THE REMSDB MACROECONOMIC DATABASE OF THE SPANISH ECONOMY

J.E. Boscá[a], A. de Bustos[b], A. Díaz[b], R. Doménech[a,c], J. Ferri[a], E. Pérez[b], and L. Puch[d]

D-2007-03

October, 2007

[a] University of Valencia, Spain
[b] Ministry of Economy and Finance, Spain
[c] Economic Bureau of the Prime Minister, Spain
[d] FEDEA, Universidad Complutense and ICAE, Spain

Financial support from European Regional Development Fund (ERDF) and Fundación Rafael del Pino is gratefully acknowledged. The authors thank V. Gómez for his assistance. J.E. Boscá, R. Doménech and J. Ferri also thank the financial support by CICYT grant SEJ2005-01365 and Luis Puch CICYT grant SEJ2004-04579. *Address for correspondence*: Luis A. Puch, FEDEA, Jorge Juan 46, 28001 Madrid, e-mail: lpuch@fedea.es.

The data set is available at:
www.sgpg.pap.meh.es/SGPG/Cln_Principal/Presupuestos/Documentación/

Abstract

This paper presents a new macroeconomic database for the Spanish economy, *REMSDB*. The construction of this database has been oriented to conducting medium-term simulations for policy evaluation with the *REMS* model, a large Rational Expectations Macroeconomic Model for Spain. The paper provides a detailed description of the data and documents its main statistical properties. The database is thought to be of major interest to related applications, whether strictly associated with the *REMS* model or, rather, with empirical macroeconomic studies.

Keywords: E32, E37.
JEL Classification: Spanish Data, Growth Data, Business Cycle Data, REMS.

1. Introduction

The construction of a macroeconomic model for simulation and policy evaluation of the Spanish economy constitutes a far-reaching endeavour. On the one hand, this project entails specifying the behavioral equations that best capture the working of the economy. On the other hand, much research effort needs to be allocated to building a macroeconomic database that takes the model to teh actual data.

In this paper we present a new macroeconomic database for the Spanish economy, REMSDB. The construction of this database has proved very demanding in terms of effort devoted to the collection and elabouration of data, not to mention the many statistical techniques applied to ensure a reasonable degree of data quality and homogeneity. Thus, REMSDB stands in its own right and may well be used independently for purposes other than conducting medium-term simulations for policy evaluation with the REMS model (see Boscá *et al.,* 2007). In this vein, this paper distinguishes the term"baseline" from the term "database", the latter being broader in scope. We will refer to the "baseline" whenever we mean a particular use of the "database", i.e., the construction of a default scenario used as a benchmark in running simulations with REMS. Moreover, the interaction between modeling requirements and updated macroeconomic information will result in subsequent vintages of REMSDB.

The purpose of this paper is to present the REMSDB database. The work is organized as follows. Section 2 describes the main features of the series included in the database. Section 3 lists the data sources. A detailed survey of the series is provided in section 4. Section 5 documents the cyclical behaviour of a selected set of series. Concluding remarks are presented in the final section.

2. The data: main features

The database is made of national aggregates. All series cover the period 1980-2010, which in turn can be divided into two sub- periods depending on the nature of the data. The first one ranges from 1980 to the last available data released by the various statistical sources, that is, 2006 in the current version of the database. The second one, which extends over a four-year period, mostly relies on the official forecasts included in the Stability and Growth Program (SGP). In the current vintage, the last available year of the forecast period corresponds to 2010. Additionally, in order to generate a baseline scenario of the REMS model, the whole set of variables are prolonged towards a 2050 horizon. Although not part of REMSDB, this forward extrapolation obviously builds on the database and further com-

plies with the requirements imposed by the *balanced-growth hypothesis.*

For simulation purposes, all series have been collected on a quarterly basis. When quarterly data were not readily available from the existing statistical sources high-frequency series have been obtained by applying the *Kalman filter and smoother* to an appropriate state-space model where the observations correspond to the low-frequency data.[1] The frequency of monthly series has also been treated and converted into quarterly data with techniques that are specific to each series. Needless to say, all series take the form of seasonally-adjusted data. Whenever the series provided by official statistical sources were not so, they have been seasonally-adjusted using TRAMO-SEATS procedures. The dataset has not been subject to any transformation other than the extraction of the seasonal component or the mere application of linking-back techniques. This is not the case of the variables used to construct a baseline scenario of the REMS model, most of which have been expressed in efficiency units so as to display a stationary pattern. Put differently, every series included in the baseline scenario exhibits a number of statistical properties that comply with the *balanced-growth hypothesis.* Section 4.2 explains in depth the treatment given to the baseline-type series.

The database considers five types of variables. While each of these groups is somewhat stylized, they gather a set of variables of a different nature. The taxonomy is as follows. The first category includes various production and demand aggregates along with their corresponding deflators. A second group brings together population and labour-market series. The third block is made of monetary and financial variables whereas the fourth one includes relevant government aggregates. A final set gathers a number of heterogeneous variables that play a role in the baseline scenario and for which no direct statistical counterpart is available in official sources.

3. Data sources

Whenever possible, the information collected for the database is taken from official sources. The key series are taken from the Spanish Quarterly National Accounts (SQNA) provided by the Instituto Nacional de Estadística (INE, National Statistical Institute). The remaining series come from various statistical sources, all of which have been chosen so as to ensure the most coherent dataset possible.

Lack of official back-linking between the two currently published SQNA series, i.e.,

[1] The tools for temporal disaggregation closely resemble the methodology described in chapter 6 of Harvey (1989). In doing so, we proceed along the lines of Chow and Lin (1971) and Fernandez (1981).

1980-2004 and 1995-2006 with base years respectively in 1995 and 2000, complete series for the period 1980-2006 have been built up on the basis of back-casted figures using the level registered in 1995Q1 and the growth rates of the SQNA 1980-2004 series. Apart from that, any treatment given to the series by the INE has been accordingly adopted and the base year 2000 equally respected.

Following the System of National Accounts methodology (SNA93), the new European System of National Accounts (ESA95) recommends the use of chain indices for quantifying growth rates in volume and price measures. The INE has duly implemented this methodology as of Spring 2005. On the plus side, according to the theory of index numbers the chaining method shows a relative superiority over the more traditional method of the fixed-base indices, as it better captures the rapid evolution of the economic reality. On the minus side, an important consequence of using chain indexes is that the associated volume measures are not additive. Empirically, it can be easily shown that the adoption of the chaining technique has a modest effect on the results obtained in terms of growth rates in the indices. That said, in REMSDB we proceed to produce series at constant market prices along the lines of the traditional methodology. In so doing, we adopt 2000 as the base year.

Details about the series, namely their statistical sources and units of measure, are presented in the Appendix. A thorough explanation of any of the five categories of series is provided below.

4. Description of the series

4.1 Historical data (1980-2006) and forecast period (2007-2010)

Production and demand variables

Production and demand variables are central to the database. They have been taken from the SQNA provided by the INE with other official sources playing a complementary role to these data.

The forecast period extends over five additional years beyond the last available data released by official sources. In particular, the demand and production series are extended in line with the forecasts included in the SGP.To enable forecasts on a quarterly basis from the macroecomic scenario reflected in the SGP, the three-month growth rates are benchmarked to their corresponding year-on-year growth rates under the assumption of similar

growth rates across all quarters. Leaving apart the demand and production block, the remaining series have been extended following specific rules as explained hereafter.

The forecast of *GDP at constant market prices* is made consistent with official forecasts for *GDP at current market prices* and the corresponding deflator. A similar strategy is also applied to the various demand aggregates, so as to warrant consistency with total GDP.

It is also assumed that *Non-Market Services' Gross Value Added* keep pace with *Government Current Expenditures*. This very much facilitates the obtention of a forecast period for both *Private GDP at market prices* and *Private GDP at basic prices*, the latter being calculated by deducting the *Net Taxes on Products* from the former.[2] In turn, *Net Taxes on Products* includes three tax categories, namely, the *Value Added Tax (VAT)*, *Taxes and duties on Imports* and *Taxes on Products, excluding VAT and Import and Export Taxes.* The latter includes, among others, taxes levied on alcoholic drinks, tobacco consumption, hydrocarbon products and the like. This means that, in order to forecast *Net Taxes on Products,* several hypothesis are made regarding its three components. First, it is assumed that both the *VAT* and *Taxes on Products, excluding VAT and Import and Export Taxes* change in line with *Private Consumption*. Second, *Taxes and duties on Imports* grow as fast as *Goods and Services Imports.*

The *Capital Stock* series has been constructed according to the perpetual inventory method. This applies both to the historical data and the forecast period. The relevant data come originally from the BDMORES database (see Daván et al., 2002). More precisely, the initial capital stock corresponds to its 1979 level in BDMORES whereas the depreciation rate stands at 5%. To estimate the depreciation rate, BDMORES assumes that the capital stock is made of private and public equipment and structures, including residential capital.

As an indicator for *Energy Demand* we use the total amount of energy commanded by any productive sector in the economy. The total amount of inter-industry energy consumption has been constructed as a weighted average of both gas-oil and electricity consumption as reported by the *Indice de Produccion Industrial (IPI, Industrial Production Index)* provided by the INE. A scale correction is successively performed in order for the total amount of inter-industry energy consumption to match its ratio to GDP that prevails in the economy in 1995 according to the system of Input - Output tables.

[2] Net taxes are calculated subtracting all subsidies to taxes of the same nature.

Prices and costs

Price deflators for *GDP at market prices*, aggregate *Private and Public Consumption, Exports and Imports* are obtained from the SQNA series at current prices and our own estimates of their counterparts at constant market prices. Similarly, the deflator of the private component of total GDP is calculated using as inputs our own estimates of private GDP both at current and constant market prices. This general procedure does not apply to *Energy Demand,* for which we select the *Indice de Precios Industriales (IPRI, Industrial Price Index).*

In general, assumptions on the forecast period rely on the SGP. This is particularly the case for *GDP* and *Private Consumption* deflators. *Public Consumption* and *Exports and Imports* deflators resemble the pattern observed in *Private Consumption*. Forward estimates of the *Energy Demand* deflator are in conformity with long-run trends in energy prices as dictated by future markets.

The data on *Compensation per Employee* are also taken from the SQNA. We subsequently use the Employees and *labour Cost per Employee* series as provided by the SGP so as to complete the forecast period.

Demographic and labour-market series

As a general rule demographic and labour-market variables are taken from the INE, which in turn resorts to a number of sources to collect the relevant series. Forecasting procedures are nevertheless specific to each series. We proceed to provide further details on the variables belonging to this block.

To start with, *Total Population* matches INE annual estimates from 1996 to 2006. Published census data available for 1980 and 1991 have subsequently been linked through the use of historical adjusted inter-census population estimates provided by the INE sometime in the past. This procedure has enabled us to compile complete series for *Total Population* covering the period 1980-2006. Forecasts rely upon population flows as captured by the so-called "Scenario 2" released by the INE.

Estimates of *Adult Population* (i.e., population over 16) are then made consistent with the *Total Population* series computed in the manner described above under the assumption that the demographic age structure remains pretty much the same as captured by official sources. As for the figures covering the forecast period 2007-2010, *Adult Population* is obtained by assuming a constant ratio of this group to *Total Population,* very much in accordance with the experts' sentiment on demographic trends.

As for the unemployment rate, the data included in REMSDB are coherent with Labour Force Survey (LFS) figures for *Total Employment, Unemployment and labour Force.* These series, which are released for two distinctive periods of time (1976-1995 and 1996-2006) imply the same unemployment rate for the last quarter of 1995 and the first quarter of 1996, meaning that no specific treatment was needed to ensure homogeneity across the two quarters of transition. The time span of the series is extended to 2010 on the basis of the forecasts reflected in the SGP.

Total Employment and *Employment Full-time Equivalents* both replicate SQNA data while forecasts are taken from the SGP. Also is SQNA the statistical source to retrieve historical data for *Employment in Non-Market Services* and *Employment Full-time Equivalents in Non-Market Services.* These two series are extended under the assumption that the *Gross Value Added of Non-Market Services* at constant market prices grows in line with public consumption as captured by the SGP and that the Services' labour productivity movements in the years to come will reflect past patterns in its growth rate.

Labour Force figures over the period 1980-2006 are endogenously determined by dividing *Total Employment* as provided by the SQNA between one minus the Unemployment rate.[3] Then the magnitude of *Total Unemployment* is computed by multiplying the unemployment rate by the labour force calculated in the manner described above.

Hours worked are computed as the product of the number of employed workers, as provided by the SQNA, and the number of hours per employed worker, as given by the *Encuesta de Población Activa (EPA, LFS).* The series obtained in this manner for the period 1980-2006 have been made the most coherent possible with the figures on hours provided by the annual National Accounts for the period 2000-2006[4]. As for the quantification of the latter series the various statistical sources significantly differ in their magnitude.[5] Beyond national data, which have been duly consulted, OECD statistics have been studied in depth as a countercheck. However, a strategy has been devised by which the EPA data covering the period 1987 to 2004 have been adjusted to allow for different employment statuses, most remarkably self-employed workers versus employees and part-time versus full-time employment. Retropolation between 1980 and 1987 has been completed by ap-

[3] In practice, this implies that the derived labor force figures significantly differ from the information on the same series provided by the *Labor Force Survey.*

[4] The release of quarterly data on hours worked by the SQN is posterior to the completion of REMSDB.

[5] The number of hours captured by surveys addressed to firms, such as the *Labor Cost Index* and the *Wage Survey,* is well below the number of hours that shows up in surveys that collect data revealed by workers, namely the LFS.

plying the percentage change in hours worked per employee taken from the labour Costs Index (LCI) to the available 1987 level. Forward extrapolation (realistically) assumes that the number of hours per employed worker will remain broadly constant.

On the other hand, REMSDB includes two labour market variables with no statistical counterpart in official sources, namely the number of vacancies and matches per quarter.[6] To construct the former we follow Antolín (1995). This methodology amounts to adjusting the series *Puestos de trabajo ofrecidos para gestión pendientes de cubrir* (job vacancies posted by firms) as reflected in the *Movimiento laboural Registrado (MRL, Recorded labour Flows)* managed by the former *Instituto Nacional de Empleo (INEM, National Employment Institute),* by introducing a scale factor which accounts for the incomplete coverage of the series and also deals with the two level shifts that show up in December 1995 and August 1999. These anomalies may be attributable to methodological changes. To estimate the quarterly number of matches we rely on the *Total de contratos registrados* (Total number of recorded contracts), also included in the MRL. Nevertheless, to provide an accurate picture of the number of matches we adjust for the chaining of successive fixed-term contracts by which workers are fired and hired simultaneously one period after another. This adjustment removes from the series any hire and fire practices not reflecting genuine matches between former unemployed workers and unfilled vacancies. More specifically, we assimilate these "virtual" contracts to those signed before 2001 and captured either by the *Contratos por obra y servicio* (works and services contracts), the *Contratos eventuales por circunstancias de la producción* (temporary contracts contingent on production) or the *Contratos a tiempo parcial* (part-time contracts).

Monetary and financial variables

The main provider of historical data on monetary and financial variables is the Bank of Spain. Two monetary aggregates have been collected, namely the *Contribution of Domestic Monetary and Financial Institutions to M1 / M3.* The two of them are monthly data rangingh from September 1997 to November 2003. Both a retropolation to 1980 and an extrapolation to 2010 have therefore been produced for them to be included in REMSDB. Also the quarterly frequency is performed at the simple arithmetic average calculated over each set of three quarters.

Two short-term interest rate series have been retained, namely the *Spanish Non trans-*

[6] For further details on the construction of these two series see Díaz (2007).

ferable three-month deposits and the *US three-month interest rates.* Originally provided as monthly data, both series of interest rates have been converted into quarterly data by calculating the simple arithmetic average over each set of three quarters.

The quarterly nominal dollar / ecu-euro exchange rate builds on the corresponding monthly series released by the Bank of Spain. The high-frequency series has been obtained by calculating the simple arithmetic average for each three-month period. The forecast period relies on the SGP estimates.

Public sector variables

A further data requirement for the implementation of REMS concerns the series on government expenditures and tax receipts. To meet our needs, the *Account of Expenditures and Receipts of the Public Administration* has been elabourated following the classification used by the *Intervención General de la Administración del Estado (IGAE, State's Accounts Department).* Although this accounting exercise is performed on an annual basis, a quarterly version has been fitted by using a quadratic interpolation, which ensures additive consistency with annual data.[7] The series obtained in this manner are compatible with the *Government Final Consumption,* one of the components of the final demand. A final remark on this set of series concerns the *Government Debt,* whose quarterly frequency of this series has been obtained by direct application of the *Kalman filter* to annual data covering the period 1980-1989. Once the quarterly counterpart is available we treat it so as to produce a complete series together with the *Public Debt according to the Excessive Deficit Procedure* as released by the *Bank of Spain.*

All tax categories in the model, namely the consumption tax, the labour income tax, the social security contributions tax and the capital income tax, are fed with the corresponding average effective tax rates estimated by Boscá *et al.* (2005).

Using behavioral equations to produce non-observable data

The present section deals with several variables which are part of the baseline scenario, not the REMSDB database. There is no direct statistical counterpart in official sources for such variables. These include the *Tobin's q,* the *Composite Capital Stock,* the $\mathrm{Marginal\ Cost}$ and a

[7] The use of quarterly data provided either by the IGAE and the SQNA has been avoided because of non-meaningful time distribution from an economic perspective owing to accountancy procedures. In general, criteria based on National Accounts rules are more relevant from a macroeconomic point of view. The related annual series have accordingly been treated to reflect these accounting criteria on a quarterly basis.

measure of *Total Factor Productivity* in the economy. To produce a sensible benchmark for them, their corresponding behavioral equations in the model have been used. For instance, time series for the *Tobin's q* has been built up on the following expression:

$$q_t = 1 + \phi\left(\frac{j_t}{k_t}\right) \tag{1}$$

This behavioral equation, which is derived optimally from the representative household's maximization problem, states that the amount of investment net of adjustment costs, j_t, is undertaken to the extent that the opportunity cost of a marginal increase in investment in terms of consumption, $1 + \phi\left(j_t/k_t\right)$, is equal to its marginal expected contribution to the household's utility, q_t.

The *Composite Capital Stock*, k_{e_t}, is modeled as a composite good made out of physical capital and energy, e_t, according to the following Constant-Elasticity-of-Substitution (CES) technology:

$$k_{e_t} = \left[a\, k_t^{-\rho} + (1-a)\, e_t^{-\rho}\right]^{-\frac{1}{\rho}} \tag{2}$$

This specification is quite general. In particular, the value of the parameter ρ is chosen such that the elasticity of substitution between physical capital and energy is smaller than the corresponding elasticity between physical capital services and labour.

As standard in the empirical literature, we use a Cobb-Douglas production function to compute residually an indicator for Total Factor Productivity, A_t, such that:

$$A_t = \left(\frac{y_t}{k_{e_t}^{1-\alpha}\left(n_t l_t\right)^{\alpha}}\right)^{\frac{1}{\alpha}} \tag{3}$$

where the product $n_t l_t$ stands for the amount of *Total Hours Worked* and y_t represents the magnitude of *Gross Output*. Not only does the latter include the aggregate *Gross Value Added* generated in the economy, but also the energy input and some (time-varying) fixed costs covered by the mark-up accruing to non-competitive firms.

Finally, we work out the value of the *Marginal Cost* series by using the inter-temporal demand for labour condition, i.e.

$$\gamma_N \lambda_t^{nd} = \frac{\pi_{t+1}}{1+r_t^n}\left[\alpha mc_{t+1}\frac{y_{t+1}}{n_t} - w_{t+1}\left(1+\tau_{t+1}^{sc}\right) l_{1t+1} + \lambda_{t+1}^{nd}(1-\sigma)\right] \tag{4}$$

Whereby optimality requires the marginal contribution of a newly created job to the firm's

profit be equal to the marginal product of labour net of the wage rate plus the capital value of the new job in $t+1$, corrected for the job destruction rate between t and $t+1$. Leaving apart the marginal cost series, any variable in the expression above is taken to the data in our calibration procedure, meaning that the former can be expressed as a function of the latter. More precisely, γ_N, π_t, r_t^n, y_t, w_t, τ_t^{sc}, l_t and σ respectively match the actual values for the labour force (gross) growth rate, (gross) inflation rate, nominal interest rate, gross output, hourly real wage, payroll tax rate, number of working hours per employee and the exogenous rate of job destruction.

The shadow value of a new worker, λ_t^{nd}, is approximated using the following equilibrium condition

$$\kappa_v \frac{v_t}{mat_t} = \lambda_t^{nd} \tag{5}$$

whereby the cost of posting a new vacancy is made equal to the shadow value of a new worker times the probability that the vacancy be filled. In this expression, κ_v denotes the nominal recruiting cost per vacancy, v_t represents the amount of available vacancies in the economy, and mat_t measures the number of new matches per period. As standard in the literature on labour market search, an empirical value for the nominal recruiting cost per vacancy κ_v has been produced under the (reasonable) assumption that, on average, the total amount of resources allocated to the posting of vacancies in the economy amounts to (at most) 1 per cent of GDP.

4.2 Long-term forecasting exercise (2011-2050)

All series included in the REMSDB cover the period 1980-2011. Among these variables, the ones which make the baseline scenario of the REMS model are prolonged further in order for the dynamic solution to achieve the *steady-state* path. In REMS, convergence to the *steady state* occurs within 40 years, i.e., the *steady state* is reached by 2050. As described in the following section, the choice of 2050 as the starting point of the *balanced-growth* path is backed by a standard statistical analysis of the relevant long-term ratios and average growth rates describing the working of the Spanish economy. Every series has accordingly been interpolated from 2011:4, the last quarter with available data, to 2050:1, when the model *steady-state* solution first occurs. The most remarkable feature of the *steady-state* dynamics is that every variable grows at a constant rate, which is equal to zero, the rate of technical progress or the latter plus the total population growth rate, depending on whether the variables are respectively expressed in efficiency units, in per capita terms or

in absolute values.

Furthermore, using the model for simulation requires the statistical data be a solution to the dynamic system, meaning that there is a need for creating a baseline scenario with its corresponding residuals. The baseline scenario is made of the data as provided by REMSDB, whether historical data (1980:1 to 2006:4), forecast data (2007:1 to 2010:4), or data resulting from the extrapolation exercise (2011:1 to 2050:1).For any endogenous variable, residuals are then computed as the difference between the actual value and the model solution obtained under no-shock conditions.

5. Long-run patterns and cyclical regularities of the Spanish economy

The relevant temporal dimension of the REMS model is the medium term. The empirical implementation of the model needs long-run features along with cyclical regularities of the Spanish economy be checked for consistency. This requires documenting the main statistical properties of the series that represent such long- and short-term behaviour. As a robustness check, the statistical properties found in this work are compared with previous related analyses. A first and thorough description of the cyclical regularities of the Spanish economy was provided by Dolado, Sebastián and Vallés (1993). Puch and Licandro (1997) later carried out an update of second-moment properties of the data together with a revision of its balanced-growth patterns in a dynamic stochastic general equilibrium setting. Other authors have subsequently produced analyses of the cyclical patterns of the Spanish macroeconomic aggregates in a Real Business Cycle framework, as reviewed by Lores (2001). The results presented here are comparable to those reported in this strand of the literature.

Figure 1 displays the Hodrick-Prescott trend (dashed line) extracted from the logarithm of real GDP per capita over the period 1980-2006. The evolution of this variable suggests a sustained growth experience over the whole time span, at an average annual growth rate of 1.6%. Two long cycles are also detected, with corresponding peaks in 1992 and 2001, whereas more muted fluctuations seem to prevail in the recent past.

If we first focus on the study of the long-run features of the Spanish economy, a refined analysis on the basis of the relevant ratios and growth rates describing the economic system, precludes the hypothesis of the Spanish economy moving along its *steady-state* path at the beginning of the sample. Several features observed in the data give support to this claim. Hours worked exhibit a marked downward trend until the mid 80s (Fig-

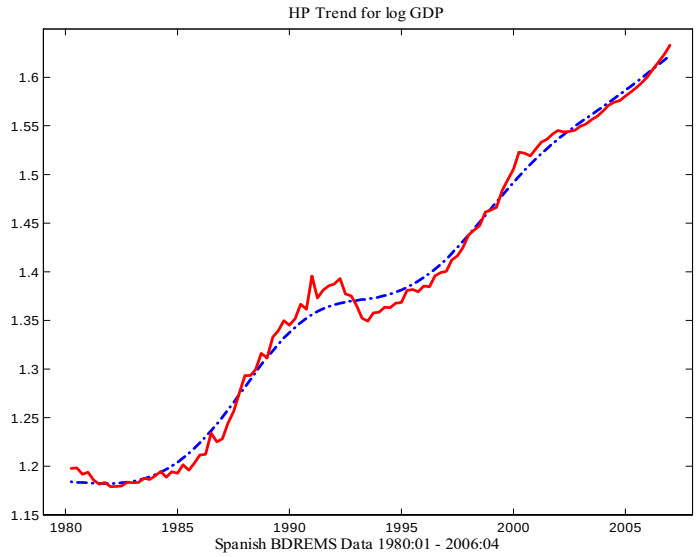

Figure 1: *HP trend of GDP in logs.*

ures 2). This is explained by an important reduction in the number of hours per in the number of hours per worker during the 70s that moderately continues during the next decade, and, to a lesser extent, by a decrease in the employment rate during the 80s. The capital-output ratio displays a slight upward trend in the last years of the sample (Figure 3. The role played by a poor measurement of capital services and the contribution of an increasing value of residential housing in recent years notwithstanding, this dynamics is molt likely drive by the impact of the new economy. In any case, the varying pattern of the capital-output ratio may suggest a very different response of the Spanish economy along the last two economic upturns during the late 80s and late 90s. The hypothesis that deep changes in the Spanish economy may be underlying heterogeneous patterns in the variables describing technology is reinforced by the vacancies series, whose trend (Figure 4) parallels the one observed in the capital-output ratio. Furthermore, while private consumption has been growing hand-in-hand with GDP, government consumption and investment have outpaced GDP. This pattern has been accompanied by a sustained increase in imports over the whole sample period.

Overall, from mid eighties onwards the growth rate of the main macroeconomic aggregates points to a movement of the Spanish economy around its balanced-growth path, where it not for the fact that the impact of an increasing openness over the last two decades has induced a remarkable growth of the foreign sector, whereby the average rate of exports

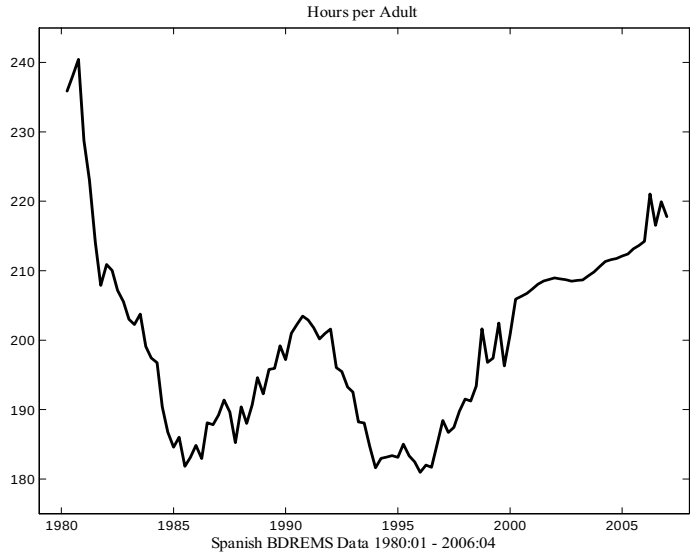

Figure 2: *Hours per adult.*

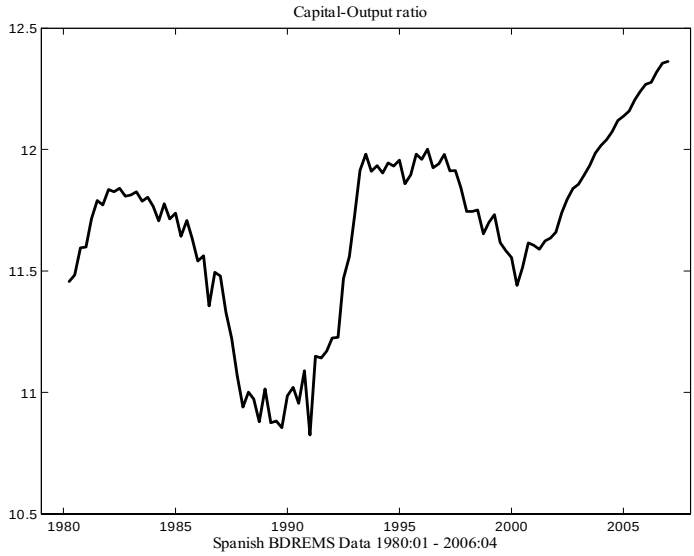

Figure 3: *Capital-output ratio.*

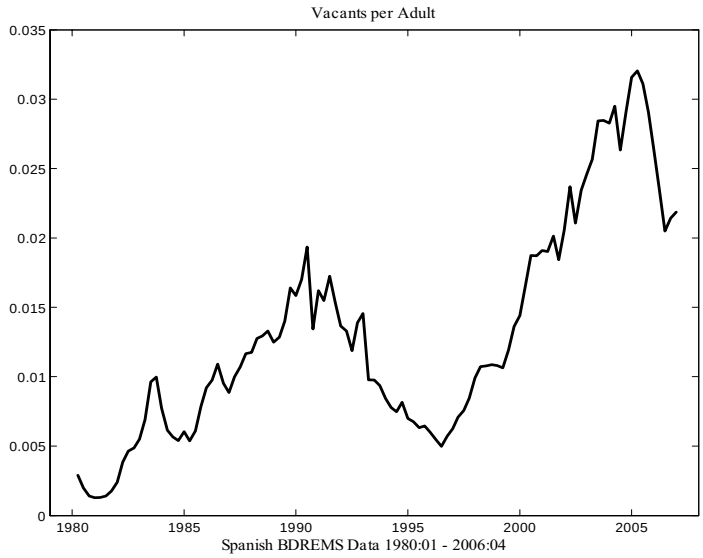

Figure 4: *Vacants per adult.*

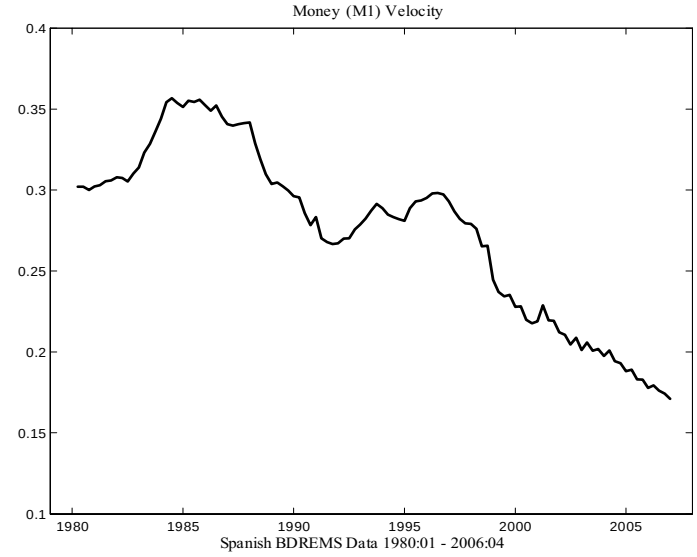

Figure 5: *Money (M1) velocity.*

and imports has respectively reached 5.3 per cent and almost 7 per cent. Likewise a mild expansion occurs with money aggregates, particularly when measured from more liquid assets (see, for instance, Figure 5, where where we represent the ratio of GDP over M1) and less intense from M3 measures, as well as government debt over the recent years.

The cyclical regularities displayed by the main macroeconomic variables (see Table 1) from 1980 to 2006 are comparable to other related studies on the short-term performance of the Spanish economy (see Figures 6 to 9 for a detail). Such regularities include the low volatility of the Spanish GDP (1.17 per cent) compared with other industrialized economies' GDP, the relatively high volatility of private consumption -in seeming contradiction with the life-cycle hypothesis- and the strong counter-cyclical behavior of net exports. Also, in conformity with international evidence, the series of vacancies shows a rather high volatile and persistent behavior, M1 and velocity are pro-cyclical, and prices appear to be counter-cyclical with the energy price index slightly leading the cycle. Finally, government debt exhibits a counter-cyclical pattern lagging the cycle. A look at cross-correlations unveils that the autocorrelation structure of the main aggregates is also in line with that reported in previous analysis of the same nature. Similar conclusions apply to the cross-correlations between the main aggregates and GDP.

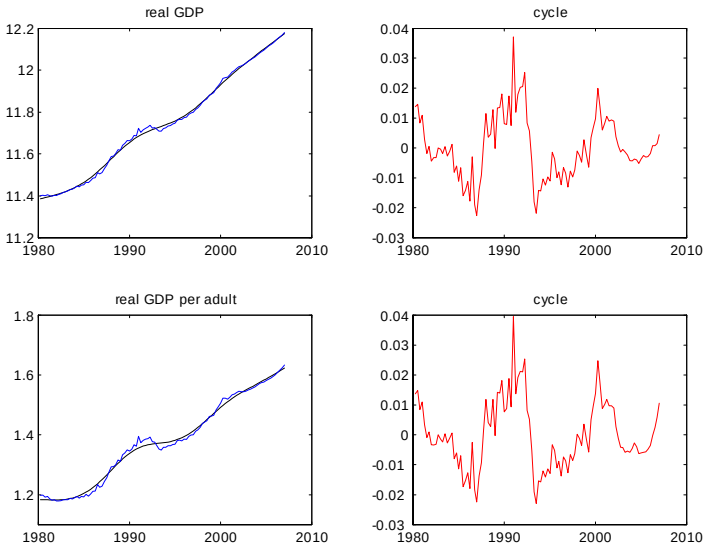

Figure 6: *Trend and cycle decomposition of GDP and GDP per adult.*

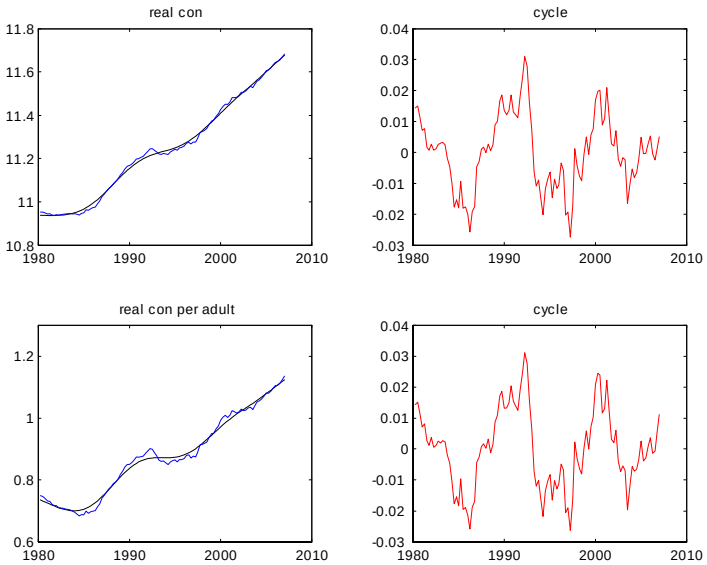

Figure 7: *Trend and cycle components of private consumption (level and per adult).*

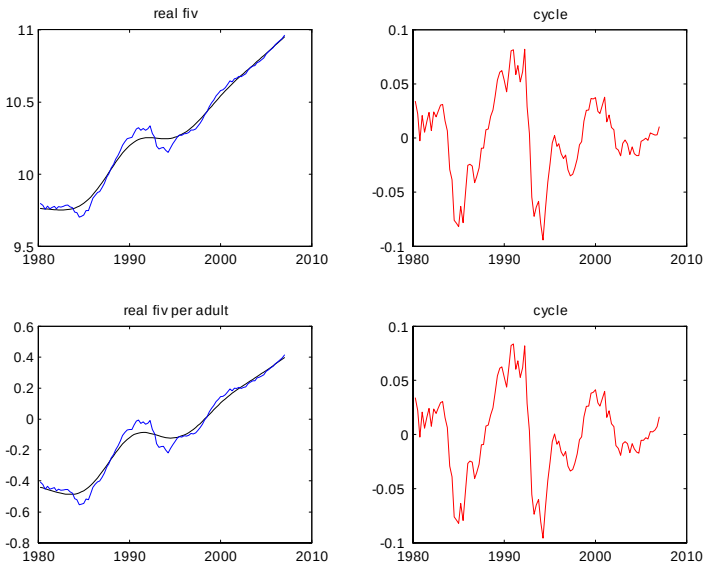

Figure 8: *Trend and cycle components of investment (level and per adult).*

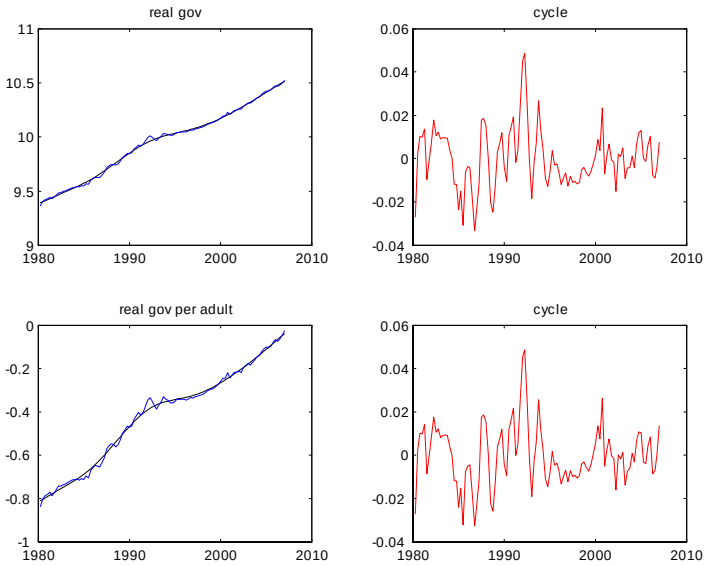

Figure 9: *Trend and cycle components of public consumption (level and per adult).*

Table 1
First, and Second Moments (HP Filtered) properties of Key Macroeconomic Aggregates

	Levels (ratio over GDP)	Growth Rates (annualized)	Relative Volatility (to GDP)	Cross-Correlations of GDP with key aggregates					Auto-correlations
				x_{t-2}	x_{t-1}	x_t	x_{t+1}	x_{t+2}	x_{t-1}
GDP	1	0.0163	1*	0.749	0.800	1	0.793	0.733	
Private consumption	0.6053	0.0145	1.1360	0.694	0.778	0.815	0.801	0.739	0.893
Government consumption	0.1670	0.0307	1.2247	0.378	0.474	0.516	0.386	0.326	0.663
Gross fixed capital formation	0.2308	0.0310	3.5038	0.785	0.815	0.821	0.722	0.648	0.926
Exports	0.2076	0.0528							
Imports	0.2145	0.0687							
Net Exports			3.371	-0.291	-0.308	-0.310	-0.316	-0.312	0.97**
Capital Stock	11.664	0.0193	0.4217	0.144	0.279	0.436	0.532	0.614	0.966
Hours Worked / Adult	199.62	-0.003	1.5828	0.640	0.680	0.729	0.647	0.632	0.766
Vacancies / Adult	$14.0*10^{-3}$	0.077	17.404	0.260	0.225	0.291	0.255	0.277	0.827
GDP Def	—	0.0609	0.6746	-0.600	-0.497	-0.440	-0.362	-0.253	0.867
Energy PI	1.1519	0.0517	4.8536	-0.237	-0.141	-0.034	0.034	0.101	0.828
M1	3.7909	0.1006	3.1178	0.566	0.636	0.684	0.707	0.705	0.891
M1 velocity	—	-0.0211	2.6680	-0.533	-0.569	-0.535	-0.621	-0.613	0.835
Debt	1.8278	0.1270	2.9936	-0.458	-0.487	-0.538	-0.567	-0.641	0.881

6. Concluding remarks

The topic of this paper is the description of a new macroeconomic database, REMSDB. The main use of REMSDB is to serve as a baseline scenario for medium-term simulation and policy evaluation with the *REMS* model, a large *R*ational *E*xpectations macroeconomic *M*odel for *S*pain. However, given the wide range of variables that make the database, as well as the high degree of homogeneity, consistency and quality of its data, the uses of *R*EMSDB are potentially manifold. Despite the various caveats expressed in this paper about the quality of some data, we therefore believe that REMSDB is a comprehensive source of aggregate data not merely for applied macroeconomic research of the Spanish economy but most importantly for medium-term policy evaluation purposes through the use of rational expectations models.

References

Antolin, P. (1995): "Job Search Behaviour And Unempoyment Benefits In Spain During The Period 1987-1991". *Investigaciones Económicas*, 19(3), 415-433.

Boscá J. E., A. Díaz, R. Doménech, J. Ferri, E. Pérez and L. Puch (2007): "A Rational Expectations Model for Simulation and Policy Evaluation of the Spanish Economy". Documento de trabajo WP-2007-03, Ministerio de Economía y Hacienda.

Chow, G.C., and A. Lin (1971): "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series," *The Review of Economics and Statistics*, 53, 372-375.

Dabán, T., F. J. Escribá, M. J. Murgui and A. Díaz (2002): "La base de datos BD.MORES". *Revista de Economía Aplicada*, 10(30), 165-86.

Díaz, A. (2007): "Obtención de las variables del mercado de trabajo en la ecuación de matching del modelo REMS". Ministerio de Economía y Hacienda. Mimeo.

Dolado, J. J., M. Sebastián and J. Vallés (1993): "Cyclical Patterns of the Spanish Economy," *Investigaciones Económicas*, XVII(3), 445-472.

Fernández, R. B. (1981): "Methodological Note on the Estimation of Time Series," *The Review of Economics and Statistics*, 63, 471-478.

Harvey, A. C. (1989), *Structural Models and the Kalman Filter*, Cambridge, Cambridge University Press.

Lores, F. X. (2001): "Growth And Cyclical Fluctuations In Spanish Macroeconomic Series," Working Paper. 2001-09. Universidad Carlos III.

Puch, L. A. and O. Licandro (1997): "Are There any Special Features in the Spanish Business Cycle," *Investigaciones Económicas*, XXI(2), 361-394.

Appendix

Group 1:

Production and demand variables and their corresponding deflators. Source: INE SQNA unless otherwise specified. Units of measure: Mill. EUR at constant prices (production and demand variables), Mill. EUR at current prices (compensation of employees and net taxes onpProducts) and index numbers with base year 2000 (deflators).

- Gross domestic product at constant market prices
- Gross value added. Non-market services
- Private gross domestic product at constant market prices*[1]
- Private consumption. Households and NPISH
- Government final consumption
- Gross capital formation
- Gross fixed capital formation
- Change in inventories
- Exports of goods and services
- Imports of goods and services
- Capital stock*[2]
- Gross domestic product deflator*[1]
- Private consumption deflator. Households and NPISH*[1]
- Public consumption deflator*[1]
- Exports of goods and services deflator*[1]
- Imports of goods and services deflator*[1]
- Energy demand*[3]
- Energy price index*[4]
- Compensation of employees
- Net taxes products

*[1]Own estimates from INE data. [2]Own estimates. [3]Own estimates from IPI. [4]INE IPRI.

Group 2:

Demographic and labour market series. Source: INE SQNA unless otherwise specified. Units of measure: 1000 persons except for hours worked (1000 hours), vacancies (total amount of vacant posts) and ratio of employment full-time equivalents to total employment.

- Total population*[1]
- Total labour force*[5]
- Population 16+*[2]
- Vacancies*[3]
- Total employment
- Employment full-time equivalents
- Employment full-time equivalents / Total employment
- Employees
- Employees full-time equivalents
- Total employment. Non-market services
- Total employment full-time equivalents. Non-market services
- Unemployment rate*[2]
- Hours worked*[4]

*[1]Population figures provided by INE. [2]LFS. [3]Own estimates. [4]Own estimates from LCI and LFS. [5] Own estimates from Total Employment and the Unemployment rate.

Group 3:

Public Sector variables. Source: IGAE. Units of measure: Mill. EUR at current prices.

- Government final consumption
- Government investment
- Total government social expenditure
- Social expenditure. Rest of categories
- Social expenditure. Unemployment subsidies
- General government net debt interest payments
- Government capital account. Residual category
- Total government receipts
- Government gross operating surplus and gross mixed income
- Net taxes on production and imports. Government
- Income and wealth taxes. Public sector
- Income and wealth taxes. Corporate sector
- Income and wealth taxes. Households
- Income and wealth taxes. Rest of the world
- Social security contributions. Government receipts
- Social security contributions. Employees
- Social security contributions. Self-employed
- Social security contributions. Employers
- Social security contributions. Unemployed workers
- Imputed social security contributions
- Government current account. Residual category
- Government net lending (+)/ net borrowing (-)
- Employees. Public Sector
- Compensation of employees. Public sector

Group 4:

Monetary and financial variables. Source: Bank of Spain. Units of measure: Mill. EUR at current prices (monetary aggregates), % (Interest rates), units of $ per EUR/ECU (Nominal $/EUR exchange rate).

- Monetary aggregate M1. Spain
- Monetary aggregate M2. Spain
- Three-month interest rate. Spain
- Three-month interest rate. US
- Nominal $ / EUR exchange rate

Group 5:

Foreign Sector Variables. Source: Bank of Spain. Units of measure: 1000 EUR at current prices.

- Net financial assets. National concept
- Current transfers with the rest of the world

Acronyms

- INE: Instituto Nacional de Estadística. National Statistical Institute.
- INEM: Instituto Nacional de Empleo. National Employment Institute.
- IPI: Indice de Produccion Industrial. Industrial Production Index.
- IPRI: Indice de Precios Industriales. Industrial Price Index.
- LCI: Labour Cost Index.
- LFS: LFS.
- NPISH: Non-Profit Institutions Serving Households.
- MORES: Regional Model for Spain.
- BDMORES: MORES database.
- REMS: Rational Expectations macroeconomic Model for Spain.
- REMSDB: REMS database.
- SQNA: Spanish Quarterly National Accounts.
- VAT: Value Added Tax.