawal of pension accrual
- Dangers to retirement savings adequacy?

# This paper

- Can I use data of Statistics Netherlands to say something about household consumption?
  - ▶ Low cash-on-hand
  - ▶ High mortgage indebtedness
  - ▶ Mandatory pension fund contributions
- Demanding data requirements: jointly expenditures, income, wealth
- Combine all information available at Statistics Netherlands
- Missing data:
  - ▶ Expenditure surveys: consumption, but no income and wealth
  - ▶ Administrative data: income and wealth, but no consumption

# This paper

- Imputation required, but which way and do the techniques work?
- Try one novel and one widely used imputation technique to fill data gaps:
  - ▶ Supplement expenditure survey with unlinkable administrative data on income and wealth using household matching
  - ▶ Impute consumption in administrative data using household budget constraint:  $c = y - \Delta w$
- Evaluate fit using perfect link in 2015
- Cautiously estimate relation between  $c$ ,  $y$  and  $h$   
(*liquidity, leverage, pension pressure*)

# Main findings

- ① Household matching works pretty well
- ② Consumption imputation implies too many hand-to-mouth households
- ③ Both imputation techniques identify higher MPC's for constrained households

# Literature

## Methodology:

- Comparison expenditure survey with consumption imputation:  
*Koijen et al. (2014), Kreiner et al. (2014), Abildgren et al. (2018)*
- Consumption imputation using budget constraint:  
*Browning & Leth-Petersen (2003), Baker et al. (2018), Eika et al. (2020), Fagereng & Halvorsen (2017)*

## Economics:

- Role of leverage and liquidity for household consumption:  
*Campbell & Cocco (2007), Bunn & Rostom (2015), Mian & Sufi (2011), Cooper (2013)*
- Similar studies with data from Statistics Netherlands:  
*Bijlsma & Mocking (2017), Ji et al. (2019), Zhang (2019)*



# Data availability

## Administrative data:

- Available from 2011-2018
- Universe of Dutch households
- Constructed from tax returns and municipality registers
- Linkable datasets on household characteristics, income, assets, liabilities, residential location
- **No consumption**

## Expenditure survey:

- Conducted in 2012, 2013 (unlinkable) and 2015 (linkable)
- 6000, 5000 and 14000 Dutch households
- Households record purchases in diaries and fill in additional questionnaires
- **No detailed information on income and wealth**

## **Income and wealth imputation**

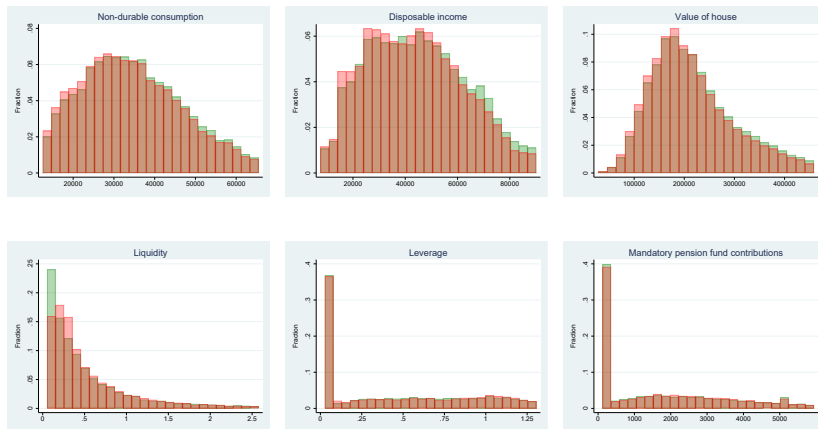
# Income and wealth imputation

- ES households are in administrative data and have highly specific information on them
- Match ES households to AD households using overlapping variables:
  - ▶ Categorical variables exactly (*e.g. composition, age cohort, municipality code, primary income source*)
  - ▶ Continuous variables with bandwidths (*e.g. income, value of house, paid interest on mortgage*)
- Different from Skinner (1987) regression prediction, where:
  - ▶ CEX to impute consumption into PSID
  - ▶ Regress consumption on overlapping variables and extrapolate

# Imputation results

- Overlap 81%-99% between two data sources per variable
- 85% of ES households matched
- 80% correct identification
- Median of two AD households per ES household

# Sample selection



**Figure:** Distribution comparison of variables in the entire expenditure survey of 2015 (in green) and in the resulting sample after the matching procedure (in red).

# Imputation errors

Median absolute errors small, but some large outliers

<i>Variable</i>	<i>Mean</i>	<i>Median</i>	<i>p95</i>
Disposable income	1,618	325	7,100
Bank account balance	17,758	3,554	76,833
Outstanding mortgage	11,608	0	58,607
Pension fund contributions	216	0	1,139

## Consumption imputation

# Household budget constraint

$$\underbrace{c_{i,t}}_{\text{total expenditures}} = \underbrace{y_{i,t}}_{\text{disposable income}} - \underbrace{\sum_j \Delta a_{i,j,t}}_{\text{return-corrected change in assets}} + \underbrace{\sum_k \Delta l_{i,k,t}}_{\text{return-corrected change in liabilities}} - \underbrace{t_{i,t}}_{\text{net transfers}}$$

Stocks example:

- Observe  $a_{i,stocks,t}$  (start of year) and  $a_{i,stocks,t+1}$  (end of year)
- $\Delta a_{i,stocks,t} \neq a_{i,stocks,t+1} - a_{i,stocks,t}$
- $\Delta a_{i,stocks,t} = a_{i,stocks,t+1} - (1 + r_{i,stocks,t})a_{i,stocks,t}$

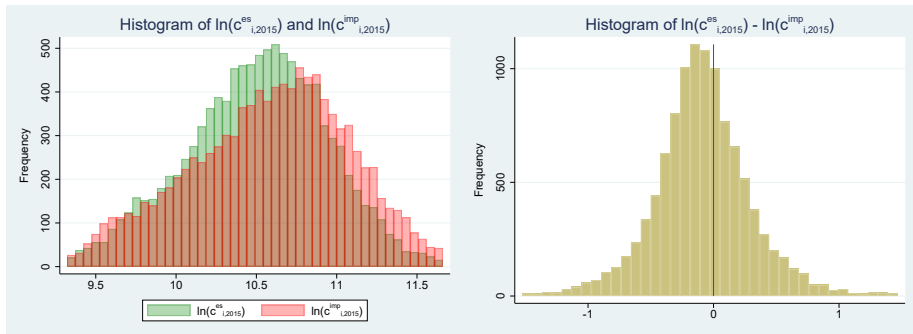
Returns  $r_{i,j,t} \neq$  consumption, quantity = consumption

**Problem:** do not observe  $r_{i,j,t}$

Asset returns require educated guesswork



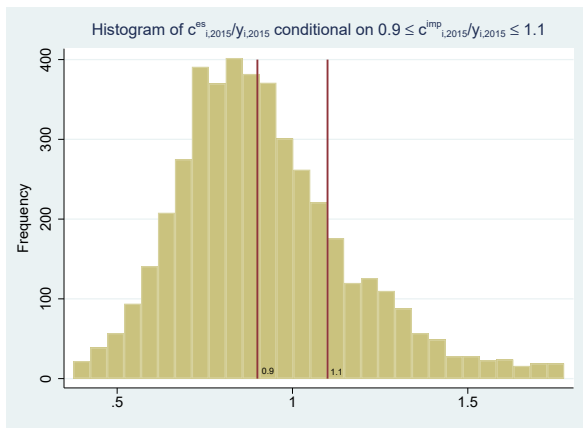
# Imputation errors



- Regressing  $\ln c^{\text{imp}}_{i,2015}$  on  $\ln c^{\text{es}}_{i,2015}$ :  $\beta = 0.83$
- Imputation error correlated with income, age, net wealth, home ownership status, family size (*even after controlling for income*)

# Implied consumption rates

30% of households hand-to-mouth consumers with  $0.9 \leq \frac{c}{y} \leq 1.1$  according to AD, while 8% according to ES:

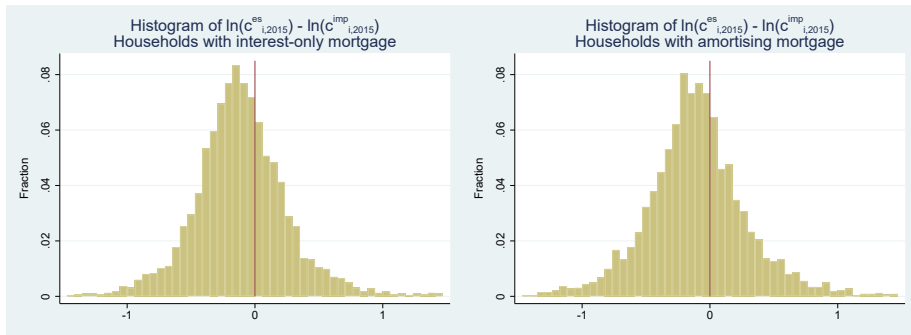


# Interest-only mortgages

- Mastrogiacomo & van der Molen (2015): 60% of Dutch households have IO-mortgage
- Often coupled with pledged savings accounts, not recorded (*will be from 2021*)
- Overestimate consumption: don't record  $a_{i,t} \uparrow$  while  $\bar{l}_{i,t}$
- Identify IO-mortgage if nominal value mortgage constant over time

# Interest-only mortgages

Overestimation of consumption does not stem from IO-mortgages:



**Discussing both imputation techniques**

# Regression results

- Regression model:

$$\Delta c_{i,t} = \alpha_i + \beta_1 \Delta y_{i,t} + \beta_2 \Delta h_{i,t} + \beta_3 Z_{i,t} + \varepsilon_{i,t}$$

- Structural form from Campbell & Cocco (2007)
- $\Delta y_{i,t}$  and  $\Delta h_{i,t}$  interacted with indicators of liquidity, leverage, pension pressure
- Similar results: higher MPC's for constrained households

# Concluding remarks

- Despite similar results, significant differences:
  - ▶ non-durable consumption  $\leftrightarrow$  total expenditures
  - ▶ repeated cross section  $\leftrightarrow$  true panel
  - ▶  $\sim$  thousands observations  $\leftrightarrow$   $\sim$  millions observations
  - ▶ small median errors, large outliers  $\leftrightarrow$  large median errors, small outliers
- Household matching works when survey information crucial
- Underreporting in spending diaries  $\leftrightarrow$  misallocating returns?
- New advances: bank account transaction data  
*Ganong & Noel (2019), Bounie et al. (2020)*

**Thank you for attending!**



# References I

- Abildgren, K., Kuchler, A., Rasmussen, A. S. L., & Sørensen, H. S. (2018). *Consistency between household-level consumption data from registers and surveys*. Danmarks Nationalbank Working Papers 131, Danmarks Nationalbank.
- Baker, S. R., Kueng, L., Meyer, S., & Pagel, M. (2018). *Measurement Error in Imputed Consumption*. Working Paper 25078, National Bureau of Economic Research.
- Bijlsma, M. & Mocking, R. (2017). The impact of house price shocks on the savings of dutch homeowners and renters. *CPB Discussion Paper 346*.
- Bounie, D., Camara, Y., Fize, E., Galbraith, J., Landais, C., Lavest, C., Pazem, T., & Savatier, B. (2020). *Consumption Dynamics in the COVID Crisis: Real Time Insights from French Transaction & Bank Data*. CEPR Discussion Papers 15474, C.E.P.R. Discussion Papers.

## References II

- Browning, M. & Leth-Petersen, S. (2003). Imputing consumption from income and wealth information. *The Economic Journal*, 113(488), F282–F301.
- Bunn, P. & Rostom, M. (2015). *Household debt and spending in the United Kingdom*. Bank of England working papers 554, Bank of England.
- Campbell, J. Y. & Cocco, J. F. (2007). How do house prices affect consumption? Evidence from micro data. *Journal of Monetary Economics*, 54(3), 591–621.
- Cooper, D. (2013). House price fluctuations: the role of housing wealth as borrowing collateral. *The Review of Economics and Statistics*, 95(4), 1183–1197.
- Eika, L., Mogstad, M., & Vestad, O. L. (2020). What can we learn about household consumption expenditure from data on income and assets? *Journal of Public Economics*, (pp. 104–163).

## References III

- Fagereng, A. & Halvorsen, E. (2017). Imputing Consumption from Norwegian Income and Wealth Registry Data. *Journal of Economic and Social Measurement*, 42(1), 67–100.
- Ganong, P. & Noel, P. (2019). Consumer spending during unemployment: Positive and normative implications. *American Economic Review*, 109(7), 2383–2424.
- Ji, K., Teulings, R., & Wouterse, B. (2019). Disentangling the effect of household debt on consumption. *CPB Discussion Paper* 395.
- Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3), 697–743.
- Koijen, R., Van Nieuwerburgh, S., & Vestman, R. (2014). Judging the quality of survey data by comparison with "truth" as measured by administrative records: Evidence from sweden. In *Improving the Measurement of Consumer Expenditures* (pp. 308–346). National Bureau of Economic Research, Inc.

## References IV

- Kreiner, C., Lassen, D., & Leth-Petersen, S. (2014). Measuring the accuracy of survey responses using administrative register data: Evidence from denmark. In *Improving the Measurement of Consumer Expenditures* (pp. 289–307). National Bureau of Economic Research, Inc.
- Mastrogiacomo, M. & van der Molen, R. (2015). *Dutch mortgages in the DNB loan level data*. DNB Occasional Studies 1304, Netherlands Central Bank, Research Department.
- Mian, A. & Sufi, A. (2011). House prices, home equity-based borrowing, and the us household leverage crisis. *American Economic Review*, 101(5), 2132–2156.
- Skinner, J. (1987). A superior measure of consumption from the panel study of income dynamics. *Economics Letters*, 23(2), 213–216.
- Zhang, L. (2019). Do house prices matter for household consumption? *CPB Discussion Paper 396*.