Data Alignment

David N. DeJong
University of Pittsburgh

Spring 2008; Revised Spring 2010

Defining Consistent Measurements

In assessing the empirical performance of a model, it is important to use appropriate data definitions.

Issues that typically require thought:

- What to include/exclude from aggregate measures of activity.
- Defining appropriate time periods.
- Variable normalization.

What to Include/Exclude from Aggregate Measurements

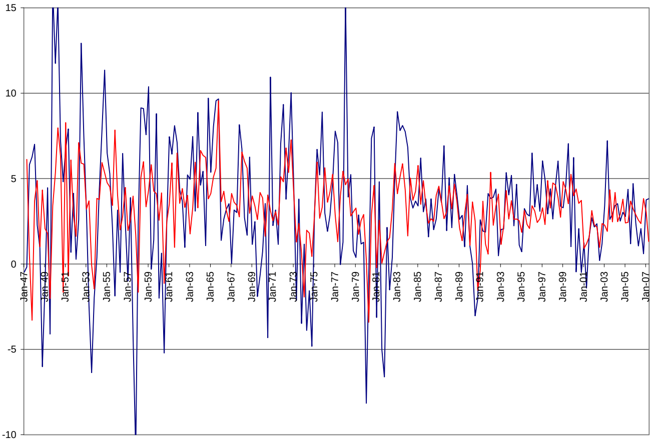
Issue: Models typically abstract from many components of the aggregate economy, and thus carry implications for only a subset of the factors included in aggregate measures of economic activity.

In assessing empirical performance, care should be taken to include in the analysis only those variables the model is specifically designed to describe.

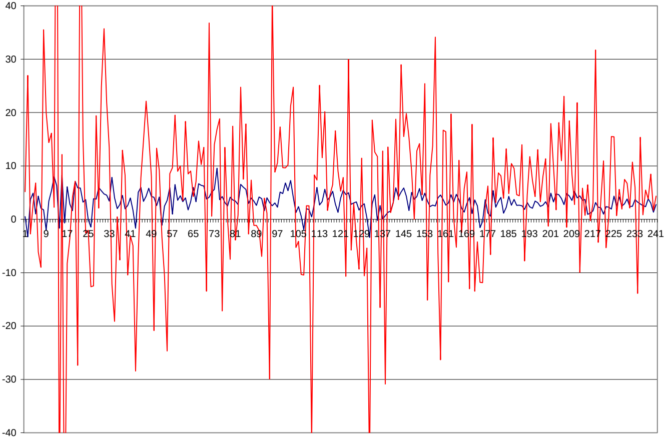
What to Include/Exclude, cont.

Examples:

- The optimal growth model abstracts from government and foreign sectors. Thus comparing its implications with GDP is inappropriate. Typically, output is measured as $C + I$.
- Unless they explicitly admit flows of services from consumption expenditures, utility functions are typically sepecified over the non-durables and services components of aggregate consumption. Thus expenditures on durable consumption goods are typically excluded from measures of aggregate consumption.



Output Growth: DGP (blue) Versus C+I



Consumption Growth: N+S (blue) Versus D

Defining Appropriate Time Periods

Issue: Depending on the question, the appropriate length of time defined for a time period can vary. Adjusting time periods is an effective means of eliminating the influence of ancillary economic phenomena on inferences regarding phenomena of direct interest in the analysis.

Defining Appropriate Time Periods, cont.

Examples:

- In studies of long-term growth, periods are typically defined as spanning 5 or 10 year intervals. Growth is measured as an average during the period. This serves to smooth over high-frequency phenomena (e.g., phases of the business cycle; seasonal surges in activity; etc.).
- For studies of business-cycle activity, periods are typically defined as quarters, and measurements are typically taken as beginning-of-period values. Typical business-cycle durations range from 6 to 40 quarters.
- Studies of long-term patterns of asset price and return behavior (e.g., Shiller 1981 *AER*; Mehra & Prescott 1984 *JME*) typically use annual time periods, and investment performance is measured using annual holding periods. This serves to eliminate, e.g., seasonal patterns of behavior. Returns are typically period averages; price and dividend levels are typically beginning-of-period values.

Defining Appropriate Time Periods, cont.

Notice that **stock** variables are typically measured at a point in time between periods, while **flow** variables are measured as period averages.

Normalization

Issue: Data alignment often requires appropriate normalization of aggregate measures of activity. This can sometimes involve subtleties related to the 'include/exclude' issue.

Normalization, cont.

Examples:

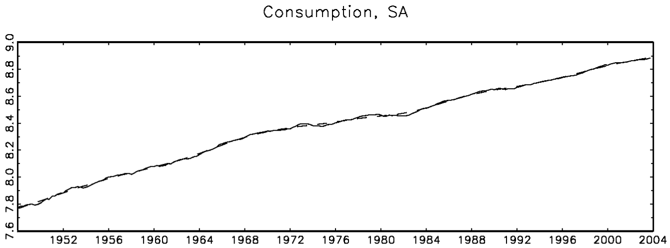
- Growth models typically seek to characterize per capita output; normalization of GDP by total population is appropriate in this case.
- Business cycle models typically seek to characterize output per worker (i.e., labor productivity); normalization by the potential aggregate labor force is appropriate in this case.
- Most models seek to characterize real economic activity, thus normalization by appropriate deflators is necessary (the GDP deflator in studies of long-term growth; the CPI for studies involving $Y = C + I$).
- However, some models seek to account explicitly for relative-price behavior (e.g., multi-sector production economies), or to describe monetary influcences on real activity. Information of relative or absolute price behavior is important to incorporate in such cases.

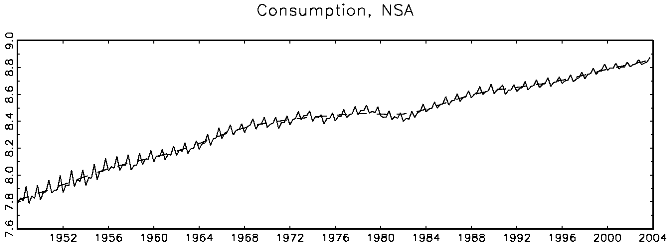
Isolating Cycles

Issue: Often, time-period adjustment is insufficient for eliminating the influence of ancillary economic phenomena on inferences regarding phenomena of direct interest in the analysis. In such cases, filters may be needed to highlight activities at certain frequencies.

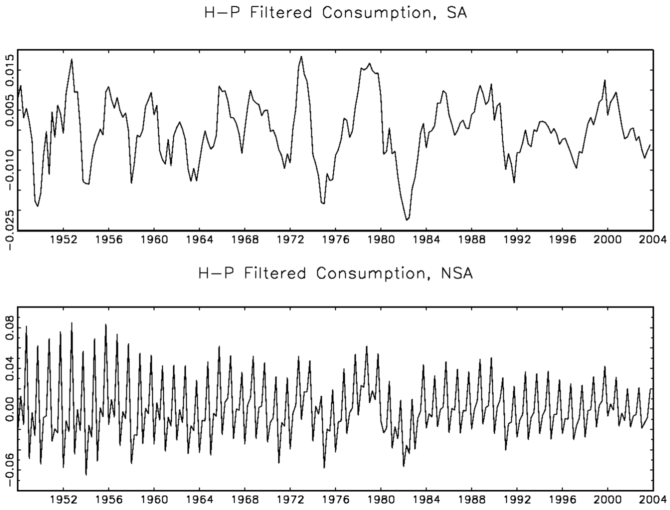
Example: Quarterly measures are typically used to study business-cycle behavior; but such measures do not eliminate the influence of seasonal activities.

Isolating Cycles, cont.





Isolating Cycles, cont.



Isolating Cycles, cont.

Most aggregate quarterly-frequency measures are available in both raw and deseasonalized form. E.g., see the rich collection of variables available at the Web site of the Federal Reserve Bank of St. Louis:
http://research.stlouisfed.org/fred2/

Deseasonalization is typically achieved using the X-11 filter (Bell and Monsell, 1992, Census Report RR-92/15) or variants.

Software is available from the Census:
http://www.census.gov/srd/www/x12a/winx12doc.html

Isolating Cycles, cont.

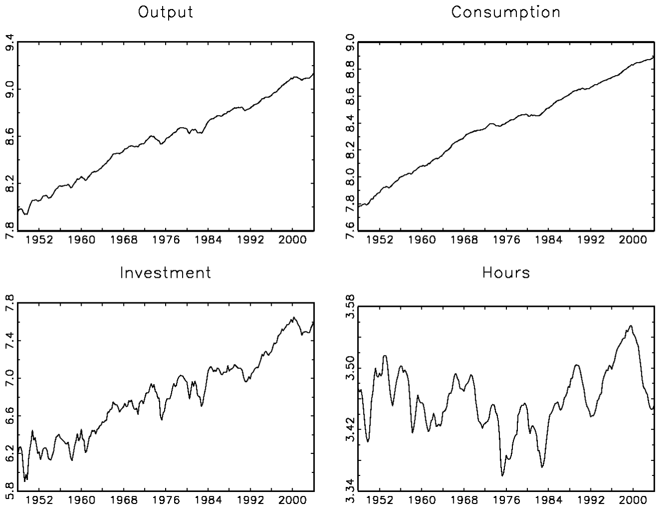
Often, additional filtering is desired. For example, filters can be used to eliminate trends, interpretable as low-frequency events. Two leading alternatives:

- Hodrick-Prescott (HP) filter
- Band Pass Filter

Removing Trends

Issue: Many facets of aggregate economic activity features trend behavior (growth). But model solution methods are typicall applicable only to stationary variables; likewise for methods used to assess these models empirically. Thus trend removal is required prior to analysis.

Removing Trends, cont.



Removing Trends, cont.

Best Practice (Symmetric Treatment): Incorporate trend behavior in the model that mimics the behavior observed in the data. Then remove trends from both the actual data and their theoretical counterparts *in the same fashion*, and compare the actual and predicted behavior of the detrended data.

Example: RBC models designed to characterize (Y, C, I, H) typically model (Y, C, I) as following constant, balanced-growth paths, and H as exhibiting stationary fluctuations.

Removing Trends,cont.

Complications, Part I

- It is difficult empirically to distinguish between leading alternative specifications regarding trend behavior.
- Trend-removal methods are specification-specific.
- Inappropriate application of trend-removal methods can induce spurious behavior.

Unit Roots Versus Trend-Stationarity

Leading representation for macroeconomic time series:

$$Z_t = Z_0 e^{gt} e^{u_t}, \qquad u_t \sim CSSP.$$

Special case (for simplicity):

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid.$$

Taking logs and quasi-differencing, we have

$$z_t = (1-\rho)\, z_0 + \rho g + (1-\rho)\, gt + \rho z_{t-1} + \varepsilon_t,$$

where $z_t = \ln Z_t$.

Unit Roots Versus Trend-Stationarity, cont.

So we have

$$z_t = (1-\rho)\, z_0 + \rho g + (1-\rho)\, gt + \rho z_{t-1} + \varepsilon_t.$$

- For $|\rho| < 1$, z_t exhibits stationary fluctuations about a linear trend (TS).
- For $\rho = 1$,

  $$z_t = g + z_{t-1} + \varepsilon_t,$$

  and thus the AR polynomial has a **unit root**. Note that

  $$\Delta z_t = (1-L)z_t = z_t - z_{t-1} = g + \varepsilon_t$$

  is stationary (DS). Note also that recursive substitution for z_{t-1}, z_{t-2} yields

  $$z_t = gt + \varepsilon_t + \varepsilon_{t-1} + ...,$$

  thus z_t is said to be **integrated** of order 1 (one, since one application of the difference operator $(1-L)$ induces stationarity).

Unit Roots Versus Trend-Stationarity, cont.

Distinguishing between TS and DS specifications:

- Unit root tests do not reject the null hypothesis for a wide range of macroeconomic time series (Nelson and Plosser, 1982 *JME*). BUT:
- Unit root tests have low power against even distant alternatives (DeJong, Nankervis, Savin and Whiteman, 1992 *Econometrica*).
- Tests of TS null hypotheses also fail to reject the null for the same series (DNSW, 1992 *J. of Econometrics*).

Unit Roots Versus Trend-Stationarity, cont.

Distinguishing between TS and DS specifications, cont.

- Bayesian procedures, designed to weigh the relative support assigned by the data to alternative specifications, assign strong support to TS specifications for macroeconomic time series (DeJong and Whiteman, 1991 *AER*, 1991 *JME*). BUT:
- Inferences supporting TS specifications may be fragile to alternative specifications of prior distributions (Sowell, 1991 *JME;* Phillips, 1991 *JAE*).

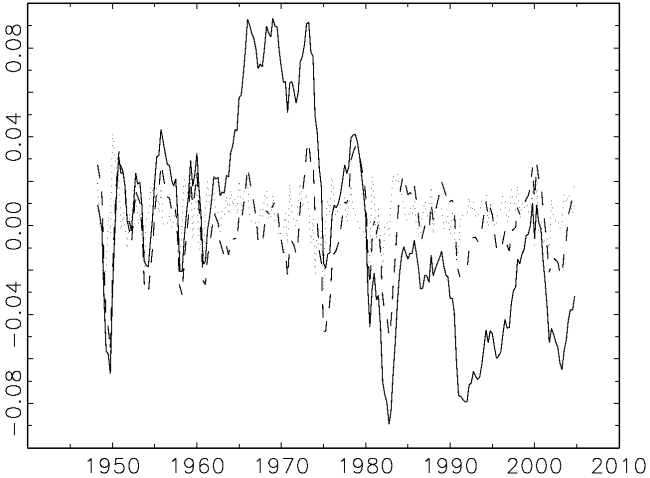
Unit Roots Versus Trend-Stationarity, cont.

Bottom line (bad news, part I): definitive and convincing statements regarding the appropriate specification of trend behavior are unavailable.

Moreover (bad news, part II): inappropriately detrended data exhibit **spurious** time series behavior. E.g., application of a filter to an i.i.d. (white-noise) process yields a new series with cyclical behavior induced by the filter.

Spuriousness

- Inappropriate application of HP and BP filters: Harvey and Jaeger (1993 *JoE*); Cogley and Nason (1995 *JEDC*); Murray (2003 REStat).
- Applying the difference operator to TS series, and detrending DS series: Chan, Hayya and Ord (1977 *Econometrica*): Nelson and Kang (1981 *Econometrica*).



Solid: Detrended; Dots: Differenced; Dashes: HP filtered

Spuriousness, cont.

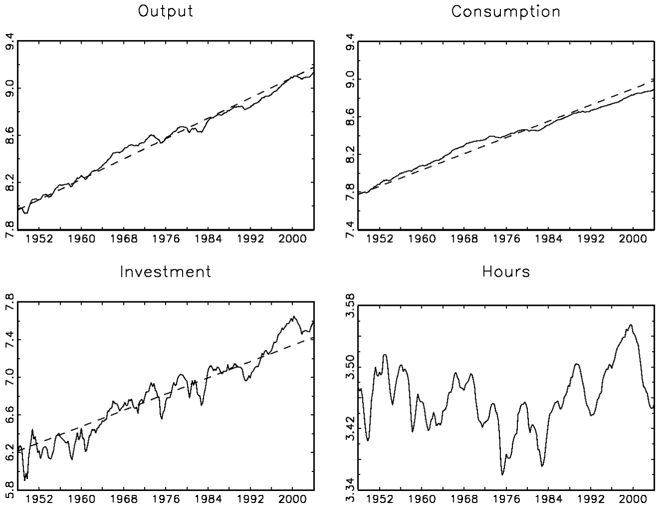
Coping with spuriousness:

- Compare results obtained using TS and DS assumptions.
  - Time consuming
  - If results differ, which set do you use?
- Hybrid approaches
  - Integrate over results obtained using TS and DS assumptions (DeJong and Whiteman, 1994 *ET*).
  - Assume an encompassing trend specification, and estimate trend jointly along with additional model parameters (Gorodnichenko and Ng, 2007 UMich WP).

Broken Trends, Unbalanced Growth

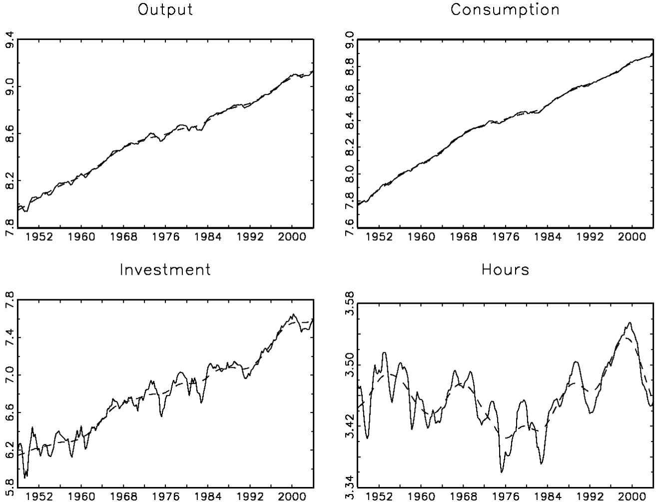
Complications, Part II: post-war observations of macroeconomic activity indicate **broken trends** (Perron, 1989 *Econometrica*) and departures from **balanced growth**.

Broken Trends, Unbalanced Growth, cont.



Growth rates: $Y, C : 1.9\%; I : 2.5\%$

Broken Trends, Unbalanced Growth, cont.



Implementing Best Practice

In the context of working with RBC models, **best practice** requires that we extend our models to incorporate:

- Stochastic trend behavior
- Unbalanced growth across consumption, investment sectors
- Example: Ireland and Schuh (2008 *Elsevier SED*)

Implementing Best Practice, cont.

Contrast with **standard practice**:

- Single-sector RBC model with deterministic growth
- HP filter actual data
- In likelihood analyses, map HP filtered data into model likelihood
- In simulation exercises, HP filter simulated (model) data, compare with HP filtered actual data.

Implementing Best Practice, cont.

Simple example for dealing with unbalanced growth: model (C, I) as distinct goods produced in distinct production sectors:

$$
\begin{aligned}
C_t &= (A_t A_{ct})^{1-\alpha} K_{ct}^{\alpha} \\
I_t &= (A_t A_{It})^{1-\alpha} K_{It}^{\alpha} \\
K_{t+1} &= K_{c(t+1)} + K_{I(t+1)} = I_t + (1-\delta)(K_{ct} + K_{It}) \\
A_t &= A_0 e^{g_a \cdot t} e^{u_{at}} \\
A_{ct} &= A_{c0} e^{u_{ct}} \\
A_{It} &= A_{I0} e^{g_I \cdot t} e^{u_{It}}
\end{aligned}
$$

Implementing Best Practice, cont.

Simple example for dealing with time-varying growth: specify as

$$g_a\left(\{\varepsilon_{at}\}\right), \qquad g_I\left(\{\varepsilon_{It}\}\right).$$

One specific possibility:

$$\begin{aligned} g_{at} &= (1-\rho_a)\,\overline{g_a} + \rho_a g_{a(t-1)} + \varepsilon_{at} \\ g_{It} &= (1-\rho_I)\,\overline{g_I} + \rho_I g_{I(t-1)} + \varepsilon_{It} \end{aligned}$$

Preserving Information in the Detrending Step

- Levels and trends in the actual data typically convey information regarding model parameterization, through implications regarding **steady state relationships**.
- This information is absent in data represented as (potentiall logged) deviations from (steady state) trends.
- How can we preserve this information, while accomplishing the detrending step?

Model Alighment in the Detrending Step

- Log-linear model representations are in terms of variables expressed as logged deviations from steady state;
- Non-linear representations are in terms of levels.
- How can we maintain direct comparability of likelihood functions across alternative representations?

Example: Optimal Growth Model

Example: consider two alternative state-space representations of the optimal growth model:

- Non-linear
- Log-Linear

Example, cont.

Log-linear:

$$
\begin{aligned}
x_t &= F x_{t-1} + e_t, \\
x_t &= \left[\widehat{y}, \widehat{c}, \widehat{i}, \widehat{k}, \widehat{z}\right]', \\
\widehat{a}_t &= \ln \frac{a_t}{\overline{a}_t}, \\
a_t &= \frac{A_t}{\left(1 + \frac{g}{1-\alpha}\right)^t}.
\end{aligned}
$$

Observation:

$$X_t = H' x_t + u_t, \qquad u_t \sim N(0, \Sigma_u).$$

Example, cont.

Non-linear (e.g., Chebyshev):

$$\begin{aligned} \left(1+\frac{g}{1-\alpha}\right)k'(\widetilde{k}_t,\widetilde{z}_t) &= i(\widetilde{k}_t,\widetilde{z}_t)+(1-\delta)k_t \\ \log z_t &= (1-\rho)\log(z_0)+\rho\log z_{t-1}+\varepsilon_t \\ \widetilde{a} &= \frac{a_t-\overline{a}}{\omega_a}\in[-1,1]\,. \end{aligned}$$

Observation (if direct comparability is to be maintained across model representations):

$$X_t = H'x_t + u_t\,, \qquad u_t \sim N(0,\Sigma_u).$$

Example, cont.

Steady state values:

$$
\begin{aligned}
\bar{k} &= \left(\frac{\alpha}{(\beta\zeta)^{-1} - (1-\delta)}\right)^{\frac{1}{1-\alpha}} \\
\bar{y} &= \bar{k}^{\alpha}, \\
\bar{i} &= \left(\delta + \frac{g}{1-\alpha}\right)\bar{k} \\
\bar{c} &= \bar{y} - \bar{i}
\end{aligned}
$$

Preserving Information

Let's take $[y, i]$ as observable. A convenient means of representing the model's characterization of these variables is in the form

$$x_t = \overline{x} e^{\left(\frac{g}{1-\alpha}\right)t} e^{\omega_{xt}}, \qquad x = y, i,$$

where ω_{xt} is a deterministic function of the stochastic component $\{\varepsilon_t\}$ of the TFP process z_t, along with (k_0, z_0). The mapping from $(\{\varepsilon_t\}, k_0, z_0)$ to ω_{xt} is obtained implicitly from the policy function approximations described above.

Preserving Information, cont.

Recall that model variables are respresented in detrended form

$$\frac{x_t}{\left(1+\frac{g}{1-\alpha}\right)^t} \approx \frac{x_t}{e^{\left(\frac{g}{1-\alpha}\right)t}}.$$

Therefore, in working with the log-linear model approximation,

$$\begin{aligned} \widehat{x_t} &= \omega_{xt} \\ &= \sigma_k \widehat{k}_{t-1} + \sigma_z \widehat{z}_{t-1} + e_{ct}; \end{aligned}$$

and in working with the non-linear approximation,

$$\begin{aligned} \frac{x_t}{\left(1+\frac{g}{1-\alpha}\right)^t} &= \overline{x} e^{\omega x} \\ &= x\left(\widetilde{k}_t, \widetilde{z}_t\right). \end{aligned}$$

Preserving Information, cont.

Preserving information regarding g.

If the actual data truly followed a stable log-linear trend, detrending could be achieved as part of the model estimation process: for each candidate parameterization, use the relevant (α, g) to construct

$$\frac{x_t}{\left(1 + \frac{g}{1-\alpha}\right)^t}.$$

However, the trend-break behavior we've noted makes this inappropriate.

Instead, we typically work with

$$\ln \frac{x_t}{HP_{xt}},$$

which has no trend or intercept term.

Preserving Information, cont.

Remedy:

Impose $g = (1 - \alpha) \cdot \widehat{g_y}$, where $\widehat{g_y}$ is the sample estimate of the growth rate of y. (0.00475 for post-war quarterly data).

Preserving Information, cont.

Preserving information regarding steady states.

As noted,

$$\ln \frac{x_t}{HP_{xt}}$$

has no trend or intercept term.

Preserving Information, cont.

Remedy:
Restore relative sample values to preserve information regarding steady states:

$$\ln y_t = \ln \frac{Y_t}{HP_{yt}},$$
$$\ln i_t = \ln \frac{I_t}{HP_{it}} + \ln \left(\overline{\frac{i}{y}}_T \right),$$

where $\left(\overline{\frac{i}{y}}_T \right)$ denotes the sample mean of $\frac{i_t}{y_t}$.

Model Alignment

Assuming interest in maintaining direct comparablility of likelihoods across models:

- For both representations, construct X_t by subtracting logged steady state ratios $\frac{\overline{x}}{\overline{y}}$, $x = y, i$ from logs of the observed data for each proposed model parameterization. The individual elements of X_t are thus

$$\begin{aligned} \widehat{y_t} &= \ln \frac{Y_t}{HP_{yt}}, \\ \widehat{i_t} &= \ln \frac{I_t}{HP_{it}} + \ln \left(\frac{\overline{i}}{y}_T \right) - \ln \frac{\overline{i}}{\overline{y}}, \end{aligned}$$

- For the log-linear approximation, X_t is fed directly into the Kalman filter.

Model Alignment, cont.

- For the non-linear approximation, compare X_t with logged values of the ratio of corresponding model variables to their steady state values:

$$\ln \frac{x\left(\widetilde{k}_t, \widetilde{z}_t\right)}{\overline{x}}, \qquad x = y, i.$$

Differences between the elements of X_t and their model counterparts are independent and normally distributed, according to the measurement equation $X_t = H' x_t + u_t$.