

The Construction of a Global trade Based Dataset

Andrea Civelli*

Abstract

This is a detailed explanation of the procedures followed to construct the dataset used in the paper "Globalization and Inflation: Evidence from a Time Varying VAR", joint with Francesco Bianchi. Once the raw data is collected from the different sources, this document would provide a complete guide to construct the final dataset. The matlab routines discussed in this note are available from the authors upon request .

1 Notes on Weights construction

Following the methodology used by the FED to construct the US real exchange rate, we create a matrix of trade weights for a sample of about 50 countries which can be applied in the construction of international measures (as the effective real exchange rate) based on the relevant international counterparts of a country. Our dataset covers the following countries: US, UK, Germany (unified after the 1992), France, Italy, Spain, Ireland, Denmark, Netherlands, Austria, Switzerland, Canada, Mexico, Australia, Japan, Korea, South Africa, New Zealand, Belgium, Luxembourg, Norway, Sweden, Finland, Greece, Iceland, Portugal, Turkey, Yugoslavia (Croatia and Slovenia after 1993), Argentina, Brazil, Colombia, Peru, Venezuela, Israel, Hong-Kong, India, Indonesia, USSR (Russia, Latvia and Lithuania after 1993), China, Czechoslovakia (Czech Republic and Slovakia after 1993), Hungary, Poland.

The IMF-DOT provides detailed quarterly imports and exports flows at country level for all of these countries starting from the first quarter of 1960. The data is available for about 130 countries, but we limit our attention to these countries which represent about 70 – 75% of the overall global volume of trade.

The dataset is elaborated in matlab. The starting point is a series of files in excel with the pairwise trade volumes in US dollars between the countries organized by exporter. The routine *cubic* imports those partial datasets to matlab and organize the data into a 3-dimensional matrix called *cube*.¹ It is important to notice

*University of Arkansas. Email: andrea.civelli@gmail.com

¹A line of zeros is inserted at the right position to virtually indicate the exports of a country toward itself. A vector with the extension of the countries names in the right order must be uploaded and can be found in the excel file *Countrycodes*.

that the missing values are set to 0, this is like assuming no trade between two countries at that particular point of time and it is a safe choice because it does not have any impact on the construction of the weights. Summing or subtracting a null share of trade is not affecting the computation of the total volume of trade.

The next step is executed by the routine *weightbox*, which runs after *cubic* without need to save the workspace yet; after *weightbox* we obtain the file *weights*. *Weightbox* prepares the weights necessary to compute any kind of relevant international measure for each country in the sample. The three 3D-matrices of partial weights are w_m , w_x and w_3rd . They respectively report: the shares of imports by country row-wise, the share of exports by country column-wise, and the third party share by country column-wise too. Notice the logical difference in the disposition along rows and columns of information here, which derives from the fact that *cube* is constructed using exports in a column-wise form. The formulas for the three shares, and in particular for the third party shares, are taken from Nico Loretan (2005) Fed Bulletin paper:

$$w_m_{i,j,t} = \frac{M_{i,j,t}}{\sum_{j=1}^{N_t} M_{i,j,t}}$$

$$w_x_{i,j,t} = \frac{X_{i,j,t}}{\sum_{j=1}^{N_t} X_{i,j,t}}$$

$$w_3rd_{i,j,t} = \sum_{k \neq j, \neq i}^{N_t} w_x_{i,k,t} \frac{w_m_{k,j,t}}{1 - w_m_{k,i,t}}$$

where $M_{i,j}$ and $X_{i,j}$ are the imports from country j to country i and exports from i to j ; the presence of N_t allows for time varying size of the pool of countries. The routine takes care of a few special cases which may arise when a country in some period does not export or import, or has only one trading partner, or imports something without exporting (and vice-versa). The final step is the aggregation of the weights according to the formula:

$$w_{i,j,t} = 0.5w_m_{i,j,t} + 0.5(0.5w_x_{i,j,t} + 0.5w_3rd_{i,j,t})$$

The weight $w_{i,j,t}$ represents the relative importance of country j for country i at period t in the international measure construction. The next step of the procedure will be described in the section about the output gaps.

2 Notes on the Construction of the Output Gaps

We describe first the construction of the domestic output gap; the next step in the matlab routine describes the computation of the foreign output gap.

2.1 Dataset

The domestic output gaps are constructed from different sources and following slightly different procedures based on the availability of data for each country.

For the main sample, starting from 1970:1 to 2006:4, we prefer to use the output gap measures provided by OECD-Economic Outlook when available. This is possible only for 8 countries: US, UK, Canada, Australia, France, Germany, Italy, and Japan; although the German, French and British series had to be partially completed.

For France and UK, a few observations at the beginning of the sample are missing. We add them computing the output gap as percentage deviation from a potential GDP defined by a HP filter of the real seasonal adjusted GDP series starting at 1970:1 (see below for a full description of the HP approach mainly used for other series). We add those observations after a small correction in order to make them consistent with the original OECD series,² keeping anyway the OECD gaps for the rest of the sample since the correlations between these and the HP-based gaps are very high (around .8) and the series basically differ only in some dates.

The problem with Germany is that IMF-DOT defines "Germany" as only West Germany before 1991 and the unified country after the 1992 unification with East Germany. Our weights for the world output gap follow this classification too. Since the output gap series relevant for our purposes must follow the same classification, we use the OECD output gap for the unified Germany after 1991 while we replace it with the West Germany output gap before 1991.

For all the other countries for which the OECD does not provide a output gap series, we construct it starting from the real GDP data we usually get from OECD-National Accounts Statistics or other sources (usually IMF or Global Financial Data) for the non-OECD countries.

The procedure is quite simple. The GDP series is first passed through a HP filter which will define our potential output for that country pot_t^s . We then compute the output gap for country s , gap_t^s , as percentage deviation of the GDP from its potential

²This correction accounts for the differences in the mean of the two series: basically we rescale the HP-based series in order to match the OECD output gap.

$$gap_t^s = \frac{gdp_t^s}{pot_t^s} - 1 \quad (1)$$

A special treatment is necessary for Belgium and Luxembourg to match the trading data to the output sources. We have separate trade data only after 1997, so we will use the separate output gaps for the two countries only after that time. Before that date, we compute a joint HP-based gap starting from the sum of the two output to construct the potential output for the region.

Austria is another special case since we do not have its quarterly data for real GDP from OECD, but only nominal data starting since 1965 and real quarterly data since 1989 that we obtain from the Global Financial Data (GFD) web database (which collects mainly national and IMF sources). However we have a deflator series covering almost the entire sample (1970-1999) from OECD, not seasonally adjusted, and we complete the deflator series with the data from GFD. After checking that the two series are consistent with each other over the sample 1989-1995, once the first deflator is seasonally adjusted, we adopt as final deflator a combination of the OECD deflator until 1995 and the GFD one after that point. The HP procedure is then applied.

For South Africa, we have nominal GDP data that we deflate by an appropriate GDP deflator.

A large set of OECD countries output gaps are easily computed with the HP-filtering with no particular corrections required. These are: Denmark, Netherlands, Norway, Sweden, Switzerland, Canada, Japan, Finland, Greece, Iceland, Ireland, Portugal, Spain, New Zealand and Mexico. Honk-Kong and South Korea must be also seasonally adjusted before applying the HP-filtering, for this adjustment we used the X12 program supplied by American Census.

For a quite large number of countries we use the HP-filtering, but we do not have data covering the entire sample or which require to be seasonally adjusted first. The series come generally from DataStream or, for some of them, the source is Global Financial Data (GFD). For Poland and Turkey the HP-filtering is used too, but the sample is incomplete (from 1990:1 and 1987:1 respectively). Argentina and Colombia's series have been adjusted first (and start at 1993:1 and 1994:1); Venezuela starts only at 1997:1; Peru and Israel data come from GFD (1980:1 and 1995:1); Brazil is from GFD and has to be adjusted as well (1991:1). Indonesia starts at 1993:1, Hungary at 1991:1.

The importance of China and India in international markets has increased in the past two decades, but unfortunately the quality of the data for these two countries is quite poor. India has quarterly data only since 1999:2 while China only starting at 1998:1, however annual data are available respectively since 1987 and 1953. In order to avoid the abrupt introduction of these countries at the end of our sample, we use the annual data to construct yearly output gaps by the HP-filtering method. Our quarterly weights will at

least partially provide the quarter-to-quarter variability of the contributions of their domestic gaps to the measure of the world gap.

The case of India is quite regular, in the sense that the series of real GDP at quarterly and yearly frequencies are already available. Therefore, we simply apply the HP-filter and get the output gaps for the quarterly data on the sample 1992:2 to the end. The yearly data are used to obtain the gaps in the same way before 1992 and we convert them into the quarterly frequency keeping the same gap for the whole year.

In the case of China, the China Data Online service (by ACMR - China Marketing Research Co.) provides quarterly data on current prices GDP and a quarterly index of constant prices GDP expressed using the same quarter of the previous year as a basis. Using these two series, we first compute the quarter to quarter relative price levels. The price at 1998:1 is standardized to 1 and the first 4 points of the series are simply initialized by a linear fit to the first observation of the relative price at 1999:1. This allows us to recursively construct a series of an implicit GDP deflator we can use to deflate the observed nominal series; the final real GDP series is then seasonally adjusted. We complete the series using the yearly data as done for India, we simply work with an index for the constant price GDP with basis 100 at 1970 to which we apply the filtering procedure.

We do not have data at the moment for Yugoslavia, USSR and Czechoslovakia. These are countries that split into several states at the beginning of the 90's and the series of the unified countries have been dropped. We decided to follow only the new countries after the separations since it seems quite difficult to find data for the period before the divisions.

Data for Czech Republic and Slovakia start since 1993:1, Slovakia GDP is from GFD. For the ex-USSR countries we have different starting dates (and sources): Russia and Lithuania since the first quarter of 1995; Latvia's data are from the GFD starting since 1993:1.³ Lithuania's data have been seasonally adjusted before applying the HP-procedure. Croatia's GDP data have been seasonally adjusted and start at 1997:1; finally Slovenia's gap is constructed by HP-filtering for data coming from GFD since 1993:1.

2.2 Matlab

We reduce the workspace *weights* to *reweight* where we keep only the final w matrix and upload the matrix with the domestic gaps (*domgap*) collected as above. We will work on the sample 1970:1 to 2006:4, 148 quarters instead of the 188 for which we have the weights. Since we have three countries for which we do not have data about the output gap at all (Yugoslavia, Czechoslovakia and URRS) and some cases in which for different reasons some observations might be missing, we decide to exclude them from the sample in

³Although data from GFD cover the sample starting at 1992:1, the first year of observations must be dropped for homogeneity with the other two sovietic countries.

computing the world gaps. This is done by the *shut/scan* loop by setting their weights to zero and rescaling the w matrix accordingly in the routine *globalgap*.

The matrix *worldgap* is eventually computed and the output of the routine saved in the workspace *gaps*. *Worldgap* is the weighted average of foreign countries output gaps, obtained weighting each foreign country gap by using for each domestic country the specific weights scheme in *w148*.

3 Notes on Exchange Rates

3.1 The dataset

To create the exchange rate matrix is a relatively easy task. Once we have the data for the bilateral exchange rates between the US and the other countries in the sample, we can create the pairwise dataset for all the countries using the US dollar as pivotal currency. The main sources of these series are the KEYIND base of Global Insight and the Global Financial Data web base. The data are originally reported as units of a currency necessary to buy 1 dollar, we keep this definition expressing every exchange rate as unit of foreign currency necessary to buy 1 unit of local currency. To avoid simply definition shifts in the accounting unit of the numeraire, we use as basis unit for all the exchange rates the most recent monetary unit used by each country. If this is not possible, because of a change in both the unit and the political definition of a country, we will adopt ad hoc solutions. All the exchange rate series to the dollar are seasonally adjusted by Census X12 before computing the overall exchange rates matrix in Matlab.

As usual some cases need a specific treatment and some extra assumptions in order to make every part of the dataset consistent with each other.

- Although we have data on the exchange rate for Belgium and Luxembourg separately, we follow the availability of trade data in computing their exchange rates. We treat them as a single country until 1997; fortunately, they share for a long time a common currency so this does not have an impact on the rest of the sample. The last observation of the "common country" series is added to the series of the separated countries and the first observations of the separated series is added to the common country's series in order to allow the index formula to work in both the directions
- The rate for Argentina is reconstructed from GFD, using as reference the most recent currency and taking into account the several shifts in the accounting unit over the sample. Same for Brazil before 1983; Peru for some shifts; Turkey before 2005; Mexico before 1993; Israel at 1982 and 1985; Poland before 1995

- For the Euro countries, we use different currencies for the period preceding the currency union occurred in 1999:1. Those countries are: Austria, Belgium, Luxembourg, Finland, France, Germany, Ireland, Italy, Netherlands, Portugal, and Spain. Greece actually enters only after 2001:1.
- Yugoslavia: the country series is used until 1991:3 from GFD. Five observations are missing for 1991:4 through 1992:4, which we replace with Slovenia's data from GFD as a proxy. Then we pass to the separate series in 1993:1 from the GI base. In order to preserve the homogeneity between the pre- and post-division, we need to initialize the series for Croatia at its most recent accounting unit. One more observation is available for Croatia and Slovenia separately, so we add it to allow the rolling over. For Yugoslavia the currency of Slovenia is chosen as reference unit
- Czechoslovakia: from GFD for the unified country, from GI for Slovakia and Czech Republic after 1993:1. No particular adjustments are required, except adding the first observations as above for Slovakia and Czech Republic (the value 28.9 for the Czech Republic is fitted from GFD data). No devaluations maintain the accounting unit the same over the sample for the two countries
- Russia: after the separation of the USSR, the three countries in our sample adopted different conversion rates from the Ruble to the new national currencies. Taking Russia as the reference country for this group, we compute the value of USSR Ruble in function of the accounting unit defined by the most recent Russian Ruble. Then, we adjust the initial values of Latvia and Lithuania's currencies according to this accounting reference. The data from GFD are preferred

3.2 Matlab

Given the pairwise exchange rate data, we can proceed in computing the indices of effective exchange rates for each country following the procedure adopted by the FED and described by Loretan (2005). This is done by the Matlab routine *exchanges*. Start again with *reweights* and upload *usexch*, then run the code and you will end up with the file *effective*.

First, we compute the matrix of *pairwise* exchange rates starting from the matrix for the US only (*usexch*). Each entry reports the price of the column country currency in terms of the row country currency. Then, we use a procedure very similar to the one described in *globalgap* to reset to zero the weights of those countries with missing observations. Given how we constructed the dataset *usexch*, this last step should not have any concrete effect here but it is only precautionary in the sense that we want to make sure that we do not improperly use observations close to the date of political redefinition of some nations and we want to have a code which might accommodate further extensions with less clean data. Fortunately, we are

able to have a complete overlapping of samples for the trade and the exchange rate data for the moment and it has always been possible to find pre-sample observations for the exchange rates of the new countries. Usually, it is easy to get that kind of observation for newly created countries. It is obviously impossible for countries already existing, since this is the nature of the missing data itself. It might be potentially a problem also for new countries for which data have never been collected before the official political creation. Our procedure takes into account all these cases. The section about CPI explains this point more deeply, since the correction does matter for the CPIs.

Finally we compute the effective nominal exchange rates and store them in *effenom*. We continue with this routine and workfile to compute the effective real exchange rates once we get the CPIs. The effective nominal exchange rate is defined as

$$I_{i,t} = I_{i,t-1} \prod_{j=1}^{N_t} \left(\frac{e_{i,j,t}}{e_{i,j,t-1}} \right)^{w_{i,j,t}}$$

where the nominal exchange rates $e_{i,j,t}$ is defined as the value (or the price) of one unit of country i currency in units of country j currency.

4 Notes on CPI

4.1 The Dataset

We set the year 2000 as the base year; the average of the indices at that year must be 100. For some countries the data provided by IMF (through GI) are already complete. The following integration are needed to complete the dataset

- For some countries, we have the percentage annual change rate; four initial points are enough to compute the whole series. This is the case for Argentina, whose initial points are taken from GFD
- In other cases, we take directly the series for the CPI from GFD, after checking its consistency with the IMF-GI data. This is the case for the group of Russian republics: Russia, Latvia and Lithuania, for which GFD has longer extended and improved data. For Lithuania there is actually a difference for a couple of years around 2003 in the common sample, otherwise GI and GFD comove quite well. We keep the longer GFD series and we standardize it so that the 2000 average is 100. As it happened for the GDP, we do not have data for URRS before 1992
- The series for Germany from IMF covers only West Germany and it stops in 1992. We can either use

the OECD_MEI source or GFD, which are basically equivalent, since they already account for the switch in definition of this nation

- The data from GFD for Brazil do not seem as reliable because the series is quite different with respect to other sources. We can use either IMF since 1980 or OECD_MEI since 1985. Standardized and compared to each other, the two series are very similar, so we take the longest one
- Slovakia and Czech Republic come from the OECD_MEI for the sample 1993:1 on (based on the trade weights scheme. For the period 1963:1 through 1992:4 we have data from GFD for the unified nation Czechoslovakia. We use it standardized with respect to the year 2000 base of the Czech Republic⁴
- Hungarian CPI is available only starting at 1976:1; the 2000 points are taken from OECD_MEI. For Poland, we can go back to 1980 at most thanks to the IMF data but the year 2000 base come from OECD again
- Slovenia and Croatia are provided by GFD; data for Slovenia can be used for Yugoslavia as well. The two sources shows a good degree of consistency
- Honk Kong: we take it from GFD for the entire sample
- China: as for GDP, data are very limited. China Data Online Service (by ACMR - China Marketing Research Co.) provides a quarterly CPI only since 1998:1, fortunately we can go back to 1987:1 using GFD or IMF data. We compare the three sources and they agree with each other over the common sample. GFD and CDOS data are exactly the same; the IMF series is very correlated with the other two but still it often differs in the levels. We use GFD data since 1987:1. Before that, given the current importance of this country, we use annual figures for CPI as we did for GDP. Annual variations are available since 1951; we get the complete series setting to 100 the year 2000 as a base
- USA is taken from OECD_MEI
- For Iceland, four observations are missing in the IMF series for year 1982. We simply fill in those observations with the OECD_MEI data since the series are exactly the same
- Finally we must aggregate Luxembourg and Belgium for the years before 1997 as we did above. Although the two series are very similar, we weight them by the relative real GDP. We do it only starting since 1970:1 because of readiness of the data, but it can be easily extended to the first 10 years of the sample as well

⁴Notice that you have to readjust it to Slovakia's base in order to give the extra initial observation to the series of this country alone.

Before exporting the series to Matlab, we seasonal adjust them using the Census X12 program. This adjustment is sensitive only for few of the countries, in particular if data come from GFD.

4.2 Matlab

The routine *exchanges* takes care of the real effective exchange rates too. Starting from *reweights*, we must upload also the matrix *cpis* with the seasonal adjusted time series of the CPI for the countries in our sample. The code begins generating the real pairwise exchange rates stored into *realwise*, simply adapting the *pairwise* matrix.

We apply the same correction to the weight matrix as done in the exchange rate section, in order to avoid positive weights in case the trade shares are available but the CPI data are not. This could not happen for the exchange rates, but there are a few countries with more limited sample availability for the CPI data.

When possible, as we did for the exchange rates, an initial pre-sample observation is added in order to make it feasible to compute the real index also in the first period for those nations which are politically created after 1970. The correction to the weights is important, however, if a country CPI data become available after 1970 and only after the trade shares data are available and for which, for this reason, it is impossible to find a pre-sample observation. Poland or Hungary are an example of this case. In those cases we set the CPI to -1 for computational reason, to prevent the code from propagating NaN values through the operations of the loop or stopping finding a division by 0 while computing the real indices, but the reweighting procedure has to set the weights of those points to 0 in the recursive computation of the indices so that the exchange rates ratios corresponding to those observations are raised to power 0 in the equation below. This is the same mechanism behind the weight correction for *we148* in the nominal exchange rates part of the code, but it also includes a necessary extra check of the object *scan* that deals with the first observation of the CPI of those particular countries.⁵

The real effective exchange rate is simply a modification of the nominal one

$$\hat{I}_{i,t} = \hat{I}_{i,t-1} \prod_{j=1}^{N_t} \left(\frac{\hat{e}_{i,j,t}}{\hat{e}_{i,j,t-1}} \right)^{\hat{w}_{i,j,t}}$$

where hats indicate the real version of each variable, with the exception of $\hat{w}_{i,j,t}$ which now indicates the appropriate weight of country j relative to country i , entry (i, j) of *wc148*. In particular the real exchange rate $\hat{e}_{i,j,t}$ is defined as

$$\hat{e}_{i,j,t} = e_{i,j,t} \frac{P_{i,t}}{P_{j,t}}$$

⁵It must be noticed that this version of the code does not take care of temporary missing observations, that is series interrupted only partially and for a few points. This problem has not occurred so far. The implementation of the right correction in this case is quite easy, but then one should re-think how to initialize the index after each break.

This is the value (or the price) of country i basket good in terms of country j basket, where $P_{i,t}$ is the CPI level of county i at period t . Currency i (good i) becomes more valuable relative to its j 's counterpart when $e_{i,j}$ ($\hat{e}_{i,j}$) increases, they devalue when they decrease. The output is saved in the workfile *effective*.

5 Nominal GDP data

Since the trade shares are expressed in dollars, it is useful to have also a comparable measure of the current GDP in US dollars. In collecting the nominal GDPs, we use different sources, especially GI and OECD_MEI. We can also construct the nominal GDP series starting from the real GDP series and using CPIs and exchange rates. The biggest limitation is to extend the sample to the 70's for many of the countries we study

- The data for France are from GFD in Francs and we covert them back to the US dollar. We compare this series to the those from IMF and GIKEY and the GFD series is very similar to the IMF one.
- Germany as usual is considered in the West Germany - United Germany form. We construct the data merging together MEI and GI data; a series from IMF-IFS is also available and it is very similar to ours
- For Switzerland, the source is the IMF-IFS which perfectly matches the other IMF source already in dollar

6 Interest Rates

Suitable times series for the interest rates going back to the seventies are more difficult to find. We select and construct series as follows. Short term interest rates are required and we prefer a three-month treasury bill yield whenever available. If this type of yield is not available for a country, we usually take a short term interbank or deposit rate. Continuous and homogenous series are obviously preferred, but in some cases it is necessary to merge more than one series in order to cover the entire sample. The next list summarizes sources and features of the series used in the VAR regressions

- Treasury bills are used for: Japan - from GFD, UK - from OECD_MEI through GI, US - GFD secondary market, France - GFD, Germany - GFD, Australia - GFD, Canada - GFD, Italy - GFD, Belgium - GFD, Netherlands - GFD, South Africa - GFD, Ireland - GFD
- Switzerland: comparing three-month deposit, treasury, and interbank rates from GFD, we can see that the interbank rate differs from the other two rates at the beginning of the sample, while the treasury

bill and the deposit rates are almost exactly the same series (.99 correlation in the common sample). The three-month deposit rate is chosen because of its longest sample

- Korea: we must use the one-month deposit rate from GFD, since it is the only one going back to the 70's
- Denmark: three-month CIBOR rate since 1972:1 from GFD. The first observations are integrated with a very short term deposit rate for bank accounts from GFD
- Spain: we merge three rates. The first is the Bank of Spain short term rate until 1973:1 from GFD,⁶ the second is the three-month interbank MIBOR until 2002:2 from GFD, and the last is the three-month interbank rate from Bank of Spain
- Austria: for 1970:1 to 1990:4 a treasury bill rate from GFD is used; the rest of the sample is covered by a three-month VIBOR rate from GI
- New Zealand: three-month treasury bill after 1978:1 and a six-month deposit rate before that date from GFD
- Mexico: the main series is the three-month treasury bill yield from GI after 1973:1. For the first years, two other three-month deposit rates are used: a rate from GI for the observations from 1974:1 to 1977:4 and a rate from GFD before 1973:4 (these two rates are the same except for a few observations due to changes in the taxation rates). Also, two observations after 1973 are missing and replaced by the GI series

7 Further Notes

The nominal GDP data are not usually seasonally adjusted. For many of them it is not necessary to proceed with the adjustment, but this should be checked each time. Of the countries of the paper, only South Korea has to be adjusted.

For the import/export data, the seasonality issue is attenuated once we annualize the data since this smooths out the seasonal effects. However, the quarterly flows show some degree of seasonality. It was not necessary to take care of these seasonal effects in constructing the weights because the flows are used in relative terms in the weights. However, if we had to consider quarterly flows for single countries, a seasonal adjustment should have been applied.

⁶This rate is a policy rate, so it is quite stable and does not vary much. It is included only to complete the series for these few years at the beginning in lack of a better alternative.

References

Loretan, Mico 2005. "Indexes of the Foreign Exchange Value of the Dollar" *Federal Reserve Bulletin*, 91 (Winter 2005): 1-8.