

The relevance of quantitative theory for historical demography

David de la Croix

Université catholique de Louvain

Quantitative theory is used by economists and by biologists

Why not much in historical demography ? the abundance of data is well suited for this methodology

I will explain what it is

How it differs from econometrics and statistics

Why it is complementary to other approaches

(1) General concepts

(2) A pedagogical example

(3) A 4-slide summary of “Fertility and Childlessness in the US”

(1) Definition

Quantitative theory uses simple, abstract models, together with data, to highlight major mechanisms.

Models are employed in three ways.

First, they are used to organize data in a new and suggestive manner - to look at data through the glass of theory.

Second, models are combined with data to display successes and failures of particular theoretical mechanisms

Third, they are used to run counterfactual experiments. What if...

(1) A theory

Hence we need to start with a theory. A formal theory.

In economics, it is usually a maximization problem:

$$\max_x f(x) \text{ subject to } \textit{constraints}$$

and x is the vector of controls

The objective $f(x)$ could be utility (economists), reproductive fitness (biologists), adherence to norms (sociologists)

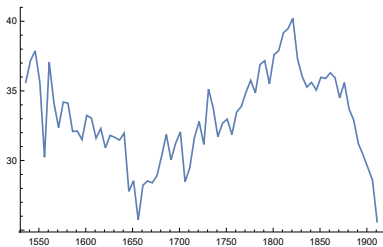
It is not that economists believe that people actually behave like this

but modelling their choice as if they were maximizing can be a good proxy, if incentives really matter

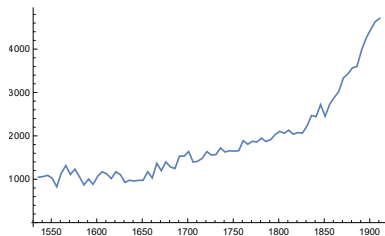
(2) Pedagogical example

Suppose we want to understand how fertility relates to income in England 1531-1911

Early CBR from Wrigley et al. in their population history of England
GDP per capita is now reconstructed since 1300 (Broadberry et al.)



Crude birth rate, ‰



GDP per capita in 1990 \$

(2) Theory example

Suppose people decide the number of children n , together with their consumption c

Problem of a couple of generation t :

$$\max_{c_t, n_t} (c_t - \bar{c})^\rho + \gamma n_t^\rho$$

subject to

$$y_t = c_t + \phi y_t n_t$$

Solving this problem leads to the optimal choice of n :

$$n_t^{optimal} = \frac{\gamma^{\frac{1}{1+\rho}} (y_t - \bar{c})}{\gamma^{\frac{1}{1+\rho}} \phi y_t + (\phi y_t)^{\frac{1}{1+\rho}}}$$

it is a function of income y_t and of the parameters \bar{c} , γ , ϕ and ρ .

(2) Mechanisms

In this simple model we have both Malthus and Becker:

Malthus (1807): The ultimate check to population appears then to be a want of food

→ modeled here with the minimum consumption \bar{c}

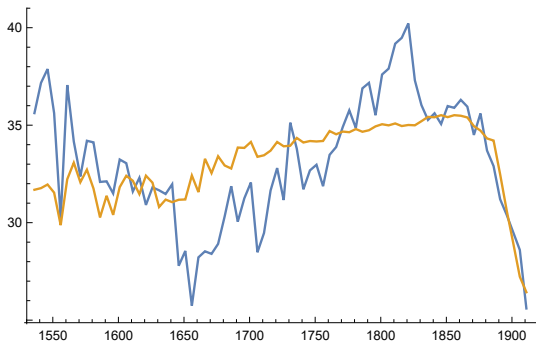
Becker (1993): Malthus neglects that the time spent on child care becomes more expensive when countries are more productive. The higher value of time raises the cost of children and thereby reduces the demand for large families.

→ modeled here by making the cost of children proportional to income $\phi y_t n_t$

(2) Calibration

The parameters can be set to get the model “reproduce” (be consistent with) some facts. Here with the CBR over time.

Approximating n_t by the CBR and y_t by GDP per capita, we can choose the parameters \bar{c} , γ , ϕ and ρ so as to minimize the distance between $n_t^{optimal}$ and $n_t^{observed}$. Here is the fit:



$$\bar{c} = 186$$

$$\gamma = 73$$

$$\phi = 0.026$$

$$\rho = 0.87$$

(2) Evaluation

Success:

picking the turning point right. For low level of income, fertility is increasing in income (Malthus). For high level of income, it is decreasing in income (Becker)

Failure:

the model does not explain that fertility actually decreased over 1531-1650

We can compute how much do we explain with the model.

Here it is 36%, computed as $R^2 = \text{corr}(n^{\text{optimal}}, n^{\text{observed}})^2$

Ideally, we should evaluate the model in its ability to predict something we did not use for the calibration

(2) Parameters

The parameters \bar{c} , γ , ϕ and ρ are “deep” or structural parameters. They have a meaning. We focus on their size, not so much on their statistical significance (although S.E. can be computed with bootstrap methods)

The minimum consumption level is $\bar{c} = 186$, that is 1/2 dollar per day, seems ok

$\phi = 0.026$ is not realistic. Indeed, it implies average share of GDP devoted to children is: $\phi n y / y = 86\%$.

Altogether, the parameters imply that fertility starts to be decreasing with income when $y_t > 2738$ (value attained in ≈ 1850).

(2) Questions that the model addresses to the data

The model is about the life of a person. Ideally we need her lifecycle fertility (=completed fertility) and her lifecycle income.

→ having the distributions of parities would help enormously

To discipline the analysis further, require the model to reproduce also fertility by social class.

→ if no data over the whole time span, reproducing cross-section data at one or two points in time would already help.

(2) Extensions to the model

Introduce mortality

will it just improve the fit, or alter our estimation of the effect of income on fertility (omitted variable bias) ?

Model marriage (it is the main margin of adjustment for Malthus) and use data on age at marriage to discipline the analysis further

Improve the modelling of the cost and benefits of children (child labor?)

Model education and health of children as investment in quality

(3) Childlessness and theory

The analysis of the reasons for being childless.

There are two main types of childlessness discussed in the literature: 1) voluntary childlessness; namely, a utility-enhancing life choice for those who decide not to have children; and 2) involuntary childlessness (natural or poverty-driven).

Poverty-driven childlessness occurs when fecundity is affected by poor living conditions and societal underdevelopment.

One main limitation of measurement without theory resides in the impossibility of distinguishing biologically and poverty-driven childlessness from voluntary childlessness in the data.

(3) Our approach

The structural model devised by Baudin et al. (AER, 2015) provides a way to use data in order to weight the various causes of childlessness.

This model is essentially based on two assumptions.

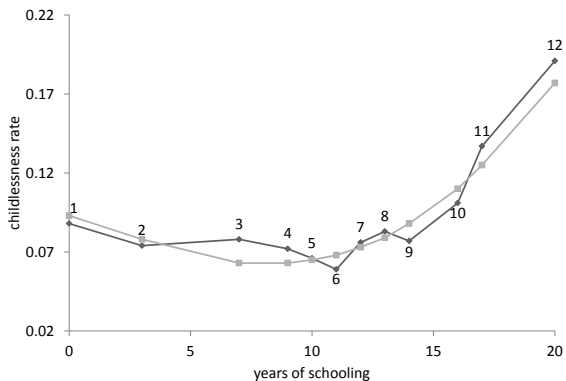
1. the ability to have children increases with the standard of living.
2. the main cost of children is the time it takes to rear them. Children are thus more expensive for highly educated people.

The presence of the first mechanism can be confirmed and its size can be measured from the fact that childlessness is more prevalent among the very poor than among the middle classes (despite the fact that fertility among the poor is higher).

The second assumption rests on a mechanism observed among the educated: namely, that the occurrence of childlessness increases with the mother's level of education.

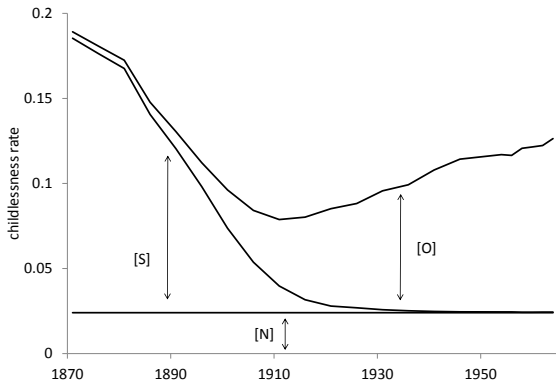
(3) Calibration

Parameters are set to match data from 1990 census



(3) Simulation

The model is then used to simulate the change over time of childlessness and of its reasons



Conclusion

To increase the quantitative content of research, many lean towards the use of regression analysis.

Not based on theory (apart from knowing which variable to include in the regression) and focuses on knowing which explanatory variable is statistically significant.

As an alternative, quantitative theory: aims at evaluating the quantitative importance of a limited number of specific mechanisms. Does not aim at providing a comprehensive understanding of the phenomenon under scrutiny, but leads to cumulative research.

“progress, do not regress”