Aggregation Over Individuals Under Polynomial Representative Draws

Frédéric VERSCHUEREN*

February 27, 2013

Abstract

To take into consideration the high degree of heterogeneity in economic agents' behaviour, modern macroeconomics has increasinly developed models with solid microfoundations. However, different constraints hinder a systematic use of these recent macromodels in empirical analysis: (i) the lack of adequate individual statistics and (ii) the high complexity of structural microeconomic models leading to not tractable forms. In addition, a limited number of studies has focused on the empirical implications of aggregation over individuals, and resulting macroeconomic models suffer from different shortcomings : (i) some individual data need to be available but may not be observed, and (ii) the models rely on not realistic assumptions (i.e. the number of economic agents is assumed to be infinite). Our paper overcomes these obstacles by proposing the use of Polynomial Representative Draws models to characterize unobserved micro data. In this context, each decision is related to the rank hold in the sample by the individual through polynomial functions. Aggregation of such decisions attractively leads to exact and observable simple macromodels, the time series estimation of which allows to assess how individual

*IWEPS (Route de Louvain-la-Neuve 2, 5001 Namur, Belgium) and EQUIPPE (University of Lille 3, Domaine Universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'ascq Cedex, France). Phone: ++32 81 468 475. Email: F.Verschueren@iweps.be

data's cumulated distribution evolves through time. Our methodology is used to evaluate employees' poverty rates in Belgium using historical aggregated data from National Accounts.

Keywords: Aggregation over individuals; Wages distribution; Macroeconomic models; Time series analysis; Unobserved micro data.

JEL Codes: C46; D31; E10.

1 Introduction

SINCE A COUPLE OF DECADES macroeconomics has increasingly been developing models with solid microfoundations. The guideline was to take at best into consideration the high degree of heterogeneity in economic agents' behaviour as well as the interactions between each decisions in a growingly complex economic world.

This research agenda has given rise to Business Cycles Models (Kydland and Prescott, 1982) and has contributed to the development of General Equilibrium Models (see a.o. Christano and al., 2005). Even macroeconometric models, which were much criticized in the mid-seventies, has been revitalized by including more elements from microeconomics (see the discussion in Fair, 1993). In addition, microsimulation models allow to reproduce each economic agent' decision-making paths so that macroeconomics boils down to the aggregation of individual paths (for an interesting paper on microsimulation, see Legendre, 2004).

However, different obstacles hinder a systematic use of recent macromodels in empirical analysis. Firstly, adequate individual data may not be available. Despite the production of more and more detailed statistics by official Institutes, the treatment of which is made easier due to rapid progresses of computer technology, data exhaustivity or up-to-date data availability are far from a reality. Consequently, some macromodels may not be estimated and need to be calibrated. Secondly, micromodels closer to reality may be too complex to deal with when aggregation is at stake. The use of an approximated (hence tractable) version of the model is then likely to generate forecasts or policy simulations of lower quality.

Although temporal aggregation has been largely documented in the literature, a more limited number of studies have analyzed the empirical implications of aggregation over individuals. About this topics, three papers can be regarded as major contributions. Stoker (1993) proposes a detailed review of empirical representations resulting from the aggregation of individual economic behaviours. Hildenbrand and Kneip (2002) obtain an explanatory model of UK aggregate consumption growth compatible with microeconomic theory and using time series of individual data (a German application is proposed by Chakrabarty and Schmalenbach, 2002). In the context of models with quantity rationing, Lambert (1988) shows that, under specific conditions, the aggregation of micro-markets in desequilibrium leads to a simple mathematical form, namely a CES function on aggregate demand and aggregate supply (see also Entorf et Sneesens, 2000).

Recognizing their strenghts for empirical research, above macroeconomic models still suffer from different shortcomings. In the first two references quoted, individual data are needed for estimation purposes but may not be observed in official statistics. As to the third study, the model assumes an infinite number of agents which may not be a realistic assumption when the stock of economic agents is fluctuating from year to year. Beyond this consideration, one can argue that the effect of the sample size on the level of aggregated variables have been somewhat neglected in the literature on individual aggregation.

Our paper can be regarded as an original attempt to overcome these empirical obstacles. While rigorously justifying the bridge between micromodels and empirical macromodels, our research advocates the use of Polynomial Representative Draws (PRD) models to characterize unobserved micro data. In this context, each individual realization is related to the rank hold in the sample by the individual through polynomial functions of order k. Substracting an error term from the realisation y_i gives the representative, or theoretical, value of y_i. Also, representative parameters of the model can be treated as endogenous and explained by a set of aggregated variables or policy instruments.

Clearly, the structure of micro data gets different properties according to the value of k. Choosing $k = 1$ assumes an equidistance between two successive representative values. The case of $k = 2$ supports a situation of microeconomic asymmetry. With $k = 3$, we come very close to the Normal distribution which can be regarded as a special case of a more general family of data generating processes.

Under weak assumptions, aggregation over individuals made in the PRD context attractively leads to exact and observable time series macromodels, the econometric estimation of which allows to assess how individual data's cumulated distribution changes over time. So the key point of our approach is to propose empirical models able to extract unobserved microeconomic information from aggregated data. Consequently, the inclusion of PRD models in larger macroeconomic models could help to evaluate redistributive effects of many economic shocks or public policy changes on specific categories of agents, providing hence a promising tool to design more efficient macroeconomic policies.

We believe that the field of potential economic applications of the PRD methodology is huge. One of them is investigated in the paper and is related to the evaluation of employees' poverty rates. The case study is conducted for Belgium using time series data from National Accounts.

The paper is organized as follows: section 2 defines PRD models and analyzes their aggregative and distributional properties, while section 3 presents the case study. Concluding remarks are drawn in section 4.

2 Definitions and aggregation of PRD models

2.1 The general setting

We define the Polynomial Representative Draws model of order k − noted PRD_k − as a generating process for the n *realizations* $\{y_i\}_{i=1}^n$ taking the form

$$\begin{aligned} y_i &= \widetilde{y}_i + \lambda_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j i^j + \lambda_i, \qquad i = 1, ..., n \end{aligned} \tag{1}$$

In this framework, $\widetilde{y}_i$ stands for the *representative, or theoretical,* value assigned to individual i, while $\left\{\beta_j\right\}_{j=0}^k$ are the *representative parameters*[1] of the PRD_k model. Unobserved errors $\{\lambda_i\}_{i=1}^n$ fill the gap between observed

[1] Representative parameters are not necessary constant over time, see later in the paper.

values $\{y_i\}_{i=1}^n$ and theoretical ones $\{\widetilde{y}_i\}_{i=1}^n$. They are assumed to be uncorrelated across individuals so that $Corr\{\lambda_i, \lambda_j\} = 0\ \forall\ i \neq j$. Also, their variance σ_λ^2 is expected to be stable over time.

Moreover, the following condition is imposed

$$\frac{1}{n}\sum_{i=1}^{n} \lambda_i[\beta] \equiv \Lambda = 0 \tag{2}$$

where β summarizes the full set of *true* representative parameters.

The idea behind (2) is to get a tractable form when (1) is aggregated.

Under (2), variance of the $\lambda's$ simplifes to

$$\sigma_\lambda^2 \equiv Var\{\lambda\} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 \tag{3}$$

It is worth clarifying the key difference between statistical laws (i.e. the Normal law) and empirical laws such as (1). A statistical law for y is described by unknown distributional parameters (mathematical expectation, variance, ...) the evaluation of which allows to draw *random realisations of* y. By constrast, our polynomial model generates *representative values for* y which in turn are used to calculate both empirical mean (i.e. the macromodel) and empirical variance of y.

The sequel of the paper analyzes important aggregative and distributional properties of PRD models with $k = 1$ and $k = 2$[2].

2.2 The PRD model of order 1

2.2.1 The model and its properties

The PRD_1 model writes

$$y_i = \beta_0 + \beta_1 i + \lambda_i, \qquad i = 1, ..., n \tag{4}$$

We assume that representative parameters β_0 and β_1 are positively signed, though a negative value for β_0 might also be allowed. To set $\beta_1 > 0$ has one

[2] A deep analysis of PRD models with $k = 3$, including a comparison with the Normal law, is the subjet of an ongoing research.

interesting implication: the representative values of y generated by (4) are *sorted in ascending order as rank i increases.* This key property will facilitate the measure of probabilities, as we shall see later.

In the above context, representative parameters get the following definitions

$$\beta_1 = \widetilde{y}_i - \widetilde{y}_{i-1} \equiv \Delta\widetilde{y}_i > 0 \tag{5}$$

and, letting $\lambda_1 = 0$ for $y_{\min} = y_1$,

$$\beta_0 = y_{\min} - \beta_1 > 0 \tag{6}$$

Equation (5) points out that β_1 is to be interpreted as the distance between two consecutive representative values of y. Equation (6) allows to extend the PRD_1 model by adding information on the minimum value of y, so that (4) becomes

$$y_i = y_{\min} + \beta_1(i-1) + \lambda_i, \qquad i = 1, ..., n \tag{7}$$

Let us examine the attractive part of PRD models. It is readily shown that aggregation of (7) over individuals leads to the mean, or macro, representation for y given by

$$\overline{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i = y_{\min} + \beta_1\left(\frac{n-1}{2}\right) \tag{8}$$

As aggregates y, n (hence $\overline{y}$) and $y_{\min}$ are very often observed, only parameter β_1 is unknown in (8) but it could be estimated using standard econometric methods.

Also, the effect of sample size on the macro variable is stable and get a positive sign as

$$\frac{\partial\overline{y}}{\partial n} = \frac{\beta_1}{2} > 0 \tag{9}$$

Next, empirical variance of y is easily proven to be

$$\sigma_y^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \beta_1^2\left(\frac{n^2-1}{12}\right) + \sigma_\lambda^2 + 2\beta_1\gamma \tag{10}$$

with $\gamma \equiv Cov[i, \lambda_i]$. Note that when $\gamma > 0$ (versus $\gamma < 0$), λ errors tend to be larger (versus lower) for higher values of i.

Parameters σ_λ^2 and γ are the two remaining unknown parameters of the PRD_1 model. Their knowledge (together with β_1) allows to find the value of σ_y^2 through (10). For a sake of simplification, we assume $\gamma = 0$.

A key result based on the PRD_1 model rests on the calculation of probabilities. It is straighforward to show that the probability for y to be less than a threshold value, say y_τ, is simply given by

$$\begin{aligned}
\Pr[y \quad < \quad y_\tau] \equiv \pi_\tau &= \frac{\#\{y < y_\tau\}}{n} \\
&= \frac{1}{n}\left(\frac{y_\tau - y_{\min} - \lambda_\tau + \beta_1}{\beta_1}\right) \\
&\simeq \frac{1}{n}\left(1 + \frac{y_\tau - y_{\min}}{\beta_1}\right)
\end{aligned} \tag{11}$$

considering $\dfrac{\lambda_\tau}{\beta_1 n}$ is small.

Consequently, any level of cumulated probability $\{\pi_k\}_{k=1}^{n}$ can be computed as soon as an estimate for β_1 becomes available. Formula (11) has many potential applications in economics, one of them being studied in section 3.

2.2.2 Estimation issues

With an exact aggregate model such as (8), we end up in a very familiar ground. Indeed, aggregated time series $\{\overline{y}[t]\}_{t=1}^{T}$, $\{y_{\min}[t]\}_{t=1}^{T}$ and $\{n[t]\}_{t=1}^{T}$ are very often available in official economic statistics. Therefore, performing an econometric regression on (8) using (say) OLS estimation methodology will provide an estimate for β_1.

We would then have

$$\widehat{y}[t] = y_{\min}[t] + \widehat{\beta}_1 \left(\frac{n[t]-1}{2} \right) + \widehat{\Lambda}[t], \qquad t = 1, ..., T \tag{12}$$

To get an estimate of σ_λ^2, we notice that

$$\begin{aligned} \sigma_{\widehat{\Lambda}}^2 &\equiv \frac{1}{T} \sum_{i=1}^{n} \left(\widehat{\Lambda}[t] \right)^2 \\ &\simeq \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{n} \lambda_i[\widehat{\beta}_1; t] \right)^2 \\ &\simeq \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \lambda_i^2[\beta_1; t] \\ &\simeq \frac{1}{T} \sum_{t=1}^{T} n[t] \sigma_\lambda^2 \end{aligned} \tag{13}$$

assuming $\lambda_i[\widehat{\beta}_1; t] \simeq \lambda_i[\beta_1; t]$ and σ_λ^2 does not depend on time.

Consequently,

$$\widehat{\sigma}_\lambda^2 \simeq \frac{\sigma_{\widehat{\Lambda}}^2}{\overline{n}} \tag{14}$$

with $\overline{n} = \frac{1}{T} \sum_{t=1}^{T} n[t]$.

Finally, substituting results obtained for $\widehat{\beta}_1$ and $\widehat{\sigma}_\lambda^2$ in equation (10) allows us to write (remember $\gamma = 0$)

$$\widehat{\sigma}_y^2[t] = \widehat{\beta}_1^2 \left(\frac{n^2[t]-1}{12} \right) + \widehat{\sigma}_\lambda^2 \tag{15}$$

2.2.3 Models with endogenous parameters

At this stage, we have to recognize our *PRD* model suffers from important limitations. Specifically, the mean value and the variance of y *only* depend on the sample size n. Should our approach be used as such for forecasting or policy simulations, it would clearly formalize a very restrictive and unrealistic economic context.

Actually, we *do not need* β_1 to be constant over time, and relaxing this assumption will simply lead to PRD models with *endogeneous representative parameters*. We could easily imagine that β_1 is sensitive to changes in some aggregated variables or policy instruments. Let $z[t]$ be one of these observed macro variables. Economic theory could help us to predict the best candidate for $z[t]$. For instance, $z[t]$ could be a trend variable to take into account indexation on wages. We may as well expect that workers' wage have a larger dispersion (i.e a greater value for $\beta_1[t]$) when economic situation improves, linking the representative parameter to economic growth. Another example can be found in the topics of aggregated consumption. Let individual consumption $c_i[t]$ be generated by (7), and let $\overline{c}[t]$ and $\overline{r}[t]$ be representing consumption per capita and revenue per capita, respectively. if $z[t]$ would depend on *mean* revenue per capita, i.e. $z[t] = \gamma \dfrac{\overline{r}[t]}{n[t]}$, we would obtain the traditional macro equation for consumption

$$\overline{c}[t] = c_{\min}[t] + \frac{\gamma}{2}\overline{r}[t]\left(1 - \frac{1}{n[t]}\right), \qquad t = 1, ..., T \tag{16}$$

with $\overline{c}[t] \longrightarrow c_{\min}[t] + \frac{\gamma}{2}\overline{r}[t]$ as n is increasing.

A more direct approach to learn more about $z[t]$ is to solve (8) for β_1

$$\widehat{\beta}_1[t] = 2\left(\frac{\overline{y}[t] - y_{\min}[t]}{n[t] - 1}\right), \qquad t = 1, ..., T \tag{17}$$

and to find out which macro variable has a similar time profile as the one observed for $\widehat{\beta}_1[t]$.

Defining $Z[t]$ as a vector of relevant macro variables, Γ a vector of stable parameters and $\beta_1[t] = f\left\{Z[t];\Gamma\right\}$, aggregation over the n individual draws taken from the PRD_1 model leads to

$$\overline{y}[t] = y_{\min}[t] + f\left\{Z[t];\Gamma\right\}\left(\frac{n[t] - 1}{2}\right), \qquad t = 1, ..., T \tag{18}$$

In this general form (18), macro variable $\overline{y}$ is explained by a set of *observed* variables which, in turn, can be defined as endogenous by adding new equations to aggregated model (18). Of course, $Z[t]$ may also include $n[t]$ so that the sign of the sample size' effect on $\overline{y}[t]$ and $\pi_\tau[t]$ will depend on the functional form assigned to f. Including result such as (11) in larger macroeconomic models could thus help the policy maker to evaluate the redistributive effects of many economic shocks or public policy changes on specific categories of agents .

Notice that a special case arises when $z[t] = \overline{y}[t]$, yealding to the following macromodel

$$\log \overline{y}[t] = \log y_{\min}[t] - \log \left\{ 1 - \gamma_1 \frac{n[t]-1}{2} \right\}, \qquad t = 1, ..., T \tag{19}$$

2.3 The PRD model of order 2

The PRD model with $k = 2$ writes

$$y_i = \beta_0 + \beta_1 i + \beta_2 i^2 + \lambda_i, \qquad i = 1, ..., n \tag{20}$$

To keep the property of an increasing value of y with i, we impose $\beta_2 > 0$ and $\frac{\partial y_i}{\partial i} | \left(y_i = y_{\min} \equiv y_1 \right) = 0$ so that $\beta_1 = -2\beta_2$. Also, the definition of $y_{\min}$ gives rise to condition $\beta_0 = y_{\min} + \beta_2$.

As a result, equation (20) simplifies to a single parameter model

$$y_i = y_{\min} + \beta_2 \left(i - 1 \right)^2 + \lambda_i, \qquad i = 1, ..., n \tag{21}$$

Due to the quadratic form, the micro data structure generated by a PRD_2 model is asymmetric. Restrictions on the parameters implies that the number of realizations below the median value outweighs the number of realizations above the median value.

It is straighforward to show that aggregation of (21) over individuals leads to the exact macromodel

$$\overline{y} = y_{\min} + \beta_2 \frac{(n-1)(2n-1)}{6} \tag{22}$$

Again, we could either estimate β_2 using time series observations on $\overline{y}[t]$, $y_{\min}[t]$ and $n[t]$ (constancy assumption of the parameter) or consider the parameter as endogenous by solving (22) for β_2.

Tedious but obvious mathematical manipulations lead to the following formula for the empirical variance of y

$$\sigma_y^2 = \beta_2^2(n-1)\left\{\frac{48n^3 - 42n^2 - 57n + 33}{540}\right\} + \sigma_\lambda^2 \tag{23}$$

An estimate for σ_λ^2 is given by (13) using the regression results based on (22) so that (23) provides an estimate for σ_y^2 (or time series for σ_y^2 when β_2 is considered to be endogenous).

Finally, the probability for y to be less than threshold value y_τ writes

$$\begin{aligned} \Pr[y &< y_\tau] \equiv \pi_\tau = \frac{\#\{y < y_\tau\}}{n} \\ &\simeq \frac{1}{n}\left(1 + \sqrt{\frac{y_\tau - y_{\min}}{\beta_2}}\right) \end{aligned} \tag{24}$$

3 Application to the estimation of poverty rates

3.1 *PRD* models and the probability of poverty

To illustrate our methodology, we study how poverty rates can be computed from aggregated empirical models such as the ones we have defined in previous section. Let $w_i[t]$ be the wage[3] earned by employee (or worker) i at time t. Let us investigate PRD models with $k = 1$ and $k = 2$ so that individual wages and mean wages are assumed to be generated respectively by

$$w_i^k[t] = w_{\min}[t] + \beta_1(i-1)^k + \lambda_i^k[t], \qquad i = 1, ..., n[t];\ k = 1, 2 \tag{25}$$

and

$$\overline{w}^1[t] \equiv w_{\min}[t] + \beta_1\left(\frac{n[t]-1}{2}\right), \qquad t = 1, ..., T \tag{26}$$

[3] In the application, income level is defined as the wage level.

$$\overline{w}^2[t] = w_{\min}[t] + \beta_2 \frac{(n[t]-1)(2n[t]-1)}{6}, \qquad t = 1, ..., T \tag{27}$$

In the PRD_2 model, the number of low-wage workers outweighs the number of high-wage workers. To take into account part-time work, we define the minimum wage level as a proportion θ of the minimum wage proposed to a full-time worker, i.e.

$$w_{\min}[t] = \theta w_{\min}^{FT}[t], \qquad t = 1, ..., T \tag{28}$$

Time series for poverty rates can now be easily calculated using the distributional properties of our polynomial models. Let poverty treshold $w_P[t]$ be defined as 60% of the median income, the latter being set to the median wage. Equation (25) allows us to write

$$w_P^k[t] = 0,6 \left\{ w_{\min}[t] + \beta_k \left(\frac{n[t]}{2} - 1 \right)^k + \lambda_P^k[t] \right\}, \qquad k = 1, 2;\ t = 1, ..., T \tag{29}$$

We notice that the median value is identical in both versions of the model, i.e. $w_P^1[t] = w_P^2[t]$, when the following restriction on the representative parameter β_1 is imposed

$$\beta_1 = \beta_2 \left(\frac{n[t]}{2} - 1 \right) \tag{30}$$

so that $\beta_1 >> \beta_2$ for large n.

At period t, the number of poor workers $\tau[t]$, i.e. the number of workers whose wage is lower than $w_P^k[t]$, will be different according to the value of k. As $\beta_k > 0$, one can define τ as the rank related to the (theoretical) worker who benefits from poverty threshold.

Putting together (25) and (29) yields to equation

$$0,6\left\{w_{\min}[t]+\beta_k\left(\frac{n[t]}{2}-1\right)^k+\lambda^k_{n/2}[t]\right\} \quad = $$
$$w_{\min}[t]+\beta_k(\tau-1)^k+\lambda^k_p[t], \qquad k \quad = \quad 1,2;\ t=1,...,T \tag{31}$$

so that, for $k=1,2$ and $t=1,...,T$

$$\tau[t]=1+\sqrt[1/k]{\frac{1}{\beta_k}\left\{0,6\beta_k\left(\frac{n[t]}{2}-1\right)^k-0,4w_{\min}[t]+0,6\lambda^k_{n/2}[t]-\lambda^k_p[t]\right\}} \tag{32}$$

Finally, the ratio of $\tau[t]$ and the number of workers $n[t]$ provides a time series for poverty rates $\pi^k[t]$ at time $t=1,...,T$

$$\pi^k[t]\simeq\frac{1}{n[t]}\left\{1+\sqrt[k]{0,6\left(\frac{n[t]}{2}-1\right)^k-\frac{0,4}{\beta_k}w_{\min}[t]}\right\}, ...k=1,2 \tag{33}$$

assuming $\dfrac{0,6\lambda^k_{n/2}[t]-\lambda^k_p[t]}{\beta_k n^2[t]}\simeq 0.$

It is easy to show that $\pi^2[t]>\pi^1[t]$.

3.2 Results with Belgian aggregated data

Data comes from the Belgian National Accounts and are observed from 1981 up to 2011. Variable $w[t]$ is defined as the "Wages and salary" aggregate while $n[t]$ is the number of employees. Time series for $w_{\min}[t]$ has been found in the *OECD* minimum wages database. We take into consideration part-time work by setting θ to $0,5$. Aggregated models (26) and (27) are regressed using the standard *OLS* estimation procedure. Three parametric versions of the models are investigated : the first one (named exogenous *PRD* model) assumes a constant β_k while the second and third ones (named endogenous *PRD* models) consider respectively $\beta_k[t]=\gamma_0+\gamma_1 t$ and $\beta_k[t]=$

$\gamma_0 + \gamma_1 GDP[t]$, with t a time trend and $GDP[t]$ the level of Belgian Gross Domestic Product observed at time t. The idea is to test wether economic growth influences the dispersion of wages. Actually, there seems to be a common trend between GDP and the time series of β_k obtained solving (8) and (22), as shown in figures 1 and 2 (see the appendix).

[Insert figures 1 and 2]

Table 1 and 2 hereafter summarize the econometric results obtained by regressing the six empirical PRD models.

[Insert tables 1 and 2]

In each model investigated, all the parameters are checked to be significant at the 99% level of confidence. Coefficients of determination get very high values in each case, supporting a near-perfect quality of fit for the PRD-based regressions. It appears that the variance of the $\lambda's$ is the smallest one when both $k = 1$ and the representative parameter is treated as trend-dependent (i.e. endogenous), although very similar results are found when GDP is selected. Also, estimated poverty rates get higher values when $k = 2$, due to the positive asymmetry captured by the quadratic form of the model together with $\beta_2 > 0$.

Table 1: PRD models of order 1, wages and poverty

$(k=1)$	Exog. PRD model	Endog. PRD models	
		$Z = Trend$	$Z = GDP$
$\overline{\beta}[t]$	$0,0107$		
$se(\beta[t])$	$(0,0026)$		
$\widehat{\beta}_k$	$0,0111$ **		
$se(\widehat{\beta}_k)$	$(0,0004)$		
$t(\widehat{\beta}_k)$	$24,21$		
$\widehat{\gamma}_1$		$0,0067$ **	$0,0051$ **
$se(\widehat{\gamma}_1)$		$(0,0004)$	$(0,0006)$
$t(\widehat{\gamma}_1)$		$15,72$	$8,39$
$\widehat{\gamma}_2$		$2,53$ x 10^{-4} **	$2,58$ x 10^{-14} **
$se(\widehat{\gamma}_2)$		$(2,17$ x $10^{-5})$	$(2,46$ x $10^{-15})$
$t(\widehat{\gamma}_2)$		$11,67$	$10,46$
R^2	$95,1\%$	$99,1\%$	$99,0\%$
$\sigma^2_{\widehat{\Lambda}}$	18.658.427	3.388.340	4.044.910
$\widehat{\sigma}^2_{\lambda}$	$5,55$	$1,01$	$1,20$
$\overline{\widehat{\sigma}^2_y}[t]$	117.561.021	122.523.286	122.318.549
$w^k_M[2011]$	39.486	43.140	43.633
$w^k_P[2011]$	23.692	25.884	26.180
$\boldsymbol{\pi}^k_P[\mathbf{2011}]$	$\mathbf{13,47\%}$	$\mathbf{17,33\%}$	$\mathbf{17,56\%}$
$N^k_P[2011]$	523.369	673.361	682.374
$\overline{w}[2011]$	43.427		
$w_{\min}[2011]$	17.868		
$\overline{n}[t]$	3.364.606		

** Significant at the 99% level of confidence.

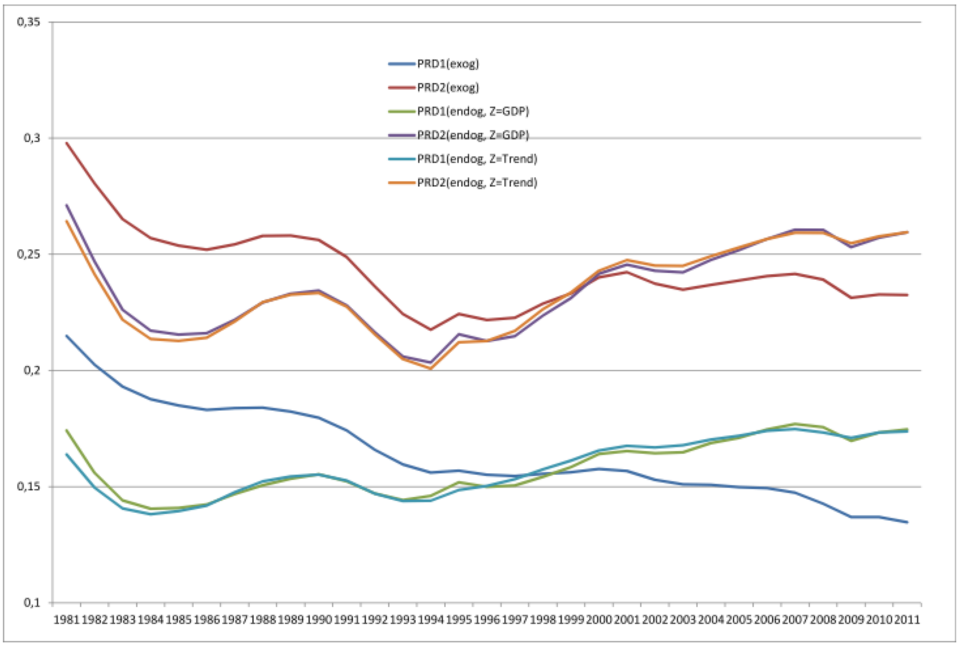
Table 2: PRD models of order 2, wages and poverty

$(k=2)$	Exog. PRD model	Endog. PRD models	
		$Z = trend$	$Z = GDP$
$\overline{\beta}[t]$	$4,76 \text{ x } 10^{-9}$		
$se(\beta[t])$	$(9,41 \text{ x } 10^{-10})$		
$\widehat{\beta}_k$	$4,93 \text{ x } 10^{-9}$ **		
$se(\widehat{\beta}_k)$	$(1,48 \text{ x } 10^{-10})$		
$t(\widehat{\beta}_k)$	$33,18$		
$\widehat{\gamma}_1$		$3,70 \text{ x } 10^{-9}$ **	$3,35 \text{ x } 10^{-9}$ **
$se(\widehat{\gamma}_1)$		$(2,48 \text{ x } 10^{-10})$	$(3,35 \text{ x } 10^{-10})$
$t(\widehat{\gamma}_1)$		$14,90$	$9,99$
$\widehat{\gamma}_2$		$6,51 \text{ x } 10^{-11}$ **	$6,40 \text{ x } 10^{-21}$ **
$se(\widehat{\gamma}_2)$		$(1,19 \text{ x } 10^{-11})$	$(1,28 \text{ x } 10^{-21})$
$t(\widehat{\gamma}_2)$		$5,45$	$4,98$
R^2	$97,3\%$	$98,69\%$	$98,57\%$
$\sigma^2_{\widehat{\Lambda}}$	10.164.499	5.191.963	5.670.914
$\widehat{\sigma}^2_\lambda$	$3,02$	$1,54$	$1,69$
$\overline{\widehat{\sigma}^2_y}[t]$	296.591.604	257.339.978	257.515.065
$w^k_M[2011]$	35.846	36.940	37.108
$w^k_P[2011]$	21.507	22164	22.264
$\boldsymbol{\pi}^k_P[\mathbf{2011}]$	$\mathbf{23,25\%}$	$\mathbf{25,92\%}$	$\mathbf{26,05\%}$
$N^k_P[2011]$	903.590	1.007.347	1.012.446
$\overline{w}[2011]$	43.427		
$w_{\min}[2011]$	17.868		
$\overline{n}[t]$	3.364.606		

** Significant at the 99% level of confidence.

Finally, econometric results allow to generate time series for poverty rates, as depicted in figure 3.

Figure 3: time series PRD poverty rates



Focusing on the versions allowing endogenous parameters, poverty rates of Belgian employees seem to be constantly rising from 1994 onwards. It is worth mentioning that the values obtained for poverty rates are higher than the ones provided by official statistics given that our definition of median income does not include other sources of income such as unemployment benefits or pensions.

4 Conclusions and future research

The aim of this paper was to provide simple and tractable aggregated models which have solid microfoundations and allow a direct calculation of distributional aspects of unobserved, or not yet available, micro data. We believe that our empirical methodology, based on Polynomial Representative Draws models for individual realizations, could be useful for the applied economists and improve significantly the quality of larger macroeconomic models in terms of forecasting or policy analysis.

The case study investigated in the paper deals with poverty rates of employees in Belgium, which appears to be constantly increasing from 1994. Such a result is obtained in a simplified context but many extensions are obvious, among others: (i) to test our methodology with aggregated data coming from other countries, (ii) to include additional equations to take into account simultaneous relationships likely to occur between the sample size and the Z-variables, and (ii) to set up a theoretical context for the PRD models to learn more about the potential Z-candidates influencing β. On this matter, we point out that PRD_2 models for wages exhibit a similar form as the one proposed in the Mincer equation for wages where both individual experience and *squared* experience play an important role (see Mincer, 1974). We anticipate that a link between the two methodologies might be found.

Finally, our approach could be helpful to evaluate the redistributive effects of a large range of economic shocks or policy changes on specific categories of economic agents, providing hence a promising tool to design more efficient macroeconomic policies. This fascinating research agenda is scheduled for the very near future.

References

[1] Chakrabarty M. and A. Schmalenbach (2002), **The Effect of Current Income on Aggregate Consumption**. The Economic and Social Review, 33(3), pp. 297-317.

[2] Christiano L., M. Eichembaum and C. Evans (2005), **Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy**. Journal of Political Economy, 113, pp. 1-45.

[3] Entorf H. and H. Sneesens (2000), **Aggregation with Quantity Constraints: the CES Aggregation Function**. Empirical Economics, 25, pp. 35-59.

[4] Fair R. (1993), **Testing Macro-econometric Models**. American Economic Review Paper and Proceedings, 83(2), pp. 287-293.

[5] Hildenbrand W. and A. Kneip (2002), **Aggregation under Structural Stability: the Change in Consumption of a Heterogeneous Population**. Bonn Econ Discussion Paper n°4, University of Bonn, 22 p.

[6] Kydland F. and E. Prescott (1982), **Time to Build and Aggregate Fluctuations**. Econometrica, 50(6), pp. 1345-1370.

[7] Lambert J-P (1988), **Disequilibrium Macroeconomic Models**. Cambridge University Press, Cambridge.

[8] Legendre F. (2004), **Micro-simulation et Evaluation des Politiques Economiques et Sociales : un Panorama des Développements Récents en France**. Revue d'Economie Politique, 114(1), pp. 17-53.

[9] Mincer J. (1974), **Schooling, Experience and Earnings**. Columbia University Press.

[10] Stoker T. (1993), **Empirical Approaches to the Problem of Aggregation over Individuals**. Journal of Economic Litterature, 31, pp. 1827-1874.

5 Appendix

Figure 1: Constant and time-varying β_1

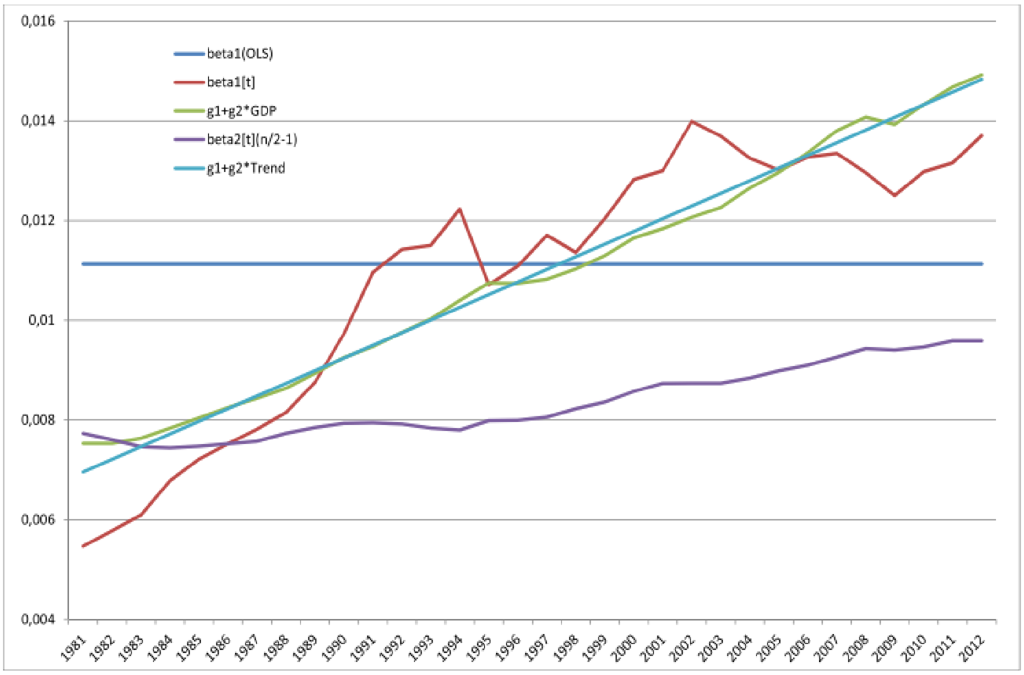


Figure 2: Constant and time-varying β_2