

# Documentation of the **nkpc** Stata database

Sophocles Mavroeidis      Mikkel Plagborg-Møller  
James H. Stock

February 16, 2013

## 1 Overview

The Stata database **nkpc.dta** contains a number of time series that the literature has used to estimate the pure or hybrid New Keynesian Phillips Curve. The amply commented executable **create.do** imports data from Excel spreadsheets (in which only minimal manipulations take place) and performs various transformations. Detrending of output gap series is carried out with the auxiliary file **detrend.do**.

Most raw series are found in the **series.xls** spreadsheet, which also collapses monthly series into quarterly ones by arithmetic averaging. The output is saved in **series.csv** to be read by Stata. Data on real-time real GDP/GNP is available in **RGDPGNP.xls** and stored in **rt\_rgdp.csv**.

Every growth rate or deviation-from-something series in the database are expressed in *logarithmic* differences, except for the few raw series with the letters “PCT” in their label. All rates are *quarterly*.

Most series have been downloaded from the FRED database.

## 2 Inflation

Please refer to **create.do** for a comprehensive list of series. All inflation series have labels starting with “**inf\_**” and are expressed as quarterly (log) growth rates of the underlying index. The database includes inflation based on the implicit GDP or GNP deflator (**inf\_pgdp**, **inf\_pgnp**), the GDP or GNP chain-type price index (**inf\_gdpctpi**, **inf\_gnpctpi**), PCE (**inf\_pce**), core PCE (**inf\_pcexfe**), CPI (**inf\_cpi**), core CPI (**inf\_cpixfe**), median inflation from the Cleveland Fed (**inf\_medcpi**), nominal hourly wage inflation (**inf\_wage**), inflation in the total nominal wage bill (**inf\_comp**) and various

PPI series from BLS. The GDP deflator and wage series are also available for the non-farm business (NFB) sector (e.g., `inf_pgdpnfb`).

### 3 Forcing variables

#### 3.1 Non-filtered

The database contains the log labor share `ls` as well as various cointegrating relationships meant to capture a similar but stationary relation between ULC and the price level. These are all available in NFB versions as well (e.g., `lsnfb`). Real per capita GDP `y` and real per capita potential GDP (as estimated by CBO) `ypot` yield the CBO output gap `ygap_cbo`. The civilian unemployment rate `UNRATE` minus the CBO NAIRU `NROU` gives the CBO unemployment gap `ugap_cbo`. The database also contains long-term unemployment `UEMP270V`, civilian employment `CE160V`, the civilian labor force `CLF160V` and the civilian noninstitutional population `CNP160V`.

#### 3.2 Filtered

Four filtered output gaps are calculated using `detrend.do`: The HP filter ( $\lambda = 1600$ ), the Baxter-King bandpass filter (preserving periods between 6 and 32 quarters and with a lead-lag length of 12 quarters), linear detrending and quadratic detrending. The full-sample detrended output series are called, respectively, `ygap_hp`, `ygap_bk`, `ygap_lt` and `ygap_qt`. Furthermore, real-time output gap series are constructed as optimal one-sided filters:

1. For every quarter since 1965Q4,<sup>1</sup> a real-time real output series from the Philadelphia Fed is loaded from `rt_rgdp.csv` (the data is for real GNP up to 1991Q3, for fixed-weight GDP from 1991Q4–1995Q4 and subsequently for chain-weighted GDP).
2. An AR(6) in growth rates is estimated on the real-time series and a 40-period forecast for real output is computed.
3. Each of the four filters mentioned above are applied to the concatenation of the real-time and forecasted series, generating a detrended series.
4. The real-time value of the output gap (i.e., the log deviation from trend) for the current quarter is stored in `ygap_XX_rt0Q`, where “XX”

---

<sup>1</sup>This is quarter 76 after 1947Q1, which explains the column headings in `RGDPGNP.xls`.

is one of the above-mentioned four filter abbreviations. Similarly, the real-time value of *last quarter's* output gap is stored in `ygap_XX.m1Q`, and so on for `ygap_XX.rtm2Q` and `ygap_XX.rtm3Q`.

For example, the value of `ygap_hp_rt0Q` in 1980Q1 contains the optimal one-sided HP filtered output gap (in log deviations) for 1980Q1 that one would have computed with real-time data available in 1980Q1. The value of `ygap_hp_rtm1Q` in 1980Q1 contains the optimally filtered output gap for 1979Q4 that one would have computed with real-time data available in 1980Q1.

Similar labor share gaps are available, both for the business and NFB series. For example, `lsgap_hp` is the full-sample HP-filtered (log) labor share gap, while `lsnfbgap_qt` is the full-sample quadratically detrended NFB labor share gap. The smoothing parameter for the HP filter is set to  $\lambda = 10000$ , larger than that for output detrending. (Pseudo-)real-time labor share gaps are calculated somewhat differently from the output gaps. First, the data used isn't actually real-time but is instead based on the latest vintage.<sup>2</sup> Second, to better capture the marked decrease in the labor share in the latter half of the sample, the forecasting regression is an AR(15) in *second* differences.<sup>3</sup> Like the real-time output gaps, the pseudo-real-time labor share gaps start in 1965Q4 (even though it would in principle be possible to compute them for earlier dates).

Almost all gap series have a corresponding trend series (the actual series minus the gap) in the database. For example, `ytrend_hp` is the full-sample HP filter log output per capita trend, while `lstrend_hp.lt0Q` is the current-quarter pseudo-real-time estimate of the linear log labor share trend. The only gap series that do not have a corresponding trend series available are the real-time output gaps because the raw real-time data undergoes renormalizations over time. The pseudo-real-time labor share trend, analogously to the pseudo-real-time labor share gap, is the current estimate of the trend level at any given point in time.

## 4 Forecasts

The dataset includes SPF median and mean forecasts, Greenbook forecasts and Michigan survey expectations. For the SPF and Greenbook forecasts a

---

<sup>2</sup>Thus, the pseudo-real-time labor share gap for 1990Q1 uses the *latest vintage* data for 1947Q1–1990Q1 to calculate optimal one-sided filters.

<sup>3</sup>For the business sector log labor share, the AIC selects 15 lags on the full sample and the BIC selects 8.

few data transformations have been performed in the relevant Excel spreadsheets in order to import the series as indices. Only growth rates matter; the level of these indices is not meaningful, as the forecasted series often changes definition or base year. As with the rest of the dataset, all inflation forecasts are expressed at *quarterly* rates (e.g., one-year inflation forecasts are divided by 4).

SPF inflation forecasts are of the form `SPF_INFXXXXYYY_ZZ`. “XXXX” is either “PGDP” (implicit GDP/GNP deflator) or “CPI.” The suffix “ZZ” is either “MN” (mean forecast) or “MD” (median forecast). “YYY” denotes the forecast horizon and span and can be either “0Q” (nowcast), “1Q” (1-quarter-ahead forecast), “2Q” (2-quarters-ahead forecast), “1Y” (forecast of average inflation over the current and following three quarters) and “10Y” (forecast of average inflation over the next 10 years). The 10-year forecast is only available for CPI.

Greenbook deflator inflation forecasts are of the form `GB_INFPGDPYY`, where “YY” is either “0Q,” “1Q,” “2Q” or “1Y.” Greenbook output gap forecasts (in log deviations) are of the form `GB_YGAPYY`, where “YY” is either “m1Q” (previous quarter), “0Q” (nowcast) or “1Q” (1-quarter-ahead).<sup>4</sup>

The average 1-year inflation survey expectation from the Michigan Survey of Consumers is labeled `MICH_INF1Y_MN`.<sup>5</sup>

## 5 Real-time inflation, output and labor share

The real-time inflation series from the Philadelphia Fed are of the form `RT_INFPGDPmXQ`, where “X” is an integer from 1 (inflation in last quarter) to 4 (inflation four quarters before the current). For example, the value of `RT_INFPGDPm2Q` in 2005Q1 is the real-time inflation rate for 2004Q3 using data available in 2005Q1. The real-time output per capita series have the similar form `RT_ymXQ`, with “X” ranging from 1 to 4.<sup>6</sup> In the calculation of

<sup>4</sup>Greenbook forecasts are compiled twice every quarter. Unlike for inflation, the Philadelphia Fed website doesn’t pair the output gap forecasts with a specific quarter but instead simply lists all eight forecasts per year. However, the chosen forecast date for inflation every quarter is listed (it is the date closest to the SPF release date), and so I have picked output gap forecasts corresponding to those same dates.

<sup>5</sup>It is the average response to the question: “By about what percent do you expect prices to go up, on the average, during the next 12 months?” The average inflation rate has been transformed into a log growth rate as always.

<sup>6</sup>Due to the U.S. government shutdown in late 1995, the BEA did not release GDP estimates for 1995Q4 in 1996Q1. Hence, the raw real-time deflator and output series have a missing observation in 1996Q1. To circumvent this problem, I have simply copied the 1996Q2 estimate of the 1995Q4 figure into the empty slot. Note also that starting from

the real-time output per capita series it is assumed that the FRED civilian noninstitutional population series `CNP160V` has not undergone revisions over time.

The real-time change in the log labor share has the form `RT_dls $m$ XQ`, with “X” ranging from 1 to 4. For example, `RT_dls2Q` is the log labor share two quarters ago minus the log labor share three quarters ago. The real-time change in the log labor share for the non-farm business sector has the similar form `RT_dlsnfb $m$ XQ`. While the dataset does contain real-time levels of the business or non-farm business labor share (of the forms `RT_L $S$  $m$ XQ` and `RT_LSNFB $m$ XQ` for “X” from 1 to 5), these are tricky to use due to changes in base years. See the document “Documentation of real-time unit labor cost and labor share data” for more information on the real-time labor share series.

## 6 Other variables

### 6.1 Interest rates

`FEDFUNDS` is the effective federal funds rate. Series with labels of the form `GSXX`, with “XX” being an integer (either 1, 5, 10, 20 or 30), are XX-year Treasury rates measured at constant maturity. The series `GS3M` and `GS6M` are, respectively, 3- and 6-month T-bill rates measured at constant maturity. The series `TB3MS`, `TB6MS` and `TB1YR` are the secondary market rates on, respectively, 3-month, 6-month and 1-year T-bills.

`ts5y90d` is the 5-year to 3-month Treasury spread. Similarly, `ts10y90d` is the 10-year to 3-month spread.

### 6.2 Supply shock variables

The series `OILPRICE`, `MCOILWTICO` and `MCOILBRETEU` are, respectively, the spot WTI oil price, the crude WTI oil price and the crude Brent oil price (in dollars per barrel).<sup>7</sup> `crudedef` is the BLS PPI crude materials index deflated by the implicit GDP deflator, while `oildef` is the similarly deflated WTI spot price.

---

the 1996Q1 vintage, the Philadelphia Fed’s definition of the real-time GDP deflator shifts from being the simple deflator to the chain-type price index.

<sup>7</sup>The first two series are almost coincident.

### **6.3 Business cycle**

The binary series `NBERRES` is 1 if the relevant quarter is categorized as a recession quarter by the NBER.