

European Data Watch

This section offers descriptions as well as discussions of data sources that are of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We focus on data from German speaking countries that allow international comparative research. While most of the data are at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) are included as well. Suggestions for data sources to be described in future columns (or comments on past columns) should be sent to: Joachim Wagner, Leuphana University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to (wagner@leuphana.de). Past “European Data Watch” articles can be downloaded free of charge from the homepage of the German Council for Social and Economic Data (RatSWD) at: <http://www.ratswd.de>.

The Bundesbank’s Macroeconomic Real-time Database for the German Economy (Gerda)

By Thomas A. Knetsch*

1. Introduction

Many economic data are subject to revision, which means that the results of economic indicators which statistical agencies publish for the same reporting period may change over time. The main reason for revisions is that early announcements are measured on the basis of incomplete statistical informa-

* The real-time database project has been the joint work of the Economics Department, the Statistics Department and the Economic Research Centre of the Deutsche Bundesbank. In particular, the input of Irene Fenzlein, Julia Greb, Gerhard Juracka, Mirjam Mathias, Matthias Klimpel and Reinhold Stahl is gratefully acknowledged. The work has benefited from comments and suggestions by Heinz Herrmann, Robert Kirchner and Patricia Staab. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Deutsche Bundesbank.

tion. However, revisions may also occur when measurement concepts and practices are altered. The time series of economic indicators obviously lack information about past revisions, as they reflect the latest stage taken from available statistical sources. Real-time databases keep this information by storing all issues that were published in the past. In other words, they collect the time-series vintages of economic indicators, each of them identified by its publication date.

Real-time databases make it possible to analyze how prone economic indicators are to revision and to track the “revision history” of the results of individual periods, for example. It also allows a precise reconstruction of the information base at given points in time in the past. With either function, real-time data can be used, for instance, to study the revision processes of economic indicators, to check the performance of forecasting tools under conditions closely conforming to reality, and to evaluate past economic policy strategies and decisions given policymakers’ information of that time.

In recent years, macroeconomic real-time databases have been established for a number of industrialized countries. Although they differ with regard to concept and coverage, the major examples share the characteristic of being maintained by a central bank or an international institution, at least if they are sufficiently large in scope. In the 1990s, the Federal Reserve Bank of Philadelphia started reconstructing historical vintages for key macro variables of the U.S. economy (Croushore/Stark, 2001), creating a prototype which has been extensively used by the academic community since then. The Bank of England publishes real-time data for the United Kingdom (Garratt/Vahey, 2006). Real-time data covering a number of European countries are presented on the member-restricted area of the website of the Euro Area Business Cycle Network (EABCN).¹ The OECD also maintains a multi-country real-time database.

In August 2009, the Deutsche Bundesbank made its macroeconomic real-time database Gerda available online (Deutsche Bundesbank, 2009). The German version of the application can be accessed at http://www.bundesbank.de/statistik/statistik_realtime.php and the English version at http://www.bundesbank.de/statistik/statistik_realtime.en.php. Containing historical vintages of more than 280 economic indicators,² Gerda is the first general and comprehensive real-time database for the German economy.³ It is permanently

¹ The EABCN is a research network of European central banks.

² The indicator counter on the website shows a total number of more than 550 indicators. This discrepancy is explained by the fact that, in the website count, indicators are defined as being separate from each other if they possess different identifiers which follows a formal statistical definition.

³ Real-time data of German economic indicators were collected and published earlier (Gerberding et al., 2005; Jacobs/Sturm, 2005, for instance). In contrast to the

up-to-date, as newly incoming data are stored immediately after their publication. At the time of its launch, the database contained roughly 18,000 vintages. Owing to the current updating procedure, the content is growing rather rapidly, with an add-on of about 1,000 vintages per quarter. National accounts data at 25 percent and short-term business indicators at 55 percent constitute the overwhelming part of the vintages stored in the database. Variables stemming from these two accounting systems usually attract most interest in revision analysis and economic studies using real-time data. The remaining part of the database is equally shared by vintages of short-term labour market indicators and price indices.⁴

The remainder of this report is organized as follows. The data content is presented in Section 2. An brief overview on codification, search functions as well as view and download options is given in Section 3. Finally, Section 4 concludes.

2. Data Content

Gerda merges historical vintages of revision-prone macroeconomic indicators stemming from separate accounting systems. The database currently comprises national accounts data, short-term business cycle and labour market indicators as well as price indices. During the initial collection process, the database was expanding simultaneously in two directions. First, incoming releases of the comprehensive list of economic indicators were continuously recorded by the computerized updating procedure which was established in 2005 (and is still in action to fill the database at the current end). Second, historical vintages of a limited set of key variables were reconstructed on the basis of printed or electronic documents.

The remainder of this section presents an overview of the content of the real-time database including a description of current publication and revision practices. The outline is organized according to the accounting systems included in the real-time database.

new database, its predecessors typically referred to specific research projects, meaning that the variables, for which real-time data were collected, were chosen selectively. Thus, such databases are rather limited in scope and have not been kept up to date, as collection stopped after the project had been finalized. A real-time dataset of German national accounts is maintained by the Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI, 2007).

⁴ The mentioned proportions do not refer to the current stock but to the quarterly increments. As the scope of the database increases dynamically, actual shares will converge to the shares in increments.

2.1 National Accounts

Approximately one-third of Gerda's content is data from the national accounts. The updating procedure took off in May 2005 when the volumes were first published as chain-weighted indices. Since then, the quarterly national accounts statistics are widely reproduced in the real-time database, including the central economic aggregates such as gross domestic product (GDP) and gross national income, the main expenditure components such as private and government consumption, gross fixed capital formation as well as export and imports, the sectoral breakdown of gross value added, and the distribution of national income. Conditional on public availability, these economic entities are found in nominal and real terms as well as with and without seasonal and working-day adjustment. The number and hours worked of persons employed and employees are also provided – both as total-economy aggregates and in sectoral breakdown. The same structure of data is available for the compensation of employees as well as gross wages and salaries, while net wages and salaries are given only as total-economy aggregates.

The set of economic indicators, for which vintages have been recorded manually, is less comprehensive, with the scope of retrograde recording differing from one indicator to another. Moreover, availability is restricted to seasonal and working-day adjusted data. A natural limit for the vintage dimension is given by the year 1995 when, for the first time (namely in September 1995), the national accounts were published for Germany as a whole. Historical vintages of GDP and the main expenditure aggregates are available in real terms, seasonally and working-day adjusted, back to this date.

The current publication practice is to present a first release of quarterly GDP including the total-economy figures of persons employed and employees about six weeks after the end of the reporting period, while the complete national accounts follow approximately ten days later. Just one or two days before the latter date, however, the number and the hours worked of the persons employed as well as the hours worked of employees are released in a sectoral breakdown and, if this has not already happened, also as total-economy aggregates. This multi-segmented schedule was established in May 2003, while there had been only a single publication date prior to this. Starting with the year 2000, the publication lag of the first release was reduced by a fortnight. Traditionally, the first release of the annual national accounts is presented in January, just a few days after the end of the reporting period.

In general, the national accounts of the current and the past three years may be revised at every publication date, as they possess a provisional status. In terms of revision practice, however, the summer release is most important because a comprehensive update of the preliminary data is regularly presented at that date as a consequence of incorporating lagging information about, for instance, the value-added tax statistics and the cost structures of firms. The data

which dates back about three years is thereby given a final status. Apart from the summer revision, there might also be revisions at the remaining regular publication dates during the year. These are typically small in magnitude, however, and refer to very recent data only.

2.2 Short-term Business Indicators

About one-half of the time series collected in Gerda belongs to the group of short-term business indicators which give a timely and detailed view of the cyclical stance of the production and trade sectors at monthly frequency. The set of indicators group into production, orders received and turnover indices. Production is measured in real terms, orders received in real and nominal terms. As far as the production sector is concerned, turnover is available only in nominal terms.

The production sector is divided into industry, energy and construction. The industrial sector is further broken down into the main industrial groups, i.e. producers of intermediate goods, producers of capital goods and producers of durable and non-durable consumer goods. Construction indices capture only the main construction industry, which is subdivided into structural and civil engineering. This breakdown is not only found in the output measures but also for the orders received and the turnover indicators. For the industrial sector, the latter two categories, however, are divided into flows from the domestic and the foreign market, with the orders and the turnover from abroad being further differentiated into intra and extra EMU. Orders and turnover of the construction sector are reported separately for housing and in a breakdown by the type of clients, i.e. industrial clients and public sector clients.

The automatic updating procedure was established in November 2005. From this date onwards, the historical vintages of the above-mentioned indicators are available. An exception is the intra/extra EMU breakdown, which was released for the first time in March 2006. The manual recording of vintages published prior to November 2005 focused on key indicators only. This lean set comprises the production indices for industry (or the manufacturing sector before September 2001), energy and construction as well as the main industrial groups. Orders are reported for industry as a whole and in the broad sectoral and regional breakdowns.

The short-term business statistics of the production sector obey a revision scheme which consists of a quarterly and an annual revision. This was established in March 1999, when a new survey method was introduced with the aim of relieving the enterprises' statistical workload. Preliminary figures are released regularly about five weeks after the end of the reference period. The quarterly revision occurs roughly two and a half months after the end of the reporting quarter when the information from the monthly and the quarterly

reports are collated. Indices are given final status in the annual revision, which coincides with the quarterly report of the fourth quarter.

Output indices of the main construction industry are revised regularly when the annual survey of the complete construction sector is available. The revision usually occurs at the beginning of a year and refers to the data of the recent summer semester. The orders received by the main construction industry are published about two weeks later than their industrial counterparts. The raw indices, however, are immediately given final status. Since May 2005, preliminary announcements of construction output have been adjusted in advance by a factor which represents the expected correction of the total annual survey.

An important part of short-term business statistics are monthly indices of the turnover in the trade sector, where the focus is on retail trade and trade with motor vehicles. The indicators are generally available in current and constant prices as well as with and without seasonal and working-day adjustment. Vintages have been stored from November 2005 onwards. Retail turnover indices are published regularly twice a month. Around the turn of the month, there is a first announcement for the reporting period finalized about 30 days ago, with a prompt revision following about 15 days later.

The benchmark revision of March 2009 called for a redefinition of indicators which include trade with motor vehicles. Only the “pure” retail trade sector has remained more or less unchanged in conceptual terms.⁵ The broader aggregate of retail trade including retail trade of motor vehicles and petrol stations, however, could not be maintained because car sales were no longer distinguishable between retail and wholesale. The database since comprises the total sales and repair of motor vehicles in nominal terms as an individual indicator and the merger of this with retail turnover.

2.3 Short-term Labour Market Indicators

Short-term labour market indicators are monthly data on employment, unemployment, hours worked, gross wages and salaries as well as (negotiated) pay rates on a monthly and hourly basis. The data are taken from several statistical sources, implying differing publication schemes and revision practices. While most indicators refer to the total economy, some of them are also available for the industrial sector and the main construction industry. Real-time data are collected for indices with and without seasonal and working-day adjustment except pay rates, for which adjusted data is generally not available.

In particular, Gerda comprises the monthly statistics of persons employed (domestic concept), employees subject to social security contributions and the

⁵ The turnover of petrol stations is now part of retail trade, while it was previously imputed to retail trade of motor vehicles.

number of unemployed persons. For the industrial sector and the main construction industry, data on the number and the hours worked of employees are available. Gross wages and salaries are found in the same sectoral structure, while pay rates are provided as total-economy aggregates and for the production sector including construction.

For employment and unemployment, the computer-automated vintage collection started in August and September 2005 respectively. The sectoral data on the number and the hours worked of employees as well as on gross wages and salaries has been recorded since November 2005. The vintages of employees subject to social security contributions are available from March 2006 onwards. Total pay rates, which include ancillary benefits and one-off payments, have been collected since November 2005. Pay rates excluding one-off payments and basic pay rates, which additionally correct for ancillary benefits, are available since they were initially issued in August 2007.

2.4 Price Indices

Price indicators condense the information of wide spread and detailed price observations. As surveys are complete prior to the first announcement, regular revisions do not emerge in price statistics. Benchmark revisions, however, even alter price indices. Base year changeovers reflect adjustments in weighting schemes because underlying commodity baskets change. Methodological and procedural innovations are often implemented in the context of base year changeovers.

The set of price indices available in the real-time database group into consumer prices, producer prices as well as export and import prices. The consumer price index is published with and without the energy component. Further subcomponents are those for food products, consumer durables and non-durables excluding energy and food, services excluding house rents, and house rents. The database contains producer prices of industrial goods with and without energy as well as a construction price index and a price index of farm products. External trade prices capture all products traded.

The price indicators are measured monthly, except the construction price index which is available only quarterly. With the same exception, raw data is complemented by seasonally adjusted indices. The seasonal adjustment of the consumer price index as well as its energy price and services price subcomponent additionally includes a correction for working days.

3. Data Use

This section gives an overview on the structure of the real-time database and the search functionalities as well as view and download options. The ex-

position here is rather brief. Readers who are interested in a comprehensive handbook of Gerda including a tutorial illustrating the various capabilities of the website application are referred to Knetsch (2009).

3.1 Concept and Codification

In general, the architecture of a macroeconomic real-time database differs from usual time series databases in two respects. First and most obviously, the time series of an economic indicator issued at distinct points in time are regarded as different entities. Second, a real-time database contains a set of indicators which is heterogenous from a statistical point of view because the proneness to revision is a characteristic shared by many economic statistics. However, not all indicators which are subject to revision face the same attention in a real-time context. Gerda is thus not a merger of revision-prone economic statistics but an appropriate selection of indicators which are both subject to revision and sufficiently interesting from the standpoint of real-time data research.

Each vintage of an economic time series has a unique identifier, representing the defining characteristics of the stored vector. Thus, the key structure comprises the identifying criteria of the indicator and the publication date. To ensure the broadest applicability of the database, economic entities are defined according to their exact statistical convention. In the case of benchmark revisions, this may imply that vintages regarded as congruent from an economic point of view may differ not only in the publication date but also in other key components. For instance, the economic entity “quarterly real output of the German economy, has been measured by a chain index of GDP in previous year’s prices since May 2005, while, before, it was measured by a volume in constant prices of a specific base year. As the concepts differ from each other in statistical terms, they possess different identifiers in the real-time database (see Figure).

The benefit of conceptual clarity is paid at the cost that the full revision history of some economic entities cannot be retrieved by downloading the full real-time data matrix which is stored for a single (statistically defined) indicator. Under these circumstances, the vintages of two or even more indicators have to be merged after individual downloads. In this respect, it may be useful to consult the metadata sheets provided on the website application or the database documentation (Knetsch, 2009) and the quoted literature therein. Knowledge of benchmark revisions is often very helpful for real-time data research. In particular, the empirical results of revision analysis may be strongly affected by whether or not benchmark revisions are dealt with appropriately (Knetsch/Reimers, 2006, 2009; Lorenz, 2008). Further useful information, such as base-year shifts which cannot be taken from the identifier, are delivered in the form of vintage-specific attributes.

Statistical indicator definition	Quarterly GDP, seasonally and calendar-adjusted, at constant prices						Quarterly GDP, seasonally and calendar-adjusted, at previous year's prices				
	Q:DE:Y:A:AG1:CA010:C:A						Q:DE:Y:A:AG1:CA010:A:1				
Database key											
Publication date	6.9.1995	7.12.1995	...	4.3.1999	8.6.1999	...	15.2.2005	12.5.2005	...	15.5.2009	13.8.2009
Base year	1991	1991		1991	1995		1995	–		–	–
Measurement unit	DM bn	DM bn		DM bn	€ bn		€ bn	2000=100		2000=100	2000=100
1991 Q1	709.4	709.4		713.3	428.56		429.86	86.04		85.95	85.94
1991 Q2	717.3	717.3		716.4	428.97		428.43	85.32		85.30	85.31
...											
1995 Q1	750.1	750.1		746.2	447.89		447.23	89.77		89.82	89.81
1995 Q2	758.1	758.1		751.1	451.37		451.26	90.77		90.73	90.73
1995 Q3		758.0		748.9	451.57		451.16	90.92		90.86	90.87
1995 Q4				749.3	451.01		451.14	90.67		90.80	90.80
...											
1998 Q3				800.2	472.99		468.18	94.88		94.96	94.97
1998 Q4				797.2	472.33		468.67	94.85		94.78	94.79
1999 Q1					474.40		473.56	95.84		95.83	95.82
1999 Q2							472.39	95.71		95.78	95.78
...											
2004 Q3							502.77	102.62		101.80	101.85
2004 Q4							501.60	102.48		101.83	101.84
2005 Q1								103.55		102.00	101.98
2005 Q2										102.72	102.57
...											
2008 Q4										107.61	107.74
2009 Q1										103.52	103.93
2009 Q2											104.26

Source: Deutsche Bundesbank (2009).

Real-time data matrix of quarterly real output of the German economy

Codification is based on GESMES / TS (abbreviating the term *Generic Statistical Message for Time Series*), which has become the preferred method in international data exchange. As the indicators contained in Gerda stem from rather heterogenous statistical sources, a special key family has been developed in order to keep the coding system as concise as possible. In particular, the key structure consists of 11 key components, where the first eight key components identify the economic indicator,⁶ while the remaining three specify the publication date.

3.2 Search, View and Download

The search for time series, which are individual vintages of an economic indicator in a real-time database, is separated into two stages. The search functionalities as such refer to the selection of economic indicators only, while vintages are chosen within download options. Indicators are found via pre-determined indicator trees, free text search or an indicator search form.

An intuitive access to the content of the real-time database is provided by a search along indicator trees. Users are guided through the key components step-by-step and in a pre-specified way. At each node, available codes are presented as branches and twigs, with the option of selecting one of them. The

⁶ An economic indicator is described by the frequency, reference area, adjustment indicator, accounting system, indicator category, indicator identification, price or unit reference and calculation basis.

tree structure deviates from the ordering of components in the key structure in order to economize search and to enhance the readability of trees. More experienced users may alternatively prefer to choose the indicator search form because, in this environment, key components can be filled by codes in quite a flexible manner. In contrast to indicator trees, search is constrained neither by the pre-specified ordering of key components nor by the restriction of selecting precisely one code. In general, it is possible to fill the fields directly by typing in the desired codelist members. As this practice requires familiarity with the codification, the standard approach is instead to select codes via codelists which contain the eligible members as acronyms accompanied by a verbal description. The indicator search form also provides a free text field where a key word search can be carried out.

Search results are summarized in a hit-list which comprises all indicators meeting the set criteria.⁷ The scope and structure of the real-time data of a selected economic indicator can be considered through the direct display which initially shows the full real-time data matrix. However, there are options for altering the range of displayed reporting periods and vintages as well as the ordering of the matrix. In addition, information on indicator and vintage attributes, data sources and benchmark revisions are provided. Detailed statistical metadata are tabulated in a separate text file.

The hit-list and the direct view offer the possibility to directly download the real-time data in matrix format which is a rather simple and fast export option. However, the website application provides a number of download formats which allows users to output the data of more than a single indicator simultaneously in a rather flexible manner and for different uses. For preparation, it is necessary to create a download list consisting of all indicators which may be chosen for further processing using the shopping cart function.

The download list can be loaded, manipulated and stored for iterative use by the bookmark option. In general, extensive download lists may be established and maintained. However, downloads should be as concise as possible. The reason for this is that real-time data is generally vast, which makes simultaneous downloads of a large number of indicators time-consuming. Due to the main uses of real-time data, two download forms are offered. First, it can be presented as a collection of spreadsheets, each of which comprises the real-time data of an indicator in the standard matrix representation. Second, spreadsheets may be ordered according to cut-off dates.

The download in matrix format sorts the time-series vintages of an economic indicator next to each other in chronological order where columns are headed by the publication date. In general, available real-time data is exported comple-

⁷ The search along the indicator tree also ends up with a hit-list. Of course, the hit-list contains only a single entry in this case.

tely with respect to reporting periods and vintages. However, the ranges in both matrix dimensions can be specified alternatively. In the case of irregularities in the pattern and timing of publications, users may additionally control the output of vintages. The standard is to output all vintages whose publication dates fall into the pre-specified range. In many applications, however, it may be advantageous for the real-time data matrices to be delivered in a format which satisfies a triangular structure at the bottom. Such a matrix will emerge if the publication schedule is regular, i.e. there is precisely one issue of data per reporting period, which implies that the observations in both matrix dimensions possess the same fixed time interval. The merit of this structure is that diagonals bear a special meaning. In particular, the diagonals counted from the bottom stack the releases in ascending order, i.e. the time series of indicator values which are published first, second, third and so on. Differences between the diagonals can thus be directly interpreted as revisions. The overwhelming majority of economic indicators, by their nature, have a regular publication scheme and, thus, no further treatment is needed. However, if either no or more than one vintage is issued within a time interval defined by the length of the reporting period, the output can nonetheless be trimmed to a triangular structure.

The download in snapshot format comes out as a spreadsheet containing the data status of a specified set of indicators as it existed at a given point in time. The aim of this download function is to offer the option of reconstructing the information set which was available to policymakers at the time when decisions were made. The choice is not restricted to a single cut-off date. Whenever the download list comprises indicators measured at different frequencies, the highest frequency is chosen as reference. The time series of the indicators, which are measured at a longer periodicity, possess missing values by construction.

4. Conclusion

In the past years, a number of real-time databases has been published. In contrast to its early predecessors, the recent generation shares the characteristics of comprehensiveness, timeliness, general (internet) availability and, in the most cases, maintenance by a central bank or an international institution. In August 2009, the macroeconomic real-time database Gerda was made available on the Bundesbank's website. For the time being, it is a comprehensive collection of historical vintages of German national accounts, short-term business and labour market indicators as well as price statistics. Apart from being permanently updated in the vintage dimension, it is planned to expand the set of revision-prone indicators from time to time.

On the one hand, Gerda enables commentators on current developments of the German economy to become aware of the extent of recent revisions in economic statistics. On the other hand, it provides the basis for several strands

of macroeconomic and statistical research, which has attracted growing interest in recent years, with a focus on German data. These include revision analysis, the evaluation of forecasts and forecasting tools and the *ex post* assessment of macroeconomic policy decisions.

Some applications of real-time data research might fail to be feasible at the current juncture, as the vintage dimension of many indicators is still too short. However, with the given configuration and especially the implementation of a current updating procedure, the foundations are set and a growing body of interesting research in these fields can be expected for the future.

References

- Croushore, D./ Stark, T.* (2001): A real-time data set for macroeconomists, *Journal of Econometrics* 105, 111–130.
- Deutsche Bundesbank* (2009): Publication of a comprehensive real-time database for the German economy, *Monthly Report* August 2009, 52–53.
- Garratt, A./ Vahey, S. P.* (2006): UK real-time macro data characteristics, *Economic Journal* 116, F119–F135.
- Gerberding, C./ Kaatz, M./ Seitz, F./ Worms, A.* (2005): A real-time data set for German macroeconomic variables, *Schmollers Jahrbuch* 125, 337–346.
- Jacobs, J./ Sturm, J.-E.* (2005): Do Ifo indicators help explain revisions in German industrial production, in: J.-E. Sturm/T. Wollmershäuser (eds), *Ifo survey data in business cycle and monetary policy analysis*, Heidelberg/New York.
- Knetsch, T. A.* (2009): The German real-time database: Content, structure, search and download forms, Economic Research Centre of the Deutsche Bundesbank: Technical documentation.
- Knetsch, T. A./ Reimers, H.-E.* (2006): How to treat benchmark revisions? The case of German production and orders statistics, Economic Research Centre of the Deutsche Bundesbank: Discussion Paper 38/2006.
- Knetsch, T. A./ Reimers, H.-E.* (2009): Dealing with benchmark revisions: The case of German production and orders statistics, *Oxford Bulletin of Economics and Statistics* 71, 2, 209–235.
- Lorenz, A.* (2008): Revision analysis and the role of metadata, Contribution to the OECD/Eurostat Task Force on “Performing Revision Analysis for Sub-Annual Economic Statistics,.”
- Rheinisch-Westfälisches Institut für Wirtschaftsforschung* (2007): Ein Echtzeit-Datensatz für Daten der VGR für Deutschland, RWI: Datenzeche #3, November 2007.