N O T A S E C O N Ó M I C A S 54

Imprensa da Universidade de Coimbra
Coimbra University Press

Faculdade de Economia da
Universidade de Coimbra

NOTAS

CONÓMICAS 54

Informações aos autores (Submission guidelines)
https://impactum-journals.uc.pt/notaseconomicas/about/submissions

Publicação resumida e indexada por (Indexing and Abstracting)
EconLit
RePEc
REDIB
latindex

Design
Imprensa da Universidade de Coimbra

Execução gráfica
BOOKPAPER

Tiragem 200 exemplares

Propriedade
Faculdade de Economia
Universidade de Coimbra
Av. Dias da Silva, 165
3004-512 Coimbra

Apoios (Sponsors)
Esta publicação é financiada por fundos nacionais através da
FCT – Fundação para a Ciência e a Tecnologia, I.P.,
no âmbito do projeto UIDB/05037/2020

FCT Fundação para a Ciência e a Tecnologia

Número avulso
Portugal 7,5 €
Outros países 15 €

Condições de assinatura bienal
Portugal: 20 €
Outros países: 40 €

ISSN 0872-4733

ISSN Digital 2183-203X

DOI https://doi.org/10.14195/2183-203X_54

Depósito Legal 64932/93

Contents

Vasco J. Gabriel
University of Victoria, Canada
vgabriel@uvic.ca
orcid.org/0000-0001-8659-8661
Young-Bae Kim
Cheongju University, South Korea
ybkim01@gmail.com
orcid.org/0000-0001-8208-7645
Luis Martins
ISCTE-LUI, Portugal
luis.martins@iscte-iul.pt
orcid.org/0000-0001-9422-7027
Paul Middleditch
University of Manchester, United Kingdom
paul.middleditch@manchester.ac.uk
orcid.org/0000-0002-2036-7809

DOI: https://doi.org/10.14195/2183-203X_54_1

The Inflation-Unemployment Trade-Off: Empirical Considerations and a Simple US-Euro Area Comparison

A Relação entre Inflação e Desemprego: Considerações Empíricas e uma Comparação Simples entre EUA e Zona Euro

Vasco J. Gabriel
Young-Bae Kim
Luís Martins
Paul Middleditch

Received for publication: December 28, 2021
Revision accepted for publication: March 7, 2022

ABSTRACT
This paper uses recently developed robust estimation methods to empirically reassess the long-standing inflation-unemployment trade-off debate. Indeed, we study to what extent unemployment-based New Keynesian Phillips Curves are informative about the relationship between inflation dynamics and labor market conditions. In particular, we attempt to quantify the 'elasticities' of inflation with respect to unemployment in two economies, the US and the Euro Area, whose labor market characteristics are admittedly very different. We find that the relevance of the inflation-unemployment trade-off and its empirical adequacy is greatly enhanced once the informational content of key labor market variables is explored in our estimations.
Keywords: Phillips curve; unemployment; model averaging.

JEL Classification: E24; E31; C26

1. Introduction

The inflation-unemployment trade-off has played a key role in the development of modern macroeconomics and policymaking. Central banks regularly monitor labour market conditions, as these are important sources of business cycle fluctuations (e.g. Zanetti, 2011), but also because they affect firms' pricing decisions through their impact on marginal costs. Thus, if there is a significant link between the structural features of labour markets and inflation, this will have significant consequences for monetary policy and how it is transmitted to the economy. Given the potential importance of labour market conditions for inflation dynamics and the recent developments in the literature, there is an opportunity to reassess the long-standing debate about the inflation-unemployment relationship.

The original Phillips curve essentially stemmed from the observation of a historical inverse relation between inflation and unemployment rates. This ad-hoc relationship has more recently been superseded by micro-founded structural versions, based on the behaviour of optimising forward-looking agents. The so-called New Keynesian Phillips curve (NKPC) describes inflation as being driven by inflationary pressures, either in the form of an output gap (Fuhrer and Moore, 1995) or conveyed by firms' marginal costs, as in Galí and Gertler (1999).

A distinct, but related, approach that has attracted a great deal of interest is one that incorporates labour market frictions into the NK theory of inflation. Blanchard and Galí (2007 and 2010), Ravenna and Walsh (2008), and Krause, et al. (2008), for example, show how standard measures of marginal cost are incorrect in the presence of these rigidities. Accounting for such frictions provides an additional and broader source of inflation persistence, giving rise to a NKPC that explicitly depends on unemployment – at first resembling a "traditional" Phillips curve, albeit stemming from a properly micro-founded framework.

The contribution of this paper is to empirically re-examine the inflation-unemployment relationship by making use of the NKPC with labour market frictions of Blanchard and Galí (2010, BG henceforth) and Ravenna and Walsh (2008, RW henceforth). In particular, we are interested in quantifying the elasticities of inflation with respect to unemployment of two economies, the US and the Euro Area (EA), which are known to have markedly distinct labour market characteristics. These specifications provide a theoretically sound, dynamic description of the inflation-unemployment trade-off, in contrast with traditional approaches investigating the empirical connection between unemployment and inflation with little formal theory. In turn, this allows us to relate the empirical relationships with the stylised facts of these two economies.

Indeed, we first explore a simple implication of the baseline BG formulation: while it is impossible to identify and estimate all the structural parameters of the model, we note that a reduced-form approach is still able to convey interesting information, as we can infer the level of labour market sclerosis and its relationship with inflation from the relative magnitude of the coefficients associated with unemployment and the change in unemployment. This, in itself, provides a check on the empirical adequacy and relevance of the inflation-unemployment trade-off. We then consider an extended specification discussed in both BG and RW, which offers a more complete description of the dynamic relationship between inflation and unemployment. To achieve this, we employ a recently developed model averaging approach

for instrumental variables estimation that allows us to circumvent some of the difficulties typically associated with inference on NKPCs. This is very convenient in our case, given that we are mainly interested in 'composite' unemployment elasticities.[1]

Our results suggest that a *stricto sensu* implementation of the theoretical models, although broadly adequate, provides mild empirical support only. However, once we consider several extensions that relax some of the limiting assumptions implied by the theoretical framework in BG and RW, we are able to provide a richer empirical description of the inflation-unemployment trade-off and the empirical adequacy of these models is strengthened. In particular, using additional information about key labour market variables (such as measures of the NAIRU, labour market tightness and separation rates) improves the fit and the precision of our estimations.

Our paper is related to, and complements, the work of Ravenna and Walsh (2008), which provides empirical tests on the connection between the structural features of the US labor market and inflation, and that of Krause et al. (2008), which focus on a structural approach to analyse the relevance of a specific form of labour market rigidity for the US economy (search-match frictions, albeit without allowing for real wage rigidities). Both these papers emphasise 'labour market augmented' definitions of real marginal costs as the driving variable for inflation. In contrast, we focus on the unemployment elasticities of inflation by studying versions of the same model in which inflation is explicitly written in terms of unemployment.[2] This then allows us to carry out feasible and meaningful comparisons between the US and the EA economies, given the lack of detailed data on relevant labour market variables for the latter.

While we acknowledge that the EA comprises countries with different labour market characteristics, for the purposes of our study it makes sense to treat them as a bloc. First, they share a common monetary policy framework with the European Central Bank targeting an EA wide measure of inflation. Our study aims precisely to understand to what extent inflation in the EA bloc is driven by aggregate labour market pressures. Second, recent cross-country evidence produced by Jolivet, et al. (2006) and Hobijn and Sahin (2009), for example, support the view that the US labour market is much more fluid – higher job finding rates and lower unemployment durations – than any of the EA countries, so it is interesting to understand how these differences are translated in terms of the inflation-unemployment relationship.

It is also important to recognise the inherent difficulties of our single-equation estimation approach. An alternative approach to investigate the empirical merits of these NKPC specifications would be to employ system Bayesian methods on a DSGE model (Galí et al., 2011), for example, who add unemployment as an observable to a DSGE model for the US). While a full-information approach is in principle more efficient, a misspecification in a bloc of the model may spillover to the whole model (Ruge-Murcia, 2007). Thus, our single-equation method may be seen as a complement to this approach, by allowing us to focus solely on the specification of the NKPC. Also, a method-of-moments framework is, in principle, more

[1] Although not strictly 'elasticities', we follow RW in designating the coefficients associated with the right-hand side variables of the unemployment-based NKPC as elasticities.

[2] Both BG and RW derive a linearized version of the NKPC with inflation as a function of unemployment, but do not study this empirically.

robust to statistical misspecifications, as it only requires minimal distributional assumptions. Moreover, our approach allows for simple comparisons between different economies to be carried out, which might otherwise prove too cumbersome with fully specified DSGE models.

The paper proceeds as follows. The next section describes briefly the theoretical framework that shows explicitly the role of unemployment in determining inflation dynamics. In Section 3, we analyse the baseline specifications using GMM and the Model Averaging procedure of Kuersteiner and Okui (2010), while in Section 4 we extend the empirical framework to allow for a more flexible description of the inflation-unemployment relationship. Section 5 concludes.

2. **NKPC Models with Unemployment**

A number of recent papers has attempted to modify the New Keynesian setup by introducing labour market frictions in addition to standard nominal rigidities. Here, we focus on two contributions that lead to very similar specifications of the NKPC with inflation written as a function of unemployment fluctuations.

2.1. A simplified nkpc with unemployment

BG construct a model with staggered price and nominal wage rigidities combined with Diamond-Mortensen-Pissarides-type search and match frictions, with the addition of real wage rigidities. This setup gives rise to explicit interactions amongst productivity shocks, unemployment fluctuations and inflation. By making some simplifying assumptions (hiring costs small relative to output and small separation rates; see BG for details), these authors first express the inflation rate π_t as a function of labour market tightness $\hat{x}_t$ (with $x \in [0,1]$ defined as the steady-state ratio of hires to unemployment) and log labor productivity $\hat{\alpha}_t$ (assumed to follow an AR(1) process)

$$\pi_t = \underset{\wedge}{\eta x_t}' - \Psi\gamma\hat{\alpha}_t \tag{1}$$

with " $\cdot$ " denoting variables in deviation-from-steady-state form, where $\gamma \in [0,1]$ reflects real wage rigidities, while η and Ψ are composite parameters that depend on frictions such as hiring costs, firms gross markup and the degree of price stickiness.

Noting that the relation between labour market tightness and unemployment $\hat{u}_t$ can be rewritten as

$$(1-u)\delta\hat{x}_t = -\hat{u}_t + (1-\delta)(1-x)\hat{u}_{t-1} \tag{2}$$

where $\delta \in (0,1)$ is the (exogenous) separation rate, BG then derive a simplified NKPC

$$\pi_t = \kappa\hat{u}_t + \kappa(1-\delta)(1-x)\hat{u}_{t-1} - \Psi\gamma\hat{\alpha}_t \tag{3}$$

with $\kappa = \eta/\delta(1-u)$, or equivalently

$$\pi_t = -\kappa(1-(1-\delta)(1-x))\,\hat{u}_t - \kappa(1-\delta)(1-x)\Delta\hat{u}_t - \Psi\gamma\hat{\alpha}_t \tag{4}$$

These simple equations are convenient in that they allow us to draw conclusions on the inflation-unemployment trade-off and its relationship with labour market rigidities. In particular, as BG (p. 16) comment "[t]he more sclerotic the labour market, the weaker the effect of the level of unemployment, and the stronger the effect of the change in unemployment."

Moreover, using $\zeta = (1-\delta)(1-x)$ as an overall measure of labour market rigidities, we expect the US economy to display high fluidity and low unemployment (i.e., large δ and x). In contrast, the Euro-area labour market is generally considered more sclerotic, i.e., with a larger ζ. If these effects are found in the data, this implies that the BG formulation not only captures the inflation-unemployment trade-off, but it also correctly reflects the nature of labour market rigidities in the economies under study.

As alluded to above, this seemingly simple formulation involves several deep parameters describing the model economy. However, notice that it suffices to estimate (3) and (4) in reduced form, i.e.

$$\pi_t = -\kappa_1\hat{u}_t + \kappa_2\hat{u}_{t-1} + \kappa_3\hat{\alpha}_t + \varepsilon_t \tag{5}$$

$$\pi_t = -\kappa_1^*\hat{u}_t + \kappa_2^*\hat{u}_{t-1} + \kappa_3^*\hat{\alpha}_t + \varepsilon_t^* \tag{6}$$

where $\kappa_1 = -\kappa$, $\kappa_2 = -\kappa\zeta$, $\kappa_1^* = -\kappa(1-\zeta)$, $\kappa_2^* = -\kappa\zeta$, and $\kappa_3 = \kappa_3^* = -\Psi\gamma$ and ε_t, ε_t^* are uncorrelated disturbances. Estimating (5) allows us to identify separately κ and ζ, while (6) is used to compare the relative magnitude of the effects of the level of unemployment and the change in unemployment for the two economies.[3]

2.2. An extended NKPC with unemployment

The simplicity of the model above is quite appealing from an empirical perspective. Nevertheless, it is also interesting to consider the richer specifications of the unemployment-based NKPC of both BG and RW.[4] The latter also incorporates a theory of unemployment into the NKPC setup with search frictions. Interestingly, both papers lead to a very similar extended NKPC of the form

$$\pi_t = -\beta E_t\pi_{t+1} - \kappa_0\hat{u}_t + \kappa_L\hat{u}_{t-1} + \kappa_F E_t\hat{u}_{t+1} - \kappa_p\hat{\alpha}_t \tag{7}$$

[3] It should be noted that κ itself is a nonlinear function of the separation rate δ and labour market tightness x, which makes identification and subsequent estimation of these parameters difficult without strong assumptions (and calibration) about the remaining structural parameters, something we wish to avoid here.

[4] The BG framework can also be viewed as an employment adjustment cost model and therefore slightly distinct from RW, but for the purpose of our analysis, these differences are not crucial.

where, again, the composite elasticities κ_0, κ_L, κ_F and κ_p are complicated functions of deep parameters describing labour market frictions and nominal rigidities (see BG and RW for details). One of the distinctive features of RW is the inclusion of a 'cost channel' effect, whereby the real interest rate has a direct impact on inflation, thus adding another channel for monetary policy to affect inflation. We will consider this effect in the empirical section, by estimating an extended version of (7), namely

$$\pi_t = -\beta E_t\pi_{t+1} - \kappa_0\hat{u}_t + \kappa_L\hat{u}_{t-1} + \kappa_F E_t\hat{u}_{t+1} - \kappa_p\hat{\alpha}_t + \kappa_p\hat{r}_t + \varepsilon_t \quad (8)$$

where $\hat{r}_t$ is the real interest rate $(i_t - E_t\pi_{t+1})$.

Equations (7)-(8) have interesting implications. First, under sensible values for the underlying structural parameters, we expect the magnitude of κ_0 to dominate κ_L and κ_F. Also, the model predicts that inflation is a lot more responsive to unemployment dynamics if the labour markets are rigid, i.e., we would expect larger (in absolute terms) κ_0, κ_L, and κ_F for the EA when compared to the US.[5] This is consistent with RW's calibration exercise, which suggests that the higher the separation and vacancy rates (higher fluidity), the lower the κ_0, while the higher the labour share of surplus (e.g. in an economy where workers have higher bargaining power), the larger the inflation elasticity with respect to unemployment. Furthermore, RW demonstrate that the magnitude of the cost channel effect depends positively on the rigidity of labour markets, so that we expect a larger κ_R for the EA.

RW estimate an equation in which inflation depends on the probability of filling a posted vacancy, denoted as $\hat{q}_t$ in their paper, rather than on unemployment. While data for $\hat{q}_t$ (the ratio between the job finding probability and labour market tightness) is available for the US, that is not the case for the Euro Area, hence the convenience of using (8) for our comparison. However, the effects go in the same direction: the higher the (contemporaneous) unemployment rate, the higher the probability of filling a posted vacancy and therefore the impact on inflation should be negative, the converse being true for expected values of $\hat{u}_{t-1}$ and $\hat{q}_{t+1}$, as predicted by the models.

Given that we are mainly interested in obtaining empirical estimates of these elasticities for the US and the EA, this reduced-form specification entails little loss of information, as discussed next.

3. **Baseline Estimations**

3.1. Data and methods

In order to empirically quantify the predictions of the models discussed in the previous section, we use quarterly data for the sample period 1970-2007. The start of the sample period is determined by the availability of reliable data for the Euro Area, while we restrict the study up to 2007 because of (public) data availability for synthetic measures of some

[5] Using BG's calibration, we would have $\kappa 0 \simeq 0.14$, $\kappa L \simeq 0.06$, and $\kappa F \simeq 0.08$ for the EA, while for the US the values would be $\kappa_0 \simeq 0.09$, $\kappa_L \simeq 0.02$, and $\kappa_F \simeq 0.06$.

important labour market variables. The sampling period is similar to RW, though, thus ensuring some degree of comparability.

US data is taken from the FRED database at the Federal Reserve Bank of St Louis, namely US inflation measured by the GDP deflator (other measures such as CPI or PCE expenditures produce similar results), the (demeaned) unemployment rate and labour productivity, measured as the log ratio of GDP and total non-farm employment.[6] The corresponding data for the EA comes from the updated Area-Wide Model database of Fagan et al. (2001), with similar transformations applied.

Endogeneity issues usually plague NKPC formulations, namely due to the presence of expectations in these models (which lead to forecast errors picked up by the error term), errors-in-variables due to the use of proxies (for marginal costs, productivity, etc.) and because shocks affecting inflation are likely to be correlated with the driving variables as well. To deal with these problems, estimation is usually carried out by GMM, although this itself entails several problems (see Martins and Gabriel, 2009; and Kleibergen and Mavroeidis, 2009). Indeed, inference concerning the NKPC is plagued with the 'many weak instruments curse', partly because any information these instruments may contain on future inflation fluctuations would have already been exploited by the central banks to contain inflation. In practice, this leads to a very limited number of (arbitrarily chosen) instruments being used, resulting in a loss in efficiency.

To attenuate these difficulties, we suggest using the procedure proposed by Kuersteiner and Okui (2010), which allows us to construct "optimal instruments" by applying a model averaging (MA) approach to the first stage of two-stage least squares (2SLS). This procedure has several advantages: i) it delivers a more favourable trade-off between bias and efficiency relative to estimators that rely on a single set of instruments; ii) no ad-hoc choice of instruments or weak instruments pre-testing must be entertained, and iii) it possesses good finite sample properties even when there are many weak instruments available, which is likely to be our case.

The weights for first-stage model averaging are chosen to minimise the asymptotic mean squared error and are found numerically as the solution of a quadratic programming problem. We use the MA-2SLS$_P$ version, where the weights are constrained to be positive, in the interval [0,1] (see Kuersteiner and Okui, 2010, for details). For comparison, however, we also present results from GMM estimation with a heteroskedasticity and autocorrelation consistent Newey-West weight matrix.

For conciseness, we focus on a baseline instrument set containing two lags of the variables appearing explicitly in each estimated model (inflation rate, unemployment, productivity and real interest rates in the case of (8)), plus the labour share, commodities price inflation, wage inflation, interest rate spread and HP-filtered real GDP, thus following the previous literature on the NKPC (e.g. Galí and Gertler, 1999; and Ravenna and Walsh, 2008). Although results vary little when more lags are included (or some variables excluded), we found that the Stock and Yogo (2005) 2SLS-bias based test usually failed to reject the null

[6] We use productivity in levels, as in RW. These authors show that it is relatively straightforward to allow for long-run productivity growth in their setup. Detrending productivity, regardless of the detrending/filtering method, results in the corresponding estimated elasticity being positive, which is not consistent with the theoretical predictions. Also, using productivity defined as output per hour (only possible for the US) delivers almost identical results.

of weak instruments (though, strictly speaking, this test is valid for IV estimation only). For the MA-2SLS estimation, in particular, changes in the instrument set had very little impact on final results, mainly due to the first-step averaging, which underlines the robustness of this procedure.

3.2. Results

Table 1 presents the estimation results for equations (5)-(6) under this baseline setting. A general assessment reveals that the MA-2SLS estimator tends to deliver slightly smaller, but more precisely estimated, elasticities (a general feature throughout the paper). Also, parameter estimates are broadly similar across the different estimators, which suggests that results are relatively robust. Under column Sargan-*J*, we see that tests of over-identifying restrictions (the usual J-test for GMM and a Sargan-type test for MA-2SLS) were also satisfied.

Table 1: Reduced-form Phillips Curve, equations (5)-(6), 1970:1-2007:4

	From eq. (5)						
	κ_1^*	κ_2^*	κ_3^*	κ	ζ	Sargan-*J*	Stock-Yogo
US							
GMM	0.184 (0.066)	0.710 (0.296)	– 0.014 (0.005)	– 0.894 (0.321)	0.794 (0.077)	0.729	8.819
MA-2SLS$_p$	0.252 (0.032)	0.460 (0.151)	– 0.003 (0.001)	– 0.511 (0.148)	0.899 (0.002)	1.000	§
Euro-area							
GMM	0.134 (0.021)	1.242 (0.334)	– 0.036 (0.004)	– 1.376 (0.337)	0.903 (0.026)	0.614	12.861*
MA-2SLS$_p$	0.191 (0.021)	0.950 (0.246)	– 0.029 (0.003)	– 1.018 (0.242)	0.933 (0.033)	1.000	§

Notes: Standard errors (s.e.) in parentheses; nonlinear Wald-type s.e. in the case of ζ; the Sargan-*J* column refers to *p*-values of tests for over-identifying restrictions; Stock-Yogo refers to the Stock and Yogo (2005) statistic (based on 2SLS bias) for the null hypothesis of weak instruments; * denotes rejection at the 10% significance level (critical value: 10.47); § denotes the same statistic.

As explained above, we are interested in comparing the magnitudes of ζ and the κ^*'s for the two economies, which should reveal the relative degree of labour market sclerosis. The results lend some support to the view that the US labour market appears to be more fluid than the Euro-area one and, therefore, the nature of the inflation-unemployment trade-off is distinct for these two economies. The point estimate for the composite parameter ζ that we derive from the estimations is lower for the US, but not substantially so. Indeed, although we cannot test whether or not the coefficients are the same for the two economies, we observe

that the implicit confidence intervals overlap, so we cannot confidently say that this measure of overall labour market rigidity is significantly different for the US and the Euro-area.

However, we find the impact of current unemployment on inflation (κ_1^*) to be stronger in the US case relative to $\Delta\hat{u}_t$, while the effect of changes in unemployment (κ_2^*) is relatively larger for the Euro-area: taking κ_1^* as a proportion of κ_2^*, the lowest value for the US is 0.26, while the highest for the Euro-area is 0.20. Interestingly, a direct comparison of the coefficients of the two economies agrees with this conclusion. In fact, κ_1^* tends to be larger for the US, while κ_2^* dominates in the Euro-area case.

Note also that the coefficient κ_3^* on labour productivity delivers the expected sign for both economies, indicating a negative relationship with inflation. This reflects in-sample underlying trends for these variables, with inflation displaying a long decline, while productivity has steadily grown. Noticeably, our estimates indicate that this variable has a smaller effect on US inflation than on European inflation (although broadly in line with the results in RW). Recall that the (composite) labour productivity elasticity is an increasing function of the degree of real wage rigidit ies, so these results reinforce the view that labour market frictions play a more significant role in European inflation dynamics. In the next section, we will attempt to appraise the role of productivity in more detail.

Table 2: BG-RW extended model, baseline case

	Model: $\pi_t = \beta E_t\pi_{t+1} + \kappa_0^*\hat{u}_t + \kappa_L^*\hat{u}_{t-1} + \kappa_F^*E_t\hat{u}_{t+1} + \kappa_p^*\hat{\alpha}_t + \kappa_R^*\hat{r}_t$						
	β	κ_0^*	κ_L^*	κ_F^*	κ_p^*	κ_R^*	Sargan-J
US							
GMM	0.877 (0.038)	– 0.982 (0.412)	0.336 (0.195)	0.694 (0.231)	– 0.001 (0.0004)		0.589
MA-2SLS$_p$	0.804 (0.058)	– 0.010$^{\#}$ (0.120)	0.132 (0.080)	0.171 (0.073)	– 0.004$^{\#}$ (0.003)		0.835
GMM	0.903 (0.057)	– 1.150 (0.382)	0.529 (0.174)	0.688 (0.240)	– 0.008 (0.003)	0.028 (0.014)	0.438
MA-2SLS$_p$	0.806 (0.058)	– 0.013$^{\#}$ (0.126)	0.127$^{\#}$ (0.085)	0.139 (0.081)	– 0.003$^{\#}$ (0.003)	0.014 (0.012)	0.985
Euro Area							
GMM	0.953 (0.089)	– 0.974$^{\#}$ (1.453)	0.131$^{\#}$ (0.663)	– 0.861$^{\#}$ (0.817)	– 0.000$^{\#}$ (0.003)		0.277
MA-2SLS$_p$	0.616 (0.110)	– 0.470$^{\#}$ (0.442)	– 0.166$^{\#}$ (0.238)	0.623 (0.261)	– 0.011 (0.004)		0.361
GMM	0.843 (0.071)	– 3.440 (1.025)	– 1.819 (0.502)	– 1.660$^{\#}$ (0.551)	0.001$^{\#}$ (0.001)	0.033 (0.015)	0.507
MA-2SLS$_p$	0.841 (0.220)	– 3.876$^{\#}$ (3.085)	– 2.003$^{\#}$ (1.480)	– 1.941$^{\#}$ (1.645)	0.006$^{\#}$ (0.007)	0.057$^{\#}$ (0.036)	0.367

Notes: See notes to Table 1; $^{\#}$ denotes not significant at the 10% significance level; $\hat{r}_t$ is measured as the short term (3-month) interest rates minus the inflation rates in period t+1; Stock-Yogo critical values are only available for models with up to three endogenous regressors.

Turning to the extended BG-RW specification discussed in section 2.2, estimates of equations (7)-(8) are presented in Table 2.[7] The most noticeable result is the fact that, although the relative magnitude of the coefficients is the anticipated one (with κ_0^* dominating κ_L^* and κ_F^* and these elasticities being larger for the EA than the US), the empirical adherence of this extended specification is somewhat weak. Indeed, several of the estimated coefficients are statistically insignificant and in some cases the signs of the coefficients are not in accordance with the model. This is particularly more severe in the case of the EA, where most parameters are not significant and κ_L^* and κ_F^* display the wrong (negative) sign. Interestingly, the productivity elasticity κ_p^* is estimated to be smaller, for both economies, than in the simpler version estimated in Table 1, while the cost channel given by κ_R^* is non-negligible, although smaller than that reported in RW for the US (ranging between 0.07 and 0.15).

[7] We follow the typical practice of parodying expected values of inflation (and unemployment here as well) by realised values, which naturally introduces an additional measurement disturbance in the error term. See Dufour et al. (2006), who consider using survey-based measures of expectations, as well as Kleibergen and Mavroeidis (2009) and discussions therein.

Also, the results differ little across GMM and the MA estimator, although the latter tends to produce more insignificant coefficients.

These difficulties are not wholly unexpected, as the strong persistence in unemployment rates, especially for the EA (autocorrelation coefficient in excess of 0.95), means that it will be difficult to distinguish between the effects of current unemployment and lags/leads of this series on inflation, thus leading to imprecision in the estimations. This is not inconsistent with the relative success of the baseline specification estimated in Table 1, as it only depends on current unemployment. In the next section, we consider ways of improving the empirical fit of these specifications.

4. **Extensions**

The baseline results discussed above focus on a simple implementation of BG, which nevertheless produces interesting insights, broadly consistent with the stylised facts about labour market rigidities in the US and the Euro-area and that are robust across different estimation methods. Nevertheless, the strict adherence to the theoretical framework of BG and RW is somewhat limiting from an empirical perspective. Thus, we next consider a few extensions that add realism to our empirical exercise, analysing their individual impact on the baseline estimations.[8]

4.1. Time-varying nairu

Albeit convenient from a theoretical point of view, the assumption of a constant steady-state level of unemployment ($\bar{u}_t$) in the specification above is restrictive and runs counter the evidence of fluctuations in estimated NAIRUs, both for the US and the Euro Area (see Staiger et al., 1997; and Fabiani and Mestre, 2000 and 2004). Thus, we explore how the baseline results are affected by allowing for this source of variation in $\bar{u}_t$.

Several methods for estimating the NAIRU are possible and there is considerable uncertainty about its true level. We considered "official" estimates of the NAIRU, as well as estimates based on our sample. In the case of the US, the Congressional Budget Office publishes a NAIRU based on their potential output estimates. In addition, the OECD also publishes annual estimates of the NAIRU for both the US and the Euro Area. In the latter case, estimates are only available from 1991 onwards, so we use estimates from Fabiani and Mestre (2004) for the earlier part of the sample. Given that these NAIRU measures only display annual variation, the resulting quarterly stepwise deviations are HP-filtered to create a smoother series. It turns out that the resulting measures of $\bar{u}_t$ are highly correlated with those obtained from simple HP-filtering or quadratic detrending of the corresponding unemployment rate series (correlations in excess of 0.8). To save space, while ensuring consistency for both economies, we report results using the OECD-based NAIRU deviations (results vary little when other measures are employed).

[8] In what follows, the instruments are naturally based on the lags of the new variables considered under each specification.

Table 3: Equations (5)-(6) with new variables

Using NAIRU as $\bar{u}_t$ (with baseline $\hat{\alpha}_t$)					From eq. (3)			
		κ_1^*	κ_2^*	κ_3^*	κ	ζ	Sargan-J	Stock-Yogo
US	GMM	0.222 (0.072)	0.683 (0.297)	– 0.014 (0.005)	– 0.905 (0.320)	0.755 (0.049)	0.736	13.988*
	MA-2SLS$_p$	0.267 (0.034)	0.485 (0.151)	– 0.004 (0.0002)	– 0.551 (0.147)	0.879 (0.002)	1.000	§
Euro Area	GMM	0.196 (0.031)	1.566 (0.314)	– 0.049 (0.002)	– 1.762 (0.324)	0.889 (0.021)	0.553	13.656*
	MA-2SLS$_p$	0.194 (0.021)	1.429 (0.178)	– 0.037 (0.003)	– 1.573 (0.531)	0.909 (0.029)	1.000	§
US only								
Using Fernald TFP as $\hat{\alpha}_t$ (with baseline $\bar{u}_t$)								
	GMM	0.203 (0.069)	0.751 (0.315)	– 0.006 (0.002)	– 0.954 (0.345)	0.787 (0.073)	0.699	13.991*
	MA-2SLS$_p$	0.257 (0.026)	0.401 (0.155)	– 0.0001 (0.0001)	– 0.559 (0.155)	0.719 (0.001)	0.999	§
NAIRU as $\bar{u}_t$ and Fernald TFP as $\hat{\alpha}_t$								
	GMM	0.237 (0.076)	0.667 (0.312)	– 0.005 (0.002)	– 0.904 (0.339)	0.738 (0.097)	0.714	14.132*
	MA-2SLS$_p$	0.194 (0.036)	0.410 (0.174)	– 0.001 (0.000)	– 0.504 (0.171)	0.813 (0.001)	1.000	§
NAIRU as $\bar{u}_t$ and Fernald TFP as $\hat{\alpha}_t$, plus separation rate and tightness								
US	κ_1^*	κ_2^*	κ_3^*	κ	ζ	tightness	sep. rate	Sargan-J
GMM	0.324 (0.161)	0.802 (0.280)	– 0.003 (0.001)	– 1.113 (0.376)	0.721 (0.080)	0.021 (0.002)	0.810 (0.132)	0.802
MA-2SLS$_p$	0.308 (0.034)	0.432 (0.164)	– 0.004 (0.001)	– 0.540 (0.103)	0.801 (0.001)	0.011 (0.003)	1.033 (0.107)	1.000
NAIRU as $\bar{u}_t$ and tightness								
Euro-area								
GMM	0.215 (0.078)	1.478 (0.259)	– 0.052 (0.005)	– 1.792 (0.256)	0.825 (0.083)	0.088 (0.028)	–	0.705
MA-2SLS$_p$	0.224 (0.088)	1.388 (0.431)	– 0.041 (0.005)	– 1.612 (0.406)	0.861 (0.072)	0.011 (0.033)	–	1.000

Notes: See notes to Table 1. The first panel reports estimates using the NAIRU as $\bar{u}_t$ and labour productivity as $\hat{\alpha}_t$ (baseline case); the middle panel reports estimates for the US using the utilisation-adjusted TFP measure, first with the unemployment rate as $\bar{u}_t$ (baseline case); the bottom panel includes labour market tightness (and separation rates for the US) as additional regressors.

The results are shown in the first panel of Table 3, with no noticeable differences when compared to the baseline results of Table 1. Indeed, the magnitude of the κ_i^*, κ and ζ coefficients is similar. This suggests that controlling for a time-varying steady-state unemployment level does not change the nature of the baseline results discussed above. However, given that it offers a more realistic interpretation of movements in steady-state levels of unemployment, we will also consider this measure for the extended BG-RW specification further below.

4.2. The role of productivity measures

The theoretical framework outlined above does not include capital and defines a production function that is linear in labour, which implies that labour productivity coincides with total factor productivity (TFP). Ideally, one would use measures of "exogenous TFP", but it is well known that TFP as conventionally calculated may be mismeasured, due to variable input utilisation, non-constant returns to scale, reallocation effects, etc. (Basu, Fernald and Kimball, 2006). These effects make measured TFP endogenous and may lead to biased coefficients in estimated Phillips curves, as we found in the baseline estimation results. Suppose cycles are primarily caused by non-technology shocks, such as demand shocks. A positive demand shock raising marginal costs induces an increase in inflation, but it also increases observed productivity due to an increase in utilisation.

To attenuate these potential biases, we re-estimate the baseline BG model using the utilisation-adjusted quarterly TFP series constructed by Fernald (2012). This variable still omits some of the corrections that may be done with annual data, but it is, as far as we know, the best proxy available for true "exogenous" TFP.

Unfortunately, neither standard nor utilisation-adjusted TFP measures are available for the Euro Area at a quarterly frequency and for our sample period. Existing measures are annual from 1980 onwards and do not include all Euro Area countries.[9] Thus, our analysis focuses on the US case only. We conjecture that the estimation results, relative to those using labour productivity, would be probably similar had a TFP measure been available for the Euro Area. Indeed, the correlation between the available annual TFP growth measure and labour productivity growth is 0.71, while the correlation between the corresponding US variables at the quarterly frequency is 0.70.

The middle panel of Table 3 contains results for estimations based on the Fernald (2012) measure of US total factor productivity. The first set reports results using unemployment rates, while the second set considers the NAIRU deviations as $\tilde{u}_t$, as discussed above. At a first glance, estimates are largely similar to the baseline results of Table 1 and the top panel of Table 3, regardless of the measure for $\tilde{u}_t$. However, the coefficient κ_3^* associated with $\hat{\alpha}_t$ is now estimated to be considerably smaller, thus suggesting that "purified" technology shocks, while relevant, appear to play a minor role in explaining the dynamics of US inflation. Also, note that ζ is now estimated to be smaller than in the baseline estimations, indicating that the degree of labour market rigidities is smaller once corrected measures of technology shocks are employed. Again, this is consistent with the idea that labour productivity will

[9] Data available from the EU KLEMS database.

include effects on the intensive margin that may lead to higher decreases in marginal costs and therefore have a more negative impact on inflation, but in a way that it also biases the coefficients associated with unemployment. This suggests that using a "purified" TFP measure is important in an empirical setup, to ensure consistency with the theory and to avoid misspecifications.

The presence of productivity in the NKPC, although desirable from a theoretical point of view, raises some empirical difficulties, given the absence of precise measures of productivity. For robustness, we also considered a restricted version of the baseline specification that allows us to obtain estimates of some parameters of interest, namely ζ, without the use of $\hat{\alpha}_t$. Indeed, assuming that $\hat{\alpha}_t$ follows an AR(1) process with autoregressive coefficient ρ, then (3) can be rewritten as

$$\pi_t = \rho\pi_{t-1} + \phi_1\hat{u}_t + \phi_2\hat{u}_{t-1} + \phi_3\hat{u}_{t-2} \tag{9}$$

where $\phi_1 = -\kappa$, $\phi_2 = \kappa(\zeta + \rho)$ and $\phi_3 = -\rho\kappa\zeta$ so that π_t is now a function of current and two lags of unemployment, with a more traditional 'intrinsic' persistence component in π_{t-1}.

Table 4: Restricted version, Eq. (9), GMM estimation

	ρ	ϕ_1	ϕ_2	ϕ_3	ζ	Sargan-*J*
Baseline						
US	0.839 (0.058)	– 0.865 (0.214)	1.449 (0.326)	– 0.607 (0.131)	0.836 (0.142)	0.244
Euro-area	0.934 (0.022)	– 0.841 (0.388)	1.465 (0.709)	– 0.635 (0.328)	0.814 (0.128)	0.237
New variables						
US	0.852 (0.037)	– 0.582 (0.221)	0.863 (0.344)	– 0.343 (0.131)	0.743 (0.114)	0.616
Euro-area	0.909 (0.021)	– 1.432 (0.368)	1.583 (0.651)	– 1.164 (0.299)	0.894 (0.093)	0.338

Notes: See notes to Table 1; MA-2SLS and Stock-Yogo tests are not available in the case of nonlinear estimation.

Table 4 reports estimation results when we directly recover the parameters using their nonlinear relationships, both for the baseline variables of Section 2 (first panel) and when we use the NAIRU and TFP measures. Again, the results, although slightly less precise, are remarkably consistent with previous estimations, namely for the relative magnitudes of κ and ζ for the two economies. Note that using the variables introduced in this section improves the fit and precision in these estimations. Moreover, it is interesting to observe that the persistence of productivity, as captured by ρ, is higher for the EA, which helps to explain the larger (negative) impact of productivity on inflation noted previously, as one could anticipate from the BG setup. Thus, it seems sensible to include measures of productivity in estimations of the NKPC, as the biases stemming from the use of proxies for $\hat{\alpha}_t$ appear to be relatively modest.

4.3. Labour market variables

As discussed above, a potential shortcoming of our baseline exercise is the fact that we focus on (constant) steady-state levels of some relevant variables. Note that from (2) there is a direct relationship between tightness, separation rates and unemployment rates, which suggests that they could be used interchangeably to study the effects on inflation dynamics. However, as explained above, there can be significant discrepancies between the theoretical framework and its empirical implementation. Indeed, it is likely that cyclical variations in labour market tightness and separation rates, for example, in addition to their influence in unemployment dynamics, may have additional explanatory power in determining the inflation-unemployment trade-off.

Indeed, as discussed in Shimer (2005), the cyclical component of labour market tightness (the vacancy-unemployment ratio) displays a great deal more volatility than predicted by standard models (and than most of the variables studies so far). On the other hand, while there has been some debate about the cyclicality of separation rates (Hall, 2005), recent literature (Fujita and Ramey, 2009 and 2012; Elsby et al., 2009) emphasises the contribution of changes in separation rates to unemployment fluctuations. Thus, it is an interesting exercise to assess how variations in these key labour market variables affect inflation dynamics, even after we control for variation in unemployment and productivity.[10]

Measures of labour market tightness and time varying separation rates are readily available for the US (see Shimer, 2005; Shimer, 2012; and the author's webpage).[11] For the Euro Area, data availability is an issue for all variables that are not present in the AWM database. Nevertheless, we were able to construct a time series for labour market tightness (the vacancies to unemployment ratio) in the Euro Area for our sample period, by constructing a vacancies index (based on data available for approximately two thirds of the EA countries, as in Christoffel et al., 2009) and, using the same methodology to construct EA-wide unemployment levels.[12] Following Shimer (2005), we extract the cyclical component of this measure by employing the HP filter with smoothing parameter 10^5 (a similar transformation is used for the US separation rates).

We then gauge the contribution of these variables in two concurrent ways. First, we add these variables to the baseline BG equation, therefore directly controlling for variation in the key labour market variables. Second, indirectly, by adding lags of labour market tightness (and separation rates in the case of the US) to the instrument set, thus utilising correlations of these variables with unemployment and productivity measures to obtain potentially better estimates of the inflation-unemployment trade-off.

Results for this exercise are displayed in the bottom panel of Table 3. Including these new variables generally improves the fit and estimation precision, but in a way that reinforces the initial conclusions. Indeed, *i*) the composite parameter ζ measuring the degree of

[10] Note that this echoes Ravenna and Walsh (2008) NKPC specification that is explicitly written in terms of the probability of filling a vacancy, itself a function of labour market tightness.

[11] For labour market tightness, US vacancies were constructed by splicing the Help Wanted Advertising Index used in Shimer (2005) and data from the Job Openings and Labor Force Turnover Survey (JOLTS).

[12] The composition of the EA is updated as data for both vacancies and unemployment levels become available for each country (source: OECD Statistics).

rigidities is lower for the US than the EA, but more so than in the baseline estimation, while the relative magnitude of the κ_i^*'s support the premise of the BG model that the more sclerotic the labour market, the stronger the elasticity of $\Delta\hat{u}_t$ relative to the effect of $\hat{u}_t$. The lowest ratio of estimated κ_1^* to κ_2^* is now 0.40 for the US, while for the EA the highest ratio is 0.16.

It is also interesting to check the direct impact of labour market tightness on inflation. The tightness coefficient, although estimated to be relatively small (but larger than productivity's in absolute terms for the US), is highly significant and positive, as expected.[13] This indicates that there may be further cyclical effects in labour market tightness influencing inflation, but that are not captured solely by deviations from the NAIRU.

This is further supported by the inclusion of time-varying separation rates in the case of the US economy. Allowing for a time-varying separation rate uncovers additional responses of inflation to labour market conditions, with a sizeable and significant estimated coefficient. This can be explained by the fact that both the US inflation and separation rates appear to share a secular decline (even after HP filtering the latter variable), particularly after the early 1980's. This decline in separation rates was noted by Shimer (2012), but not previously related to inflation dynamics.

4.4. The extended RW-BG specification with new variables

We now assess how the empirical modifications discussed above impact on estimates of the extended RW-BG specification. In Table 5 we present results with NAIRU as $\hat{u}_t$ for both the US and EA, Fernald's (2012) TFP as $\hat{\alpha}_t$ for the US and adding labour market tightness (and separation rates for the US) to the instrument set. Doing so delivers a significant improvement in terms of fit compared with Table 2, particularly for the US case, with all coefficients now statistically significant, with the correct signs and with sensible magnitudes, and similar for both the GMM and MA estimators. Indeed, we find a one-to-one relationship between inflation and current deviations from the NAIRU, with $\kappa_0 > \kappa_F > \kappa_L$, as expected. As before, the effects of productivity and real interest rate fluctuations are mostly significant, but relatively small. The results are also consistent with the RW estimations based on $\hat{q}_t$, as discussed in section 2.2.

[13] This is consistent with the finding of Ravenna and Walsh (2008), who consider the (negative) effects on inflation of the probability of filling a posted vacancy, itself inversely proportional to labour market tightness.

Table 5: BG-RW extended model, new variables

	Model: $\pi_t = \beta E_t\pi_{t+1} + \kappa_0^*\hat{u}_t + \kappa_L^*\hat{u}_{t-1} + \kappa_F^*E_t\hat{u}_{t+1} + \kappa_p^*\hat{\alpha}_t + \kappa_R^*\hat{r}_t$						
	β	κ_0^*	κ_L^*	κ_F^*	κ_p^*	κ_R^*	Sargan-J
US							
GMM	0.931 (0.030)	– 0.943 (0.364)	0.350 (0.180)	0.481 (0.203)	– 0.001 (0.0001)	–	0.820
MA-2SLS$_p$	0.787 (0.053)	– 1.015 (0.117)	0.252 (0.081)	0.415 (0.070)	– 0.002 (0.001)	–	0.986
GMM	0.861 (0.024)	– 1.089 (0.262)	0.424 (0.121)	0.681 (0.145)	– 0.001 (0.0001)	0.007 (0.004)	0.716
MA-2SLS$_p$	0.785 (0.052)	– 1.008 (0.121)	0.411 (0.085)	0.512 (0.078)	– 0.002 (0.001)	0.017 (0.012)	0.975
Euro Area							
GMM	0.913 (0.076)	– 2.865 (1.148)	1.10 (0.541)	1.803 (0.636)	– 0.002$^{\#}$ (0.003)	–	0.404
MA-2SLS$_p$	0.693 (0.115)	– 0.899 (0.266)	0.114$^{\#}$ (0.174)	0.749 (0.144)	– 0.011 (0.006)	–	0.862
GMM	0.759 (0.088)	– 3.846 (1.301)	1.563 (0.653)	2.305 (0.678)	– 0.010 (0.004)	0.007$^{\#}$ (0.011)	0.389
MA-2SLS$_p$	0.753 (0.111)	– 0.709 (0.325)	0.103$^{\#}$ (0.202)	0.502 (0.161)	– 0.013 (0.005)	0.055 (0.017)	1.000

Note: See notes to Table 2.

In the case of the EA, the empirical adequacy of this model is also improved, although some coefficients are occasionally insignificant (in particular κ_L with the MA estimator). The unemployment elasticities are estimated to be larger than those of the US, as expected, given that the κ's are proportional to the degree of labour market rigidities. Also, both their relative magnitudes and their 'net effect' on impact (i.e. $\kappa_0 - \kappa_F - \kappa_L$) are in accordance with the calibration exercises in BG and RW. The improvements in fit appear to be due to the use of additional information concerning labour markets and the fact that deviations from the NAIRU are slightly less persistent than unemployment rates, thus helping to discriminate the dynamic effects of $\hat{u}_t$ on inflation.

5. **Conclusion**

Although the inflation-unemployment trade-off is at the heart of modern macroeconomics, most empirical studies consider alternative drivers of inflation dynamics in competing specifications, namely measures of real activity such as (proxies for) marginal cost

or output gaps. However, empirical support for these relationships is mixed, at best. This paper attempts to shed light on the empirical usefulness of unemployment-based NKPCs stemming from a fully micro-founded framework, more particularly in explaining observed differences between the US and the Euro Area economies. Although our study focuses on a simple implementation of Ravenna and Walsh (2008) and Blanchard and Galí (2010), it nevertheless produces interesting insights.

We find that these specifications are broadly consistent with the stylised facts about inflation persistence and labour market rigidities in the US and the Euro-area. More precisely, we found that, once appropriate adjustments are made and the informational content of relevant labour market variables (such as the NAIRU, time-varying separation rates and labour market tightness) is explored, an unemployment-based NKPC, both in its simple and extended form, produces results that are in line with the theoretical predictions. Indeed, our results are robust across different estimation methods and show that unemployment and productivity elasticities are larger for the EA compared to the US. Given that these elasticities reflect underlying labour markets characteristics, this, in turn, is consistent with the view that the US labour market is considerably more fluid that the Euro-Area one, as discussed by Jolivet et al. (2006), Hobijn and Sahin (2009), and Elsby et al. (2013).

It is important to recognise that several of the limitations identified in empirical studies of the NKPC are also present here. First, there are clear identification difficulties, with the specifications studied here depending on a considerable number of deep parameters that cannot be recovered in a single-estimation setup. The fact that our interest is on 'reduced-form' elasticities mitigates, but cannot obviate, this fact. Second, the models studied here depend on expectations for inflation and unemployment and also measures of productivity, all of these difficult to measure or observe accurately.

Nevertheless, we suggest that exploring the additional information provided by labour market variables may help to understand inflation rate dynamics and, therefore, to better inform economic policy. This is, in fact, consistent with the current practice of some central banks, such as the "forward guidance" principle, which puts labour market conditions at the centre of monetary policy decisions.

An interesting challenge would be to study the inflation-unemployment relationship in the aftermath of the Great Recession. Indeed, inflation has remained relatively stable, despite the increase in unemployment, though this has also been accompanied by a decline in productivity. Given that the models studied here hinge on the relationship among inflation, unemployment and productivity, they will provide a useful tool in disentangling the contribution of these effects to the recent modest and protracted decline of inflation rates. Thus, once more data becomes available, further empirical research on this topic would clearly be a worthwhile pursuit.

References

Basu, S.; Fernald, J. G.; Kimball, M. S. (2006) Are technology improvements contractionary? *American Economic Review*, 96, 1418-1448.

Blanchard, O.; Galí, J. (2007) Real wage rigidities and the New Keynesian model. *Journal of Money, Credit and Banking*, 39, 35–65.

Blanchard, O.; Galí, J. (2010) Labor markets and monetary policy: A New Keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2, 1-30.

Dufour, J-M.; Khalaf, L.; Kichian, M. (2006) Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis. *Journal of Economic Dynamics and Control*, 30, 1707-1727.

Elsby, M. W. L.; Michaels, R.; Solon, G. (2009) The ins and outs of cyclical unemployment. *American Economic Journal: Macroeconomics*, 1, 84-110.

Elsby, M. W. L.; Hobijn, B.; Sahin, A. (2013) Unemployment dynamics in the OECD. *Review of Economics and Statistics*, 95, 530-548.

Fabiani, S.; Mestre, R. (2000) Alternative measures of the NAIRU in the euro area: Estimates and assessment, European Central Bank Working Paper no. 17.

Fabiani, S.; Mestre, R. (2004) A system approach for measuring the euro area NAIRU. *Empirical Economics*, 29, 311-341.

Fagan, G.; Henry, J.; Mestre, R. (2001) An Area-Wide Model (AWM) for the Euro Area, European Central Bank Working Paper no. 42.

Fernald, J. (2012) A quarterly utilization-adjusted series on total factor productivity, Federal Reserve Bank of San Francisco Working Paper 2012-19.

Fuhrer, J. C.; Moore, G. R. (1995) Inflation persistence. *Quarterly Journal of Economics*, 440, 127-159.

Fujita, S.; Ramey, G. (2009) The cyclicality of separation and job finding rates. *International Economic Review*, 50, 415-430.

Fujita, S.; Ramey, G. (2012) Exogenous vs. endogenous separation. *American Economic Journal: Macroeconomics*, 4, 68-93.

Galí, J.; Gertler, M. (1999) Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44, 195-222.

Galí, J.; Smets, F.; Wouters, R. (2011) Unemployment in an Estimated New Keynesian model. *NBER Macroeconomics Annual 2011*, University of Chicago Press, 329-360.

Hall, R. E. (2005) Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 98, 1692-1706.

Hobijn, B.; Sahin, A. (2009), Job-finding and separation rates in the OECD. *Economics Letters*, 104, 107-111.

Jolivet, G.; Postel-Vinay, F.; Robin, J-M. (2006) The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US. *European Economic Review*, 50, 877-907.

Kleibergen, F.; Mavroeidis, S. (2009) Weak instrument robust tests in GMM and the New Keynesian Phillips Curve. *Journal of Business Economics and Statistics*, 27, 293-311.

Kuersteiner, G.; Okui, R. (2010) Constructing optimal instruments by first-stage prediction averaging. *Econometrica*, 78, 697-718.

Krause, M.; López-Salido, D.; Lubik, T. (2008), Inflation dynamics with search frictions: A structural econometric analysis, *Journal of Monetary Economics*, 55, 892-916.

Martins, L.; Gabriel, V. (2009), New Keynesian Phillips curves and potential identification failures: A generalized empirical likelihood analysis. *Journal of Macroeconomics*, 31, 561-571.

Ravenna, F.; Walsh, C. (2008) Vacancies, unemployment, and the Phillips Curve. *European Economic Review*, 52, 1494-1521.

Ruge-Murcia, F. J. (2007) Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, 31, 2599-2636.

Shimer, R. (2005), The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95, 25-49.

Shimer, R. (2012), Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, 15, 127-148.

Staiger, D. O.; Stock, J. H.; Watson, M. W. (1997) How precise are estimates of the natural rate of unemployment? In C. Romer and D. Romer (eds), *Reducing Inflation: Motivation and Strategy*. Chicago, University of Chicago Press.

Stock, J. H.; Yogo, M. (2005) Testing for weak instruments in linear IV regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.* New York, Cambridge University Press, 80-108.

Zanetti, F. (2011) Labor market institutions and aggregate fluctuations in a search and matching model. *European Economic Review*, 55, 644-658.

Ana Afonso
University of Porto, Faculty of Economics
ana.catarina.afonso1995@gmail.com
Óscar Afonso
University of Porto, CEFUP, Faculty of Economics
oafonso@fep.up.pt
orcid.org/0000-0002-9597-9850
Ana Maria Bandeira
Instituto Politécnico do Porto, Instituto Superior de Contabilidade e Administração do Porto, CEOS.PP
bandeira@iscap.ipp.pt
orcid.org/0000-0003-2324-9119

DOI: https://doi.org/10.14195/2183-203X_54_2

Effects of In-School Marketing Actions on Consumption Behavior

Efeito das Ações de Marketing na Escola sobre o Comportamento de Consumo

Ana Afonso
Óscar Afonso
Ana Maria Bandeira

Received for publication: September 15, 2020
Revision accepted for publication: November 3, 2020

ABSTRACT
Our study aims to analyze the effect of marketing actions in schools, driven by health-food companies, on the consumption behavior of students to understand whether they (a) show higher preference for healthy food and drinks; (b) choose to take such goods home; are aware that consuming these goods is a healthy habit; and (c) realize that sales of the companies that produce such health foods and drinks improve. We wish to contribute to the literature by using an experimental exercise in three schools in Porto district, Portugal. Our sample included 307 students aged from 6 to 16 years old, with 153(154) students belonging to the experimental (control) group. The regression results confirm our hypotheses. In particular, we found that the level of education of parents did not seem to have an impact, the results across male and female students are very close and that students' age did not affect the results between groups, but did affect score values.
Keywords: Health-food firms; in-school marketing actions; children and adolescents; consumption behavior; business performance.

JEL Classification: I12; L66; L83; M3; Z21

RESUMO
Este estudo tem como objetivo analisar o efeito das ações de marketing nas escolas, conduzidas por empresas de alimentos saudáveis, sobre o comportamento de consumo dos estudantes. O objetivo é entender se essas ações levam os estudantes a preferir alimentos e bebidas saudáveis, se passam a aconselhar esses alimentos e bebidas em casa, se estão cientes de que o consumo desses bens melhora a saúde, e se têm a perceção de que as empresas de alimentos e bebidas saudáveis melhoram as vendas com essas ações. Para o efeito conduziu-se um exercício experimental em três escolas do distrito do Porto, Portugal. A amostra incluiu 307 estudantes dos 6 aos 16 anos. Os estudantes pertencentes ao grupo experimental (controle) foram 153 (154) e foi usado um questionário para obter os dados quantitativos necessários. Os resultados, obtidos por meio de técnicas estatísticas e econométricas, confirmam as hipóteses. Os resultados são independentes da escola, do nível de educação dos pais, do sexo e da idade dos estudantes.
Palavras-chave: Empresas de alimentos saudáveis; ações de marketing na escola; crianças e adolescentes; comportamento de consumo; desempenho empresarial.

1. Introduction

Nowadays, obesity in childhood and youth has come to be considered a serious public health problem, which affects the future health of the population and, thus, health costs and productivity (WHO 2003, 2016; Wang and Lobstein, 2006). The consumption of healthy (i.e., nutritious) goods and drinks[1] during childhood and youth are crucial for current health (Story et al., 2002), and long-run health (Taliaferro, 2010; Swayne and Dodds, 2011). There are several factors that affect the behavior and, thus, the food choices of children and young students (hereinafter students). One of these factors is undoubtedly the marketing activity (Story et al., 2002; Hastings et al., 2006; Nestle, 2006).

Since schools have a powerful effect on how students eat (Kubik, 2003; Taliaferro, 2010; Swayne and Dodds, 2011), this study intends to contribute to the literature on nutrition and health, marking activity and eating behaviors, marketing and youth consumers, and the power of in-SMA (e.g., Levine, 1999; Valkenburg, 2000; Gentile and Walsh, 2002; Story et al., 2002; Coon and Tucker, 2002; Hastings et al., 2006; Larson et al., 2008; Swayne and Dodds, 2011; Mintz and Currim, 2013; Roberts et al., 2013) by analyzing if, indeed, in-school marketing actions (hereinafter in-SMA), driven by health-food firms, contribute to reducing the risk of chronic disease.

In-SMA are usually understood as a case in which schools authorize, usually via a contract, firms to carry out marketing activities in-school, namely advertising. A school can then allow only a particular brand to be sold on the school and, as a result, the firm would provide some compensation to the school. To reduce the possible negative impact of marketing programs on unhealthy food, firms should contribute to an advance in healthy eating habits, which does not seem to be happening.

Our aim is to verify if the effectiveness of these in-SMA affect the behavior of the students, increasing the consumption of the respective goods/services and, by this way, if they contribute to fighting obesity in childhood and youth. The growing epidemic of childhood (and youth) overweight and obesity can compromise the future health of the population, penalize the productivity, and create relevant health care costs (WHO, 2016). Thus, nutrition during childhood and youth is crucial for current and long-run health and productivity (Swayne and Dodds, 2011; Chadwick and Burton, 2011).

Looking at the food industry, several factors affect the food choices of students and one of these factors, with significant relevance, is food marketing (Coon and Tucker, 2002; Story et al., 2002; Hastings et al., 2006; Larson et al., 2008). Thus, students have been the target of multiple food marketing actions. Marketers have been using a wide variety of channels since they are interested directly in these consumers from their expenditures and indirectly since they affect the consumption of their family and are future adult consumers (Valkenburg, 2000).

The use of schools as spaces closed and controlled for in-SMA has grown with the desire to raise sales and ensure students' loyalty as consumers. The in-SMA happen since firms take advantage of the financial fragility of schools (the lack of resources) to obtain permission

[1] Healthy goods and drinks are those that include a variety of relevant elements to improve health status or reduce the risk (non-prevention) of the disease. The functional ingredients are useful for the treatment of obesity and associated comorbidities, surpassing even the benefits of medication (Baboota et al., 2013).

for their realization (Levine, 1999). Yet, due to the epidemic of childhood overweight and obesity, the food sector is increasingly regulated, but in-SMA aimed at health-food firms will always be permitted and still encouraged, regardless of the restrictions imposed by the law.

As the in-SMA towards health-food goods and drinks firms are permitted and encouraged, it is expected that they positively affect the attitudes and, thus, the willingness to consume them. We consider that it makes sense to answer the questions: I have preference for nutritious/healthy food products (food and drink)? I believe that health relies on the consumption of nutritious/healthy food? I advise my parents to buy nutritious/healthy food products? Healthy food brands sell better quality products? Healthy food brands are more reputable and therefore more advisable? Healthy product brands sell more? Moreover, using statistic and econometric techniques, we also intend to observe if the in-SMA has indeed a strong explanatory power in the results and if sociodemographic data is important.

With children and adolescent obesity very much in focus, marketing activity on the need for a new healthier physical reality make sense. This activity can be characterized as a management task, which includes many tasks, such as doing market research, identifying the needs in the market, analyzing the market trends, planning the conception of goods and services, branding, price, promotions, and distributions to meet consumer demand (Wilkie and Moore, 2007). It is a set of entrepreneurial actions that are part of the management process that connects a business with its consumers and can be designed in some ways (Finkelstein and Peteraf, 2007).

The effectiveness of marketing in terms of revenue and profitability, for example, is a topic that brings together the interests of professionals and researchers from the business management field (Mintz and Currim, 2013; Roberts et al., 2013). Because we do not know – to the best of our knowledge – the effect of in-SMA carried out by health-food goods and drinks firms on students' consumption behaviors and then on productivity, sales and profits of the firms, and on health and productivity in general within the economy, we also contribute to close this gap in the literature. Indeed, while not responding directly to these questions, we will get a partial answer by noticing the attitude change in the involved students and bearing in mind that an in-SMA is a small task in the enormous task that is the marketing activity.

In short, we intend to contribute to the literature by precisely considering the effect of in-SMA on student's (health-food) consumption behaviors. By increasing the consumption of health food in a hasty manner, in-SMA can contribute to improving sales of the respective firms and, although there are costs involved, we anticipate that they can also contribute to increase profits (Mintz and Currim, 2013; Roberts et al., 2013; Porto et al., 2017). Moreover, healthier eating habits will contribute to improve the health of the general population, thereby lowering the costs of health expenditure and increasing productivity, which, in turn, positively affects economic growth (Kuhn and Prettner, 2016). As Naylor et al. (2009) observe, health is one of the main variables in the study of consumer behavior towards health-food: consumers present a more positive attitude and increase the intention to buy goods when they are presented as healthy and have favorable nutritional information.

Due to the operational limitations involving how to measure the effect of marketing actions and how to relate them to business performance indicators, researchers have made few advances in the scientific knowledge that could improve the effectiveness of marketing

management (Porto et al., 2017). Part of what is researched regarding this subject is related to the marketing costs, especially advertising, described in accounting information such as the firm's income statement and balance sheet (Raman et al., 2012). This line of research seems to be rewarding, leading to general conclusions on the effect of marketing costs without identifying the specific cost responsible for differences in rates of returns. These studies tend not to indicate the impact of the execution of an individual marketing action itself, but rather to emphasize the connected costs (Ward, 2013). Thus, the true sources of the firm's good and service sales, those that directly affect consumers, tend to remain unknown.

The reason for the lack of studies directly investigating marketing actions and their financial performance involves the differences between what is recorded and focused on firm customer management and what is recorded for accounting and financial purposes (Ward, 2013). Marketing activities are almost exclusively intended for managers to decide on sale and accounting costs (McDonald et al., 2014). As stated by Porto et al. (2017), the effects of implementing marketing activities to raise firm sales and increase profitability for the owners of or partners in micro-businesses remains open for scientific research. When trying to evaluate the effectiveness of in-SMA on student's (health-food) consumption behaviors, through an experimental exercise, the expected impact in sales and revenues is also perceived.

To answer the research questions, that resulted from the objective of the study, the desired contribution to the literature and the characteristics of the respondents, an experiment was carried out in three schools of the Porto district – two of the Matosinhos county and one of the Maia.

Experiments have two groups of subjects (e.g., Bailey, 2008). One group is the experimental group (half of the classes per school), which, in our case, was exposed to an in-SMA, whereas the other group is the control group (the other half classes), which was not exposed to an in-SMA. Our sample included 307 students aged from 6 to 16 years old: (i) 129 students from school 1, (ii) 57 students from school 2, and (iii) 154 from school 3. Students belonging to the experimental (control) group were 153 (154). Then, to collect data, we used a questionnaire (see Section 3.5), properly explained according to the age of the students, by applying six closed-end questions to test the hypotheses that underlie our investigation, bearing in mind the Likert-type scale of five levels (Likert, 1932).

The results allow us to emphasize some main conclusions due to the in-SMA. First, the experimental group obtained higher median scores for all the proposed questions, with significant results, confirming our hypotheses. Hence, in-SMA has affected student consumption behavior. Second, the different schools did not establish any noticeable difference in the results, with higher median scores for all the proposed questions on the experimental group. Third, results were very close between male and female students. Fourth, parent's level of education and the students' age did not seem to influence the results.

2. Hypotheses

From the literature, six hypotheses emerge to be tested. For example, in line with Ajzen (1991), the attitude towards a behavior is the degree of acceptance or rejection of the behavior; i.e., a positive (negative) attitude encourages (discourages) the purpose to execute

that behavior. Thus, as occurs for example in Urala and Lähteenmäki (2003), the attitudes are affected by biological, psychological, socio-cultural, and economic factors, which, in our view, can be influenced by in-SMA. In turn, attitudes (and beliefs) determine how information is processed, adapted and used, and consequently the behavior in the choice of foods and drinks. Hence, as emphasized by the literature (Story et al., 2002; Hastings et al., 2006; Larson et al., 2008), one powerful influencer force of students' food choices is clearly advertising and, in particular, in-SMA:

H1: *Children/students exposed to in-SMA towards nutritious food and drinks show a higher preference for nutritious food and drinks*.

Since the beginning of the 1990s, there has been growing interest in the study of attitudes and beliefs related to healthy eating; i.e., that increase the likelihood of consuming healthy food (Taliaferro, 2010; Swayne and Dodds, 2011). Some authors highlight how health knowledge, which can be transmitted through in-SMA, can have an effect on the attitude that increases the likelihood of healthy food selection (Naylor et al., 2009). Indeed, in-SMA can raise the students' awareness about how and why to eat suitably, as eating and nutrition are basic requisites to promote good health conditions (Story et al., 2009). Hence, health actions through in-SMA focused on students should be a priority for all social sectors, particularly in the school context. The school environment is the adequate space for dietary education activities focused on students since it affects students to achieve autonomy, construct personal values, beliefs, concepts and ways to know the world. Thus, the following hypothesis emerges:

H2: *Children/students exposed to in-SMA towards nutritious food and drinks are aware that consuming these goods tends to be healthier.*

The literature also evidences that the family played a fundamental role in the control of the choices, purchase, and preparation of the students' foods. However, the in-SMA are expected to lead students to influence family purchases of healthy foods since factors like lack of time and excessive dedication to work are aspects that interfere in the family's participation in the process of encouraging healthy eating (WHO, 2016; Story et al., 2002; Boutelle et al., 2006; Wang and Lobstein, 2006; Pearson and Biddle, 2011). As stated before, marketers are thus interested in students as consumers, either directly from their consumption expenditures or indirectly since they affect the consumption of their family and they are future adult consumers (Valkenburg, 2000), and the hypothesis 3 emerges:

H3: *Children/students exposed to in-SMA towards nutritious food and drinks tend to choose to bring these goods home thus affecting family habits*.

Mitchell and Olson (1981, p. 318) define attitude toward the brand as an "individual's internal evaluation of the brand." Spears and Singh (2004, p. 55) define attitude towards the brand as "a relatively enduring, unidimensional summary evaluation of the brand that presumably energizes behaviour." Empirical evidence supports the existence of a strong relationship between attitude towards the brand and purchase intention, stressing the relevance of this attribute in the consumer's preference for a brand (Haley and Case, 1979). Also significant empirical research indicates that attitude towards the brand is vital to affect brand loyalty (Oliver, 1999). Indeed, consumers tend to avoid risks by preferring familiar

brands (Oliver, 1999). Thus, by stressing some brands in carried out in-SMA it is expected that students show superior levels of attitude towards the explicit brands, as suggested by H4:

H4: *Children/students exposed to in-SMA towards brands of nutritious food or drinks show superior levels of attitude towards these brands.*

Brand reputation means how some brand is viewed and perceived by the consumers, stakeholders, and the market as a whole (Brunk, 2010). The brand reputation is the most valuable intangible asset for the organization (Peloza, 2005; Roberts and Dowling, 2002) and its importance is the result of its value as a differentiating attribute and vital tool for creating competitive advantage (Keller, 2008). Empirical evidence supports the belief that reputation is vital in consumer choice and thus in brand performance materialized in sales (Levine, 1999; Gentile and Walsh, 2002; Roberts and Dowling, 2002; Ataman et al., 2010; Jandaghi et al., 2011) and on revenue, profitability, and value in the financial market (Mizik, 2014; Siddhanta and Banerjee, 2014; Feng et al., 2015). Moreover, we expect that in-SMA help to improve brand reputation and we consider two new hypotheses:

H5: *Children/students exposed to in-SMA towards brands of nutritious food or drinks show superior levels of perceived brand reputation.*

H6: *Children/students exposed to in-SMA towards nutritious food and drinks have the perception that nutritious food and drinks firms want to improve sales.*

Finally, to answer all research questions we also intend to confirm, using econometric techniques, if the response given by the students following the in-SMA has indeed a strong explanatory power in the results, and if socio-demographic data is relevant. In this way, we try to verify if the results obtained vary according to the: group (experimental versus control); age; sex/gender; parents' educational level; the students' school.

3. **Methodology**

The adopted research methodology is now exposed. Since interpretation of results from numerous qualitative data is complex and is time consuming (e.g., Saunders et al., 2009), and there are many students, the quantitative method is more suitable to reach conclusions about the effect of in-SMA driven by health-food firms. We resorted to using a questionnaire, detailed below and presented in the Appendix A, to collect quantitative data by applying closed-end questions to test the hypotheses. Through this way we can correct any misunderstanding of some questions that are being asked, and any missing information that has not been filled in (e.g., Zikmund et al., 2010). We performed an experiment and thus data from an experimental group is compared with data from a control group. Only the former group was subject to an in-SMA. The purpose of having a control group is to rule out other factors that may affect the results of an experiment.[2]

[2] Through this procedure, we isolate the in-SMA's effects on the experiment and can help rule out other reasons of the experimental results (e.g., Bailey, 2008).

Secondary data was also collected, from the existing literature, which contributed to identify problems and formulate hypotheses, and in designing the survey questions.

3.1. The experimental design

We now use the experimental method that analyses the causal relationship between two variables, considering two homogeneous groups – the experimental group in which the explanatory or independent in-SMA variable is introduced, and the control group in which it is not introduced the independent variable –, and then comparing the two groups is possible to conclude if the independent variable is indeed explanatory (e.g., Bailey, 2008). Students were randomly divided into two groups – see Table 1 below: the experimental group (half of the classes per school); the control group (the other half classes).

Procedure. All students responded to the same questionnaire, but those belonging to the experimental group were subject to intervention before its application (i.e., to an in-SMA, which was conducted one week before). Indeed, following Podsakoff et al. (2003), to reduce the bias of the proposed methodology related to short-run memory, it was necessary to wait a period of time – and we have considered one week – between the in-SMA and the data collection phase with an objective instrument, the proposed questionnaire. To reach this stage, it was first necessary to obtain authorization from the authorities so that schools could be involved in the investigation, which led to the request for authorization form for schools, and then the consent of the parents. In this intervention were emphasized the benefits of healthy eating, were addressed some brands of healthy products, and students were asked for maximum involvement in the in-SMA. The experimental session was composed of some activities to: (i) make known the meaning and advantages of a healthy education at every meal and every day based on some specific contents and on available videos in the Youtube, suggesting that students repeat a version of the experimental activity at home with their parents; (ii) make known some healthy food brands and the videos viewed; (iii) indicate the expected trend of increased consumption of healthy food by firms that produce and / or market them, given the benefits associated with these products and taking into account actions such as the one carried out in class.

Intervention. The experimental session, authorized by the schools, was realized at the end of one normal school class and was lectured by the authors. At the beginning of class, the teacher informed the students that at the end of the class there would be an education session about healthy life, habits, and food. The teacher explained to the students that the experimental session intended to contribute to helping them to have better-eating standards, making known or emphasizing healthy food brands, and to make them as messengers of good eating practices. The experimental session was then based on the transmission of the some of the contents, as well as on the presentation of two videos on the benefits of health education. We selected videos from Youtube, which had to meet the following three criteria: the contents of the videos were as appropriate as possible to the action; the videos were as appropriate as possible to the ages and thus they are different between types of schools – the

two schools of the first cycle, schools 1 and 2, and the school of the second and third cycles, school 3; they had a duration not exceeding 10 minutes to capture the students' full attention.[3]

The procedure in data collection serves the purpose of meeting our aims; i.e., of testing the hypotheses. Thus, by subjecting only the experimental group to some precise information allows us to test specific hypothesis. Hypotheses H1, H2, and H3 are related to the impact of in-SMA on students, H4-H5 stress the effect of in-SMA on brands and so indirectly on the respective firms, and H6 highlight the direct effect of in-SMA on firms. To test if the differences in responses are motivated by the group to which the student belongs, the age, the sex and/or the parental education, we also use econometric techniques – a logit model – and the specifications:

$$Answer_i = \beta_1 Gr_i + \beta_2 A_i + \beta_3 Sex_i + \beta_4 FEL_i + \beta_5 MEL_i + \beta_6 Sc1_i + \beta_7 Sc2_i + \beta_8 Sc3_i + \varepsilon_i, \quad (1)$$

$$Answer_i = \alpha_1 Gr_i + \alpha_2 A_i + \alpha_3 Sex_i + \alpha_4 FEL_i + \alpha_5 MEL_i + \alpha_6 (Sc1_i \& Sc3_i) + \alpha_7 Sc2_i + \vartheta_i, \quad (2)$$

$$Answer_i = \delta_1 Gr_i + \delta_2 A_i + \delta_3 Sex_i + \delta_4 FEL_i + \delta_5 MEL_i + \delta_6 (Sc1_i \& Sc2_i) + \delta_7 Sc3_i + \varphi_i, \quad (3)$$

where: (i) $Answer_i$ is the dependent (or explained) variable that assumes the value 0 if the answer of the student i in the questionnaire is "Totally disagree", "Partly disagree" or "Neither agree nor disagree" and the value 1 if the answer of the student i is "Partly agree" or "Totally agree"; (ii) the explanatory or independent variable Gr_i assumes the value 0 if the student i belongs to the control group and the value 1 if she/he belongs to the experimental group; (iii) A_i corresponds to the age of the student i; (iv) Sex_i assumes the value 1 if the student i is male and 0 if the student is female; (v) FEL_i measures the father education level of student i, which can be 1st cycle (1), 2nd cycle (2), 3rd cycle (3), secondary (4), or superior (5); (vi) MEL_i measures the mother education level of student i, which can be 1st cycle (1), 2nd cycle (2), 3rd cycle (3), secondary (4), or superior (5);[4] (vii) Scj assumes the value 1 if the student i belongs to the school j=1, 2, 3 and the value 0 if not; thus, in (1) we consider each school separately, in (2) we consider the public schools ($Sc1$ and $Sc3$) on the one hand and the private school ($Sc2$) on the other,[5] and in 3 we consider primary or first cycle schools with children mostly from the 3rd stage ($Sc1$ and $Sc2$) on the one hand and post-primary (second and third cycles) school with adolescents mostly from the 4th stage ($Sc3$) on the other;[6] (viii) ε_i, ϑ_i and φ_i are the usual unexplained error term; (ix) β_i (i = 1, … , 8), α_i (i = 1, …, 7), δ_i (i = 1, … , 7) are the coefficient for which an estimate will be obtained.

[3] Thus, from the many existing and free available videos into the Youtube, the videos in the links https://www.youtube.com/watch?v=mbfwYTBtlmQ (duration 3m34s) and https://www.youtube.com/watch?v=bd4-g76Fni4 (duration 3m34s) were chosen for all experimental groups, the video in the link https://www.youtube.com/watch?v=NZgK8e1zzHQ (duration 5m13s) for students of the primary schools (i.e., of the first cycle), and the video of the link https://www.youtube.com/watch?v=AvW1sk93Qc4 (duration 6m35s) for students of the second and third cycles. These videos were also evaluated by the teachers of the classes of the respondents. All students were advised to show the videos at home for the whole family.

[4] As variables FEL and MEL are correlated, only one should be included in the estimates.

[5] ($Sc1$&$Sc3$) assumes the value 1 if the student i belongs to the schools 1 or 3 and the value 0 if not.

[6] ($Sc1$&$Sc2$) assumes the value 1 if the student i belongs to the schools 1 or 2 and the value 0 if not.

The probability that student i chooses alternative 1 is given by $Prob\ [Answer_i = 1|X] = \frac{e^{x'i\beta}}{1+e^{x'i\beta}}$, which is the reduced form of the logit model and X' is the row vector of explanatory variables.

3.2. The sample

To test our hypotheses (H1-H6), we have used a questionnaire from three urban schools, whose names are omitted for reasons of confidentiality, in the Porto district, north Portugal:

- two schools, school 1 and school 2, are located in the Matosinhos county. In these schools are taught the first cycle of studies (primary schools).[7] School 1 is a public school frequented by students belonging to the middle and middle-low social classes, while school 2 is a private one in which students belongs to the upper and middle-upper social classes;
- school 3 is located in the Maia county,[8] is public, teaches for the second and third cycles of studies and is frequented by students belonging to low and middle-low social classes.

Bearing in mind the Piaget's theory of cognitive development (Piaget and Inhelder, 1972), which considers that children move through four mental stages – i.e., 1st sensorimotor stage, birth to 2 years old; 2nd preoperational stage: ages 2 to 7 years old; 3rd concrete operational stage: ages 8 to 11 years old; 4th formal operational stage: ages 12 and up years old – depending on understanding how children obtain knowledge and on understanding the nature of intelligence, are considered, in a separated way:[9]

- children mostly from the 3rd stage – those in-schools 1 and 2 (7-10/11 years old);
- adolescents mostly from the 4th stage – those in-school 3 (>10/11 years old).

[7] The Portuguese Educational System is organized at various levels. The basic education is from 6 to 15/16 years old and is organized in three cycles. The 1st cycle includes four years (1º, 2º, 3º and 4º years and usually denominated primary), the 2nd cycle includes two years (5º and 6º years) and the 3rd includes three years (7º, 8º and 9º years). Secondary education is only one cycle of three years (10º, 11º and 12º years). Finally, higher education includes university and polytechnic education.

[8] This school signed a contract with the Ministry of Education, recognizing that it needed concerted intervention to solve four major interrelated problems: absenteeism, school abandonment, indiscipline, and school failure.

[9] The main features in the 3rd stage are: children start by thinking reasonably about concrete events, they start by understanding the concept of conservation; their thinking is more logical and organized, but is also very concrete; and they start by using inductive logic, or reasoning from specific information to a general principle. In turn, the main features in the 4th stage are: adolescents/teens/young adults start to think abstractly and reason about hypothetical problems; abstract thought emerges; they start by thinking more about moral, philosophical, ethical, social, and political issues that require theoretical and abstract reasoning; they start by using deductive logic, or reasoning from a general principle to specific information (e.g., Piaget and Inhelder, 1972).

A school consent form was sent to the direction of each school and a set of 423 consent forms were sent to the children's / teens' parents (school 1: 174; school 2: 71; school 3: 178), resulting in 307 authorizations and responses (school 1: 129; school 2: 57; school 3: 121); i.e., the total response rate was 73% (school 1: 74%; school 2: 80%; school 3: 68%).

The total sample of 307 students aged, from 6 to 16 years old, had a mean age of 10.13 (standard deviation = 2.54), being 145 (47.2%) males and 162 (52.8%) females.

The origin of the students was 129 (42.0%) from school 1, 57 (18.6%) from school 2 and 121 (39.4%) from school 3. As mentioned earlier, students were randomly divided into 2 groups, the experimental group (number, *n* = 153, 49.8%) and the control group (*n* = 154; 50.2%). Table 1 shows the sample characteristics divided into those groups.

Table 1. Sample characteristics by experimental and control groups

	Experimental group, EG (n=153)		Control group, CG (n=154)	
	n	%	*n*	%
Sex				
Male	76	49.7%	69	44.8%
Female	77	50.3%	85	55.2%
Total	153	100%	154	100%
School				
1	66	43.1%	63	40.9%
2	29	19.0%	28	18.2%
3	58	37.9%	63	40.9%
Total	153	100%	154	100%
Father education				
Preparatory/basic level (1º, 2º, 3ºcycles)	63	41.2%	75	48.7%
Secondary	54	35.3%	44	28.6%
Higher education	36	23.5%	35	22.7%
Total	153	100%	154	100%
Mother education				
Preparatory/basic level (1º, 2º, 3ºcycles)	42	27.5%	59	38.3%
Secondary	62	40.5%	44	28.6%
Higher education	49	32.0%	51	33.1%
Total	153	100%	154	100%
Age				
≤7	24	15.6%	17	11.0%
8-11	87	56.9%	88	57.2%
≥12	42	27.5%	49	31.8%
Total	153	100%	154	100%
Mean	10.03		10.24	
Standard deviation	2.43		2.64	

3.3. Ethical and legal considerations on the intervention

As obliged by the Belmont Report, conducted by the National Commission (1979) for the Protection of Human Subjects of Biomedical and Behavioral Research, this study has taken precautions over ethical principles. In terms of ethical and legal considerations, the work was carried out in line with Portuguese law. Thus, for private school the authorization forms consist in an authorization from the direction of the school, while for public schools consist in an authorization from both the Portuguese Ministry of Education and to the schools.

Furthermore, to guarantee full protection of the respondent students there was strict compliance with the ethical guidelines proposed by (Unicef, 2013; Greig et al., 2007) since respondents should not be at risk (Fowler, 2009). Parental consent is needed, and also the involved students were asked if they desired (or not) to participate given that is obviously not a mandatory task. In addition, to avoid bias in the method related to the eventual social desirability, it was explained to them that there were no right or wrong answers and that the information collected was absolutely anonymous and that, therefore, the individual results would not be distributed or explained (Podsakoff et al., 2003).

During the statistical data analysis in the next Section 4, all information that has been gathered is pooled so individuals' response remains confidential. Data will not be altered, or specifically selected to reach some specific results since it is considered unethical and bias due to using data to the advantage of the study. Finally, data that has been collected should be analyzed based on the original design. Additionally, care is taken during the analysis stage and in the presentation of data to avoid respondents being identifiable (Fowler, 2009).

On the day of the questionnaire response, prior to its distribution, following McNeal (1992) it has been taken into account the need to create a surrounding environment that guarantees positive feedback from the students, which facilitates the process of the students expressing themselves (Greig et al., 2007). Then, after the data was collected, it was said to the students that the experimental intervention was promoted by the authors. Hence, it became clear for the respondents that they were misled to the extent that they were involved in an investigation that included the marketing session.

3.4. Measures and structure of the questionnaire

To correctly evaluate what is desired and to respond safely to the hypotheses, it requires great care in the questionnaire design and delivery (e.g., Zikmund et al., 2010). It is suitable for a survey whose purpose is to measure the parameters of a particular group of people (e.g., students) or to make comparisons between groups (e.g., students subject to different interventions). The questionnaire, in appendix A, is composed by two parts: part one, as usual, containing personal details of respondents including: age, gender, and parents' education level, and part two includes six quantitative sentences, which intend to assess the considerations underlying the hypotheses.

In the responses, students are required to rate their decisions bearing in mind the Likert-type scale of five levels (Likert, 1932). This is a common tool in social surveys or questionnaires and is a unidimensional method of assessing responses: respondents indicate

the level of agreement or disagreement with a particular statement or question using an ordinal scale (Boone and Boone, 2012). The most common Likert-type scale uses a 5-point response system (e.g., Braun and Clarke, 2006; Guest et al., 2011; Boone and Boone, 2012), and we follow this scale: totally disagree (1), partly disagree (2), neither agree nor disagree/neutral (3), partly agree (4), and totally agree (5). Like in our case, the numerical code usually begins with 1 and has 1 value increments with each level (e.g., Braun and Clarke, 2006; Guest et al., 2011; Boone and Boone, 2012). Technically, a Likert-type scale is the sum of responses to several Likert items.

To evaluate the impact of the in-SMA on *students' preference for nutritious food and drinks* (H1), it is proposed the question "1. *I have a preference for nutritious / healthy food products (food and drink)*", based on "Q14. I prefer fast food over any other meal" of the "QuestionPro: Fast Food Survey Questions Template".[10] In our case, the question was formulated positively. It is also formulated in line with questions in (Krause et al., 2018, Table 2). To analyze the impact of the in-SMA on *students' awareness of the positive effects of healthy eating* (H2), it is proposed the question "2. *I believe that health relies on the consumption of nutritious/healthy food*", which is mainly based on "Q12. What are your thoughts on fast food consumption in relation to its health effects?" of the "QuestionPro: Fast Food Survey Questions Template". This question is also closely related with some questions in (Krause et al., 2018, Table 2).

To evaluate the impact of the in-SMA on *students' choice to bring healthy food home* thus affecting family habits (H3), it is proposed the question "3. *I advise my parents to buy nutritious/healthy food products*", which was based on questions in Doustmohammadian et al. (2017, Table 4): "Q37. I talk to my friends and family about healthy eating", "Q38. If I have any questions about food and nutrition issues, I'm able to get information and advice from parents, teachers, etc", "Q44. If my family were overweight and eating a high-fat diet, I would tell them to change their eating habits", "Q24_1. When I go shopping with my mother or father, I buy healthy snacks such as nuts, raisins, and dried chickpeas instead of chips, snacks, chocolate, and sweets", and "Q24_6 When I go shopping with my mother or father, I buy foods that are certified as healthy", among others.

To analyze the effect of the in-SMA on *students' superior levels of attitude towards brands of healthy goods* (H4), it is used the question "4. *Healthy food brands sell better quality products*", based on questions found in the following studies (Van Kooten et al., 2007; Kurdi, 2016); e.g., questions on "Behavioral consequences" in Kurdi (2016, pp. 344-348), which also use the Likert-type scale of five levels, aim to reveal the best quality of healthy food brands' products. To analyse the effect of the in-SMA on *students' superior levels of perceived brand reputation* (H5), it is used the question "5. *Healthy food brands are more reputable and therefore more advisable*", based on Stanaland et al. (2011), and Kurdi (2016, pp. 344-348). Finally, to evaluate the effect of the in-SMA on *students' perception that nutritious food and drinks firms want to improve sales* (H6), it is used the question "6. *Healthy product brands sell more*", based on Kurdi (2016, pp. 344-348).

[10] Available at https://www.questionpro.com/a/editSurvey.do .

3.5. Data analysis

We have conducted the statistical data analysis in the form of IBM SPSS Software 25 with the help of Microsoft Excel. In addition, the analysis of the impact of the group (experimental or control), the age, the gender, the school(s) and the parental education on the answers to each question is also performed through the use of econometric techniques suitable for binary dependent variables, and we have used the Eviews 11 software.

4. **Results and Discussion**

The experimental group obtained higher median scores in all proposed questions being these differences statistic significant ($p<.001$) according to the Mann-Whitney test in Table 2.

Table 2. Comparison of responses by group

Question	Experimental group, EG (n=153)	Control group, CG (n=154)	M-W test
Q1	5.0 (4.0-5.0)	4.0 (2.0-4.0)	$p<.001$
Q2	5.0 (5.0-5.0)	4.0 (3.0-5.0)	$p<.001$
Q3	5.0 (4.0-5.0)	4.0 (2.0-4.0)	$p<.001$
Q4	4.0 (4.0-5.0)	3.0 (2.0-4.0)	$p<.001$
Q5	5.0 (4.0-5.0)	3.0 (2.0-4.0)	$p<.001$
Q6	4.0 (3.0-5.0)	3.0 (2.0-3.0)	$p<.001$

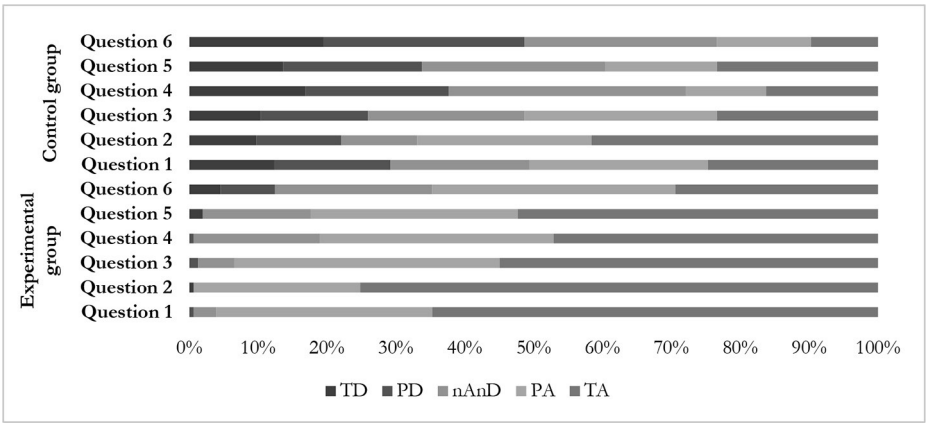
Notes: Results are presented in the format medians (percentile 25-percentile75); i.e., Mdn (P_{25}-P_{75}). M-W test is the Mann-Whitney test, also called the Wilcoxon rank sum test, which is a nonparametric test that compares two unpaired groups. If the *p*-value, *p*, is small, you can reject the null hypothesis that the difference is due to random sampling, and conclude instead that the groups are distinct; i.e., the score values of the groups, experimental and control, are statically different. If the *p*-value is large, the data do not give you any reason to reject the null hypothesis. This is not the same as saying that the two groups are the same. We just have no compelling evidence that they differ. $p<.001$ means for a significance level smaller than 0.1%.

Results confirm the formulated hypotheses H1-H6, showing that in-SMA had a positive effect on all of them. In-SMA contribute positively so that students have a preference for healthy food (H1), are aware that the consumption of these goods makes them healthier (H2), bias food and drink habits at home towards nutritious food and drinks (H3), show superior levels of attitude towards healthy food and drink brands (H4) and that these brands have superior levels of perceived reputation (H5), and have the perception that nutritious food and drinks firms will increase sales (H6). Thus, marketing actions affect the students' food choices (Story et al., 2009) and, when it comes to good eating practices, should be the school's priority to disseminate the best dietary practices (Valkenburg, 2000).

Figure 1 shows the distribution of the answers to the six questions divided into both groups, experimental and control. Results in Figure 1 visually confirm the ones presented in Table 2 with a higher prevalence of concordances obtained in the experimental group

(Mdn between 4 and 5), and discordances obtained in the control group (Mdn between 3 and 4). As stated above, the score value differences between groups are statistically significant.

Figure 1. Responses' distribution by group



Notes: TD = Total disagreement; PD = Partial disagreement; nAnD = Nor agree neither disagree; PA = Partial agreement; TA = Total agreement.

The results in Table 3 are relatively similar in the three different schools, with higher median scores for all questions in the experimental group. However, from the Mann-Whitney test: (i) for schools 1 and 3 (the public ones, attended by students of lower social classes) all questions had statistically significant differences ($p<.001$), thus approving the differences between groups, (ii) for school 2 (the private school, attended by students of higher social classes) Q4 ($p=.054$) and Q5 ($p=.514$) did not present statistically significant differences. Q4 is statistically significant for a significance level higher than 5.4% and Q5 is not statistically significant. Hence, at school 2 there was no statistical difference between groups in the answer to the Q5, probably because it is the school where students have best eating practices. Although the results are relatively similar in the three different schools, with higher scores in all proposed questions in the experimental group, in a thinner analysis, we observe slight differences in the score values of Table 3 among schools for questions 1, 3, 4, 5 and 6, which can be statistically captured by the econometric estimation of the logit specification in equation (1) and probably explained by the social group to which school students belong.

Table 3. Comparison of responses by group and school

	School 1 (n=129)			School 2 (n=57)			School 3 (n=121)		
Question	EG (n=66)	CG (n=63)	M-W	EG (n=29)	CG (n=28)	M-W	EG (n=58)	CG (n=63)	M-W
Q1	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	4.0 (4.0-5.0)	4.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	3.0 (3.0-4.0)	p<.001
Q2	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (5.0-5.0)	5.0 (4.0-5.0)	p<.003	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001
Q3	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	4.0 (4.0-5.0)	4.0 (3.0-4.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001
Q4	4.0 (4.0-5.0)	2.0 (1.0-4.0)	p<.001	5.0 (4.0-5.0)	3.5 (2.5-5.0)	p=.054	4.0 (4.0-5.0)	3.0 (2.0-3.0)	p<.001
Q5	4.5 (4.0-5.0)	2.0 (1.0-4.0)	p<.001	4.0 (3.0-5.0)	3.5 (3.0-5.0)	p=.514	5.0 (4.0-5.0)	3.0 (3.0-4.0)	p<.001
Q6	4.0 (3.0-5.0)	2.0 (2.0-4.0)	p<.001	4.0 (3.0-4.0)	3.0 (1.0-3.0)	p<.001	4.0 (3.0-5.0)	3.0 (2.0-3.0)	p<.001

Notes: See Table 2. p<0.001. p=0.054 and p=0.514 mean for a significance level smaller than 0.1%. equal to 5.5% and equal to 51.4%.

Looking at results by sex, Table 4 reveals that results were very close, with no noticeable differences between male and female students – there are higher median scores in all proposed questions in the experimental group. In particular, the Mann-Whitney test shows that results are all statistically significant for a significance level of 0.1%. Hence, gender does not introduce differences in eating habits. Moreover, Table 4 shows that the score values between males and females are relatively similar for all questions, which can also be apprehended and confirmed statistically by the econometric estimation of the logit model.

Table 4. Comparison of responses by group and sex

	Male (n=145)			Female (n=162)		
Question	EG (n=76)	CG (n=59)	M-W	EG (n=77)	CG (n=85)	M-W
Q1	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	4.0 (2.0-4.0)	p<.001
Q2	5.0 (5.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001
Q3	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	4.0 (2.0-4.0)	p<.001
Q4	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-3.0)	p<.001
Q5	5.0 (4.0-5.0)	*3.0 (2.0-4.0)*	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001
Q6	4.0 (3.0-5.0)	3.0 (2.0-4.0)	p<.001	4.0 (3.0-5.0)	2.0 (2.0-3.0)	p<.001

Notes: See Table 2. Remember that EG means Experimental group and CG means Control group.

Parent's level of education did not seem to affect the results of the in-SMA that was effective across all types of parent's education. Results are relatively similar across all types of parent's (father's or mother's) education, with higher median scores in the experimental group. Moreover, the Mann-Whitney test shows that results are all statistically significant.

Table 5. Comparison of responses by group and Parent's degree of education

	Preparatory/basic level (n=93)			Secondary education (n=104)			Higher education (n=110)		
Question	EG (n=38)	CG (n=55)	MM-W	EG (n=59)	CG (n=45)	M-W	EG (n=56)	CG (n=54)	MM-W
Q1	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	4.0 (2.0-4.0)	p<.001
Q2	5.0 (4 .0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (5.0-5.0)	4.0 (4.0-5.0)	p<.001
Q3	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	4.0 (3.0-4.0)	p<.001
Q4	4.0 (4.0-5.0)	3.0 (2.0-3.0)	p<.001	4.0 (4.0-5.0)	3.0 (1.0-3.0)	p<.001	5.0 (4.0-5.0)	3.0 (3.0-4.0)	p<.001
Q5	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.5 (3.0-5.0)	3.0 (3.0-5.0)	p=.006
Q6	4.0 (3.0-5.0)	3.0 (2.0-3.0)	p<.001	4.0 (4.0-5.0)	2.0 (2.0-3.0)	p<.001	4.0 (3.0-4.0)	3.0 (2.0-4.0)	p<.001

Notes: See Tables 2, 3, and 4. Preparatory/basic level means 1º (1-4 school years), 2º (5-6 school years), or 3º cycles (7-9 school years), Secondary education includes 10th, 11th, and 12th years.

Table 6. Comparison of responses by group and Father's level of education

	Preparatory/basic level (n=138)			Secondary education (n=98)			Higher education (n=71)		
Question	EG (n=63)	CG (n=75)	M-W	EG (n=54)	CG (n=44)	M-W	EG (n=36)	CG (n=35)	M-W
Q1	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	4.0 (2.0-4.0)	p<.001
Q2	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (5.0-5.0)	4.0 (4.0-5.0)	p<.001
Q3	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.5-4.0)	p<.001	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p=.002
Q4	4.0 (4.0-5.0)	3.0 (2.0-3.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	3.0 (3.0-5.0)	p<.001
Q5	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.5 (4.0-5.0)	3.0 (2.0-4.5)	p<.001	5.0 (3.5-5.0)	3.0 (3.0-5.0)	p=.005
Q6	4.0 (3.0-5.0)	2.0 (2.0-3.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	3.5 (3.0-4.5)	3.0 (1.0-3.0)	p<.001

Notes: See Tables 2, 3, 4, and 5.

Table 7. Comparison of responses by group and Mother's level of education

	Preparatory/basic level (n=101)			Secondary education (n=106)			Higher education (n=100)		
Question	EG (*n*=42)	CG (n=59)	M-W test	EG (n=62)	CG (n=44)	M-W test	EG (n=49)	CG (n=51)	M-W test
Q1	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (5.0-5.0)	3.0 (2.0-4.0)	p<.001	5.0 (4.0-5.0)	4.5 (2.5-5.0)	p<.001
Q2	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	3.5 (2.0-4.5)	p<.001	5.0 (5.0-5.0)	4.5 (3.0-5.0)	p<.001
Q3	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	4.0 (4.0-5.0)	2.5 (1.5-4.0)	p<.001	4.0 (3.0-5.0)	4.0 (2.5-5.0)	p<.001
Q4	4.0 (4.0-5.0)	3.0 (2.0-3.0)	p<.001	4.0 (4.0-5.0)	2.0 (1.0-3.0)	p<.001	4.0 (4.0-5.0)	3.5 (3.0-4.5)	p<.001
Q5	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.0 (4.0-5.0)	5.0 (4.0-5.0)	p=.014
Q6	4.0 (3.0-5.0)	3.0 (2.0-3.0)	p=.003	4.0 (4.0-5.0)	2.0 (1.0-3.0)	p<.001	2.0 (2.0-4.0)	3.5 (2.5-4.0)	p<.001

Notes: See Tables 2, 3, 4, and 5.

Tables 5, 6, and 7 reveal clearly that the score values between groups, experimental and control, divided by Parent's (father's or mother's) level of education are relatively similar for all questions, being the score values higher in the experimental group. The Mann-Whitney test reveals that the score values between groups are statistically different in the different levels of education: (i) basic, secondary and higher education levels, or (ii) without and with higher education, depending on the division considered. However, for example, looking at the Mother's level of education perhaps the biggest difference is in Q5, where, contrary to what is anticipated, the score values in the higher education level of the control group (Mdn = 5.0; P_{25} = 4.0; P_{75} = 5.0) exceed those of the experimental group (Mdn = 4.0; P_{25} = 4.0; P_{75} = 5.0), thus suggesting that the Mother's level of education is not a factor influencing the answer to Q5, maybe due to all Mothers, regardless of school background, are well-informed about the brands.

Tables 8 and 9 also confirm that age does not affect results. It proves to be effective for children with age smaller than 10 years old and for students with age higher (or equal) than 10 years old – see Table 8. Indeed, the results are different between groups, experimental and control, in both cases (students with age smaller than 10 years old and students with age higher (or equal) than 10 years old) and, from the Mann-Whitney test, they are statistic significant. However, from Table 9, when groups are divided according the Piaget's theory of cognitive development, the Mann-Whitney test shows that, for age ≤ 7, results are all statistically significant only for a significance level of 10% – e.g., the threshold level of significance for Q5 is 9.1%.

Table 8. Comparison of responses by group and age

	<10 (n=154)			≥10 (n=153)		
Question	EG (n=76)	CG (n=78)	M-W test	EG (n=77)	CG (n=76)	M-W test
Q1	5.0 (4.0-5.0)	4.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001
Q2	5.0 (5.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001
Q3	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001
Q4	4.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-3.5)	p<.001
Q5	4.0 (3.0-5.0)	3.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001
Q6	4.0 (3.0-5.0)	2.0 (2.0-4.0)	p<.001	4.0 (3.0-5.0)	3.0 (2.0-3.0)	p<.001

Notes: See Tables 5, 6 and 7.

Tables 8 and 9 show that the biggest differences in the responses between groups, experimental and control, bearing in mind the age is in Q1, Q3, and Q5. This observation will be confirmed statistically later on by the logit econometric estimations. Thus, by performing a finer, more disaggregated analysis, considering, for example, age differences, we can capture details that would otherwise not be perceived.

Table 9. Responses comparison by group

	Age ≤ 7 (n=41)			Age 8-11 (n=175)			Age >11 (n=91)		
Question	EG (n=24)	CG (n=17)	M-W test	EG (n=87)	CG (n=88)	M-W test	EG (n=42)	CG (n=49)	M-W test
Q1	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p=.024	5.0 (4.0-5.0)	4.0 (2.0-4.5)	p<.001	5.0 (4.0-5.0)	3.0 (3.0-4.0)	p<.001
Q2	5.0 (5.0-5.0)	5.0 (4.0-5.0)	p=.028	5.0 (5.0-5.0)	4.0 (3.0-5.0)	p<.001	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001
Q3	5.0 (4.0-5.0)	4.0 (4.0-5.0)	p=.020	5.0 (4.0-5.0)	4.0 (3.0-5.0)	p<.001	4.5 (4.0-5.0)	3.0 (2.0-4.0)	p<.001
Q4	4.0 (3.0-5.0)	2.0 (2.0-3.0)	p<.001	5.0 (4.0-5.0)	3.0 (2.0-4.0)	p<.001	4.0 (4.0-5.0)	3.0 (2.0-3.0)	p<.001
Q5	3.0 (3.0-4.0)	3.0 (2.0-4.0)	p=.091	5.0 (4.0-5.0)	3.0 (2.0-5.0)	p<.001	5.0 (4.0-5.0)	3.0 (3.0-4.0)	p<.001
Q6	4.0 (3.0-4.0)	3.0 (2.0-4.0)	p=.029	4.0 (3.0-4.0)	2.0 (1.5-3.0)	p<.001	4.0 (3.0-5.0)	3.0 (2.0-3.0)	p<.001

Notes: See Tables 5, 6 and 7.

From Table 9 the score values are higher in the experimental group – i.e., close the response "totally agree" – than in the control group – in which the "totally disagree" response prevails. Interesting and as expected, taking into account the Mann-Whitney test – see the p-values –, results are more robust for higher age levels in Piaget's theory of cognitive development. It is also worth mentioning that the answer to Q5 is the only one that is not significant (p=0.91>0.05) for children aged 7 years and under. Thus, in particular, younger children, after an in-SMA, are convinced that healthy product brands have higher quality, Q4, but this does not change their perception of brand reputation, Q5, since the results report that it is not statistically certain that they consider the brands to be most reputable and, therefore, more advisable.

Estimation results by maximum likelihood of the econometric logit models (1), (2) and (3), which allow us to analyze the effect of the explanatory variables on the probability of choosing the responses "Partly agree" or "Totally agree", are in Table 10. The eighteen estimated models – three for each of the six Questions – are significant, overall considered, but this is not true for the majority of the explanatory variables, individually considered. Each Likelihood Ratio Statistic (LRS) value is used to evaluate the overall significance of the model, and results show that all models are statically significant at 1% of significance level. Alike, from the McFadden R^2 the predictive capability of the models, measured by the percentage of successes, is around 25%.

Individually, the explanatory variable *Gr*, group, is positive and statistically significant in all models, being, as expected, the most significant and relevant variable. The positive signal tells us that assisting the in-SMA contributes positively to the responses "Partly agree" or "Totally agree". Secondly, the explanatory variable *A*, students' age, is statistically significant in Q1, Q3 and Q5. As the coefficient associated with this variable is negative in Q1 and Q3, *ceteris paribus*, *A* contributes to worsen the probability of responding "Partly agree" or "Totally agree"; in turn, in Q5, *A* improves this probability. Thirdly, the explanatory variable *Sex* does not have a significant effect on the probability, and the education level of parents

– Fathers, *FEL*, and Mothers, *MEL* – only have significance in Q3. However, *FEL* improves the probability and *MEL* punishes it.

Finally, *ceteris paribus*, variable schools (individually or not) affects positive and significantly to the probability of an affirmative answer in Q1 and Q3, whereas it contributes negatively and significantly to the probability of affirmative answer in Q4, Q5 and Q6. Given the signs and statistical significance of the impact when specification (1) is considered; i.e., when individual schools are considered, then when are taken (i) public schools (schools 1 and 3) *versus* private school (school 2) – i.e., specification (2) – and (ii) primary or first cycle schools (schools 1 and 2) *versus* post-primary (second and third cycles) school (school 3) – i.e., specification (3);– the sign and statistical significance of the impact remain, although softened, as all schools, when considered individually, contribute – in sign and significance – in the same sense. To sum up, belonging to (i) schools 1, 2 or 3 individually, (ii) schools 1 and 3 (public) *versus* school 2 (private), and (iii) schools 1 and 2 (first cycle) *versus* school 3 (second and third cycles), affects positively and significantly to the probability of an affirmative answer in Q1 and Q3, whereas it contributes negatively and significantly to the probability of affirmative answer in Q4, Q5, and Q6.

Table 10. Estimated results from the logit model

	Q1			Q2			Q3			Q4				Q5		Q6		
	Answer			*Answer*			*Answer*			*Answer*				*Answer*		*Answer*		
Variable	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Gr	3.274	3.262	3.245	4.316	4.315	4.287	2.863	2.862	2.867	2.408	2.396	2.392	2.076	2.048	2.076	1.815	1.794	1.821
	0.457*	0.452*	0.448*	1.014*	1.011*	1.025*	0.391*	0.391*	0.390*	0.283*	0.281*	0.279*	0.272*	0.268	0.272	0.259*	0.257*	0.258*
A	-0.249	-0.146	-0.249	-0.064	-0.083	-0.059	-0.244	-0.243	-0.244	0.118	0.041	0.109	0.272	0.128	0.271	0.107	-0.012	0.116
	0.117**	0.060**	0.116*	0.135	0.067	0.133	0.125***	0.067*	0.125***	0.102	0.058	0.102	0.108**	0.056**	0.108**	0.100	0.056	0.099
Sex	-0.151	-0.180	-0.160	-0.151	0.062	0.032	-0.365	-0.365	-0.362	0.066	0.084	0.061	-0.136	-0.103	-0.135	0.292	0.309	0.291
	0.311	0.309	0.310	0.311	0.351	0.347	0.309	0.306	0.309	0.279	0.276	0.279	0.269	0.266	0.268	0.257	0.254	0.256
FEL	-0.013	-0.036	0.016	-0.094	-0.090	-0.040	0.430	0.430	0.421	0.067	0.078	0.095	-0.102	-0.076	-0.101	-0.095	-0.075	-0.119
	0.178	0.182	0.177	0.203	0.201	0.200	(0.197**	0.197**	0.195**	0.160	0.157	0.154	0.155	0.150	0.153	0.171	0.166	0.168
MEL	-0.126	-0.101	-0.098	0.215	0.209	0.268	-0.415	-0.415	-0.423	0.178	0.175	0.196	0.049	0.035	0.049	0.061	0.060	0.048
	0.178	0.181	0.174	0.198	0.194)	0.196	0.190**	0.190**	0.187**	0.161	0.158	0.159	0.166	0.160	0.164	0.179	0.176	0.178
Sc1	2.815			0.807			2.784			-2.964			-2.630			-1.977		
	1.127**			1.278			1.218**			1.047*			1.055**			1.023***		
Sc2	3.212	2.355		1.706	1.880		2.662	2.651		-2.632	-2.056		-2.624	-1.521		-2.261	-1.388	
	1.255**	1.020**		1.488	1.111***		1.311**	1.018*		1.148**	0.912**		1.153**	0.892***		1.094**	0.891	
Sc3	3.466		3.290	0.678		0.288	2.792		2.841	-3.443		-3.477	-3.531		-3.532	-2.714		-2.727
	1.589**		1.594**	1.873		1.827	1.702***		1.709***	1.422**		1.434**	1.467**		1.462	1.388***		1.386**
Sc1&Sc3		2.037			0.963			2.775			-2.416			-1.581			-1.126	
		0.902**			0.923			0.942*			0.836*			0.814***			0.832	
Sc1&Sc2			2.725			0.604			2.807			-2.962			-2.629			-2.006
			1.128**			1.262			1.224**			1.054*			1.054**			1.026***
McFadden R²	0.283	0.281	0.282	0.290	0.290	0.283	0.277	0.277	0.276	0.234	0.232	0.232	0.173	0.166	0.173	0.141	0.136	0.139
LR Statistic	100.98*	99.981*	100.32*	81.037*	81.005*	78.988*	100.32*	100.32*	100.26*	99.211*	98.458*	98.583*	71.215*	68.427*	71.214*	59.326*	57.310*	58.788*
Log likelihood	-127.68	-128.18	-128.01	-99.137	-99.152	-100.16	-130.96	-130.96	-130.99	-162.00	-162.38	-162.31	-169.82	-171.21	-169.81	-180.89	-181.90	-181.16
Observations	307	307	307	307	307	307	307	307	307	307	307	307	307	307	307	307	307	307

Notes: (i) the dependent variable $Answer_i$ assumes the value 0 if the answer of the student i in the questionnaire is "Totally disagree", "Partly disagree" or "Neither agree nor disagree" and the value 1 if the answer is "Partly agree" or "Totally agree"; (ii) Gri assumes the value 0 if the student i belongs to the control group and the value 1 if she/he belongs to the experimental group; (iii) Ai corresponds to the age of the student i; (iv) Sexi assumes the value 1 if the student i is male and 0 if she is female; (v) FELi measures the father education level of student i, which can be 1st cycle (1), 2nd cycle (2), 3rd cycle (3), secondary (4), or superior (5); (vi) MELi measures the mother education level of student i, which can be 1st cycle (1), 2nd cycle (2), 3rd cycle (3), secondary (4), or superior (5); (vii) in specifications type (1), Scj assumes the value 1 if the student i belongs to the school j=1,2,3 and the value 0 if not; (viii) in specifications type (2), (Sc1&Sc3) assumes the value 1 if the student i belongs to the schools 1 or 3 and the value 0 if not; (ix) in specifications type (3), (Sc1&Sc2) assumes the value 1 if the student i belongs to the schools 1 or 3 and the value 0 if not; (x) significance levels of 1% (*), 5% (**) and 10% (***); (xi) the corresponding standard error is reported below the estimated coefficients.

5. **Concluding Remarks and Future Research**

Obesity in childhood and youth became a serious public health problem and the consumption of healthy (i.e., nutritious) goods and drinks during childhood and youth can improve the long-run health. There are several factors that affect the food choices of students, and one of these factors is marketing activities. In-SMA aimed at health-food firms will always be permitted and still encouraged, regardless of the restrictions that the law imposes now or in the future on the food sector, which is increasingly regulated. In this context, schools, as a closed and controlled space, have a powerful effect on how students eat, and we have analyzed if, indeed, in-SMA, driven by health-food agents, affect the behavior of the students, ensuring their loyalty as consumers and increasing the consumption of the respective goods/services in a sustained way. Hence, in-SMA should contribute to fight students' obesity and, thus, to rise health (and, in a near future, the productivity of all firms) and sales (and profits) of the respective firms, given that in-SMA also involve costs.

As in-SMA towards health-food goods and drinks firms are permitted and encouraged since they should positively affect the attitudes toward health-food goods and drinks and, thus, the willingness to consume them, our research questions make sense.

To answer the research questions, we have performed an experiment in three schools of the Porto district, being two of them from the Matosinhos county and one of them from the Maia county. Before the collection of the data by a questionnaire, the experimental group, which included 153 students and accounted for half of the classes per school, was exposed to an in-SMA. Our sample included 307 students aged from 6 to 16 years old: 129 from school 1, 57 from school 2, and 154 from school 3. After that, the data were treated on the basis of appropriate statistical and econometric techniques.

We have shown that school feeding plays an essential role in the students' lives because to perform any activity and ensure healthy growth, children and adolescents need to consume nutrients that contribute to the proper functioning of their body. Implementing public policies that alleviate or correct such problems is always needed and the schools are very important in this process. They accompany the student at various stages of their development, from early childhood to late adolescence and early adulthood. It, therefore, has a perfect environment to contribute to the generation of healthy eating habits that will be reflected throughout life. In addition, schools have the opportunity to meet the nutritional needs of individuals who may not have their needs met at home. Thus, the school needs to find strategies to feed

its students correctly, contributing to their growth, biopsychosocial development, learning, school performance and the formation of healthy eating habits. This can be done through food and nutrition education actions using in-SMA. These actions should be regulated by public authorities to assure that they contribute to encouraging healthy eating by suggesting balanced, healthy and safe meals that promote the health of students, avoiding sugars, fats (especially saturated) and salt.

The empirical part of our work allows us to emphasize five main conclusions. Owing to the in-SMA, the experimental group has obtained higher median scores for all the proposed questions, with significant results, confirming our hypotheses. Hence, in-SMA had effects on student consumption behavior and, thus, should be encouraged (and supported) by the authorities as a way to improve health. Secondly, the three different schools did not establish any noticeable difference on the results, with higher median scores for all the proposed questions on the experimental group; thus, there is apparently no need to adjust these actions relying on the school involved – private versus public and first cycle versus other cycles. However, there are slight differences in score values in both groups, experimental and control, among schools; e.g., in public schools, all questions had statistic significant differences, confirming the differences between groups, while in the private school, Q4 and, mainly, Q5 did not present statistically significant: differences. Moreover, belonging to (i) schools 1, 2 or 3 individually, (ii) schools 1 and 3 (public) versus school 2 (private), and (iii) schools 1 and 2 (first cycle) versus school 3 (2nd and 3rd cycles), contributes positively and significantly to the probability of an affirmative answer in Q1 and Q3, whereas it contributes negatively and significantly to the probability of affirmative answer in Q4, Q5, and Q6. Thirdly, results were very close between male and female students, so there is also no need to adjust the actions according to gender. Fourthly, the parent's level of education did not seem to strongly affect the results of the in-SMA, so that the action can be standard. Fifth, the students' age did not seem to strongly affect the results of the in-SMA between groups, although it affected some score values, which also implies that, generically, in-SMA should be independent of the students' age. The same can be said when the division attends Piaget's theory of cognitive development, although the results are statistically more robust at higher levels of development.

Therefore, the results are broadly in line with those found in the existing literature. In particular, we confirm that marketing actions affect the students' food choices, as suggested by, for example, Story et al. (2009), and, when it comes to good eating practices, they should be a priority in the school context, also as a means of disseminating the good dietary practices (e.g., Valkenburg, 2000). Furthermore, also in line with the literature which emphasizes that the attitude towards the brand is vital to affect the brand loyalty (e.g., Oliver 1999), our results reveal that in-SMA contributes to improving students' attitude towards the brands' associates to the action. In this sequence, the results also show that in-SMA can improve the brand reputation, which is a very valuable intangible asset for firms (e.g., Peloza, 2005; Roberts and Dowling, 2002) and, thus, relevant for the creation of competitive advantage, sales, revenue, and value in the financial market (Keller, 2008; Jandaghi et al., 2011; Feng et al., 2015).

Hence, it can be specifically said that our work allows us to confirm the existing results in a sample composed by three Portuguese schools, in a context that accommodates all the

six hypotheses considered together, and in a scenario of the specific in-SMA described. By detailing the aggregated data by socio-demographic variables, we have observed, in a finer analysis, results that would not be otherwise captured and, as far as we know, that are not (still) stressed by the literature; thus, we have observed: (i) results are statistically more robust for higher levels of Piaget's theory of cognitive development; (ii) the Mother's level of education is not a factor influencing the answer to the Q5, probably because all Mothers, regardless of school background, are well-informed on the brands and influence their children equally; (iii) there are slight differences in the score values among schools for Q1, Q3, Q4, Q5 and Q6, probably explained by the social group to which school students belong.

From the econometric analysis, in-SMA should namely be conducted in private schools if the concern is to improve the perception of the positive relationship between health and consumption of healthy goods, Q2, and the perception of better quality branded products associated with healthy products, Q4, and in public schools if students are intended to affect their parents' consumption, Q3. It also allows us to state that the coefficients associated with the age variable show that, at an early age, the in-SMA is more successful when it is intended to influence the preference for healthy products, Q1, and, through children, affect parents, Q3. However, if the aim is to affect the reputability of the brands, Q5, then acting on adolescents seems to be preferable. Moreover, more educated parents are more influenced by their children and the opposite happens with mothers, Q3.

This finer analysis should help the authorities to calibrate the in-SMA that, as we have seen, make perfect sense to ensure a sustainably healthy society.

The results of our study should be interpreted carefully, and only as indicating directions or trends that should be explored in the future. The need to build an in-SMA to apply only during one session, in a short period of time, the unique intervention with students in a classroom context and the difficulty in accessing parents should have created quite a few limitations. In the future, therefore, more activities should be carried out to produce changes in eating behavior and preferences.

Moreover, it would be interesting to extend the study to other schools in other districts of the country to verify the sensitivity of the results to the geographical location; e.g., comparing littoral districts with interior ones and urban versus rural areas. The small geographical area considered is not representative of the students' population and, as such, does not allow generalizations. In turn, since our study has focused on the short-run effect of an in-SMA, it would be necessary to analyze the effect of a similar initiative on the medium and long-run; it seems to us that it would be important to do a follow up a few months after the intervention to see if the changes last. Finally, another limitation that may be pointed out has to do with the fact that there are very few students in the sample who were visibly obese. It would have been interesting to analyze the pattern of responses between obese and non-obese students, considering, for example, an additional item in the questionnaire that had considered the relationship between age/height/weight. Finally, since our study is based on single items to measure the dependent variables, in future research multiple measures should be taken into account.

References

Ajzen, I. (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.

Ataman, M.B.; Van Heerde, H.J.; Mela, C.F. (2010) The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research*, 47(5), 866-882.

Baboota, R.K.; Bishnoi, M.; Ambalam, P.; Kondepudi, K.K.; Sarma, S.M.; Boparai, R.K.; Podili, K. (2013) Functional food ingredients for the management of obesity and associated co-morbidities: A review. *Journal of Functional Foods*, 5(3), 997-1012.

Bailey, R.A. (2008) *Design of Comparative Experiments*. Cambridge, University Press.

Boone, H.N.J.; Boone, D. A. (2012) Analyzing Likert Data. *Journal of Extension*, 50(2), 2TOT2.

Boutelle K.N.; Fulkerson, J.A.; Neumark-Sztainer, D.; Story, M.; French, S.A. (2006) Fast food for family meals: Relationships with parent and adolescent food intake, home food availability and weight status. *Public Health Nutrition*, 10(1), 16-23.

Braun, V.; Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 77-101.

Brunk, K.H. (2010) Reputation building: beyond our control? Inferences in consumers' ethical perception formation. *Journal of Consumer Behaviour*, 9, 275-292.

Chadwick, S.; Burton, N. (2011) The evolving sophistication of ambush marketing: A typology of strategies. *Thunderbird International Business Review*, 53(6), 709-719.

Coon, K.A.; Tucker, K.L. (2002) Television and children´s consumption patterns. A review of the literature. *Minerva Pediatric*, 54, 423-436.

Doustmohammadian, A.; Omidvar, N.; Keshavarz-Mohammadi, N.; Abdollahi, M.; Amini, M.; Eini-Zinab, H. (2017) Developing and validating a scale to measure Food and Nutrition Literacy (FNLIT) in elementary school children in Iran. *PLoS ONE*, 12(6), e0179196.

Feng, H.; Morgan, N.A.; Rego, L.L. (2015) Marketing department power and firm performance. *Journal of Marketing*, 79(5), 1-20.

Finkelstein, S.; Peteraf, M.A. (2007) Managerial activities: A missing link in managerial discretion theory. *Strategic Organization*, 5(3), 237-248.

Fowler, F.J. (2009) *Applied Social Research Methods: Survey Research Methods* (4th ed.). Thousand Oaks, CA, SAGE Publications, Inc.

Gentile, D.A.; Walsh, D.A. (2002). A normative study of family media habits. *Applied Developmental Psychology*, 23, 157-178.

Greig, A.; Taylor, J.; Mackay, T. (2007) *Doing Research with Children*. London, Sage Publications.

Guest, G.; MacQueen, K.M.; Namey, E.E. (2011) *Applied Thematic Analysis*. New York, Sage Publications.

Haley, R.I.; Case, P.B. (1979) Testing Thirteen Attitude Scales for Agreement and Brand Discrimination. *Journal of Marketing*, 43, 20-32.

Hastings, G.; McDermott, L.; Angus, K.; Stead, M.; Thomson, S. (2006) *The Extent, Nature and Effects of Food Promotion to Children: A Review of the Evidence*. Geneva, WHO, Heslin.

Jandaghi, G.; Amiri, A.N.; Amini, A.; Darani, M. (2011) Assessment and ranking the effects of marketing mix on flower and plant sale volume. *International Journal of Academic Research in Business and Social Sciences*, 1(3), 207-219.

Keller, K.L. (2008) *Strategic Brand Management: Building, Measuring, and Managing Brand Equity* (3rd edition). Upper Saddle River, NJ, Pearson Prentice Hall.

Krause, C.G.; Beer-Borst, S.; Sommerhalder, K.; Hayoz, S.; Abel, T. (2018) A short food literacy questionnaire (SFLQ) for adults: Findings from a Swiss validation study. *Appetite*, 20, 275-280.

Kubik, M. (2003) The association of the school food environment with dietary behaviors of young adolescents. *American Journal of Public Health*, 93, 1168-1173.

Kuhn, M.; Prettner, K. (2016) Growth and welfare effects of health care in knowledge based economies. *Journal of Health Economics*, 46, 100-119.

Kurdi, B.H (2016) Healthy-Food Choice and Purchasing Behaviour Analysis: An Exploratory Study of Families in the UK. Thesis submitted in fulfilment of the Requirements for the degree of Doctor of Philosophy, Durham Business School, Durham University.

Larson, N.; Neumark-Stainer, D.; Harnack, L.J.; Wall, M.M.; Story, M.; Eisenberg, M.E. (2008) Fruit and vegetable intake correlates during the transition to young adulthood. *American Journal of Preventive Medicine*, 35(1), 33-37.

Levine, J. (1999) Food industry marketing in elementary schools: implications for school health professionals. *Journal of School Health*, 69, 290-291.

Likert, R. *(1932) A technique for the measurement of attitudes. Archives of Psychology, 140, 1-55.*

McDonald, M.; Mouncey, P.; Maklan, S. (2014) *Marketing Value Metrics: A New Metrics Model to Measure Marketing Effectiveness*. London, Kogan Page Publishers.

McNeal, J. (1992) Kids as Customers: A Handbook of Marketing to Children. New York, Lexington Books.

Mintz, O.; Currim, I.S. (2013) What drives managerial use of marketing and financial metrics and does metric use affect performance of marketing-mix activities? *Journal of Marketing*, 77(2), 17-40.

Mitchell, A.A.; Olson, J.C. (1981) Are product attribute beliefs the only mediator of advertising effects on brand attitude? *Journal of Marketing Research*, 18, 318-332.

Mizik, N. (2014) Assessing the total financial performance impact of brand equity with limited time-series data. *Journal of Marketing Research*, 51(6), 691-706.

Naylor, R.W.; Droms, C.M.; K.L. Haws (2009) Eating with a purpose: Consumer response to functional food health claims in conflicting versus complementary information environments. *Journal of Public Policy & Marketing*, 28(2), 221-233.

Nestle, M. (2006) Food industry and health: mostly promises, little action. *Lancet*, 368, 564-565.

Oliver, R. (1999) Whence consumer loyalty? *Journal of Marketing*, 63, 33-44

Pearson, N.; Biddle, S.J.H. (2011) Sedentary behavior and dietary intake in children, adolescents, and adults. A systematic review. *American Journal of Preventive Medicine*, 41(2), 178-188.

Peloza, J. (2005) "Corporate Social Responsibility as Reputation Insurance" presented at the 2nd Annual Corporate Social Performance Conference, Haas School of Business, University of California, Berkeley.

Piaget, J.; Inhelder, B. (1972) *The Psychology of the Child*. New York, Basic Books.

Podsakoff, P.; MacKenzie, S.B.; Lee, J. (2003) Common method biases in behavioural research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88(5), 879-903.

Porto, R.B.; Costa, R.R.; Watanabe, E.A.M. (2017) The multilevel effect of marketing activities on sales, revenue and profitability in a micro-enterprise. *Revista Brasileira de Gestão de Negócios*, 19(65), 432-452.

Raman, K.; Mantrala, M.K.; Sridhar, S.; Tang, Y.E. (2012) Optimal resource allocation with time-varying marketing effectiveness, margins and costs. *Journal of Interactive Marketing*, 26(1), 43-52.

Roberts, J.H.; Kayande, U.; Stermersch, S. (2013) From academic research to marketing practice: Exploring the marketing science value chain. *International Journal of Research in Marketing*, *31*(2), 127-140.

Roberts, P.W.; Dowling, G. (2002) Corporate reputation and sustained superior financial performance. *Strategic Management Journal*, 23, 1077-1093.

Saunders, M.; Lewis, P.; Thornhill, A. (2009) *Research Methods for Business Students* (5th edition). Harlow, Pearson Education Limited.

Siddhanta, S.; Banerjee, N. (2014) The impact of promotional mix on profit in the B2B sector. *Marketing Intelligence & Planning*, 32(5), 600-615.

Silverman D. (1993) *Interpreting the Qualitative Data: Methods for Analyzing Talk, Text and Interaction*. London, Sage.

Silverman, D. (2000) *Doing Qualitative Research: A Practical handbook*. London, Sage.

Spears, N.; Singh, S.N. (2004) Measuring attitude toward the brand and purchase intentions. *Journal of Current Issues and Research in Advertising*, 26(2), 53-66.

Stanaland, A.; Lwin, M.; Murphy, P. (2011) Consumer perceptions of the antecedents and consequences of corporate social responsibility. *Journal of Business Ethics*, 102(1), 47-55.

Story, M.; Neumark-Sztainer, D.; French, S. (2002). Individual and environmental influences on adolescent eating behaviors. *Journal of the American Dietetic Association*, 102, S40-51.

Story, M.; Marilyn, S.; Schwartz, M.S. (2009) Schools and obesity prevention: Creating school environments and policies to promote healthy eating and physical activity. *Milbank Quarterly*, 87(1), 71-100.

Swayne, L.E.; Dodds, M. (2011) *Encyclopedia of Sports Management and Marketing*. Thousand Oaks, Calif, Sage.

Taliaferro, L.A. (2010) Relationships between youth sport participation and selected health risk behaviors from 1999 to 2007. *Journal of School Health*, 399-410.

Taylor, R.W.; Droms, C.M.; Haws, K.L. (2009) Eating with a purpose: Consumer response to functional food health claims in conflicting versus complementary information environments. *Journal of Public Policy & Marketing*, 28(2), 221-233.

UNICEF (2013) *Children Participating in Research, Monitoring and Evaluation (M&E): Ethics and Your Responsibility as a Manager*. Geneva, UNICEF.

Urala, N.; Lähteenmäki, L. (2003) Reasons behind consumers' functional food choices. *Nutrition & Food Science*, 33(4), 148-158.

Valkenburg, P.M. (2000) Media and youth consumerism. *Journal of Adolescent Health*, 27, 52-56.

Van Kooten, T.; Perssona, L.; Roos, A. (2007) Population dynamical consequences of gregariousness in a size-structured consumer–resource interaction. *Journal of Theoretical Biology*, 245, 763–774.

Wang, Y.; Lobstein, T. (2006) Worldwide trends in childhood overweight and obesity. *International Journal of Pediatric Obesity*,1(1), 11-25.

Ward, K. (2013) *Financial Aspects of Marketing*. New York, Routledge.

WHO (2003) Diet, Nutrition and the Prevention of Chronic Diseases. Report of a joint WHO/FAO Expert Consultation, Who Technical Report Series, 916.

WHO (2016) Consideration of the evidence on childhood obesity for the Commission on Ending Childhood Obesity. Available at (accessed December 2018): http://apps.who.int/iris/bitstream/10665/206549/1/9789241565332_eng.pdf?ua=1.

Wilkie, W. L.; Moore, E.S. (2007) What does the definition of marketing tell us about ourselves? *Journal of Public Policy and Marketing*, 26(2), 269-276.

Zikmund, W.G.; Babin, B.J.; Jon, C.; Griffin, M. (2010) *Business Research Methods* (8th edition). South-Western Cengage Learning.

Appendix

Questionnaire:

1. Personal data: Please fill in the missing information.

I am a boy: _____ ; I'm a girl: _____.

I'm _____ years old.

My father completed the following degree:

1st Cycle (1st, 2nd, 3rd and 4th years): ___; 2nd Cycle (5th and 6th years): ___; 3rd Cycle (7th, 8th and 9th years): ___; Secondary education (10th, 11th and 12th years): ___; Higher education: ___.

My mother completed the following degree:

1st Cycle (1st, 2nd, 3rd and 4th years): ___; 2nd Cycle (5th and 6th years): ___; 3rd Cycle (7th, 8th and 9th years): ___; Secondary education (10th, 11th and 12th years): ___; Higher education: ___ .

2. Here are 6 statements to evaluate what you think. For each one you should choose the answer that best fits what you think

	Disagree Totally	Disagree Partially	Neither agree nor disagree	Partially Agree	Totally Agree
1. I have preference for nutritious / healthy food products (food and drink).					
2. I believe that health relies on the consumption of nutritious / healthy food.					
3. I advise my parents to buy nutritious / healthy food products.					
4. Healthy food brands sell better quality products.					
5. Healthy food brands are more reputable and therefore more advisable.					
6. Healthy product brands sell more.					

António Afonso
Universidade de Lisboa, ISEG, REM/UECE
aafonso@iseg.ulisboa.pt
orcid.org/0000-0002-6926-2653
Emanuel Reis Leão
ISCTE-IUL, Department of Economics, ISCTE Business School
emanuel.leao@iscte-iul.pt
orcid.org/0000-0002-9586-7365
Dilson Tiny
Central Bank of São Tomé and Príncipe, Research Department
tinydilson@hotmail.com
orcid.org/0000-0003-4539-7067
Diptes C. P. Bhimjee
ISCTE-IUL, Department of Economics, ISCTE Business School,
DINÂMIA'CET-IUL
Diptes_Bhimjee@iscte-iul.pt
orcid.org/0000-0003-4065-5805

DOI: https://doi.org/10.14195/2183-203X_54_3

Fiscal Sustainability in the PALOP Economies

Sustentabilidade Fiscal nas Economias PALOP

António Afonso
Emanuel Reis Leão
Dilson Tiny
Diptes C. P. Bhimjee

Received for publication: May 11, 2021
Revision accepted for publication: September 12, 2021

ABSTRACT
The Global Financial Crisis has typically led to a significant widening of fiscal positions (i.e. higher budget deficits and public debt), an issue that has been addressed by many researches on fiscal sustainability. Our aim in this study is to extend the extensive knowledge about fiscal policy sustainability in advanced economies by analysing the sustainability in the African continent. In particular, we examine the sustainability of public finance in Portuguese-speaking African countries (PALOP), using stationarity tests of external public debt-to-GDP ratios and cointegration tests between public revenue and public expenditure (as a percentage of GDP). Our findings for the period 1975-2019 suggest that some of the PALOP countries have endangered the sustainability of their corresponding fiscal positions. In fact, in our sample the solvency requirement seems to have been met only by Angola. In the context of a financial crisis, stemming from the COVID-19 pandemic, such sustainability issues are bound to be further challenged.
Keywords: Debt sustainability; global financial crisis; fiscal policy; PALOP; stationarity; cointegration; time series.

JEL Classification: C22; E62; H62

RESUMO
A Crise Financeira Global conduziu a um aumento significativo das posições fiscais (isto é, défices orçamentais e dívida pública mais elevados), dando origem a inúmeras pesquisas académicas relativas à sustentabilidade fiscal. Muito embora existam bastantes estudos sobre a sustentabilidade fiscal nas economias mais avançadas, a nossa contribuição académica amplia este esforço de investigação ao continente Africano. Mais especificamente, analisamos a sustentabilidade das finanças públicas nos Países Africanos de Língua Oficial Portuguesa (PALOP's), através da aplicação de testes de estacionaridade aos rácios de dívida externa em percentagem do PIB, bem como a aplicação de testes de cointegração às receitas e despesas públicas (em percentagem do PIB) destes Países. Os nossos resultados empíricos para o período 1975-2019 sugerem que alguns PALOP apresentam uma posição fiscal comprometida do ponto de vista da sustentabilidade fiscal. Com efeito, a condição de solvência apenas é observada para o caso Angolano. Tendo em consideração o impacto económico-financeiro da presente situação pandémica, a questão da sustentabilidade fiscal irá constituir um relevante desafio futuro.
Palavras Chave: Sustentabilidade da dívida; crise financeira global; política fiscal, PALOP; estacionaridade; cointegração; series temporais.

1. Introduction

We examine the sustainability of fiscal policies in Portuguese-speaking African countries (hereafter PALOP's). This important community is composed of the following African countries: Angola, Cape Verde, Guinea-Bissau, Mozambique, and Sao Tome and Príncipe. We conduct econometric sustainability testing on macroeconomic data associated with these PALOP economies.

There are several reasons that justify the importance of this research topic. First, this is a pressing macroeconomic problem for these economies, in view of the PALOP's wide sphere of influence[1] within the African continent. In fact, given that some PALOP economies yield a significant influence over the African continent's economy as a whole (e.g., Angola), addressing this specific set of economies becomes an important research issue. Second, taking into consideration the fallout related to previous macroeconomic shocks (namely the Global Financial Crisis and the more recent global oil price collapse), there is some clear indication that fiscal policies in PALOP economies might presently be on unsustainable trajectories. This potential lack of a fiscal sustainability dimension should be further aggravated by the present COVID-19 pandemic. As of September 2020, developing economies collectively require USD 2.5 trillion to counter the impact of the latest global health crisis and associated macroeconomic fallout (Stubbs et al., 2021). Third, fiscal unsustainability is presently a clear source of concern for policy makers globally, and for African policy makers in particular. However, there has been a significant knowledge gap in the academic literature between advanced economies and African economies regarding potential fiscal policy sustainability issues over the long-term. Since the Global Financial Crisis, the potential negative link between excessive public debt and lower economic growth has been extensively examined in most advanced economies, due to its implications to the design/implementation of fiscal policies. This highly relevant policy issue gains further relevance after we take into consideration the impact brought about by the present COVID-19 global pandemics on fiscal policies. The same reasoning is applicable to African economies, but a significant knowledge gap related to fiscal policy sustainability issues still prevails.

The research question associated with this article critically analyses whether fiscal policy trajectories might be unsustainable for these PALOP economies. We employ multiple advanced unit root testing procedures to assess fiscal sustainability and we use publicly available data to critically examine this policy issue for the 1975-2019 period. Therefore, this empirical article contributes to the literature by filling an important gap associated with fiscal sustainability issues in African economies, more specifically the PALOP economies. To the best of our knowledge, this is the first research to address this relevant policy issue in the PALOP economies.

Our key findings suggest that fiscal policies in these African economies are presently unsustainable, and that the capacity to fiscally accommodate future economic or financial shocks might be seriously compromised. We are in fact facing a sovereign shock of global magnitude due to the COVID-19 pandemics, as the fiscal impact of this shock might be quite

[1] For example, Angola has been a top supplier to international oil markets, and fiscal sustainability issues might have an indirect impact on the long-term performance of globally important energy markets.

difficult to deal with, given existing fiscal weaknesses. Ultimately, this fiscal unsustainability issue might have some spillover impact to some segments of the financial markets (e.g., this might be the case of Angola's influence on international oil markets).

The present article is structured as follows: section 2 addresses the main literature review; section 3 addresses the methodology and data used in this research; section 4 presents and discusses the main empirical findings, while section 5 concludes.

2. Literature Review

Historically, advanced European economies have been severely impacted by sovereign default episodes, as multiple European nations identifiable to present-day Euro Area Member States have been in sovereign distress (i.e., they either defaulted or heavily restructured their public debt profiles) (Reinhart and Rogoff, 2009). More recently, the Euro Area Sovereign Debt Crisis has led to extensive research trying to answer whether the Eurozone's fiscal policies are sustainable (or not). For example, (Panizza and Presbitero, 2013; Reinhart et al., 2012) survey the academic literature associated with this fundamental policy topic, most notably addressing the case of advanced economies. Some seminal contributions highlight the role of sovereign debt thresholds, which essentially reference a point beyond which fiscal policy instruments (such as public debt) become unsustainable and lead to slower economic growth. The emergence of a post-crisis literature addressing the sustainability of fiscal policies globally, as well as the need to strengthen fiscal discipline in fiscally overburdened economies has been the main concern of policy makers globally. Moreover, the Euro Area Sovereign Debt Crisis has revealed how important this global research topic has become following the Global Financial Crisis (GFC). For example, Afonso and Jalles (2012) observe that the sustainability of public finances might be threatened in a wide sample of 18 OECD countries, most especially after the GFC; while Afonso and Jalles (2013) suggest that the implementation of stricter fiscal policy rules should contribute to strengthen economic growth in economies where fiscal policies have been overused in order to counter the more recent financial shocks.

In the specific case of African countries, there are some previous researches connected with either the African continent as a whole or with specific parts of the continent (e.g., Sub-Saharan Africa). Elbadawi et al. (1997) examine the link between high debt and economic growth for multiple developing countries (including Sub-Saharan countries) between 1960 and 1994. The authors conclude that there is a negative nonlinear link between the variables, especially once the 97% threshold (of the debt-to GDP ratio) is surpassed. Morever, Fosu (1999) further addresses the impact of high external debt on the economic growth trajectories of Sub-Saharan economies in the 1980's, a period marked by many sovereign debt defaults and depressed economic growth. The authors also find sufficient empirical evidence in favour of the 'direct effect of debt hypothesis' (DEDH), according to which external debt becomes a burden to economic growth.

More recent research, such as Ndoricimpa (2017), addresses the following important issues in relation to the specific link between public debt and economic growth in African economies: (i) the heterogeneity in the public debt – economic output nexus throughout the African continent; (ii) the presence of nonlinearities which are quite influential in the

determination of public debt thresholds; and (iii) the sensitivity of the estimation of thresholds to the choice of empirical method. In addition, Ncube and Brixiová (2015) observe that there are some upward risks associated with the recent increase in public debt levels, especially in the context of rising interest rates; prior to the pandemic, there was nevertheless a relative stability of the fiscal sustainability outlooks for some African economies. On the other hand, Lopes da Veiga, et al. (2016) observe that: (i) public debt and economic growth referring to Sub-Saharan economies in the 1950-2012 period are negatively related; and (ii) economic growth rates are optimal in these economies when the corresponding public debt-to-GDP ratios are in the 30%-60% interval, and this interval might also be a sound reference point for the PALOP sample, notwithstanding the 97% threshold earlier proposed by Elbadawi et al. (1997).[2]

However, and to the best of knowledge, the literature on African public debt sustainability issues is not as extensive as in the case of Eurozone countries[3] (where researchers can benefit from the existence of historical datasets). Therefore, the main motivation associated with the present research is to fill this literature gap and contribute to the critical review of fiscal sustainability issues in these young PALOP democracies, namely where public debt management is concerned. Furthermore, fiscal sustainability discipline in African economies in a post-GFC environment is thus an important source of policy concern. The present COVID-19 pandemic should reinforce the need for further research on fiscal sustainability issues, especially considering the fiscal fallout of the global health crisis on the African continent, as public resources are directed towards reinforcing PALOP national health systems in the face of this global health threat. Nevertheless, this important research topic associated with post-GFC fiscal sustainability has been largely overlooked by the academic literature. To the best of our knowledge, this is the first such attempt to address this relevant issue using a sample of the PALOP sub-set of African economies.

Our paper thus addresses a relevant research issue, as (i) tighter global financial conditions and (ii) the sustained decrease in commodity prices have drawn significant attention to the need to correct recent fiscal policy trajectories in African economies. However, creating conditions for solid economic growth in these economies might be challenging in the near future, because of the pressure to rightfully capitalise African national health systems due to the global pandemic, and the need to adequately fund these systems, adding further pressure to overused fiscal policy instruments in African economies.

The present article focuses on fiscal sustainability issues in PALOP economies, which collectively have approximately 60 million citizens. It should be observed that data limitations impact our findings. This is because the sustainability of the intertemporal budget constraint is difficult to assess in stochastic environments where the level of economic uncertainty is

[2] On the potential downside effects of government debt, Afonso and Ibraimo (2020) report that debt service components in Mozambique, in the period 2000Q1-2016Q4, depressed real output, increased the general price level and accounted for the depreciation on the domestic currency.

[3] A potential explanation for this might be related to the fact that multiple modern-day African nations have gained their sovereign independence in the second half of the 20th century, and previous historical data is not necessarily available in connection to these autonomous States. This lack of historical data naturally constrains the estimation of long-run sustainability researches. In effect, this is the case with the countries included in our sample, as the PALOP countries are effectively young democracies, having gained their corresponding independent Nation statuses mostly in the 1970's onwards. This means that these countries' autonomous fiscal policies and corresponding data did not exist before their independence date.

significant (Bohn, 1998); this is further aggravated because emerging market economies typically exhibit a more unstable access to capital markets during painful economic adjustments (i.e., during economic/financial crisis periods). This further incentivises the reversion of high public debt trajectories within sustainable limits (Mendoza and Ostry, 2008), a lesson that should also be applicable to PALOP economies.

Lastly, it should also be important to assess the specific type of fiscal sustainability issues faced by these economies. In order to provide a more accurate answer as to the specific source of fiscal stress, Afonso and Rault (2009) implement a panel data empirical approach for European Union countries, in order to assess the specific type of fiscal sustainability faced by the economies in their sample. Our research approach will also benefit from the research design implemented by these authors. That is, the empirical method used here closely follows the methodological design first proposed by Afonso and Rault (2009), which carefully addresses this important sustainability question.

3. Methodology and Data

The framework for critically analysing the sustainability of public finances in PALOP economies is based on the assessment addressing the state of each country's intertemporal budget constraints. Following Afonso (2005), complying with the intertemporal budget constraint requires that the present value of future government surpluses might be sufficient to pay the existing stock of public debt. This can be represented as follows:[4]

$$D_{t-1} = \Sigma_{s=0}^{\infty} \frac{1}{(1+i)^{s+1}} (R_{t+s} - E_{t+s}) + \lim_{s \to \infty} \frac{D_{t+s}}{(1+i)^{s+1}} \quad (1)$$

where D is public debt; R represents public revenues; E are public expenditures and i is the interest rate, and t and s are time parameters.

We implement our analysis via a two-step approach in order to assess fiscal policy sustainability in PALOP economies. The first step entails the application of stationarity tests to our sample set of PALOP economies. This is done for the variable representing external public debt as a percentage of GDP, a procedure suggested by Trehan and Walsh (1991). Specifically, the article uses both the Augmented Dick Fuller (ADF) and Phillips and Perron (1988) (PP) unit root testing procedures across the sample. In order to verify the potential existence of structural breaks in the time series, the Zivot & Andrews (1992) and Perron (1989) testing procedure are also applied.

The second step involves the application of cointegration tests (individual/country time-series and panel data) between public expenditures and public revenues (both variables are expressed as a percentage of GDP). These tests are implemented according to the Johansen and Engle-Granger (1991) procedures. Before applying the proposed tests, the integration orders of each series have to be determined, so that the second step is only applied when both series are non-stationary (e.g., integrated of order one I(1), or higher). Accordingly, the following cointegration regression is estimated:

[4] The use of the constant interest rate is an implicit assumption, which in itself is not crucial for the empirical analysis (this can be confirmed in the existing literature).

$$R_t = c + bE_t + u_t \quad (2)$$

where R and E represent public revenues and public expenditures respectively.

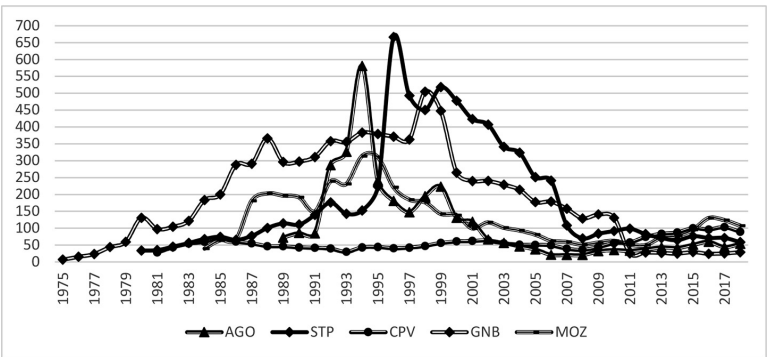
The data set that we use comes from three sources: i) the International Monetary Fund's *World Economic Outlook*; ii) the World Bank's *World Development Indicator*; and iii) Sao Tome and Príncipe's Ministry of Finance. The time span of the data covers the 1975-2019 period, and country-specific testing depends on available data for each country comprised in our PALOP sample.

4. **Empirical Findings**

As we take a first visual inspection of the debt-to-GDP ratio (Figure 1), it is easy to see that such series are hardly stationary.

Therefore, in the first step, we initially applied stationarity tests to external public debt as a percentage of GDP. We used the Augmented Dick Fuller to perform unit root testing and the Zivot and Andrews (1992) tests in order to apply unit root testing in the presence of structural breaks in the corresponding series.

Figure 1: Debt-to-GDP (GNI) ratios



Source: World Bank, IMF.

According to Table 1, our findings indicate that the series are nonstationary. This suggests that the solvency condition associated with the sustainability of public finances is not adequately met. Our sample's time series related to external public debt-to-GDP ratios are I(1). This means that we can safely reject the null hypothesis of a unit root present in the majority of the countries in our sample, with the exception of Angola (which is stationary). Similar results have been obtained using the stationarity testing procedures suggested by Zivot and Andrew (1992), and Perron's testing procedures.[5]

[5] With the exception of Guinea Bissau (Perron test).

Table 1: Unit root tests for external public debt as a percentage of GDP

			ADF						PP			Zivot and Andrews			Perron			
Country	Period	Lag	t-statistic	Test critical values: 1%	Test critical values; 5%	Test critical values; 10%	p-Value	stationarity	(Z) t-statistic	p-Value	stationarity	Break Date	ADF Break Point Test	stationarity	Lag	Break Date	ADF Break Point Test	stationarity
ANG	1989-2018	7	-11.77	-2.6648	-1.9556	-1.6087	0	yes	-1.4	0.146	no	2000	-6.101	yes	2	1994	-5.22	yes
CV	1981-2018	9	0.842	-2.6289	-1.9501	-1.6113	0.889	no	0.506	0.820	no	2007	-4.318	no	9	2010	-3.82	no
GB	1975-2018	9	-0.674	-2.6198	-1.9486	-1.6120	0.621	no	-0.695	0.409	no	2000	-3.135	no	4	2010	-8.69	yes
MZ	1984-2018	8	-0.56	-2.6347	-1.9510	-1.6109	0.652	no	-0.586	0.456	no	1996	-3.551	no	4	1998	-3.36	no
STP	1980-2018	9	-1.051	-2.6272	-1.9498	-1.6114	0.260	no	-0.975	0.289	no	1995	-3.176	no	4	1994	-3.31	no

Notes: ANG=Angola; CV=Cape Verde; GB=Guinea Bissau; MZ=Mozambique; STP=Sao Tome and Príncipe. Following the procedure adopted by Afonso (2005) and Afonso and Jalles (2012) for the Zivot & Andrews and Perron tests, "no" denotes we do not reject the null hypothesis of a unit root using t-statistics and critical values.

Moreover, we follow the approach suggested by Hakkio and Rush (1991). These authors apply cointegration testing between public revenues and expenditures (as a percentage of GDP). Accordingly, the Engle Granger and Johansen tests are applied in order to determine whether there is a long-term relationship between the two variables. Prior to applying cointegration testing, we initially assess the stationarity of the time series, using the ADF and PP tests. Accordingly, in Table 2 most country series are nonstationary in levels.[6] These findings are also confirmed by performing stationarity tests that test for the potential existence of structural breaks (Zivot and Andrews, 1992; Perron,1989) in the time series.[7]

[6] With the exception of public revenue in Guinea–Bissau, and Sao Tome and Principe for the ADF and PP test, and Public Expenditure for Guinea Bissau for the PP test.

[7] With the exception of public revenue of Sao Tome and Principe (Perron test).

Table 2: Unit root tests for public revenue and public expenditures (as a percentage of GDP)

				ADF			PP			Zivot and Andrews			Perron			
Country	Dep. variable	*Lag*	Period	*t- statistic*	*p-value*	*stationarity*	*(Z) t-statistic*	*p-Value*	*stationarity*	*Break date*	*Break point test*	*stationarity*	Lag	*Break date*	*Break point test*	*stationarity*
ANG	R	5	1997-2019	-1.361	0.58	no	-1.361	0.58	no	2005	-4.032	no	5	2014	-4.22	no
	E	5	1997-2019	-1.472	0.53	no	-1.472	0.53	no	2008	-4.107	no	5	2007	-4.043	no
CV	R	4	1995-2019	(*)-3.67	0.012	(*)no	(*)-2.986	0.05	(*)no	2011	-4.361	no	4	2003	-4.68	no
	E	4	1995-2019	(**)-3.24	0.029	(**)no	(**)-3.36	0.02	(**)no	2001	-4.384	no	4	2003	-4.262	no
GB	R	6	1993-2019	-4.842	0	yes	-3.773	0.008	yes	1997	-6.286	no	6	1998	-4.227	no
	E	6	1993-2019	(**)-3.18	0.033	(**)no	-4.463	0.0015	yes	2011	(*)-4.65	(*)no	6	2004	-4.777	no
MZ	R	9	1981-2019	-1.612	0.467	no	-1.612	0.467	no	1995	-3.85	no	9	2005	-3.814	no
	E	9	1981-2019	-2.209	0.206	no	-2.277	0.184	no	2009	(*)-4.86	(*)no	9	2008	-3.078	no
STP	R	3	1990-2019	-4.934	0.005	yes	-4.934	0.005	yes	2009	-6.135	no	3	2008	-11.443	yes
	E	7	1990-2019	-1.417	0.56	no	-1.67	0.431	no	2001	-3.79	no	7	2000	-3.74	no

Notes: R and E represent public revenue and public expenditure respectively. * and ** denote that we do not reject the null hypothesis of a unit root, at the 5% and 1% level, respectively. Following the procedure adopted by Afonso (2005) and Afonso and Jalles (2012) for the Zivot and Andrews and Perron tests, "no" denotes we do not reject the null hypothesis of a unit root using t-statistics and critical values.

The existence of breaks might be explained by two types of determinants: i) external (e.g. debt reliefs, which occurred in Guinea-Bissau, Mozambique, and Sao Tome and Principe around certain corresponding break dates); and ii) internal (e.g., strong inflationary pressures that eroded the market value of sovereign debt and public debt-to-GDP ratios, as was the case with Sao Tome and Principe in 1995). These country-specific determinants justify further research on a country-by-country basis.

To confirm the order of integration of the public revenues- and expenditures-to-GDP ratios, we conduct stationarity testing in first differences. The test results presented in Table 3 suggest that the series associated with public revenues and expenditures are stationary in first differences. This further suggests that the original series are I(1) in levels, given that the differentiated series are I(0) in first differences.

Table 3: Unit root tests for public revenues and public expenditures as a percentage of GDP (first differences)

				ADF			PP		
Country	Dependent variable	*Lag*	Period	*t-statistic*	*p-value*	*stationarity*	*(Z) t-statistic*	*p-value*	*stationarity*
ANG	R	5	1997-2019	-4.869	0	yes	-4.869	0	yes
	E	5	1997-2019	-5.673	0	yes	-5.673	0	yes
CV	R	5	1995-2019	-6.447	0	yes	-6.447	0	yes
	E	5	1995-2019	-7.237	0	yes	-7.237	0	yes
GB	R	6	1993-2019	-7.047	0	yes	-5.02	0	yes
	E	6	1993-2019	-7.314	0	yes	-6.27	0	yes
MZ	R	9	1981-2019	-6.36	0	yes	-6.36	0	yes
	E	9	1981-2019	-5.97	0	yes	-5.97	0	yes
STP	R	7	1990-2019	-9.557	0	yes	-10.545	0	yes
	E	7	1990-2019	-6.76	0	yes	-5.76	0	yes

Therefore, the Engle-Granger and the Johansen tests were only applied to Angola, Cape Verde, Mozambique and Sao Tome and Principe, indicating the same order of integration (1). Our findings in Table 4 thus suggest that in Cape Verde, Mozambique, and Sao Tome and Príncipe, the public revenues and public expenditures time series as a percentage of GDP are not cointegrated, as opposed to the case of Angola.

Table 4: Cointegration test of public revenue and public expenditure as a percentage of GDP

			Engle-Granger		Johansen	
Country	Dependent Variable	Period	*p-value*	*Cointegration*	*p-value*	*Cointegration*
ANG	R	1996-2019	0.0000	yes	0.0005	yes
	E	1996-2019	0.0000			
CV	R	1994-2019	0.1623	no	0.003	yes
	E	1994-2019	0.1718			
GB	R	1991-2019				
	E	1991-2019				
MZ	R	1980-2019	0.0760	no	0.0630	no
	E	1980-2019	0.0302			
STP	R	1989-2019	0.0006	no	0.5147	no
	E	1989-2019	0.8031			

The reduced number of observations associated with the revenues- and expenditures-to-GDP ratios for some of the countries included in our sample may limit the explanatory power of the individual cointegration tests (as per the empirical analysis presented). A more robust examination of the PALOP's fiscal sustainability assessment is conducted in this subsection. This is done through the implementation of cointegrating tests associated with a panel data framework.

In effect, Tables 5 and 6 present the findings associated with the panel cointegrating tests involving both the revenues- and expenditures-to-GDP panel data. In Table 5, the findings suggest the rejection of the null hypothesis of no cointegration for the whole of the PALOP dataset. In this case, the estimated coefficients associated with the panel dataset are statistically significant (at a 5% level).

Table 5: Pedroni Residual Cointegration Test (public expenditure and public revenue as a percentage of GDP 1980 -2019)

		Statistic	Prob.	Weighted Statistic	Prob.
Alternative hypothesis: common AR coefficients (within-dimension)	Panel v-Statistic	0.007020	0.4972	1.008817	0.1565
	Panel rho-Statistic	-9.939263	0.0000	-6.840556	0.0000
	Panel PP-Statistic	-6.286731	0.0000	-5.144135	0.0000
	Panel ADF-Statistic	-3.498435	0.0002	-4.171382	0.0000
		Statistic	Prob.		
Alternative hypothesis: individual AR coefficients. (between-dimension)	Group rho-Statistic	-3.850772	0.0001		
	Group PP-Statistic	-5.838433	0.0000		
	Group ADF-Statistic	-5.358273	0.0000		

However, when the analysis is conducted at a country-level, as in Table 6, these findings suggest that the solvency condition is met by 3 of the 5 countries of our sample (Angola, Cape Verde, and Guinea-Bissau). In this case, the *p* value (less than 0.05) clearly demonstrates that revenues and expenditures as a percent of GDP are clearly cointegrated. These findings confirm the previously-mentioned country-individual results (with the exception of Guinea-Bissau).

Although the PALOP economies are geographically dispersed and heterogeneous, our findings should support subsequent academic research addressing deeper economic and financial integration within the PALOP community. This relevance should be emphasized, as this community represents a set of countries that possess significant influence within the African continent, which is likely to grow in the forthcoming years.

Table 6: Johansen Fisher Panel Cointegration Test (Public Expenditure and Public Revenue as a percentage of GDP 1980 -2019)

	Unrestricted Cointegration Rank Test (Trace and Maximum Eigenvalue)				
	Hypothesized	Fisher statistic*		Fisher statistic**	
	No. of CE(s)	(from trace test)	Prob.	(from max-eigen test)	Prob.
	None	63.89	0.0000	44.94	0.0000
	At most 1	42.05	0.0000	42.05	0.0000
	Individual cross section results				
		Trace Test		Max-Eign Test	
	Cross Section	Statistics	Prob.**	Statistics	Prob.**
Hypothesis of no cointegration	ang	23.1228	0.0029	22.3015	0.0022
	cv	33.3824	0.0000	26.1449	0.0004
	gb	44.1225	0.0000	24.5079	0.0009
	mz	10.1969	0.2659	6.7411	0.5202
	stp	8.2948	0.4344	7.8702	0.3920
Hypothesis of at most 1 cointegration relationship	ang	0.8213	0.3648	0.8213	0.3648
	cv	7.2376	0.0071	7.2376	0.0071
	gb	19.6146	0.0000	19.6146	0.0000
	mz	3.4558	0.0630	3.4558	0.0630
	stp	0.4245	0.5147	0.4245	0.5147

Notes: * Probabilities are computed using asymptotic Chi-square distribution; **MacKinnon-Haug-Michelis (1999) p-values.

5. **Conclusion**

The present article addresses the critical issue of the sustainability of public finances in PALOP economies. This is quite an important research question associated with the sustainability of the fiscal policies of this important set of African economies, given these countries influence in Africa. The article applies econometric testing procedures to the time series associated with the major public finance variables related to this specific sub-set of African economies (the PALOP's). The applied empirical methodologies range from individual country-by-country time series econometrics analysis, to more advanced panel data testing, ensuring the completeness and robustness of the analyses. Our research design provides a more holistic perspective on the issue of fiscal sustainability in the PALOP economies.

Despite the existence of some data limitations, the individual country findings (both stationarity testing and cointegration testing) suggest that the sustainability of public finances in PALOP economies might present some sustainability issues. That is, some PALOP countries public debt trajectories are relatively unsustainable. The solvency condition of the corresponding economies' intertemporal budget constraints are not adequately respected in most

of the PALOP countries. Angola is the only PALOP economy that passes both stationarity and cointegration testing procedures, most likely due to the Angolan economy's structural role as a major oil exporter to global energy markets.

Lastly, panel data for the PALOP set fully confirms the individual country results, although there is a small discrepancy associated with Guinea-Bissau.

Therefore, the present article's overall findings suggest that further research is needed on this fundamental research topic for the African continent, as some PALOP economies (e.g., Angola) play an important role in international oil markets. These markets might be impacted by any serious fiscal unsustainability issues. As the COVID-19 pandemic impacts the African continent, there is a greater likelihood of more significant and expressive fiscal and public debt unsustainability issues, as greater financial resources will be inevitably needed to counter the global pandemic health crisis, thus adding significant pressure on the sustainability of fiscal policies associated with these African economies.

Finally, given the impact of previous financial crises, the expressive downward dynamics in the price of oil, and the required capitalizations of these countries' national health systems in order to address the COVID-19 pandemics, one would need more sustainable fiscal policy management practices, as well as the design of future corrective fiscal policy measures in the PALOP countries, with the goal of ultimately ensuring the long-term sustainability of their public finances.

References

Afonso, A. (2005) Fiscal sustainability: The unpleasant European case. *FinanzArchiv*, 61(1), 19-44.

Afonso, A.; Ibraimo, Y. (2020) The macroeconomic effects of public debt: An empirical analysis of Mozambique. *Applied Economics*, 52(2), 2012-226.

Afonso, A.; Jalles, J. (2012) Revisiting fiscal sustainability panel cointegration and structural breaks in OECD countries. ECB WP 1465.

Afonso, A.; Jalles, J. (2013) Do fiscal rules matter for growth? *Applied Economics Letters*, 20, 34-40.

Afonso, A.; Rault, C. (2009) Spend -and and-tax: A panel data investigation for the EU. *Economics Bulletin*, 29(4), A32.

Bohn, H. (1998) The behavior of U. S. public debt and deficits. *Quarterly Journal of Economics*, 113(3), 949-963.

Elbadawi, I.; Ndulu, B.J.; Ndung'u, N. (1997) Debt overhang and economic growth in sub-Saharan Africa. *External Finance for Low-Income Countries*, 49-76.

Engle, R.; C. Granger (1991) Cointegration and error correction: Representation, estimation, and testing. In Engle and Granger (Eds.), *Long Run Economic Readings in Cointegration*, Oxford University Press, New York, pp. 81-113.

Fosu, A.K. (1999) The external debt burden and economic growth in the 1980s: Evidence from sub-Saharan Africa. *Canadian Journal of Development Studies*, 20(2), 307-318.

Hakkio, G.; Rush, M. (1991) Is the budget deficit "Too large"? *Economic Inquiry*, 29(3), 429-445.

Lopes da Veiga, J.; Ferreira-Lopes, A.; Sequeira, T.N. (2016) Public debt, economic growth and inflation in African economies. *South African Journal of Economics*, 84(2), 294-322.

Mendoza, E.G., and Ostry, J.D. (2008) International evidence on fiscal solvency: Is fiscal policy "responsible"? *Journal of Monetary Economics*, 55, 1081-1093.

Ndoricimpa, A. (2017) Threshold effects of debt on economic growth in Africa. *African Development Review*, 29 (3), 471-484.

Ncube, M.; Brixiová, Z. (2015) Public debt sustainability in Africa: Building resilience and challenges ahead. *Development Policy Review*, 33(5), 555-580.

Panizza, U.; Presbitero, A.F. (2013) Public debt and economic growth in advanced economies: A survey. *Swiss Journal of Economics and Statistics* 149, 175-204.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361-1401.

Phillips, P.; Perron, P. (1988) Testing for a unit root in time series regression. *Biometrika*, 75(2), 335-346.

Reinhart, C.M.; Reinhart, V.R.; Rogoff, K.S. (2012) Public debt overhangs: Advanced-economy episodes since 1800. *Journal of Economic Perspectives*, 26(3), 69-86.

Reinhart, C.M.; Rogoff, K.S. (2009) *This Time is Different: Eight Centuries of Financial Folly*. Princeton, Princeton University Press.

Stubbs, T.; Kring, W.; Laskaridis, C.; Kentikelenis, A.; Gallagher, K. (2021) Whatever it takes? The global Financial safety net, Covid-19, and developing countries. *World Development*, 137, 1-8.

Trehan, B.; Walsh, C. (1991) Testing intertemporal budget constraints: Theory and applications to U. S. Federal budget and current account deficits. *Journal of Money, Credit, and Banking*, 23(2), 206-223.

Zivot, E.; Andrews, D. (1992) Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, 10(3), 251-270.

Aleksandar Vasilev
University of Lincoln, Lincoln International Business School,
Accountancy, Finance and Economics Department
avasilev@lincoln.ac.uk
orcid.org/0000-0002-3956-6314

DOI: https://doi.org/10.14195/2183-203X_54_4

A Real-Business-Cycle Model with Endogenous Discounting and a Government Sector

Um Modelo de Ciclos Económicos Reais com Taxa de Desconto Endógena e Estado

Aleksandar Vasilev

Received for publication: October 30, 2020
Revision accepted for publication: April 21, 2021

ABSTRACT
We introduce an endogenous discount factor as in Uzawa (1968) and Schmitt-Grohe and Uribe (2003) into a real-business-cycle setup with government sector and Greenwood et al. (1988) preferences. We calibrate the artificial economy to Bulgarian data for the period after the currency board arrangement was introduced (1999-2018). In particular, we look into the quantitative importance of endogenous discounting for the propagation of cyclical fluctuations in Bulgaria. We conclude that the presence of an endogenous discount factor improves the model, and that the extended setup performs better than the standard RBC model framework with a constant discount factor (e.g. Vasilev, 2009).
Keywords: Business cycles; Uzawa preferences; endogenous discounting; Bulgaria.

JEL Classification*:* E24; E32.

Acknowledgement: The author would like to thank an anonymous referee and Prof. Teixeira for their comments and suggestions, although the usual disclaimer applies.

1. Introduction and Motivation

In the wide class of dynamic stochastic general-equilibrium models (also known as "real-business-cycle" models), a representative one-member household is faced with the dynamic choice of consumption, investment and hours worked sequences, which are optimal with respect to the individual's utility function, the budget constraint, and the rate of time preference, or the discount rate. More specifically, the discount factor, i.e., the inverse of the discount rate plus unity, in those models is assumed to be constant across time periods, and thus independent of the time profile of the utility stream associated with the set of optimal decision sequences. One shortcoming of the class of models with a constant discount factor, is that when calibrated to Bulgarian data, model predicted consumption volatility is too high relative to that observed in data, even when the consumption variable includes consumption of durables as well.

In order to address that shortcoming of these models, we utilize an idea, first proposed in Uzawa (1968), who analyzed the time preference structure of a representative household through the lens of neoclassical choice theory and derived a formulation specifying the rate by which an individual discounts future levels. Alternative ways to decrease consumption volatility in the model is to allow for consumption habits (Vasilev, 2018a), or Epstein-Zin preferences featuring a desire for "early resolution of uncertainty" on the household's side (Vasilev, 2018b).

According to his study, the rate by which future utility is discounted, should importantly depend on present consumption. More specifically, the higher the level of real income today, the lower is the rate by which the individual discounts tomorrow's real income (or consumption). That is, it seems more realistic to think that the wealthier a person is, the more impatient the household should be, and thus the higher the agent's preference is for immediate consumption. We take those analytical results seriously and extend the standard model with an endogenous discount factor. The modelling choice follows the modification presented in (Schmitt-Grohe and Uribe (2003) due to its computational simplicity, as compared to the original representation in (Uzawa,1968), which is written in continuous time. Other papers using such preferences include (Obstfeld, 1980; Mendoza, 1991; Schmitt-Grohe, 1998; Uribe, 1997; Kim and Kose, 2003) compare the business-cycle properties of this model to those of the standard model with a constant discount factor.

Another important difference from (Uzawa, 1968) is that the endogenous discount factor will also depend positively on hours worked. In other words, a higher labor supplied is associated with a lower wealth, and hence the more patient the household is. In addition to the endogenous discount factor, the utility function in the extended model uses the formulation utilized in (Greenwood et al., 1988). Many authors, such as Mendoza (1991) and Correia et al. (1995), have demonstrated that these preferences improve the ability of these setups to capture business cycle facts: e.g., Benhabib et al. (1991) show that such preferences can be interpreted as reduced-form ones for an economy with home production. Furthermore, these preferences generate a labor supply response that is independent of the intertemporal consumption-saving decision, and the inter-temporal substitution (income) effect - which is a central mechanism in a large class of dynamic macroeconomic models - is thus eliminated; In other words, the elasticity of intertemporal substitution associate with leisure is zero. This

form of the utility function then allows us to study and emphasize different transmission mechanism in this paper.

The augmented model setup in this paper will be then used to quantitatively investigate the effect of endogenous discounting on the business cycle fluctuations in Bulgaria after the introduction of the currency board arrangement (1999-2018), which was a period of macroeconomic stability. Beside the second-moment matching exercise, the model will be validated using the methodology suggested in Canova (2007). In addition, the model is able to address the criticism of (Nelson and Plosser,1982; Cogley and Nason, 1995; Rotemberg and Woodford, 1996b), who argue that the RBC class of models do not have a strong internal propagation mechanism besides the strong persistence in the total factor productivity (TFP) process. We show those critiques are unfounded in the Bulgarian case, where the persistence of the TFP is much lower than that in the US, for example.

The rest of the paper is organized as follows: Section 2 describes the model framework and describes the decentralized competitive equilibrium system, Section 3 discusses the calibration procedure, and Section 4 presents the steady-state model solution. Sections 5 proceeds with the out-of-steady-state dynamics of model variables, and compared the simulated second moments and auto- and cross-correlation functions of theoretical variables against their empirical counterparts. Section 6 concludes the paper.

2. **Model Setup**

There is a representative household, which derives utility out of consumption and leisure. The time available to households can be spent in productive use or as leisure. The government taxes consumption spending and levies a common tax on all income, in order to finance non-productive purchases of government consumption goods, and government transfers. On the production side, there is a representative firm, which hires labor and capital to produce a homogenous final good, which could be used for consumption, investment, or government purchases.

2.1. Representative household

There is a representative household, which maximizes its expected utility function, which features an instantaneous felicity function, as in (Greenwood et al., 1988; Schmitt-Grohe and Uribe, 2003):

$$E_0 \Sigma_{t=0}^{\infty} \theta_t \frac{\left[c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{1-\sigma}}{1-\sigma}, \tag{2.1}$$

where

$$\theta_0 = 1, \tag{2.2}$$

$$\theta_{t+1} = \beta(c_t, h_t)\theta_t,\ t \geq 0, \tag{2.3}$$

$$\beta(c_t, h_t) = \left[1 + c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{-\psi}. \tag{2.4}$$

Parameter $\sigma > 1$ captures the curvature of the utility function, $\nu > 0$ is the labor supply elasticity, and $\psi > 0$ is the curvature parameter of the discount factor function. The novelty in the setup is that the subjective discount factor θ_t is no longer a fixed scalar, but rather a function of current individual consumption and labor supply, and thus an endogenous variable. Furthermore, the discount factor is time-varying and assumed to be decreasing in consumption, and increasing in hours. In other words, as pointed out in Uzawa (1968), agents become much more impatient the more they consume. The household will internalize the effects of consumption and hours worked on the discount factor when choosing optimally how much to consume and how much labor to supply in each period.

Next, the household starts with an initial stock of physical capital $k_0 > 0$, and has to decide how much to add to it in the form of new investment. The law of motion for physical capital is

$$k_t+1 = i_t + (1 - \delta)k_t, \tag{2.5}$$

where $0 < \delta < 1$ is the depreciation rate. Next, the real interest rate is r_t, hence the before-tax capital income of the household in period t equals $r_t k_t$. In addition to capital income, the household can generate labor income. Hours supplied to the representative firm are rewarded at the hourly wage rate of w_t, so pre-tax labor income equals $w_t h_t$. Lastly, the household owns the firm in the economy and has a legal claim on all the firm's profit, π_t.

Next, the household's problem can be now represented as

$$E_0 \sum_{t=0}^{\infty} \theta_t \left\{ \frac{\left[c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{1-\sigma}}{1-\sigma} + \lambda_t \left[-(1+\tau^c)c_t - k_t + (1-\delta)k_t + (1-\tau^y)\left[r_t k_t + w_t h_t\right] + g_t^t\right] \right\}$$

$$+ \eta_t \left[\theta_{t+1} - \beta(c_t, h_t)\theta_t\right], \tag{2.6}$$

where λ_t is the Lagrangean multiplier of the period-t budget constraint, while η_t is the Lagrangean multiplier associated with the evolution of the endogenous discount factor. Note also that λ_t is discounted, while η_t is not. The first-order optimality conditions (FOCs) are as follows:

$$c_t: \left[c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{-\sigma} + \psi\eta_t \left[1 + c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{-\psi-1} = \lambda_t(1+\tau^c), \tag{2.7}$$

$$h_t: h_t^{\upsilon-1}\left[c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{-\sigma} = \lambda_t(1-\tau^y)w_t - \psi\eta_t h_t^{\upsilon-1}\left[1 + c_t - \frac{h_t^{\upsilon}}{\upsilon}\right]^{-\psi-1}, \tag{2.8}$$

$$k_{t+1}: \lambda_t = \left[1 + c_t - \frac{h_t^{\nu}}{\nu}\right]^{-\psi} E_t \lambda_{t+1}\left[1 - \left(1 - \tau^{y}\right) r_{t+1} - \delta\right], \tag{2.9}$$

$$\theta_{t+1}: \eta_t = -E_t\left[c_t - \frac{h_t^{\nu}}{\nu}\right]^{-\sigma} + E_t \eta_{t+1}\left[1 + c_{t+1} - \frac{h_{t+1}^{\nu}}{\nu}\right]^{-\psi}, \tag{2.10}$$

$$TVC: \lim_{t\to\infty} \theta^t \lambda_t k_{t+1} = 0. \tag{2.11}$$

The interpretation of the FOCs above is standards: for each household, the marginal utility of consumption (taking into consideration the effect of consumption on the discount factor) equals the marginal utility of wealth, corrected for the consumption tax rate. With endogenous discounting, the marginal utility of consumption contains an additional term, which reflects the fact that an increase in consumption this period lowers the discount factor. More specifically, a unit decline in the discount reduces t-period utility by the value of the Lagrange multiplier, η_t, which is now an additional state variable in the system. Alternatively, η_t can be regarded as the present discounted value of utility from $t + 1$ onward (or the "continuation value"). Substituting forward in (2.10) yields:

$$\eta_t = -E_t \Sigma_{j=1}^{\infty} \left(\frac{\theta_{t+j}}{\theta_{t+1}}\right) \frac{\left[c_t - \frac{h_t^{\nu}}{\nu}\right]^{1-\sigma}}{1-\sigma}, \tag{2.12}$$

Similarly, the second equation states that when choosing labor supply optimally, at the margin, each hour spent by the household working for the firm should balance the benefit from doing so in terms of additional income generates, and the cost measured in terms of lower utility of leisure, plus the correction from the effect of work on the discount factor. The third equation is the so-called "Euler condition," which describes how the household chooses to allocate physical capital over time. The difference here is that it features an endogenous discount factor. The last condition is called the "transversality condition" (TVC): it states that at the end of the horizon, the value of physical capital should be zero. The evolution of the continuation value is not presented, as it was discussed above.

2.2. Firm problem

There is a representative firm in the economy, which produces a homogeneous product. The price of output is normalized to unity. The production technology is Cobb-Douglas and uses both physical capital, k_t, and labor hours, h_t, to maximize static profit

$$\pi_t = A_t k_t^{\alpha} h_t^{1-\alpha} - r_t k_t - w_t h_t, \tag{2.13}$$

where A_t denotes the level of technology in period t. Since the firm rents the capital from households, the problem of the firm is a sequence of static profit maximizing problems. In equilibrium, there are no profits, and each input is priced according to its marginal product, i.e.:

$$k_t: \alpha \frac{y_t}{k_t} = r_t, \tag{2.14}$$

$$h_t: (1-\alpha)\frac{y_t}{k_t} = w_t, \tag{2.15}$$

2.3. Government

In the model setup, the government is levying taxes on labor and capital income, as well as consumption, in order to finance spending on wasteful government purchases, and government transfers. The government budget constraint is as follows:

$$g_t^c + g_t^t = \tau^c c_t + \tau^y \left[w_t h_t + r_t k_t \right]. \tag{2.16}$$

Tax rates and government consumption-to-output ratio would be chosen to match the average share in data, and government transfers would be determined residually in each period so that the government budget is always balanced.

2.4. Dynamic competitive equilibrium (DCE)

For a given process followed by technology $\{A_t\}_{t=0}^{\infty}$ average tax rates $\{\tau^c, \tau^y\}$, initial capital stock $\{k_0\}$, the decentralized dynamic competitive equilibrium is a list of sequences $\{c_t, i_t, k_t, h_t, \theta_t\}_{t=0}^{\infty}$ for the household, a sequence of government purchases and transfers $\{g_t^c, g_t^t\}_{t=0}^{\infty}$ and prices $\{w_t, r\}_{t=0}^{\infty}$ such that (i) the representative household maximizes utility; (ii) the firm maximizes profit; (iii) the government budget is balanced; (iv) all markets clear.

3. **Data And Model Calibration**

To characterize business cycle fluctuations with an endogenous depreciation rate in Bulgaria, we will focus on the period following the introduction of the currency board (1999-2018). Quarterly data on output, consumption and investment was collected from National Statistical Institute (2020), while the real interest rate is taken from Bulgarian National Bank Statistical Database (2020). The calibration strategy described in this section follows a long- established tradition in modern macroeconomics: first, the steady-state value of the discount factor, $\beta(c, h) = 0.982$, is set to match the steady-state capital-to-output ratio in Bulgaria, $k/y = 13.964$, in the steady-state Euler equation. The labor share parameter, $1 - \alpha = 0.571$, is set equal to the average value of labor income in aggregate output over the period 1999-2016. Next, the average income tax rate was set to $\tau^y = 0.1$, which is the average effective rate. Similarly, the tax rate on consumption is set to its value over the period, $\tau^c = 0.2$.

Next, the curvature of the utility function is set to $\sigma = 2$, which is a standard value in the literature (e.g. Hansen and Singleton 1983), while the curvature of the endogenous discount factor function, ψ, is calibrated to match the steady-state value of the discount factor. In turn, the labor supply elasticity, ν, is calibrated to match that in steady-state consumers would supply one-third of their time endowment to working. This is in line with the estimates for Bulgaria as well over the period studied. Next, the steady-state depreciation rate of physical capital in Bulgaria, $\delta = 0.013$, was estimated as the average quarterly depreciation rate over the period 1999-2018. Finally, the processes followed by total factor productivity (TFP) is estimated from the detrended series by running an AR(1) regression and saving the residuals. Table 1 summarizes the values of all model parameters used in the paper.

Table 1: Model parameters

Parameter	Value	Description	Method
β	0.982	Discount factor	Calibrated
α	0.429	Capital Share	Data average
$1-\alpha$	0.571	Labor Share	Calibrated
σ	2.000	Curvature of the utility function	Set
ν	1.400	Labor supply elasticity	Calibrated
ψ	0.110	Curvature of the discount factor function	Calibrated
Δ	0.013	Depreciation rate on physical capital	Data average
τ^y	0.100	Average tax rate on income	Data average
τ^c	0.200	VAT/consumption tax rate	Data average
ρ	0.701	AR(1) persistence coefficient, TFP process	Estimated
σ	0.044	standard error, TFP process	Estimated

3.1. Steady-state

Once the values of model parameters were obtained, the steady-state equilibrium system solved, the "big ratios" can be compared to their averages in Bulgarian data. The results are reported in Table 2 on the next page. The steady-state level of output was normalized to unity (hence the level of technology A differs from one, which is usually the normalization done in other studies), which greatly simplified the computations. Next, the model matches consumption-to-output and government purchases ratios by construction; The investment ratios are also closely approximated, despite the closed-economy assumption and the absence of foreign trade sector. The shares of income are also identical to those in data, which is an artifact of the assumptions imposed on functional form of the aggregate production function. The after-tax return, where $\overline{r} = (1-\tau^y)r - \delta$ is also relatively well-captured by the model.

Lastly, given the absence of debt, and the fact that transfers were chosen residually to balance the government budget constraint, the result along this dimension is understandably not so close to the average ratio in data.

Table 2: Data averages and long-run solution

Variable	Description	Data	Model
y	Steady-state output	N/A	1.000
c/y	Consumption-to-output ratio	0.648	0.674
i/y	Investment-to-output ratio	0.201	0.175
k/y	Capital-to-output ratio	13.96	13.96
g^c/y	Government consumption-to-output ratio	0.151	0.151
wh/y	Labor income-to-output ratio	0.571	0.571
rk/y	Capital income-to-output ratio	0.429	0.429
h	Share of time spent working	0.333	0.333
$\bar{r}$	After-tax net return on capital	0.014	0.015

4. Out Of Steady-State Model Dynamics

Since the model does not have an analytical solution for the equilibrium behavior of variables outside their steady-state values, we need to solve the model numerically. This is done by log-linearizing the original equilibrium (non-linear) system of equations around the steady- state. This transformation produces a first-order system of stochastic difference equations. First, we study the dynamic behavior of model variables to an isolated shock to the total factor productivity process, and then we fully simulate the model to compare how the second moments of the model perform when compared against their empirical counterparts.

4.1. Impulse response analysis

This subsection documents the impulse responses of model variables to a 1% surprise innovation to technology. The impulse response functions (IRFs) are presented in Fig. 1 and 2 on the next page, for the case of an endogenous discount factor, and with a constant discount factor. The IRFs are qualitatively very similar across setups: In both models, as a result of the one-time unexpected positive shock to total factor productivity, output increases upon impact. This expands the availability of resources in the economy, so uses of output – consumption, investment and government consumption also increase contemporaneously. The only major difference between the two models is that with an endogenous discount factor, the response in consumption is smoothed ("excess smoothness" in consumption), while the response in investment is increased. This "excess sensitivity" in investment behavior is due to

the fact that the consumer internalizes the effect of consumption on the discount factor. As a result, consumption volatility drastically decreases. In turn, with smooth consumption, the adjustment happens with saving (i.e., physical capital accumulation in the model). In turn, physical capital becomes more volatile, and exhibits a hump-shaped behavior.

Figure 1: Impulse responses to a 1 percent surprise innovation in technology (endogenous discount factor)

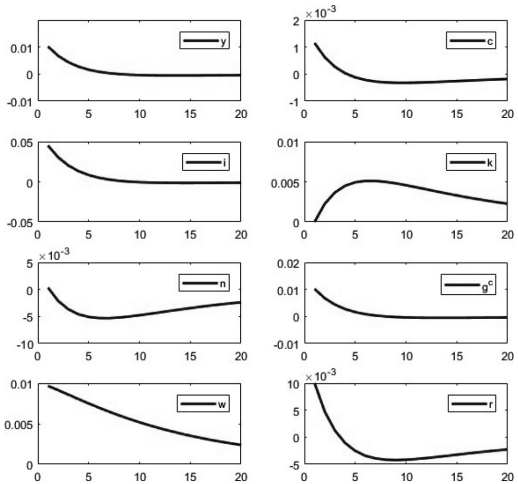
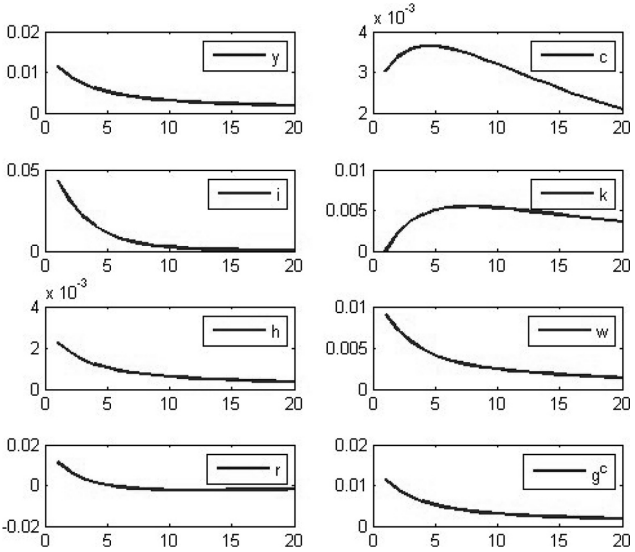


Figure 2: Impulse responses to a 1 percent surprise innovation in technology (constant discount factor)



At the same time, in both models the increase in productivity increases the after-tax return on the two factors of production, labor and capital. The representative households then respond to the incentives contained in prices and start accumulating capital, and supplies more hours worked. In turn, the increase in capital input feeds back in output through the production function and that further adds to the positive effect of the technology shock. In the labor market, the wage rate increases, and the household increases its hours worked. In turn, the increase in total hours further increases output, again indirectly.

Over time, as capital is being accumulated, its after-tax marginal product starts to decrease, which lowers the households' incentives to save. As a result, physical capital stock eventually returns to its steady-state, and exhibits a hump-shaped dynamics over its transition path. The rest of the model variables return to their old steady-states in amonotone fashion as the effect of the one-time surprise innovation in technology dies out.

4.2. Simulation and moment-matching

We will now simulate the model 10,000 times for the length of the data horizon. Both empirical and model simulated data is detrended using the Hodrick and Prescott (1980) filter. Table 3 on the next page summarizes the second moments of data (relative volatilities to output, and contemporaneous correlations with output) versus the same moments computed from the model-simulated data at quarterly frequency. Against the model with endogenous dis- counting ("Uzawa model"), we present a model with constant discount factor, and a standard log-log RBC model. All models match quite well the absolute volatility of output. By construction, government consumption in the model varies as much as output. The model with endogenous discount factor underestimates the variability in consumption, but predicted volatility of investment is too large. The other models overestimate both consumption and investment volatility. Still, all the models are qualitatively consistent with the stylized fact that consumption generally varies less than output, while investment is more volatile than output.

With respect to the labor market variables, the variability of employment and wages predicted by the model is much higher than that in data, but the variability of unemployment in the model is not that far away from the observed volatility in data. This is yet another confirmation that the perfectly-competitive assumption, (e.g. Vasilev, 2009), as well as the benchmark calibration here, does not describe very well the dynamics of labor market variables. The models with constant discount factor, and the standard RBC model underestimate hours volatility, but are not far off in terms of wage variability.

Next, in terms of contemporaneous correlations, aside from the Uzawa model (1968?), the other setups systematically over-predicts the pro-cyclicality of the main aggregate variables – investment, and government consumption. This, however, is a common limitation of this class of models. The puzzle from the endogenous discount factor (containing consumption and hours) is that the predicted contemporaneous consumption correlation with output is too low. Also, along the labor market dimension, the contemporaneous correlation of employment with output, and unemployment with output, are a bit too weak. With respect to wages, the models predict moderate to perfect cyclicality, while wages in data are acyclical. This shortcoming is well-known in the literature and an artifact of the wage being equal to the labor productivity in the model.

Table 3: Business cycle moments

	Data	Uzawa Model	Constant β Model	Standard RBC model
σy	0.05	0.05	0.05	0.05
σc/σy	0.55	0.37	0.78	0.82
σi/σy	1.77	4.90	2.34	2.35
σg/σy	1.21	1.00	1.00	1.00
σh/σy	0.63	2.69	0.20	0.28
σw/σy	0.83	3.52	0.80	0.86
σy/h/σy	0.86	3.52	0.80	0.86
σu/σy	3.22	2.80	0.20	0.28
corr(c, y)	0.85	0.15	0.91	0.90
corr(i, y)	0.61	0.95	0.87	0.83
corr(g, y)	0.31	1.00	1.00	1.00
corr(h, y)	0.49	0.14	1.00	0.59
corr(w, y)	-0.01	0.42	1.00	0.96
corr(u, y)	-0.47	-0.14	-0.99	-0.59

In the next subsection, as in Vasilev (2017), we investigate the dynamic correlation between labor market variables at different leads and lags, thus evaluating how well the model matches the phase dynamics among variables. In addition, the autocorrelation functions (ACFs) of empirical data, obtained from an unrestricted VAR (1) are put under scrutiny and compared and contrasted to the simulated counterparts generated from the model.

4.3. Auto- and cross-correlations

This subsection discusses the auto-(ACFs) and cross-correlation functions (CCFs) of the major model variables. The coefficients empirical ACFs and CCFs at different leads and lags are presented in Table 4 below against the averaged simulated AFCs and CCFs. For the sake of brevity, we only perform results for the Uzawa model specification. Following Canova (2007), this is used as a goodness-of-fit measure.

Table 4: Autocorrelations for Bulgarian data and the model economy

		k			
Method	Statistic	0	1	2	3
Data	$corr(u_t, u_{t-k})$	1.000	0.765	0.552	0.553
Model	$corr(u_t, u_{t-k})$ (s.e.)	1.000 (0.000)	0.958 (0.024)	0.910 (0.047)	0.856 (0.068)
Data	$corr(n_t, n_{t-k})$	1.000	0.484	0.009	0.352
Model	$corr(n_t, n_{t-k})$ (s.e.)	1.000 (0.000)	0.958 (0.024)	0.910 (0.047)	0.856 (0.068)
Data	$corr(y_t, y_{t-k})$	1.000	0.810	0.663	0.479
Model	$corr(y_t, y_{t-k})$ (s.e.)	1.000 (0.000)	0.953 (0.028)	0.896 (0.054)	0.829 (0.078)
Data	$corr(a_t, a_{t-k})$	1.000	0.702	0.449	0.277
Model	$corr(a_t, a_{t-k})$	1.000	0.955	0.901	0.837
	(s.e.)	(0.000)	(0.027)	(0.051)	(0.075)
Data	$corr(c_t, c_{t-k})$	1.000	0.971	0.952	0.913
Model	$corr(c_t, c_{t-k})$ (s.e.)	1.000 (0.000)	0.958 (0.025)	0.909 (0.052)	0.851 (0.076)
Data	$corr(i_t, i_{t-k})$	1.000	0.810	0.722	0.594
Model	$corr(i_t, i_{t-k})$ (s.e.)	1.000 (0.000)	0.955 (0.027)	0.899 (0.052)	0.834 (0.076)
Data	$corr(w_t, w_{t-k})$	1.000	0.760	0.783	0.554
Model	$corr(w_t, w_{t-k})$ (s.e.)	1.000 (0.000)	0.958 (0.025)	0.909 (0.048)	0.854 (0.070)

As seen from Table 4 above, the model compares relatively well vis-a-vis data. Empirical ACFs for output and investment are slightly outside the confidence band predicted by the model, while the ACFs for total factor productivity and household consumption are well-approximated by the model. The persistence of labor market variables is also relatively well-described by the model dynamics. Overall, the model with habits in consumption generates too much persistence in output and both employment and unemployment, and is subject to the criticism in (Nelson and Plosser,1982; Cogley and Nason, 1995; Rotemberg and Woodford, 1996), who argue that the RBC class of models do not have a strong internal propagation

mechanism besides the strong persistence in the TFP process. In those models, and in the current one, labor market is modelled in the Walrasian market-clearing spirit, and output and unemployment persistence is low.

Next, as seen from Table 5 below, over the business cycle, in data labor productivity leads employment. The model, however, cannot account for this fact. As in the standard RBC model a technology shock can be regarded as a factor shifting the labor demand curve, while holding the labor supply curve constant. Therefore, the effect between employment and labor productivity is only a contemporaneous one.

Table 5: Dynamic correlations for Bulgarian data and the model economy

		k						
Method	Statistic	-3	-2	-1	0	1	2	3
Data	$corr(n_t, (y/n)_{t-k})$	-0.342	-0.363	-0.187	-0.144	0.475	0.470	0.346
Model	$corr(n_t, (y/n)_{t-k})$ (s.e.)	-0.091 (0.346)	-0.135 (0.304)	-0.198 (0.259)	-0.947 (0.047)	-0.514 (0.272)	-0.425 (0.312)	-0.348 (0.347)
Data	$corr(n_t, w_{t-k})$	0.355	0.452	0.447	0.328	-0.040	-0.390	-0.57
Model	$corr(n_t, w_{t-k})$ (s.e.)	-0.091 (0.346)	-0.135 (0.304	-0.198 (0.259)	-0.947 (0.047)	-0.514 (0.272)	-0.425 (0.312)	-0.348 (0.347)

5. Conclusions

We introduce an endogenous discount factor as in Uzawa (1968) and Schmitt-Grohe and Uribe (2003) into a real-business-cycle setup with Greenwood et al. (1988) preferences and augment the model with a detailed government sector. We calibrate the artificial economy to Bulgarian data for the period following the introduction of the currency board arrangement (1999-2016). We investigate the quantitative importance of endogenous discounting for the propagation cyclical fluctuations in Bulgaria. The presence of an endogenous discount factor improves the model performance against data, and in addition this extended setup dominates the standard RBC model framework with a constant discount factor (Vasilev, 2009).

References

Benhabib, J.; Rogerson, R.; Wright, R. (1991) Homework in Macroeconomics: Household Production and Aggregate Fluctuations. *Journal of Political Economy*, 99(6), 1166-1187.

Bulgarian National Bank (2020) *Bulgarian National Bank Statistics*. Available on-line at www.bnb.bg. Accessed on Oct. 30, 2020.

Canova, F. (2007) *Methods for Applied Macroeconomic Research*. Princeton, Princeton University Press.

Correia, I.; Neves, J. C.; Rebelo, S. (1995) Business cycles in a small open economy. *European Economic Review*, 39(6), 1089-1113.

Cogley, T.; Nason, J. (1995) Output dynamics in Real-Business-Cycles. *American Economic Review*, 85(3), 492-511.

Greenwood, J.; Hercowitz, Z.; Huffman, G. (1988) Investment, capacity utilization, and the Real Business Cycle. *American Economic Review*, 78(3), 402-17.

Hansen, L.; Singleton, K. (1983) Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy*, 91(2), 249-265.

Hodrick, R.; Prescott, E.C. (1980) Post-war US business cycles: An empirical investigation, Unpublished manuscript (Carnegie-Mellon University, Pittsburgh, PA).

Kim, S.H.; Kose, A. (2003) Dynamics of open economy business cycle models: Understanding the role of the discount factor. *Macroeconomic Dynamics,* 7, 263-270.

Mendoza, E. (1991) Real business cycles in a small-open economy. *American Economic Review,* 81, 797-818.

Nelson, C.; Plosser, C. (1982) Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139-162.

National Statistical Institute (2020) *Aggregate Statistical Indicators*. Available on-line at www.nsi.bg. Accessed on Oct. 30, 2020.

Obstfeld, M. (1980) Intertemporal dependence, impatience, and dynamics. *Journal of Monetary Economics,* 26, 45-75.

Rotemberg, J.; Woodford, M. (1996) Real-Business-Cycle models and the forecastable movements in output, hours, and consumption. *American Economic Review*, 86, 71-89.

Schmitt-Grohe, S.; Uribe, M. (2003) Closing small open economy models. *Journal of International Economics,* 61, 163-185.

Schmitt-Grohe, S. (1998) The international transmission of economic fluctuations: Effects of US business cycles on the Canadian economy. *Journal of International Economics,* 44, 257-287.

Uribe, M. (1997) Exchange rate-based inflation stabilization: the initial real effects of credible plans. *Journal of Monetary Economics,* 39, 197-221.

Uzawa, H. (1968) Time preference, the consumption function and the optimum asset holdings. In: Wolfe, J.N. (Ed.), *Value, Capital and Growth: Papers in Honor of Sir John Hicks*. Edinburgh, The University of Edinburgh Press, 485-504.

Vasilev, A. (2018a) Are habits important for the propagation of business cycle fluctuations in Bulgaria? *Central European Journal of Economic Modelling and Econometrics*, 11(3), 133-151.

Vasilev, A. (2018b) An RBC model with Epstein-Zin (non-expected-utility) recursive preferences: Lessons from Bulgaria (1999-2016). *EconStor Preprints* 182577, ZBW - German National Library of Economics.

Vasilev, A.Z. (2017) A Real-Business-Cycle model with efficiency wages and a government sector: The case of Bulgaria. *Central European Journal of Economic Modelling and Econometrics*, 9(4), 359-377.

Vasilev, A. (2009) Business cycles in Bulgaria and the Baltic countries: An RBC approach. *International Journal of Computational Economics and Econometrics*, 1(2), 148-170.

IMPRENSA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS