OVERHAULING INDUSTRIAL PRODUCTION: THE 2002 HISTORICAL AND ANNUAL REVISION

John J. Stevens
Board of Governors of the Federal Reserve System, Washington, D.C. 20551
John.J.Stevens@frb.gov

Presented August 7, 2003

KEY WORDS: Industrial production; capacity utilization; NAICS; SIC; industry classification

ABSTRACT. The Federal Reserve Board's monthly indexes of industrial production, capacity, and capacity utilization, are principal indicators of economic activity in the United States' industrial sector. In December 2002, the Federal Reserve Board issued a historical revision of these indexes going back to 1972. This revision, unprecedented among statistical agencies, is the first significant historical restatement of economic time series under the new North American Industrial Classification System (NAICS). This paper reviews the history of the industrial production indexes, including the NAICS revision; the organizing principles underlying NAICS; and the implementation of NAICS at other statistical agencies.

1. Introduction

Each year in the late fall, the Federal Reserve Board (FRB) publishes a comprehensive revision of its monthly statistical release on industrial production and capacity utilization.[1] These annual revisions are an opportunity to incorporate new and revised source data, improve underlying statistical methodologies, and expand the set of covered industries. The 2002 annual revision was unprecedented in its scope: 30 years of published data were retabulated from the bottom up using the new North American Industry Classification System (NAICS).[2] Previously, the Standard Industrial Classification (SIC) system was used. Industrial production and capacity utilization (IP/CU) were the first major economic indicators issued with substantial history on a NAICS basis. The issuance of these revised indicators marked a major advance within the statistical community. The historical continuity afforded by consistent time series data is particularly important to economists concerned about low frequency events, like business cycles, where each observation is quite important. Corrado (2003) provides additional information about the revision and its implications for users of the data; for example, despite the scope of the revisions, the dates of business cycle peaks and troughs were essentially unchanged.

This paper first provides some background on the IP/CU indexes, what they measure, and their primary applications; readers familiar with IP may wish to skip ahead. Sections 3 and 4 provide an introduction to the major steps involved in the 2002 annual revision; further details are available in Bayard and Klimek (2003), Corrado (2003), and Morin (2003). For illustrative purposes, the "motor vehicles and parts" industry (NAICS 3361-3) is frequently used to help fix ideas. The economic significance of this industry is well known; it is a highly cyclical industry with strong upstream and downstream linkages. What makes this industry useful as an illustrative tool, is that the IP/CU indexes for its sub-industries span the set of data types and methods that are used more generally to construct the nearly 300 individual IP series (plus another 150 or so aggregate series) and the 88 capacity series that comprise the IP/CU system.[3] The paper concludes by

[1] The "G.17 Statistical Release of Industrial Production and Capacity Utilization" is available at www.federalreserve.gov/releases/G17. This release reports monthly measures of output (IP), capacity, and capacity utilization for the industrial sector, which the FRB defines as the manufacturing, mining, and electric and gas utilities industries. The production indexes and utilization rates are widely reported in the media and are used by analysts in government agencies, businesses, and universities to measure current developments and trends in real output and operating rates in the industrial sector. The production indexes are also used by the Bureau of Labor Statistics to estimate manufacturing productivity.

[2] NAICS was jointly developed by Mexico's Instituto Nacional de Estadistica, Geografia e Informatica (INEGI), Statistics Canada, and the U.S. Economic Classification Policy Committee. It is a production-oriented approach to classifying establishments into industries based on the establishment's primary activity. Particular attention is paid to new and emerging industries, service industries, and industries producing advanced technologies. Among North American countries, the system facilitates cross-country comparisons of detailed industry-level statistics; at higher levels of aggregation it strives to maintain comparability with the United Nations' International Standard Industrial Classification System of All Economic Activities. See Executive Office of the President of the U.S. (2002) for more information on NAICS.

[3] IP is constructed at a much finer degree of detail than capacity, hence the difference in the number of series. However, the coverage is the same for both IP and capacity—each capacity series has a corresponding IP individual or aggregate series from which capacity utilization (the ratio of IP to capacity) is calculated.

placing the FRB efforts in the context of the broader statistical community.

2. What does IP/CU measure?

The IP index is a measure of *real output* as opposed to a nominal (i.e., dollar value) output measure. In a simple world with a single product of constant quality, the IP index would equal the number of units produced. Of course, the real world is not so simple. For a fairly homogeneous good like iron, the corresponding IP index roughly tracks the actual tonnage that is mined. For industries with heterogeneous goods, the IP index also reflects shifts over time in the mix of goods produced: Producing one more Jeep Grand Cherokee and one less Ford Escort, will, everything else equal, raise the IP index for motor vehicles due to the higher value of the Cherokee and its greater parts content. For goods that improve over time, the IP index also reflects embodied technological improvements: Today's Pentium IV microprocessor unit handles far more computations per second than yester-year's 386 microprocessor and thus represents an increase in real output.

The FRB also constructs indexes of industrial capacity to measure the productive capability of the country. Specifically,

> The FRB's capacity indexes attempt to capture the concept of *sustainable maximum output*—the greatest level of output a plant can maintain within the framework of a realistic work schedule, after factoring in normal downtime and assuming sufficient availability of inputs to operate the capital in place.

In economics jargon, this concept of capacity generally conforms to that of a full-input point on a production function, i.e., a point at which all resources are fully utilized, with the qualification that capacity represents a sustainable maximum, rather than some higher unsustainable short-term maximum. This economic concept of capacity encompasses both engineering considerations (How deep is the mineshaft? How big is the blast furnace? How wide is the pipe?) as well as operating practices (Is it a 2 or 3 shift plant? What are the prevailing work rules and how is the workplace organized?). In practice, engineering considerations and operating practices are jointly determined and may depend upon the type of production process a firm employs. For example, a continuous processor (e.g., a refinery) will face a slightly different menu of options than an assembly line (e.g., an auto plant).

Depending upon the question being asked, the IP and capacity indexes may be of independent interest. Part of this stems from the fact that the two indexes are measuring very different, yet related, concepts. They are different in that capacity is often a less tangible notion than IP. Nonetheless, the indexes must be consistent with one another, as we are also interested in their ratio—capacity utilization—that attempts to measure the proportion of the available productive capacity that was actually used in a given time period. The rate of *aggregate* capacity utilization can be a useful predictor of inflationary pressures in the economy. High rates of utilization may reflect strong demand (and hence pressure on prices in output markets). High rates of utilization may also put pressure on input prices, such as wages, due to the use of overtime. Rates of utilization are also of interest at the *industry level* as they help identify potential bottlenecks in production. For example, before the enormous fall in output by the producers of semiconductors, computers, and communications equipment in 2001, utilization rates among chip producers reached as high as 108 percent, which reflected enormous demand by computer and communications equipment makers and was consistent with reported shortages of certain chip types. Similarly, a sharp increase in utilization rates among steel producers might reflect steel shortages, which would disrupt the supply chains of steel consumers like automakers. As a result, automakers would pay higher prices for steel, and these prices would likely be passed on, at least in part, to auto buyers.

3. In search of an IP roadmap

Given the significance of the economic issues addressed using the IP/CU data, it was critical that the restatement of the production and capacity indexes on a NAICS-basis be completed accurately. The construction of historical NAICS-based IP/CU indexes would have been straightforward if each detailed SIC industry corresponded with only one NAICS industry; we could have easily renamed and reorganized the existing structure to match the new system. Unfortunately, the mapping between SIC and NAICS is often not one-for-one. Table 1 illustrates the mapping between NAICS 3361-3, motor vehicle and parts, and SIC 371, motor vehicles and motor vehicle equipment; this mapping is based on the 1997 Census of Manufactures.

Note first that the coverage for the two classifications is not identical. For example, military armored vehicles (a part of SIC 3711) is not included in the NAICS version of the motor vehicle and parts industry. Conversely, travel trailers and campers

Table 1. A NAICS-based decomposition of the motor vehicle and parts industry.

NAICS	NAICS Industry Description	SIC	SIC Industry Description
3361	Motor vehicle		
336111	Automobile mfg.	3711 pt.	Motor vehicles & car bodies
336112	Light truck & utility vehicle mfg.	3711 pt.	Motor vehicles & car bodies
336120	Heavy duty truck mfg.	3711 pt.	Motor vehicles & car bodies
3362	Motor vehicle body & trailer mfg.		
336211	Motor vehicle body mfg.	3711 pt. + 2 series	Motor vehicles & car bodies
336212-4	Truck trailer, motor home, travel trailer & camper mfg.	4 series	
3363	Motor vehicle parts mfg.		
33631-6,9	Motor vehicle parts mfg. excluding motor vehicle metal stamping	16 series	
33637	Motor vehicle metal stamping	3465	Automotive stampings

Note. SIC 3711, "Motor vehicle and car bodies," also goes into NAICS 336992, "Military armored vehicle, tank, & tank component mfg". SIC 3465, "Automotive stampings" previously was not in the motor vehicle and motor vehicle equipment industry, SIC 371.

(formerly contained in "Miscellaneous transportation equipment industry, SIC 379") and automotive stampings (formerly in "Metal forgings and stampings, SIC 346" a fabricated metal products industry) are now included in the motor vehicle and parts industry. The most complicated aspect of this mapping is the splitting up of SIC 3711 into multiple NAICS industries.

The key to developing historical NAICS indexes, therefore, was to determine the share of each SIC industry that should be assigned to each of its corresponding NAICS industries. The Census Bureau took care of this problem for 1997 by publishing industry-level statistics, such as shipments and value-added, on both an SIC and a NAICS basis. The underlying establishment level data for 1997 were dual-coded. For example, Census reported statistics for motor vehicles and passenger car bodies (SIC 3711) as a single aggregate. From 1997 on, under NAICS, the data are split into the five industries. The 1997 Census data gave us the share of total output in SIC 3711 that should be attributed to its NAICS counterparts for automobiles, for light trucks, and so on. We could have simply applied those 1997 shares to our historical motor vehicle IP aggregate to produce estimates of the value going into the five NAICS industries. However, the assumption that the shares derived from the 1997 data also reflected the distribution of economic activity in earlier years seemed too strong.

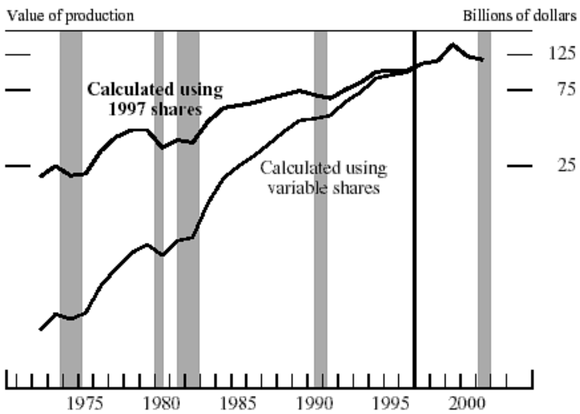
Instead, a joint project between the FRB and the Census Bureau's Center for Economic Studies calculated, for earlier years, the actual shares for individual NAICS industries from the plant-level records available at the Census Bureau (see Bayard and Klimek, 2003). Bayard and Klimek converted the industry assignment from SIC to NAICS for the hundreds of thousands of plant records covered in each of the Census of Manufactures, back to 1963, and then aggregated the plant level data to generate the NAICS shares in each census year. The conversion techniques were a combination of "exact matching" and "statistical matching". Specifically, in roughly 75 percent of the cases, NAICS codes were assigned directly using product code data. For the remaining cases they employed multinomial logit models to assign plants to NAICS industries using plant-specific probabilities derived from the available plant-level data.

The advantage of their method is that it allows, for example, the share of motor vehicles attributed to autos, light trucks, and other industries, to vary in each Census year. This variable-share approach captures fundamental changes in the structure of economic activity at very basic levels. In Figure 1 (see also Corrado (1993), p. 157) we illustrate the benefit of the variable-shares approach, which shows the value of light truck production from 1972 to 2001 under the two approaches to constructing shares. The lower line shows the value of shipments of light trucks on a NAICS basis using the variable-shares method and the upper line traces the output we would have obtained if we only had access to the shares from the 1997 Census. As you can see, if we had simply applied the 1997 structure back in time, we would have significantly overestimated light truck production for earlier years, particularly prior to the early 1990s.

4. Under the hood of the IP/CU system

The mappings provided by Bayard and Klimek (2003) formed the basis for rebuilding the inner

Figure 1. Light trucks (NAICS 336112)



Note. The shaded areas are periods of business recession as defined by the National Bureau of Economic Research (NBER).

workings of the IP/CU system. Six primary databases were reconstructed with these mappings:

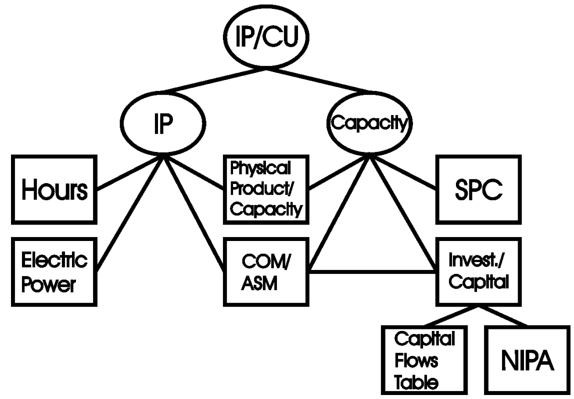
1. *Hours.* Production worker hours by industry (Bureau of Labor Statistics).
2. *Electric Power.* Power usage by industry (Federal Reserve System).
3. *COM/ASM.* Census of Manufactures and Annual Survey of Manufactures (Census Bureau).
4. *SPC.* Survey of Plant Capacity (Census Bureau).
5. *CFT.* Capital Flows Tables (Bureau of Economic Analysis).
6. *Invest./Capital.* Investment/capital database by industry (Federal Reserve Board).

Other data, especially physical product/capacity data, could essentially be mapped one-for-one between SIC and NAICS.

Figure 2 displays the primary data sources that are used to construct industrial production and capacity utilization. Solid lines denote the links between the various data. For example, the line between "Hours" and "IP" indicates that the production worker hours data are used to construct IP. Similarly, the Census of Manufactures and Annual Survey of Manufactures, are used both for industrial production and for capacity.

The balance of this section provides a cursory look at the changes to data and methods that were required to complete the revision. These changes are discussed in greater detail in Corrado (2003) and Morin (2003).

Figure 2. Database structure of IP/CU system



Notes. *Hours* is the Bureau of Labor Statistics data on production worker hours by industry. *Electric Power* is kilowatt usage by industry. *COM/ASM* is the Census of Manufactures/Annual Survey of Manufactures; includes value-added, shipments, and investment by industry. *SPC* is the Census Bureau's annual Survey of Plant Capacity by industry. *Capital Flows Table* is the Bureau of Economic Analysis' data on capital expenditures by industry in various asset categories. *NIPA* is the Bureau of Economic Analysis's National Income and Product Accounts.

4.1 Constructing IP

Each series in the IP/CU system is either an individual industry or an aggregate of individuals. The methods and data used to construct individual IP series vary by industry and by the point in time. Monthly estimates are typically based on physical product data, hours data, or electric power use. Some series are based on physical data, such as motor vehicle assemblies. Others, like motor vehicle metal stamping, are estimated using data on electric power usage. Still other series, such as motor vehicle parts, are estimated using production worker hours (although physical data on motor vehicle assemblies also guide the estimates). Together, the high frequency physical data (weekly, monthly, quarterly), electric power data, and production worker hours data, provide the monthly patterns to the various IP series. The monthly patterns are benchmarked to annual data from the COM/ASM, Current Industrial Reports from the Census Bureau, and other comprehensive data sources from other government agencies and trade associations.

The mappings produced by Bayard and Klimek (2003) provided the bridge needed to translate each database. Some data, like the production worker hours, were translated directly using the shares. The electric power data was partially translated, and some incoming data continues to be translated; generators may report on either an SIC or a NAICS basis to ease reporting burdens. For the mining industries, many of which have physical data, the translations were one-for-one, as NAICS did not significantly reorganize the classification structure of mining.[4]

In addition to translating databases, this revision was an opportunity to apply current statistical methods to previously unrevised years. For example, the published indexes now reflect current methods of seasonal adjustment[5] for all years, and each aggregate series is now calculated as a chain-type index from 1972-on.[6]

4.2 Constructing Capacity Utilization

Our estimates of industrial capacity for a year are developed from (1) capacity data reported in physical units from government and trade sources and (2) models that smooth the implied capacity measures from the Survey of Plant Capacity by using estimates of capital input by industry. The models in (2) primarily use measures of industry capital input to determine the annual changes in capacity from one year to the next. The capital input figures are estimates of the flow of services from an industry's net stocks of physical assets; the net stocks are developed principally from investment data, which lag a couple of years. Manufacturing investment for recent years is forecast using a variety of capital spending indicators.

Most of the capacity data reported in physical units was for industries that mapped one-for-one from SIC to NAICS (as was the case for IP). For industries based on the SPC and capital input data, we first translated the appropriate databases (see figure 2) and then recomputed our estimates of capacity using the restated data. Our translation of the SPC involved a non-trivial use of the shares from Bayard and Klimek (2003).[7] To derive the new measures of capital input by industry, we first translated the BEA's capital flows tables (which provide investment by asset category); we already had investment by industry from the restated versions of the COM/ASM. The rest of the work was standard in that we applied the same techniques as before. Specifically, we used bi-proportional matrix balancing to calculate industry-level investment in each asset category (for years without a capital flows table); capital stocks were calculated using a perpetual inventory model[8] and measures of capital services were derived from these stocks; finally, the implied capacity estimates from the SPC were smoothed using the capital input measures and utilization rates were calculated.

4.3. Constructing Analytical Aggregates

The IP system is intended to serve as an analytical tool for understanding the behavior of prices and output both for individual industries and for aggregates of industries. Over time, the FRB has developed useful analytical aggregates that 1) group industries by major market group, and 2) group industries by their stage of processing. Prior to the recent revision, these aggregates were based primarily upon a judgmental allocation of each IP industry to a market group and a stage of process. While all industries were assigned to a single stage of process, some were assigned to multiple market groups. For example, autos were split between the consumer and business market groups.

The 2002 revision significantly revised the market groupings by using the 1992 input-output (I-O) tables published by the Bureau of Economic Analysis to divide the IP indexes into multiple market groups where appropriate.[9] The resulting revisions are discussed in Corrado (2003). Similarly, the I-O tables were used to assign each industry to a *single* stage of process—crude, primary, semifinished, and finished. (Previously there were two stages—primary and advanced—that lacked a formal I-O explanation for their composition.) The stages are designed so that the output of each stage serves as the primary input for the subsequent stage; the output of finished processing goes to final demand.[10]

The market groups and stages of process focus attention on particular segments of the economy. Their role is similar to the earlier discussion about bottlenecks in production at the industry level. If IP drops in a given month, we can use the market group indexes to separate the effects of production for consumers from production for businesses. The interpretation of the decline is then colored by our

---

[4] An exception was the newly created "support activities for mining (NAICS 213)" industry that consists largely of the former "oil and gas field services" industry but that also includes pieces from many of the other SIC mining industries.

[5] Individual IP series are seasonally adjusted using the Census X-12 ARIMA; all seasonally adjusted aggregate indexes are calculated by aggregating the seasonally adjusted indexes of the individual series.

[6] Documentation of current methods is available at www.federalreserve.gov/releases/g17/ip_notes.htm.

[7] Morin (2003) reviews the method used to translate the SPC.

[8] See Gilbert and Mohr (1996) for a detailed discussion of the FRB's capital stock estimates.

[9] The I-O tables express total production as the sum of its use as an intermediate input in each industry and its absorption by the various components of final demand (consumption, investment, government expenditures, and exports). The 2003 annual revision will use the newly published 1997 I-O tables.

[10] More specifically, the industry composition of the four stages was primarily intended to minimize "backflows" from one stage to previous stages. A secondary objective was to minimize the number of cases where an industry's output "skips" a subsequent stage. The approach was modeled on Gaddie and Zoller (1988).

understanding of the various market forces that are impacting these two groups. Similarly, a jump in the market group for business equipment may suggest an increase, or an anticipated increase, in economic activity on the business side. Likewise, if utilization rates in the crude processing stage rise noticeably, the increase may have implications for commodity prices (now or later) and hence impact production in later stages. Thus, the industry, market group, and stage-of-process indexes all help identify potential choke-points in the industrial sector.

5. Restated, Revised, Rebased

The "3-R" mantra that lends a title to this section is also a useful summary of the 2002 annual revision of industrial production and capacity utilization. IP was *restated* on a NAICS basis; *revised* to reflect current methods and updated source data for production, capacity, price deflators, annual benchmarks; and *rebased* so that all IP indexes equal 100 in 1997 (previously 1992 was the base year). Without a doubt, it is the restatement of 30 years of data on a NAICS basis that has drawn the most attention from the statistical community. The absence of breaks in the time series for IP and CU simplifies seasonal adjustment, forecasting, and research.

Other statistical agencies throughout North America have made similar shifts toward implementing NAICS as a data organizing principle. However, none of these agencies have restated such a long time series; many have simply drawn a line at a particular date, forever separating the SIC and NAICS worlds with little or no overlap. Table 2 at the end of this paper summarizes the status of these efforts for a number of statistical programs. For example, as discussed earlier, the Census of Manufactures is available on both a NAICS and SIC basis for just one year (1997); this short NAICS history was the impetus for the joint effort between the FRB and the Census Bureau's Center for Economic Studies. The data source with the longest history on a NAICS, other than IP/CU, is the Census Bureau's M3 data on manufacturing shipments, inventories, and orders, which was restated back to 1992. Statistics Canada and Mexico's INEGI have not been any more successful than U.S. programs in restating historical data. Nonetheless, while this has hampered time series analysis, cross-country analysis within North America has benefited from the harmonized system.[11]

Going forward, economists and other data users will continue to reap significant benefits from the cross-country data organized on a consistent basis and they will begin to reap the benefits of time series data accumulated on a NAICS basis. In the meantime, the revised IP/CU data provides researchers with the only timely data with a lengthy NAICS-based history.

6. References

Bayard, Kimberly and Shawn Klimek, "Reclassifying the Census of Manufactures from the Standard Industrial Classification System to the North American Industry Classification System, 1963 to 1992," *2003 Proceedings of the American Statistical Association*, Business & Economic Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association.

Corrado, Carol, "Industrial Production and Capacity Utilization: The 2002 Historical and Annual Revision," *Federal Reserve Bulletin* (April 2003), pp. 151-176.

Executive Office of the President of the United States, *North American Industry Classification System, United States, 2002*, Office of Management and Budget (2002).

Gaddie, Robert and Maureen Zoller, "New Stage of Process Price System Developed for the Producer Price Index," *Monthly Labor Review* 111 (April 1988), pp. 3-16.

Gilbert, Charlie E. and Michael F. Mohr, "Capital Stock Estimates for Manufacturing Industries: Methods and Data," Board of Governors of the Federal Reserve System, www.federalreserve.gov/releases/g17/capital_stock_doc-latest.pdf (March 1996).

Holmes, Thomas J. and John J. Stevens, "Spatial Distribution of Economic Activities in North America," in *Handbook of Regional and Urban Economics, Volume 4,* edited by Vernon Henderson and Jacques-François Thisse (forthcoming 2003).

Morin, Norman, "NAICS and the 2002 Historical Revision of Industrial Production, Capacity, and Capacity Utilization," *2003 Proceedings of the American Statistical Association*, Business & Economic Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association.

[11] For example, Holmes and Stevens (2003) compare the geographic concentration of industries in Canada and the United States using NAICS-based census data.

Table 2. NAICS Status of Select Data Releases by Government Agency

	Data year(s) released	Release date
Federal Reserve Board		
Indexes of Industrial Production, Capacity, and Capacity Utilization	1972-2002	December 2002
Census Bureau		
Quinquennial Economic Census	1997	2000
Annual Survey of Manufactures	1998	2000
Annual Wholesale and Retail Trade Surveys	1992-1999	2001
Monthly Manufactures Shipments, Inventories, and Orders (M3)	1992-2001	2001
Monthly Wholesale and Retail Trade Surveys	1992-2001	2001
Internal Revenue Service		
Statistics of Income	1998	2000-2001
Bureau of Labor Statistics		
Unemployment Insurance—Covered Employment and Wages (ES-202)	2001	Fall 2002
Current Employment Statistics (BLS-790)	May 2003	June 2003
Current Population Survey	February 2003	January 2003
Producer Price Indexes	January 2004	February 2004
Bureau of Economic Analysis		
Benchmark Input-Output Accounts	1997	2002
NIPAs	2000-2002	2003
Capital Flow Input-Output Table	1997	2003
Annual Input-Output Accounts	1998-2002	2004
GDP by Industry	1998-2002	2004
USDIA balance of payments	1999-2003	2004