Ensaio

Evaluating the Forecast Quality of GDP Components: An Application to G7[1]

Pedro M. Esperança[2], Paulo Júlio[3]

1. Introduction

The literature has focused primarily on the quality of forecasts for real Gross Domestic Product (GDP) and for other macroeconomic variables (*e.g.* inflation or unemployment), but has not satisfactorily addressed the forecast accuracy of the major expenditure components of GDP – private consumption (C), government consumption (G), investment (I), exports (X) and imports (M)[4]. In a recent article, Júlio et al. (2011) have analyzed, for the first time, the quality of forecasts for GDP expenditure components. The authors showed that overpredictions in investment and exports explain most of Portuguese GDP overpredictions at 1-year horizons. GDP forecast bias diminishes significantly for same-year predictions, a fact that is mostly explained by canceling out effects in component prediction errors rather than by accurate component predictions. The authors have also proposed two new statistics – Mean of Total Weighted Absolute Error (MTWAE) and Mean of Total Weighted Squared Error (MTWSE) – to objectively evaluate the overall accuracy of component predictions.

This article uses similar techniques to analyze the forecast quality of GDP expenditure components for G7 countries. Three dimensions of forecast quality are addressed here: bias, accuracy, and efficiency. We use forecast data issued by the Organization for Economic Cooperation and Development (OECD) and by the International Monetary Fund (IMF) for the 1993-2010 period, and evaluate both 1-year ahead and same-year predictions. Our focus lies on the overall quality of institutions' forecasts, and thus we pool evaluation statistics across countries in order to obtain an aggregate overview of the main features driving these forecasts. In addition, we propose panel versions of two types of efficiency tests presented in the literature, and analyze the effects of the 2008 crisis on the quality of forecasts[5].

This article is organized as follows. The next section introduces the statistical methodology used to evaluate forecast quality. Section 3 describes the data. Section 4 analyzes the results. Section 5 evaluates the effects of the 2008 crisis on the quality of forecasts. Section 6 summarizes the results of efficiency tests. Section 7 concludes.

2. Methodology

2.1. Notation

Let the subscript j index the country and the subscript t index the forecast period (the period for which the forecast was produced), $j = 1, \ldots, N$ and $t = 1, \ldots, T$. Define $F_{jt}(s)$ as the s-period (or s-step) ahead forecast for the target variable A_{jt}. The variable s is known as the forecast horizon or time span: the number of periods between the production of the forecast $F_{jt}(s)$ and the actual realization A_{jt}. The forecast error – the difference between actual and forecasted values for a given variable in country j – is:

$$e_{jt}(s) = A_{jt} - F_{jt}(s) \quad (1)$$

[1] The opinions expressed in this article represent the views of the authors and do not necessarily correspond to those of the Portuguese Ministry of Economy and Employment.
[2] Currently a postgraduate student at the University of Oxford.
[3] Office for Strategy and Studies, Portuguese Ministry of Economy and Employment. Corresponding author: paulo.julio@gee.min-economia.pt.
[4] In this article, we always refer to real growth rates, even if not explicitly stated.
[5] Our paper (Júlio and Esperança, 2012) also provides an evaluation of the forecast quality at the country level, in addition to a more detailed literature review.

for $j, t, s \in \mathbb{N}_0$. From (1) it is clear that a positive forecast error implies an underestimation, whereas a negative error implies an overestimation, of A_{jt}. Henceforth the forecast horizon s will be suppressed for notational convenience, if not strictly needed.

We analyze 1-, 2-, 3- and 4-step ahead forecasts, so that $s \in \{1,2,3,4\}$: 1-step (3-step) ahead forecasts are those issued on the Autumn of the same (previous) year, and 2-step (4-step) ahead forecasts are those issued on the Spring of the same (previous) year.

2.2. Standard Evaluation Statistics

To evaluate the quality of forecasts, we start with the pooled versions of the standard measures of forecast evaluation. These are termed Pooled Mean Error (PME), Pooled Mean Absolute Error (PMAE), and Root of Pooled Mean Squared Error (RPMSE), and are respectively given by

$$PME := \frac{1}{\mathrm{NT}} \sum_{j=1}^{N} \sum_{t=1}^{T} e_{jt} \qquad (2)$$

$$PMAE := \frac{1}{\mathrm{NT}} \sum_{j=1}^{N} \sum_{t=1}^{T} |e_{jt}| \qquad (3)$$

$$RPMSE := \sqrt{\frac{1}{\mathrm{NT}} \sum_{j=1}^{N} \sum_{t=1}^{T} e_{jt}^2} \qquad (4)$$

PME is the average forecast error (across time and across countries), thus providing a simple measure of central tendency. A negative value means that forecasts overpredict actual values, whereas a positive value indicates an underprediction. PMAE provides a measure of the average total forecast error, regardless of the direction of the error (how much, on average, the forecasts are off-target). Hence, a lower PMAE reflects more accurate forecasts. RPMSE also provides a measure of total forecast error, but attributes disproportionally higher contributions to larger deviations from target. Thus, whereas PME measures how biased forecasts are on average, PMAE and RPMSE evaluate forecast accuracy.

2.3. Scaled Statistics

The previous statistics are only valid when comparing a variable's forecast coming from different institutions or forecasting methods. If one aims to compare the accuracy of institutions' forecasts across a group of variables, these statistics are inadequate, as they do not take into account the intrinsic level of volatility of each series. A more volatile series is naturally harder to predict and thus forecast errors tend to be larger; however, this does not necessarily mean that forecast models perform worse predicting that series. A comparative statistic which addresses this issue can be obtained by scaling each series' errors with the inverse of the corresponding in-sample average absolute difference between the actuals of consecutive periods (a measure of volatility). Let K be the sample size. The scaled errors

$$v_{jt} = e_{jt} \left(\frac{1}{K} \sum_{k=1}^{K} \left| A_{jk} - A_{jk-1} \right| \right)^{-1} = e_{jt} V_j^{-1} \qquad (5)$$

can thus be used in (3), with v_{jt} replacing e_{jt}, to obtain the Pooled Mean Absolute Scaled Error (PMASE). Table 1, which presents the volatility of each series measured by V_j for G7 countries, shows that investment, exports and imports are much more volatile than the remaining series.

Table 1 – Volatility measured by V_j.

	Canada	France	Germany	Italy	Japan	UK	US	Pooled
GDP	1.70	1.30	1.97	1.85	2.20	1.76	1.47	1.75
Priv. Cons.	1.20	0.78	0.79	1.05	1.12	1.45	0.99	1.05
Gov. Cons.	1.13	0.95	1.05	0.99	1.00	1.26	0.93	1.04
Investment	6.63	3.53	4.70	4.28	4.00	6.35	4.02	4.79
Exports	4.87	5.98	6.78	7.33	10.10	5.68	6.02	6.68
Imports	5.96	5.75	5.74	6.97	6.89	5.63	5.68	6.09

2.4. Contributions Analysis

It is also possible to decompose the GDP forecast error into the individual contributions of the corresponding expenditure components. This exercise enables one to identify which components contribute the most to the GDP forecast error and whether errors in forecasted expenditure components tend to add up or to cancel out. Let z_{jt} denote the effective real growth rate of variable Z in country j at year t, and z_{jt}^f the corresponding forecasted real growth rate, $Z = \{GDP, C, G, I, X, M\}$; and define $w_{jt}^Z = Z_{jt}/GDP_{jt}$ – variable Z's share on GDP in country j at t. The effective real GDP growth rate in country j can therefore be decomposed into the corresponding component contributions

$$gdp_{jt} \equiv c_{jt}w_{j,t-1}^C + g_{jt}w_{j,t-1}^G + i_{jt}w_{j,t-1}^I + x_{jt}w_{j,t-1}^X - m_{jt}w_{j,t-1}^M + \epsilon_{jt} \qquad (6)$$

where ϵ_{jt} is a discrepancy term which accounts for the non-additivity of component contributions resulting from chain-linked data[6]. There are two additional discrepancy sources when using forecasted data: neither the weights used by institutions nor the base year for those weights are known. Instead, we use effective weights, and thus[7]

$$gdp_{jt}^f \equiv c_{jt}^f w_{j,t-1}^C + g_{jt}^f w_{j,t-1}^G + i_{jt}^f w_{j,t-1}^I + x_{jt}^f w_{j,t-1}^X - m_{jt}^f w_{j,t-1}^M + \epsilon_{jt}^f \qquad (7)$$

Let e_{jt}^Z denote the forecast error of variable Z's growth rate, *i.e.* $e_{jt}^Z = z_{jt} - z_{jt}^f$, and define the following vectors

$$\boldsymbol{e}_{jt} = (e_{jt}^c, e_{jt}^g, e_{jt}^i, e_{jt}^x, e_{jt}^m)' \text{and} \boldsymbol{w}_{jt} = (w_{jt}^C, w_{jt}^G, w_{jt}^I, w_{jt}^X, w_{jt}^M)'$$

Subtracting (7) from (6) and taking the average across time yields

$$\tfrac{1}{T}\textstyle\sum_{t=1}^T\left(e_{jt}^{gdp} - v_{jt}\right) = \tfrac{1}{T}\sum_{t=1}^T \boldsymbol{e}_{jt}'\boldsymbol{w}_{j,t-1} \qquad (8)$$

where $v_{jt} = \epsilon_{jt} - \epsilon_{jt}^f$. In equation (8), $T^{-1}\sum_{t=1}^T e_{jt}^z w_{j,t-1}^Z$ represents the average contribution of the forecast error arising from variable Z, in percentage points, to the GDP growth forecast error in country j. However, we are instead interested in averaging equation (8) across j in order to evaluate the average contributions of component forecast errors to the average GDP forecast error across countries

$$\tfrac{1}{NT}\textstyle\sum_{j=1}^N\sum_{t=1}^T\left(e_{jt}^{gdp} - v_{jt}\right) = \tfrac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T \boldsymbol{e}_{jt}'\boldsymbol{w}_{j,t-1} \qquad (9)$$

This equation allows one to analyze global trends regarding the decomposition of GDP forecast errors. A negative value in $(NT)^{-1}\sum_{j=1}^N\sum_{t=1}^T e_{jt}^z w_{j,t-1}^Z$ means that the component is overestimated in general, whereas a positive value has the opposite interpretation. It follows from (9) that, even if GDP forecast errors are small on average, this can result from large *cancelling out effects* in component contribution errors.

For this reason, we propose an additional statistic to evaluate the forecast accuracy of component predictions. This statistic – termed Pooled Mean of Total Weighted Absolute Error (PMTWAE) – is an extension to the panel environment of the MTWAE statistic originally proposed in Júlio et al. (2011). It

[6] Chain-linked data is the rule followed by most statistical offices in developed countries. Non-additivity of chain-linked data is explicitly recognized by Statistical Offices when computing contributions to GDP growth.

[7] An additional source of forecast errors may arise from positive or negative contributions of the statistical discrepancy ϵ_{jt} to growth. In practice, institutions have to deal with ϵ_{jt} explicitly in order to obtain GDP growth directly through the sum of component contributions. This can be done by distributing the statistical discrepancy's weight on GDP across components. The term ϵ_{jt} usually affects additivity of component contributions up to the second decimal place, thus having a negligible effect on conclusions.

evaluates the mean of the sum across components of absolute errors, weighted by the corresponding shares on GDP, averaged across all countries. Its purpose is to coherently aggregate individual measures of forecast accuracy for GDP expenditure components, thus evaluating the overall accuracy of component predictions, regardless of the *canceling out effects*. The PMTWAE is defined as

$$PMTWAE := \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\left|\boldsymbol{e}_{jt}\right|'\boldsymbol{w}_{j,t-1} \quad (10)$$

where $\left|\boldsymbol{e}_{jt}\right|$ is a vector whose entries are the absolute values of the entries in $\boldsymbol{e}_t$. The weights reflect the relative importance of each component: those components with higher shares on GDP are naturally more important from the forecaster's point of view and should be weighted heavily. These statistics are computed for each institution and forecast horizon. Those institutions whose forecasts are associated with higher values in these statistics issue less accurate component predictions, even if GDP is accurately forecasted. Naturally, PMTWAE can be decomposed into the corresponding component contributions, $(NT)^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\left|e_{jt}^{z}\right|w_{j,t-1}^{Z}$.

2.5. Efficiency Tests

The previous statistics do not attest whether it would be possible to improve issued forecasts. If one could issue a more accurate forecast with the information currently available, then improvements in quality would be possible. The tests which evaluate this feature are known as efficiency tests (Wallis, 1989; Fildes and Stekler, 2002).

A systematic bias signals that forecasts are either tendencially pessimistic or optimistic, and thus forecast accuracy could be permanently improved by adjusting predictions upwards or downwards, respectively. An unbiased forecast is a necessary condition for "weak informational efficiency". However, it is not sufficient, since efficiency also requires that forecast errors contain only unpredictable effects, *i.e.,* forecast errors cannot contain systematic information that could have been used to improve forecast accuracy. In other words, forecast errors cannot be serially correlated.

In what follows, we propose panel versions of two types of efficiency tests. Panel tests allow for a considerable gain in power as compared to the corresponding time series versions. This is particularly important in this context due to the reduced time series dimension of forecast data.

To test for bias and serial correlation, we start by regressing the forecast errors on a constant and several lagged terms[8]

$$e_{jt} = \gamma_0 + \sum_{l=1}^{p}\gamma_l e_{j,t-l} + \varepsilon_{jt} \quad (11)$$

The residuals are assumed to be serially uncorrelated after p is properly selected, but they may be heteroskedastic and contemporaneously correlated over j, $E(\varepsilon_{jt}\varepsilon_{kt}) = \sigma_{jkt}^2$, $\forall j,k,t$. Notice that forecast errors may be positively correlated across j, since unforeseen changes in GDP or any of its components in a large economy affects macroeconomic aggregates in other countries as well. We do not include individual-specific effects, as they are not supported by the Hausman test.

The model is estimated by OLS and parameter estimates are consistent as long as the underlying process is stationary. This requirement is obviously satisfied, since any disturbance to the forecast error at t should not influence forecasts errors in the long run. Bias is evaluated by performing a z-test on the non-linear hypothesis that $\gamma_0/\left(1-\sum_{l=1}^{p}\gamma_l\right) = 0$. Serial correlation, in turn, is evaluated through a χ^2 test on the null hypothesis that $\gamma_l = 0$, $\forall l$.

Another framework used to test for bias in institutions' forecasts dates back at least to Theil (1966), and is applied for instance in Joutz and Stekler (2000), Loungani (2001) and Vuchelen and Gutierrez (2005). The test (adapted to the panel framework) consists in evaluating whether the coefficients α and β in the following regression

$$A_{jt} = \alpha + \beta F_{jt} + u_{jt}, \quad u_{jt} = \rho u_{j,t-1} + \varepsilon_{jt} \quad (12)$$

[8] A similar framework is presented in Öller and Barot (2000). The authors, however, perform a cross section analysis for each country, thus avoiding some panel data complications.

do not differ significantly from 0 and 1 respectively. We assume again that residuals, ε_{jt}, are heteroskedastic and contemporaneously correlated over j. The Hausman test does not support individual-specific effects, and hence we do not include them in (12).

A drawback of this approach is that serial correlation is not tested, but instead modeled by assuming that u_{jt} follows a common autoregressive process of order 1. Modeling autocorrelation is necessary, since autocorrelated residuals inflate the tests for bias, making any inference invalid.

The two tests for bias are conceptually different. The test resulting from equation (11) evaluates whether forecast errors have zero mean, whereas that resulting from equation (12) evaluates whether a regression line representing unbiased forecasts can fit the data. The conclusions of these tests may differ, for instance, if forecast errors have zero mean, but there is a tendency to overestimate when actual data takes high values and a tendency to underestimate when actual data takes low values. In this case, the former test may not reject the null of unbiasedness, whereas the latter might. Thus, the test resulting from (12) imposes stronger conditions, as it requires that forecast errors exhibit no specific patterns of over or underestimation.

3. Data

Our dataset contains information on forecasts for Gross Domestic Product, Private Consumption, Government Consumption, Investment (namely gross fixed capital formation), Exports and Imports (all in volume percentage change), issued by OECD and IMF for G7 countries: Canada, France, Germany, Italy, Japan, UK, and US. Forecasts were retrieved from OECD's Economic Outlook and from IMF's World Economic Outlook. These institutions issue forecasts twice a year: OECD issues them on May/June and November/December and IMF on April/May and September/October. Although institutions' forecasts are not issued exactly in the same month, they are done in the same time horizon and should thus use similar information sets (Pons, 2000). We thus classify forecasts according to the season in which they are issued. For convenience, forecasts are labeled as 1-, 2-, 3- and 4-step ahead forecasts. Table 2 summarizes the terminology.

Table 2 – Forecast horizon and issue date.

Forecast Period	Forecast Horizon	Issue Date
t	1-step	Autumn t
	2-step	Spring t
	3-step	Autumn t-1
	4-step	Spring t-1

Actual values, which were also used to compute expenditure component shares on GDP, were taken from National Statistical Offices. The period scrutinized is 1993-2010 for same-year forecasts, and 1994-2010 for 1-year ahead forecasts[9]. The choice of the realization is not consensual in the literature, as one should find a compromise between the argument that forecasters do not know the nature of data revisions and the argument that the realization should reflect exact economic outcomes (Vuchelen and Gutierrez, 2005)[10]. First releases do not incorporate all information about economic activity. When information is not available, Statistical Offices use imputation and forecasting methods, and econometric models, to issue an estimate

[9] Prior to 1992, OECD and IMF reported GDP forecasts for some G7 countries and GNP forecasts for others. Only from 1993 onwards forecasted variables were harmonized, with GDP and the corresponding expenditure components being reported for all G7 countries.

[10] For instance, in an often cited article, Keane and Runkle (1990) argue that forecasters aim at predicting first releases, since they do not know in advance the nature of data revisions occurring after the date on which they make their forecasts. Similar arguments are used in Zarnowitz and Braun (1993) and Joutz and Stekler (2000). On the opposite direction, Ash et al. (1998) and Öller and Barot (2000) use data published 6 months and 12 months after the event, respectively, as these are neither flash estimates nor late revisions.

of economic outcomes. Thus, comparing forecasts with first releases or even with intermediate releases is equivalent to compare forecasts with an estimation of economic outcomes. Even though there might exist revisions between first and final releases that forecasters could not be aware of, most revisions derive from the incorporation of new and updated information, whose sources are usually known to forecasters. For this reason, we evaluate forecasts against final releases and not first or intermediate releases.

4. Results

4.1. Gross Domestic Product

Table 3 presents the pooled statistics for GDP growth forecasts. The PME statistic shows that OECD and IMF overestimate GDP growth at 1-year spans – a fact that is explained by overestimations for France, Germany, Italy and Japan. Same-year forecasts are more accurate, although characterized by minor underpredictions. Bias decreases consistently as the forecast horizon shortens (except for the 1-period horizon), suggesting that forecast performance improves as more recent information is incorporated into predictions. Moreover, OECD's and IMF's forecasts display similar biases except at the 3-period span, in which OECD takes a small lead.

Table 3 – Pooled Statistics: GDP.

	OECD				IMF			
	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
PME	-0.49	-0.15	0.09	0.12	-0.46	-0.29	0.12	0.13
RPMSE	2.14	1.61	1.00	0.78	2.13	1.81	1.06	0.82
PMAE	1.60	1.23	0.76	0.61	1.57	1.36	0.82	0.62
PMASE	0.93	0.71	0.44	0.35	0.91	0.79	0.48	0.36

The RPMSE and PMAE statistics point towards a negative relationship between forecast accuracy and the forecast horizon[11]. This fact illustrates the role that new and updated information has on the quality of forecasts. The most accurate 1-year ahead forecasts are issued for France, whereas the least accurate are issued for Germany and Japan. For same-year predictions, GDP forecasts display the highest accuracy in France and the lowest accuracy in Japan and the UK. Forecasts issued by OECD are slightly more accurate than IMF's at the 3-period span, but for other forecast horizons the difference is marginal.

4.2. GDP Expenditure Components

Table 4 presents an evaluation of forecasts for GDP expenditure components. The PME statistic suggests that forecasts for private consumption have the smallest biases at almost all time spans. Government consumption also displays a comparatively small bias and is underestimated at all horizons.

At 1-year spans, the largest biases occur in predictions for investment, exports and imports. Investment and exports are systematically overestimated, and the same holds for imports with OECD's forecasts. On the opposite direction, IMF's forecasts for imports have a comparatively small bias at the 4-period span and are downwardly biased at the 3-period span. For same-year predictions, the largest biases occur in exports and imports, as these components are clearly underestimated by both institutions. Investment is overestimated by IMF at the 2-period span, but in other cases bias is small.

[11] Since PMASE is a rescaling of PMAE, it draws exactly the same conclusions.

Table 4 – Pooled Statistics: GDP expenditure components.

		OECD				IMF			
		4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
C	**PME**	-0.15	0.15	0.14	0.09	-0.19	-0.03	0.18	0.07
	RPMSE	1.53	1.30	1.05	0.80	1.67	1.41	1.09	0.85
	PMAE	1.20	1.01	0.75	0.61	1.28	1.10	0.83	0.62
	PMASE	1.17	0.96	0.67	0.55	1.23	1.07	0.75	0.56
G	**PME**	0.31	0.31	0.21	0.14	0.34	0.24	0.32	0.16
	RPMSE	1.39	1.33	1.33	1.14	1.51	1.55	1.19	1.15
	PMAE	1.07	1.04	0.98	0.81	1.15	1.17	0.95	0.85
	PMASE	1.03	0.99	0.90	0.75	1.11	1.12	0.88	0.80
I	**PME**	-1.32	-0.55	-0.09	0.08	-1.17	-0.88	-0.37	-0.04
	RPMSE	5.45	4.29	3.27	2.57	5.54	4.66	3.36	2.74
	PMAE	3.86	3.04	2.44	1.84	3.99	3.49	2.47	2.01
	PMASE	0.82	0.64	0.51	0.38	0.85	0.75	0.53	0.42
X	**PME**	-1.56	-0.79	0.37	0.75	-0.31	-0.15	0.93	1.08
	RPMSE	7.38	5.84	3.69	2.35	7.35	6.59	4.31	2.89
	PMAE	5.42	4.35	2.83	1.73	5.63	5.00	3.39	2.23
	PMASE	0.83	0.66	0.44	0.28	0.86	0.77	0.53	0.36
M	**PME**	-0.85	-0.18	0.35	0.77	0.11	0.59	0.76	0.92
	RPMSE	6.64	5.64	3.73	2.46	6.55	6.14	4.35	3.13
	PMAE	5.10	4.32	2.87	1.78	5.10	4.92	3.30	2.26
	PMASE	0.84	0.71	0.47	0.30	0.84	0.81	0.55	0.37

The RPMSE and PMAE statistics suggest that forecasts for investment, exports and imports have the lowest accuracy. However, this does not imply that forecast models perform comparatively worse predicting these components, as they are also more volatile. By scaling the errors with the inverse of the volatility of each series, PMASE becomes more appropriate to make inferences about the predictive quality of institutions' forecast models across GDP components.

PMASE demonstrates that forecast models perform comparatively worse when predicting private consumption and government consumption at all horizons. Possibly, institutions' forecast models cannot accurately predict revisions in the consumption bundle carried out by households when new macroeconomic information becomes available. Government consumption is a policy-making tool, often facing unexpected increases, particularly in election years. The performance of forecast models is similar across the remaining GDP components – investment, exports and imports.

4.3. Decomposing GDP Forecast Errors into Component Contributions

Pooled contributions of expenditure components to the average GDP forecast error are detailed in Table 5. Notice that average discrepancies, υ, originating from the non-additivity of component contributions and from the difference between actual component shares on GDP and the shares used by institutions in forecast models, are small.

Table 5 – Pooled contributions of expenditure components (in percentage points) to the average GDP forecast error.

	OECD				IMF			
	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
C	-0.08	0.10	0.08	0.06	-0.10	-0.01	0.11	0.04
G	0.05	0.05	0.03	0.02	0.06	0.04	0.06	0.03
I	-0.27	-0.11	-0.02	0.03	-0.25	-0.18	-0.07	0.00
X	-0.45	-0.26	0.06	0.17	-0.17	-0.04	0.20	0.22
M	-0.30	-0.13	0.07	0.18	-0.04	0.08	0.18	0.20
υ	-0.04	-0.06	0.00	0.02	-0.04	-0.02	0.02	0.04
GDP	-0.49	-0.15	0.09	0.12	-0.46	-0.29	0.12	0.13

Notes: *(i)* GDP in the table represents the mean (across time and across countries) of e_{jt}^{gdp}, and equals the Pooled Mean Error statistic in Table 3, by definition; *(ii)* The sum of the contributions in the table differs from GDP by an error υ whose source is explained in Section 2.4.

Institutions overestimate private consumption (with the exception of OECD's 3-step ahead forecasts) and underestimate government consumption at 1-year spans. Nevertheless, the contributions of these components to the GDP forecast error are small relative to other components. The remaining components are overestimated at 1-year spans (with the exception of IMF's 3-step ahead forecasts for imports). At the 4-period span, around 85% of the GDP forecast error is explained by investment and net exports. However, despite net export's similar contribution (-0.15 for OECD and -0.13 for IMF) to the GDP forecast error, OECD overestimates exports and imports by a larger magnitude on average. At the 3-period span, component contributions are smaller, leading to a less biased GDP forecast relative to the 4-period span. Overestimations in investment and net exports still explain the largest fraction of GDP overpredictions in this case. For OECD's forecasts, exports present the largest contribution to the GDP forecast error, whereas for IMF's forecasts the largest contribution comes from investment. Specific conclusions vary across countries: whereas component contribution errors for France, Germany, Italy and Japan tend to add up, contributing to larger biases in forecasted GDP, for Canada, the UK and the US they tend to cancel out, resulting in smaller biases.

Contributions of private consumption, government consumption and investment to the GDP forecast error are small at same-year spans, when compared with other time spans or components. The largest contributions are displayed by exports and imports, both underestimated. However, the effects of these components tend to cancel out, as imports enter the GDP equation with a negative sign. This results into relatively accurate GDP predictions. The cancel out effect is stronger for Canada, Germany and Japan. For France, Italy and the US, GDP forecasts are obtained with relatively unbiased component predictions. On the opposite direction, UK's component forecast errors tend to add up, originating a significant GDP overestimation as compared with other countries.

The PMTWAE statistic, presented in Table 6, summarizes the overall accuracy of component predictions. A lower value means that GDP forecasts are assembled with more accurate component predictions, whereas a higher value has the opposite interpretation. At 1-year horizons, PMTWAE fluctuates between 3.7 (OECD's forecasts at the 3-period span) and 4.6 (IMF's forecasts at the 4-period span) percentage points. The components which most significantly contribute to this outcome (*i.e.*, whose predictions, weighted by the component's share on GDP, are least accurate) are, by descending order, exports, imports and investment. On the opposite direction, the contribution of government consumption to the statistic is marginal. This ordering is highly correlated with the volatility of the variables. For same-year spans, the overall accuracy of component predictions increases, explaining more accurate GDP forecasts. The decomposition of PMTWAE leads to similar conclusions as for 1-year spans. The highest overall accuracy in component predictions is achieved for the US and Japan, whereas the poorest performance is attained for Canada and Germany. This diversity is explained by the different accuracy levels of exports and imports across countries.

Table 6 – PMTWAE statistic and its decomposition (in percentage points).

	OECD				IMF			
	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
PMTWAE	4.46	3.69	2.61	1.89	4.60	4.18	2.93	2.14
	Component contributions to PMTWAE							
C	0.73	0.61	0.45	0.36	0.78	0.67	0.50	0.37
G	0.21	0.20	0.19	0.16	0.23	0.23	0.19	0.17
I	0.97	0.81	0.62	0.53	1.01	0.90	0.64	0.55
X	1.34	1.05	0.68	0.41	1.37	1.21	0.81	0.52
M	1.21	1.01	0.67	0.42	1.21	1.17	0.78	0.52

It is also evident from Table 6 that OECD's component forecasts are more accurate than IMF's component forecasts, even though the latter displays smaller biases for some components. This is explained by cancel out effects that occur across years, which are not captured by the first moment: positive errors in some years are canceling out negative errors in other years, leading to lower biases, but also to lower accuracy. These effects are more acute for IMF's forecasts.

5. The Effects of the 2008 Crisis on the Quality of Forecasts

To evaluate the effects of the 2008 crisis on the quality of forecasts, we compute the same measures of forecast quality until 2007, and analyze how these have changed relatively to the complete time period. Results are displayed in Table 7[12].

In general, the crisis contributed to increase bias (evaluated by PME), particularly at 1-year spans. However, this conclusion does not hold for all variables. Prior to 2008, 4-period ahead forecasts for GDP were downwardly biased, and the crisis strengthened this bias. OECD's 3-step ahead forecasts for GDP were upwardly biased, and IMF's 3-step ahead forecasts nearly unbiased. In the former case the crisis changed the sign of the bias, but presented no relevant effect on its absolute magnitude, whereas in the latter case the crisis originated a bias of around -0.3 percentage points. Private consumption became slightly biased (downwards) at the 4-step span, but bias decreased at the 3-period span as a result of the crisis. Government consumption remained overestimated as before the crisis. Different conclusions hold for the remaining GDP components, with PME changing between -0.8 and -0.5 percentage points at 1-year spans. Investment was overestimated before 2008, and the crisis contributed to foster this tendency. Exports and imports were underestimated by IMF and overestimated by OECD at 1-year horizons prior to 2008 (with the exception of OECD's 3-period ahead forecasts for imports). For these components, the crisis led to a substantial increase in bias for OECD's forecasts (*i.e.* overestimation increased), but to a decrease in bias for IMF's forecasts (*i.e.* underestimation decreased). The exception is OECD's 3-step ahead forecasts for imports, for which bias was reduced.

The crisis also led to a substantial decrease in the accuracy of 1-year ahead predictions, but not of same-year predictions. The decrease in accuracy (measured by PMAE) was higher for more volatile components – investment, exports and imports – and affected OECD's and IMF's forecasts. The accuracy of forecasts for private consumption and government consumption were only marginally affected. Somewhat surprisingly, PMASE shows that the quality of institutions' prediction models increased in recent years, albeit marginally. The opposing signs displayed by the changes in PMAE and PMASE are due to large increases in volatility after 2008. Thus, the fall in accuracy after the triggering of the crisis is explained by an increase in uncertainty, rather than by a decline in the quality of forecast models.

[12] For brevity, we only present the major statistics.

Table 7 – Effects of the 2008 financial crisis: GDP and GDP expenditure components.

		OECD				IMF			
		4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
GDP	PME	-0.16	0.15	0.18	0.22	-0.17	0.01	0.20	0.23
	ΔPME	-0.33	-0.30	-0.09	-0.10	-0.29	-0.30	-0.08	-0.10
	ΔPMAE	0.41	0.22	0.03	0.01	0.42	0.31	0.01	0.01
	ΔPMASE	-0.14	-0.19	-0.19	-0.17	-0.12	-0.16	-0.22	-0.16
C	PME	0.04	0.32	0.25	0.19	-0.03	0.13	0.24	0.14
	ΔPME	-0.19	-0.17	-0.11	-0.10	-0.16	-0.16	-0.06	-0.08
	ΔPMAE	0.19	0.08	0.05	0.03	0.18	0.11	0.03	0.03
	ΔPMASE	-0.08	-0.18	-0.11	-0.09	-0.11	-0.14	-0.15	-0.10
G	PME	0.35	0.36	0.25	0.17	0.36	0.28	0.40	0.23
	ΔPME	-0.04	-0.05	-0.04	-0.03	-0.01	-0.05	-0.08	-0.07
	ΔPMAE	-0.04	-0.02	0.00	-0.02	0.01	0.00	-0.03	-0.02
	ΔPMASE	-0.05	-0.04	-0.01	-0.03	0.00	-0.01	-0.04	-0.03
I	PME	-0.58	-0.03	-0.09	0.17	-0.53	-0.35	-0.36	-0.02
	ΔPME	-0.73	-0.52	0.00	-0.08	-0.64	-0.52	-0.01	-0.02
	ΔPMAE	0.87	0.59	0.09	0.04	1.00	0.73	0.17	0.11
	ΔPMASE	-0.09	-0.10	-0.19	-0.16	-0.07	-0.10	-0.16	-0.15
X	PME	-0.98	-0.08	0.21	0.94	0.37	0.83	0.91	1.26
	ΔPME	-0.59	-0.71	0.15	-0.19	-0.68	-0.68	0.02	-0.18
	ΔPMAE	1.40	0.83	0.17	-0.03	1.22	0.97	0.11	0.00
	ΔPMASE	-0.03	-0.08	-0.14	-0.11	-0.08	-0.10	-0.19	-0.13
M	PME	-0.22	0.56	0.31	0.90	0.85	1.20	0.71	0.90
	ΔPME	-0.63	-0.74	0.05	-0.14	-0.74	-0.61	0.05	0.02
	ΔPMAE	1.20	0.74	0.00	-0.06	1.17	0.88	0.06	0.00
	ΔPMASE	-0.02	-0.08	-0.15	-0.10	-0.04	-0.09	-0.17	-0.11

Notes: PME is the value of the Mean Error for the 1993-2007 period. Δ corresponds to the difference in the value of the statistic between the 1993-2010 period and the 1993-2007 period. Thus, a positive (negative) value means that the financial crisis originated a positive (negative) change in the value of the statistic.

Table 8 – Effects of the 2008 financial crisis: pooled contributions of expenditure components to the average GDP forecast error.

	OECD				IMF			
	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut t–1	2–step Spr t	1–step Aut. t
	Average contribution until 2007							
C	0.04	0.20	0.15	0.11	0.00	0.10	0.15	0.09
G	0.06	0.06	0.04	0.03	0.06	0.05	0.07	0.04
I	-0.09	0.02	0.01	0.07	-0.08	-0.05	-0.05	0.03
X	-0.26	-0.05	0.02	0.22	0.04	0.15	0.19	0.25
M	-0.12	0.07	0.05	0.20	0.14	0.23	0.15	0.19
u	-0.03	-0.01	0.00	0.00	-0.04	0.00	0.00	0.01
GDP	-0.16	0.15	0.18	0.22	-0.17	0.01	0.20	0.23
	Changes in contribution relative to the complete period							
ΔC	-0.12	-0.10	-0.07	-0.05	-0.10	-0.11	-0.04	-0.05
ΔG	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.01
ΔI	-0.18	-0.13	-0.03	-0.04	-0.17	-0.13	-0.02	-0.03
ΔX	-0.19	-0.21	0.04	-0.05	-0.21	-0.19	0.01	-0.03
ΔM	-0.18	-0.20	0.02	-0.02	-0.18	-0.15	0.03	0.01
Δu	-0.01	-0.05	0.00	0.02	0.00	-0.02	0.02	0.03
ΔGDP	-0.33	-0.30	-0.09	-0.10	-0.29	-0.30	-0.08	-0.10

Notes: Same as in Table 5.

Table 8 displays the contributions of expenditure components to the average GDP forecast error prior to the crisis. When compared with the complete time period, contributions are substantially smaller, mainly those from investment, exports and imports. The crisis also led to substantial decreases in the overall accuracy of component predictions, as shown by the change in PMTWAE in Table 9, particularly at 1-year spans. Forecasts for investment, exports and imports were the prime sources of this result.

Table 9 – Effects of the 2008 financial crisis: PMTWAE statistic and its decomposition.

	OECD				IMF			
	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
	PMTWAE and component contributions until 2007							
PMTWAE	3.36	3.00	2.43	1.81	3.54	3.37	2.74	2.02
C	0.61	0.56	0.42	0.34	0.66	0.60	0.49	0.35
G	0.22	0.21	0.19	0.17	0.23	0.23	0.19	0.17
I	0.75	0.65	0.55	0.48	0.77	0.71	0.57	0.48
X	0.94	0.81	0.63	0.42	1.02	0.94	0.77	0.51
M	0.84	0.77	0.64	0.41	0.86	0.89	0.73	0.51
	Changes relative to the complete period							
ΔPMTWAE	1.10	0.69	0.18	0.08	1.06	0.81	0.19	0.12
ΔC	0.12	0.05	0.03	0.02	0.12	0.07	0.01	0.02
ΔG	-0.01	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00
ΔI	0.22	0.16	0.07	0.05	0.24	0.19	0.07	0.07
ΔX	0.40	0.24	0.05	-0.01	0.35	0.27	0.04	0.01
ΔM	0.37	0.24	0.03	0.01	0.35	0.28	0.05	0.01

6. Testing for "Weak Informational Efficiency"

As shown in Section 5, the 2008 crisis led to a significant decrease in the quality of forecasts. The large forecast errors for the 2008-2010 period constitute atypical (influential) observations, which greatly affect OLS estimates in regressions (11) and (12). As such, we restrict the tests for "weak informational efficiency" to the subsample period 1993-2007[13].

Table 10 presents the efficiency tests for the model in equation (11). We included only one lagged term, since there was no evidence of higher order serial correlation. Recall that forecasts are efficient in this context if and only if they are unbiased and serially uncorrelated. The former requires that $\gamma_0/(1-\gamma_1)=0$, whereas the latter imposes $\gamma_1=0$. Evidence suggests that forecasts are, in general, unbiased. The main exceptions are forecasts for government consumption at several time spans, and 1-step ahead forecasts for imports and exports. Serial correlation characterizes forecasts for private consumption and government consumption at all horizons, and forecasts for GDP at most time spans. Investment forecasts are also serially correlated in some cases.

An alternative and more robust test for bias, using model (12), is displayed in Table 11. The χ^2 test on the joint hypothesis α = 0 and β = 1 indicates that forecasts are in general inefficient. In particular, the null hypothesis of efficiency is rejected for government consumption at all time spans (with 1 exception), and for GDP at all but the 2-period span. One also rejects the null hypothesis of efficiency at 1-year spans for private consumption, exports and imports (with 1 exception). Results for GDP are consistent, for instance, with those in Loungani (2001), who finds that Consensus Forecasts are biased for 1-year ahead predictions.

[13] In what follows, we use a 5% significance level unless otherwise stated.

Table 10 – Tests for “weak informational efficiency” (1993-2007 period): Model in equation (11).

		OECD				IMF			
		4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
GDP	$\gamma_0/(1-\gamma_1)$	-0.21	0.08	0.20	0.23 ***	-0.18	-0.08	0.26	0.25 **
		(0.40)	(0.27)	(0.19)	(0.09)	(0.43)	(0.35)	(0.20)	(0.12)
	γ_1	0.35 **	0.18	0.32 **	0.28 **	0.43 ***	0.35 **	0.20	0.33 ***
		(0.14)	(0.15)	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.12)
C	$\gamma_0/(1-\gamma_1)$	-0.02	0.23	0.29 *	0.18	0.03	0.05	0.28	0.15
		(0.37)	(0.22)	(0.15)	(0.11)	(0.40)	(0.29)	(0.20)	(0.12)
	γ_1	0.57 ***	0.30 **	0.28 **	0.35 ***	0.66 ***	0.51 ***	0.27 **	0.33 ***
		(0.12)	(0.13)	(0.13)	(0.12)	(0.11)	(0.12)	(0.13)	(0.13)
G	$\gamma_0/(1-\gamma_1)$	0.46 **	0.41 **	0.27	0.18	0.46 **	0.32	0.45 ***	0.21 **
		(0.23)	(0.18)	(0.19)	(0.15)	(0.19)	(0.22)	(0.14)	(0.10)
	γ_1	0.42 ***	0.42 ***	0.39 ***	0.30 ***	0.28 **	0.33 ***	0.29 ***	0.28 **
		(0.10)	(0.11)	(0.10)	(0.11)	(0.11)	(0.12)	(0.11)	(0.11)
I	$\gamma_0/(1-\gamma_1)$	-0.52	-0.03	0.15	0.37	-0.52	-0.41	-0.14	0.22
		(0.73)	(0.53)	(0.41)	(0.33)	(0.65)	(0.63)	(0.32)	(0.34)
	γ_1	0.29 **	0.15	0.06	0.15	0.30 **	0.28 **	0.04	0.18
		(0.13)	(0.14)	(0.13)	(0.13)	(0.13)	(0.14)	(0.12)	(0.11)
X	$\gamma_0/(1-\gamma_1)$	-0.34	-0.34	0.23	0.86 ***	-0.04	0.44	1.00	1.20 **
		(1.16)	(0.75)	(0.70)	(0.32)	(1.27)	(1.10)	(0.67)	(0.49)
	γ_1	0.05	-0.19	0.05	0.22	0.15	0.10	-0.11	0.24 *
		(0.18)	(0.16)	(0.15)	(0.15)	(0.17)	(0.17)	(0.15)	(0.13)
M	$\gamma_0/(1-\gamma_1)$	-0.06	0.32	0.53	0.90 ***	0.49	0.87	1.00	1.15 ***
		(1.15)	(0.80)	(0.57)	(0.33)	(1.23)	(1.09)	(0.65)	(0.44)
	γ_1	0.15	-0.06	-0.12	0.27 **	0.23	0.12	-0.12	0.16
		(0.17)	(0.16)	(0.13)	(0.14)	(0.17)	(0.17)	(0.14)	(0.11)
Observations		91	91	98	98	91	91	98	98

Notes: *(i)* Efficiency tests for the model $e_{jt} = \gamma_0 + \gamma_1 e_{j,t-1} + \varepsilon_{jt}$; *(ii)* Bias is evaluated by testing the null hypothesis $\gamma_0/(1-\gamma_1) = 0$, whereas serial correlation is evaluated by testing $\gamma_1 = 0$; *(iii)* Higher orders of serial dependence were insignificant and thus not included in the final specification; *(iv)* Panel-corrected standard errors in parenthesis; *(v)* *, ** and *** represent rejections at 10, 5 and 1 percent significance levels, respectively.

Table 11 – Tests for “weak informational efficiency” (1993-2007 period): Model in equation (12).

		OECD				IMF			
		4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t	4–step Spr. t–1	3–step Aut. t–1	2–step Spr. t	1–step Aut. t
GDP	χ^2 test	24.97	7.58	3.18	11.54	17.84	7.63	2.74	7.97
	p-value	0.00	0.02	0.20	0.00	0.00	0.02	0.25	0.02
C	χ^2 test	12.66	10.95	3.84	4.90	21.85	10.58	2.53	2.05
	p-value	0.00	0.00	0.15	0.09	0.00	0.01	0.28	0.36
G	χ^2 test	16.13	12.04	12.14	5.78	34.49	41.81	13.99	15.86
	p-value	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
I	χ^2 test	12.28	3.14	0.06	6.14	7.98	1.29	0.97	2.85
	p-value	0.00	0.21	0.97	0.05	0.02	0.53	0.61	0.24
X	χ^2 test	11.12	1.80	0.13	15.99	21.09	6.34	2.11	9.58
	p-value	0.00	0.41	0.94	0.00	0.00	0.04	0.35	0.01
M	χ^2 test	22.85	10.99	0.39	12.63	16.07	19.59	1.75	3.64
	p-value	0.00	0.00	0.82	0.00	0.00	0.00	0.42	0.16
Observations		98	98	105	105	98	98	105	105

Notes: *(i)* Efficiency tests for the model $A_{jt} = \alpha + \beta F_{jt} + u_{jt},\ u_{jt} = \rho u_{j,t-1} + \varepsilon_{jt}$; *(ii)* The χ^2 test evaluates the null hypothesis that $\alpha = 0$ and $\beta = 1$ in the equation above.

There are several reasons which explain inefficiency in forecasts for government consumption. First, this variable is often used by policy-makers to manipulate the economic cycle and to boost GDP. These changes in policy are often unexpected and difficult to predict, even in the short run. Second, the supply of public goods is often chosen by bureaucrats, whose decisions may be driven by self-interests, such as power or reputation, rather than by an optimal allocation rule (*i.e.* Samuelson rule). This may originate an unpredictable increase in government expenditure. Third, legislators and officials may grant numerous favors to interest groups in response to rent-seeking efforts. Several of these favors may lead to a higher government size than previously anticipated. Finally, policy-makers may deliberately increase public expenditures in election years beyond the budgeted, in order to impress voters with an increase in the provision of public goods or publicly-financed goods.

7. Conclusions

This article analyzed the quality of forecasts for real GDP growth and for the corresponding expenditure components in the G7. We concluded that forecast accuracy is lower for investment, exports and imports, but forecast models perform comparatively worse predicting private consumption and government consumption. This is explained by different volatility levels of GDP components. At 1-year horizons, GDP overpredictions are mostly explained by investment and net exports. GDP forecast bias diminishes substantially at same-year horizons, a fact that is explained by canceling out effects in component prediction errors – mainly exports and imports – rather than by accurate component predictions.

The overall accuracy of component predictions is evaluated through the extension of a recently proposed statistic to the panel environment: the Pooled Mean of Total Weighted Absolute Errors. This statistic suggests that the accuracy of component predictions is substantially low, meaning that GDP forecasts are assembled with rather inaccurate component predictions. Investment, exports and imports are the major contributors to this outcome. The Pooled Mean of Total Weighted Absolute Errors also suggests that OECD's forecasts for GDP components are more accurate than IMF's forecasts.

We also show that forecasts are in general inefficient, both for GDP and its components, but inefficiency is more acute in government consumption predictions, which are both biased and serially correlated. The 2008 crisis had a large negative effect on the quality of forecasts – mainly for investment, exports and imports – but this was due to an increase in volatility rather than to a decrease in the performance of forecast models.

References

Ash, J., D. Smyth, and S. Heravi (1998). *Are OECD forecasts rational and useful? A directional analysis,* International Journal of Forecasting, 14, 381-391.

Fildes, R. and H. O. Stekler (2002). *The state of macroeconomic forecasting*, Journal of Macroeconomics, 24, 435-468.

Joutz, F. and H. O. Stekler (2000): *An evaluation of the predictions of the Federal Reserve,* International Journal of Forecasting, 16, 17-38.

Júlio, P. and P.M. Esperança (2012): *Evaluating the forecast quality of GDP components: An application to G7,* GEE Papers No. 47, Office for Strategy and Studies, Portuguese Ministry of Economy and Employment.

Júlio, P., Esperança, P. M. and J. C. Fonseca (2011): *Evaluating the forecast quality of GDP components,* GEE Papers No. 41, Office for Strategy and Studies, Portuguese Ministry of Economy and Employment.

Keane, M. and D. Runkle (1990): *Testing the rationality of price forecasts: new evidence from panel data,* American Economic Review, 80, 714-735.

Loungani, P. (2001). *How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth,* International Journal of Forecasting, 17, 419-432.

Öller, L. E. and B. Barot (2000). *The accuracy of European growth and inflation forecasts,* International Journal of Forecasting, 16, 293-315.

Pons, J. (2000). *The accuracy of IMF and OECD forecasts for G7 countries,* Journal of Forecasting, 19, 53-63.

Theil, H. (1966). Applied economic forecasting, Amsterdam: North-Holland Pub. Co.

Vuchelen, J. and M. I. Gutierrez (2005). *A direct test of the information content of the OECD growth forecasts,* International Journal of Forecasting, 21, 103-117.

Wallis, K. F. (1989). *Macroeconomic forecasting: a survey,* Economic Journal, 99, 28-61.

Zarnowitz, V. and P. Braun (1993): *Twenty-two years of the NBER-ASA Quarterly Economic Outlook Surveys: aspects and comparisons of forecasting performance,* in Business Cycles, Indicators and Forecasting, National Bureau of Economic Research, NBER Chapters, 11-94.