The R-package phtt: Panel Data Analysis with Heterogeneous Time Trends

Oualid Bada
University of Bonn

Dominik Liebl
Universities of Cologne and Bonn

Abstract

A slightly modified version of this introduction to the **phtt** package is submitted to the Journal of Statistical Software.

The R-package **phtt** provides estimation procedures for panel data with large dimensions n, T, and general forms of unobservable heterogeneous effects. Particularly, the estimation procedures are those of Bai (2009) and Kneip, Sickles, and Song (2012), which complement one another very well: both models assume the unobservable heterogeneous effects to have a factor structure. The method of Bai (2009) assumes that the factors are stationary, whereas the method of Kneip *et al.* (2012) allows the factors to be non-stationary. Additionally, the **phtt** package provides a wide range of dimensionality criteria in order to estimate the number of the unobserved factors simultaneously with the remaining model parameters.

Keywords: Panel data, unobserved heterogeneity, principal component analysis, factor dimension.

1. Introduction

One of the main difficulties and at the same time appealing advantages of panel models is their need to deal with the problem of unobserved heterogeneity. Classical panel models, such as *fixed effects* or *random effects*, try to model unobserved heterogeneity using dummy variables or structural assumptions on the error term (see, e.g., Baltagi (2005)). In both cases the unobserved heterogeneity is assumed to remain constant over time within each cross-sectional unit—apart from an eventual common time trend. This assumption might be reasonable for approximating panel data with fairly small temporal dimensions T; however, for panel data with large T this assumption becomes implausible.

Nowadays, the availability of panel data with large cross-sectional dimensions n and large time dimensions T has triggered the development of a new class of panel data models. Recent discussions by Ahn, Lee, and Schmidt (2006), Pesaran (2006), Bai (2009), Bai, Kao, and Ng (2009), and Kneip *et al.* (2012) have focused on advanced panel models for which the unobservable individual effects are allowed to have heterogeneous time trends that can be approximated by a factor structure. The basic form of this new class of panel models can be presented as follows:

$$y_{it} = \sum_{j=1}^{P} x_{itj}\beta_j + \nu_{it} + \epsilon_{it}, \quad \text{for } i \in \{1, \ldots, n\}, \quad \text{and } t \in \{1, \ldots, T\}, \tag{1}$$

where y_{it} is the dependent variable for each individual i at time t, x_{itj} is the jth element of the vector of explanatory variables $x_{it} \in \mathbb{R}^P$, and ϵ_{it} is the idiosyncratic error term. The time-varying individual effects $\nu_{it} \in \mathbb{R}$ of individual i for the time points $t \in \{1, \ldots, T\}$ are assumed to have a d-dimensional factor structure. The following two specifications of the time-varying individual effects ν_{it} are implemented in the R package **phtt**:

$$\nu_{it} = \begin{cases} v_{it} &= \sum_{l=1}^{d} \lambda_{il} f_{lt}, & \text{for the model of Bai (2009),} \\ v_i(t) &= \sum_{l=1}^{d} \lambda_{il} f_l(t), & \text{for the model of Kneip } et\ al. \text{ (2012).} \end{cases} \tag{2}$$

Here, λ_{il} are unobserved individual loadings parameters, f_{lt} are unobserved common factors for the model of Bai (2009), $f_l(t)$ are the unobserved common factors for the model of Kneip *et al.* (2012), and d is the unknown factor dimension. We consider the standard case of iid error terms ϵ_{it} with $\mathrm{E}(\epsilon_{it}) = 0$ and $\mathrm{V}(\epsilon_{it}) = \sigma^2$.

Note that the explicit consideration of an intercept in model (1) is not necessary but may facilitate interpretations. If x_{it} includes an intercept, the time-varying individual effects ν_{it} are centered around zero. If x_{it} does not include an intercept, the time-varying individual effects ν_{it} are centered around the overall mean.

Model (1) includes the classical panel data models with additive time-invariant individual effects and common time-specific effects. Consider the case in which $d = 2$ with a first common factor $f_{1t} = 1$ for all $t \in \{1, \ldots, T\}$ that has individual loadings parameters λ_{i1}, and a second common factor f_{2t} that has the same loadings parameter $\lambda_{i2} = 1$ for all $i \in \{1, \ldots, n\}$.

An intrinsic problem of factor models lies in the fact that the true factors are only identifiable up to rotation. In order to ensure the uniqueness of these parameters, a number of d^2 restrictions are required. The usual normalization conditions are given by

$$\begin{aligned} &\text{(a)} \quad \tfrac{1}{T} \textstyle\sum_{t=1}^{T} f_{lt}^2 = 1 \quad && \text{for all } l \in \{1, \ldots, d\}, \\ &\text{(b)} \quad \textstyle\sum_{t=1}^{T} f_{lt} f_{kt} = 0 \quad && \text{for all } l, k \in \{1, \ldots, d\} \text{ with } k \neq l, \text{ and} \\ &\text{(c)} \quad \textstyle\sum_{i=1}^{N} \lambda_{il} \lambda_{ik} = 0 \quad && \text{for all } l, k \in \{1, \ldots, d\} \text{ with } k \neq l. \end{aligned}$$

For the model of Kneip *et al.* (2012), f_{lt} in conditions (a) and (b) has to be replaced by $f_l(t)$.

Kneip *et al.* (2012) consider the case in which the common factors $f_l(t)$ show relatively smooth patterns over time. This includes strongly positive auto-correlated stationary as well as non-stationary factors. The authors propose to approximate the time-varying individual effects $v_i(t)$ by smooth functions $\vartheta_i(t)$. In this way (1) becomes a semi-parametric model and its estimation is done using a two-step estimation procedure, explained in more detail in Section 2.

Alternatively, Bai (2009) proposes an iterated least squares approach to estimate (1) for stationary time-varying individual effects v_{it} such as ARMA or white noise processes. The estimators are the result of an iterative procedure solving a system of non-linear equations. However, Bai (2009) assumes the factor dimension d to be a known parameter, which is usually not the case. Therefore, the **phtt** package uses an algorithmic refinement of Bai's method proposed by Bada and Kneip (2010) in order to estimate the number of unobserved common factors d jointly with the remaining model parameters; see Section 4 for more details.

Besides the implementations of the methods proposed by Kneip *et al.* (2012), Bai (2009), and Bada and Kneip (2010) the R package **phtt** comes with a wide range of criteria (13 in total)

for estimating the factor dimension d. The main functions of the **phtt** package are given in the following list:

- `KSS()` : Computes the estimators of the model parameters according to the method of Kneip *et al.* (2012); see Section 2
- `Eup()` : Computes the estimators of the model parameters according to the method of Bai (2009) and Bada and Kneip (2010); see Section 4
- `OptDim()` : Allows for a comparison of the optimal factor dimensions $\hat{d}$ obtained from many different panel criteria; see Section 3
- `checkSpecif()` : Tests whether to use a classical fix effects panel model or a panel model with individual effects ν_{it}; see Section 5.1

The functions are provided with `print()`-, `summary()`-, `plot()`-, `coef()`- and `residuals()`-methods.

Standard methods for estimating models for panel and longitudinal data are also implemented in the R R Development Core Team (2012) packages **plm** (Croissant and Millo 2008), **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Core team 2012), and **lme4** (Bates, Maechler, and Bolker 2012); see Croissant and Millo (2008) for an exhaustive comparison of these packages. Recently, Millo and Piras (2012) published the R package **splm** for spatial panel data models. The **phtt** package further extends the toolbox for statisticians and econometricians and provides the possibility of analyzing panel data in the case when the unobserved heterogeneity is time-varying.

To the best of our knowledge, the **phtt** package is the first software package that offers the estimation methods of Bai (2009) and Kneip *et al.* (2012). Regarding the different dimensionality criteria (in total 13) that can by accessed via the function `OptDim()` only those of Bai and Ng (2002) are publicly available as MATLAB codes (The MathWorks Inc. 2012) from the homepage of Serena Ng (`http://www.columbia.edu/~sn2294/`).

To demonstrate the use of our functions, we re-explore the well known `Cigar` dataset, which is frequently used in the literature of panel models. The panel contains the amounts of cigarette consumption of $n = 46$ American states from 1963 to 1992 ($T = 30$) as well as data about the income per capita and cigarette prices in the same states during the same period (see, e.g., Baltagi and Levin (1986) for more details on the dataset).

We follow Baltagi and Li (2004), who estimate the following panel model:

$$\ln(\texttt{Consumption}_{it}) \quad = \quad \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + e_{it}. \tag{3}$$

Here, $\texttt{Consumption}_{it}$ presents the sales of cigarettes (packs of cigarettes per capita), $\texttt{Price}_{it}$ is the average real retail price of cigarettes, and $\texttt{Income}_{it}$ is the real disposable income per capita. The index $i \in \{1, \ldots, 46\}$ denotes the single states and the index $t \in \{1, \ldots, 30\}$ denotes the year.

Baltagi and Li (2004) assume the error term e_{it} to be affected by time-varying spatial correlations between neighboring states. To estimate the model, the authors use a pre-defined $n \times n$ spatial weights matrix $W = \{\omega_{ij}\}_{i,j=1,\ldots,n}$, where ω_{ij} is equal to one if state i and state j are neighboring states and zero else. If a state i has more than one neighboring state then the corresponding ω_{ij}'s are normalized to sum up to one.

However, the model of Baltagi and Li (2004) is very restrictive, since the assumptions on the structure of the error term e_{it} are fixed a priori. Instead, we apply the panel methods introduced above and allow for the state-cross-correlations in the error term e_{it} to be approximated from the data by a multidimensional factor structure such that

$$e_{it} = \sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it}.$$

The `Cigar` dataset can be obtained from the **phtt** package using the function `data("Cigar")`. The panels of the variables $\ln(\texttt{Consumption}_{it})$, $\ln(\texttt{Price}_{it})$, and $\ln(\texttt{Income}_{it})$ are shown in Figure 1.

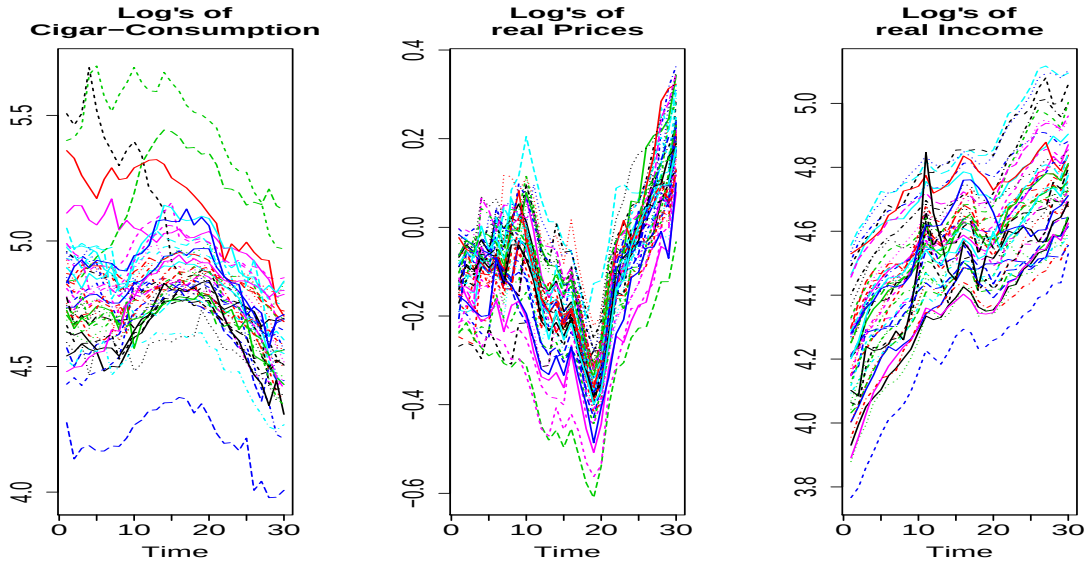

Figure 1: Plots of the dependent variable $\ln(\texttt{Consumption}_{it})$ and regressor variables $\ln(\texttt{Price}_{it})$ and $\ln(\texttt{Income}_{it})$.

Section 2 is devoted to a short introduction of the method of Kneip *et al.* (2012), which is appropriate for relatively smooth common factors $f_l(t)$. Section 3 presents the usage of the function `OptDim()`, which provides access to a wide range of panel dimensionality criteria recently discussed in the literature on factor models. Section 4 deals with the explanation as well as application of the panel method proposed by Bai (2009), which is appropriate for stationary and relatively unstructured common factors f_{lt}.

2. Panel models for heterogeneity in time trends

The panel model proposed by Kneip *et al.* (2012) can be presented as follows:

$$y_{it} = \sum_{j=1}^{P} x_{itj}\beta_j + v_i(t) + \epsilon_{it}, \tag{4}$$

where the time-varying individual effects $v_i(t)$ are parametrized in terms of common non-parametric basis functions $f_1(t), \ldots, f_d(t)$ such that

$$v_i(t) \quad = \quad \sum_{l=1}^{d} \lambda_{il} f_l(t). \tag{5}$$

The asymptotic properties of this method rely on second order differences of $v_i(t)$, which apply for continuous functions as well as for classical discontinuous stochastic time series processes such as (S)AR(I)MA processes. Therefore, the functional notation of the time-varying individual effects $v_i(t)$ and their underlying common factors $f_1(t), \ldots, f_d(t)$ does not restrict them to a purely functional interpretation. The main idea of this approach is to approximate the time series of the time-varying individual effects $v_i(t)$ by smooth functions $\vartheta_i(t)$.

The estimation approach proposed by Kneip *et al.* (2012) relies on a two-step procedure: first, estimates of the common slope parameters β_j and the time-varying individual effects $v_i(t)$ are obtained semi-parametrically. Second, functional principal component analysis is used to estimate the common factors $f_1(t), \ldots, f_d(t)$, and to re-estimate the time-varying individual effects $v_i(t)$ more efficiently. In the following we describe both steps in more detail.

Step 1: The unobserved parameters β_j and $v_i(t)$ are estimated by the minimization of

$$\sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \sum_{j=1}^{P} x_{itj}\beta_j - \vartheta_i(t) \right)^2 + \sum_{i=1}^{n} \kappa \int_1^T \frac{1}{T} \left(\vartheta_i^{(m)}(s) \right)^2 ds, \tag{6}$$

over all $\beta_j \in \mathbb{R}$ and all m-times continuously differentiable functions ϑ_i, where $\vartheta_i^{(m)}$ denotes the mth derivative of the function ϑ_i. A first approximation of $v_i(t)$ is then given by $\tilde{v}_i(t) := \hat{\vartheta}_i(t)$. Spline theory implies that any solution $\hat{\vartheta}_i(t)$ possesses an expansion in terms of a natural spline basis $z_1(t), \ldots, z_T(t)$ such that $\hat{\vartheta}_i(t) = \sum_{s=1}^{T} \hat{\zeta}_{is} z_s(t)$; see, e.g., De Boor (2001). Using the latter expression, we can rewrite (6) to formalize the following objective function:

$$S(\beta, \zeta) = \sum_{i=1}^{n} \left(||Y_i - X_i\beta - Z\zeta_i||^2 + \kappa \zeta_i^\top A \zeta_i \right), \tag{7}$$

where $Y_i = (y_{i1}, \ldots, y_{iT})^\top$, $X_i = (x_{i1}^\top, \ldots, x_{iT}^\top)^\top$, $\beta = (\beta_1, \ldots, \beta_p)^\top$, $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{iT})^\top$, Z and A are $T \times T$ matrices with elements $\{z_s(t)\}_{s,t=1,\ldots,T}$ and $\{\int z_s^{(m)}(t) z_k^{(m)}(t) dt\}_{s,k=1,\ldots,T}$ respectively. κ is a preselected smoothing parameter to control the smoothness of $\hat{\vartheta}_i(t)$. We follow the usual choice of $m = 2$, which leads to cubic smoothing splines.

The semi-parametric estimators $\hat{\beta}, \hat{\zeta}_i = (\hat{\zeta}_{i1}, \ldots, \hat{\zeta}_{iT})^\top$, and $\tilde{v}_i = (\tilde{v}_{i1}, \ldots, \tilde{v}_{iT})^\top$ can be obtained by minimizing $S(\beta, \zeta)$ over all $\beta \in \mathbb{R}^p$ and $\zeta \in \mathbb{R}^{T \times n}$.

The solutions are given by

$$\hat{\beta} \quad = \quad \left(\sum_{i=1}^{N} X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i^\top (I - \mathcal{Z}_\kappa) Y_i \right), \tag{8}$$

$$\hat{\zeta}_i \quad = \quad (Z^\top Z + \kappa R)^{-1} Z^\top (Y_i - X_i \hat{\beta}), \text{ and} \tag{9}$$

$$\tilde{v}_i \quad = \quad \mathcal{Z}_\kappa \left(Y_i - X_i \hat{\beta} \right), \text{ where } \mathcal{Z}_\kappa = Z \left(Z^\top Z + \kappa R \right)^{-1} Z^\top. \tag{10}$$

Step 2: The common factors are obtained by the first d eigenvectors $\hat{\gamma}_1, \ldots, \hat{\gamma}_d$ that correspond to the largest eigenvalues $\hat{\rho}_1, \ldots, \hat{\rho}_d$ of the empirical covariance matrix

$$\hat{\Sigma} = \frac{1}{n} \tilde{v}_i \tilde{v}_i^\top. \tag{11}$$

The estimator of the common factor $f_l(t)$ is then defined by the lth scaled eigenvector

$$\hat{f}_l(t) = \sqrt{T} \hat{\gamma}_{lt} \quad \text{for all } l \in \{1, \ldots, d\}, \tag{12}$$

where $\hat{\gamma}_{lt}$ is the tth element of the eigenvector $\hat{\gamma}_l$. The scaling factor $\sqrt{T}$ yields that $\hat{f}_l(t)$ satisfies the normalization condition $\frac{1}{T} \sum_{t=1}^{T} \hat{f}_l(t)^2 = 1$ as listed above in Section 1. The estimates of the individual loadings parameters λ_{il} are obtained by ordinary least squares regressions of $\left(Y_i - X_i \hat{\beta}\right)$ on $\hat{f}_l$, where $\hat{f}_l = (\hat{f}_l(1), \ldots, \hat{f}_l(T))'$. Recall from conditions (a) and (b) that $\hat{\lambda}_{il}$ can be calculated as follows:

$$\hat{\lambda}_{il} = \frac{1}{T} \hat{f}_l^\top \left(Y_i - X_i \hat{\beta}\right). \tag{13}$$

The time-varying individual effects $v_i(t)$ are re-estimated by $\hat{v}_i(t) := \sum_{l=1}^{d} \hat{\lambda}_l \hat{f}_l(t)$, where the factor dimension d can be determined, e.g., by the sequential testing procedure of Kneip *et al.* (2012) or by any other dimensionality criteria. In Section 3 we introduce several such criteria.

To determine the optimal smoothing parameter κ_{opt}, the authors propose the following cross validation (CV) criterion:

$$CV(\kappa) = \sum_{i=1}^{n} ||Y_i - X_i \hat{\beta}_{-i} - \sum_{l=1}^{d} \hat{\lambda}_{-i,l} \hat{f}_{-i,l}||^2. \tag{14}$$

Unfortunately, this criterion is computationally very costly and requires determining the factor dimension d in advance. To overcome this disadvantage, we propose a plug-in smoothing parameter that is discussed in more detail in Section 2.1.

Kneip *et al.* (2012) show the consistency of the estimators as $n, T \to \infty$ and derive the asymptotic distribution of common slope estimators $\hat{\beta}$ as $(\hat{\beta} - \mathrm{E}(\hat{\beta}|\epsilon)) = \mathbf{N}(0, \hat{\Sigma}_\beta)$, where

$$\hat{\Sigma}_\beta = \sigma^2 \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1} \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa)^2 X_i \right) \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1}. \tag{15}$$

A consistent estimator of σ^2 can be obtained by

$$\hat{\sigma}^2 = \frac{1}{(n-1)T} \sum_{i=1}^{n} ||Y_i - X_i \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{i,l} \hat{f}_l(t)||^2. \tag{16}$$

2.1. Computational details

A problem that remains to be discussed is the determination of the smoothing parameter κ in (8), (12), and (13). Generally, it is possible to determine κ by the CV criterion in (14); however, for relatively large dimensions T and n cross validation is computationally very

costly. Moreover, Kneip *et al.* (2012) do not explain how the factor dimension d is to be specified during the optimization process, which is critical since $\hat{d}$ itself depends on κ; see (21) in Section 3.

We propose to determine the smoothing parameter κ by generalized cross validation (GCV). However, we cannot apply the classical GCV formulas as proposed, e.g., in Craven and Wahba (1978) since we do not know the parameters β and $v_i(t)$. Our computational algorithm for determining the GCV smoothing parameter κ_{GCV} is based on the method of Cao and Ramsay (2010), who propose optimizing objective functions of the form (7) by updating the parameters iteratively in a functional hierarchy. Formally, the iteration algorithm can be described as follows:

1. For given κ and β, we optimize (7) with respect to ζ_i to get

$$\hat{\zeta}_i = (Z'Z + \kappa R)^{-1} Z^\top (Y_i - X_i \beta). \tag{17}$$

2. By using (17), we minimize (7) with respect to β to get

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i^\top X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i^\top (Y_i - Z\hat{\zeta}_i) \right) \tag{18}$$

3. Once (17) and (18) are obtained, we optimize the following GCV criterion to calculate κ_{GCV}:

$$\kappa_{GCV} = \arg\min_{\kappa} \frac{1}{\frac{n}{T} tr(I - \mathcal{Z}_\kappa)^2} \sum_{i=1}^{n} ||Y_i - X_i\hat{\beta} - \mathcal{Z}_\kappa(Y_i - X_i\hat{\beta})||^2. \tag{19}$$

The program starts with initial estimates of β and κ and proceeds with steps 1, 2, and 3 in recurrence until convergence of all parameters, where the initial value $\hat{\beta}_{start}$ is defined in (45) and the initial value κ_{start} is the GCV-smoothing parameter of the residuals $Y_i - X_i\hat{\beta}_{start}$.

The advantage of this approach is that the inversion of the $P \times P$ matrix in (18) does not have to be be updated during the iteration process. Moreover, the determination of the GCV-minimizer in (19) can be easily performed in R using the function `smooth.spline()`, which calls on a rapid C-routine.

The GCV smoothing parameter κ_{GCV} in (19) does not explicitly account for the factor structure of the time-varying individual effects $v_i(t)$ as formalized in (2). However, given the assumption of a factor structure, the goal is not to obtain optimal estimates of $v_i(t)$ but rather to obtain optimal estimates of the common factors $f_l(t)$, which implies that the optimal smoothing parameter κ_{opt} will be smaller than κ_{GCV}; see Kneip *et al.* (2012).

We use the GCV smoothing parameter κ_{GCV} as an upper bound for κ_{opt} and approximate the optimal smoothing parameter by the following naive plug-in estimator:

$$\kappa_{plug\text{-}in} = 0.75 \cdot \kappa_{GCV}.$$

Alternatively, it is possible to optimize the CV criterion (14). In this case, the optimal smoothing parameter κ_{opt} is selected from the interval $(0, \kappa_{GCV})$ and the factor dimension d in (14) is estimated by (21) using the plug-in estimator $\kappa_{plug\text{-}in}$.

2.2. Application

This section is devoted to the application of the method above discussed, which is accessible through the function `KSS()`. In total, the function `KSS()` has the following arguments:

```
R> args(KSS)

function (formula, additive.effects = c("none", "individual",
    "time", "twoways"), consult.dim.crit = FALSE, d.max = NULL,
    sig2.hat = NULL, factor.dim = NULL, level = 0.01, spar = NULL,
    CV = FALSE, convergence = 1e-06, restrict.mode = c("restrict.factors",
        "restrict.loadings"), ...)
NULL
```

The argument `additive.effects` makes it possible to extend the model (4) for additional additive `individual`, `time`, or `twoways` effects as discussed in Section 5. If the logical argument `consult.dim.crit` is set to `TRUE` all dimensionality criteria discussed in Section 3 are computed and the user is asked to choose one of their results.

The arguments `d.max` and `sig2.hat` are required for the computation of some dimensionality criteria discussed in Section 3. If their default values are maintained, the function internally computes `d.max`$= \left\lfloor \min\{\sqrt{n}, \sqrt{T}\} \right\rfloor$ and `sig2.hat` as in (16), where $\lfloor x \rfloor$ indicates the integer part of x. The argument `level` allows us to adjust the significance level for the dimensionality testing procedure (21) of Kneip *et al.* (2012); see Section 3.

The factor dimension d can also be pre-specified by the argument `factor.dim`. Recall from restriction (a) that $\frac{1}{T}\sum_{t=1}^{T} \hat{f}_l(t)^2 = 1$. Alternatively, it is possible to standardize the individual loadings parameters such that $\frac{1}{n}\sum_{i=1}^{n} \hat{\lambda}_{il} = 1$, which can be done by setting `restrict.mode = "restrict.loadings"`.

As an illustration we estimate the Cigarettes model (3) introduced in Section 1:

$$\begin{aligned} \ln(\texttt{Consumption}_{it}) &= \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + e_{it} \\ \text{with} \quad e_{it} &= \sum_{l=1}^{d} \lambda_{il} f_l(t) + \epsilon_{it}, \end{aligned} \tag{20}$$

where $v_i(t) = \sum_{l=1}^{d} \lambda_{il} f_l(t)$. In the following lines of code we load the `Cigar` dataset and take logarithms of the three variables, $\texttt{Consumption}_{it}$, $\texttt{Price}_{it}/\texttt{cpi}_t$ and $\texttt{Income}_{it}/\texttt{cpi}_t$, where $\texttt{cpi}_t$ is the consumer price index. Please note that we store the variables as $T \times n$-matrices. This is necessary, because the `formula` argument of the `KSS()`-function takes the panel variables as matrices in which the number of rows has to be equal to the temporal dimension T and the number of columns has to be equal to the individual dimension n.

Note that the function `KSS()` is written for balanced panels, and eventually missing values have to be replaced in a pre-processing step by appropriate estimates.

```
R> library("phtt")
R> data("Cigar")
R> N <- 46
R> T <- 30
```

```
R> l.Consumption   <- log(matrix(Cigar$sales, T, N))
R> cpi             <- matrix(Cigar$cpi,       T, N)
R> l.Price         <- log(matrix(Cigar$price, T, N)/cpi)
R> l.Income        <- log(matrix(Cigar$ndi,   T, N)/cpi)
```

The model parameters β_1, β_2, the factors $f_l(t)$, the loadings parameters λ_{il}, and the factor dimension d can be estimated by the `KSS()`-function with its default arguments. Inferences about the slope parameters can be obtained by using the method `summary()`.

```
R> Cigar.KSS <- KSS(formula = l.Consumption ~ l.Price + l.Income)
R> (Cigar.KSS.summary <- summary(Cigar.KSS))

Call:
KSS.default(formula = l.Consumption ~ l.Price + l.Income)

Residuals:
   Min     1Q Median     3Q    Max
 -0.11  -0.01   0.00   0.01   0.12


 Slope-Coefficients:
            Estimate  StdErr z.value    Pr(>z)
(Intercept)   4.0600  0.1770   23.00 < 2.2e-16 ***
l.Price      -0.2600  0.0223  -11.70 < 2.2e-16 ***
l.Income      0.1550  0.0382    4.05  5.17e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type:  none

Used Dimension of the Unobserved Factors: 6

Residual standard error: 0.000725 on 921 degrees of freedom
R-squared: 0.99
```

The effects of the log-real prices for cigarettes $\ln(\texttt{Price}_{it})$ and the log-real incomes $\ln(\texttt{Income}_{it})$ on the log-sales of cigarettes $\ln(\texttt{Consumption}_{it})$ are highly significant and in line with results in the literature. The summary output reports an estimated factor dimension of $\hat{d} = 6$. In order to get a visual impression of the six estimated common factors $\hat{f}_1(t), \ldots, \hat{f}_6(t)$ and the estimated time-varying individual effects $\hat{v}_1(t), \ldots, \hat{v}_n(t)$, we provide a `plot()`-method for the `KSS`-summary object.

```
R> plot(Cigar.KSS.summary)
```

The left panel of Figure 2 shows the six estimated common factors $\hat{f}_1(t), \ldots, \hat{f}_6(t)$ and the right panel of Figure 2 shows the $n = 46$ estimated time-varying individual effects $\hat{v}_1(t), \ldots, \hat{v}_n(t)$.

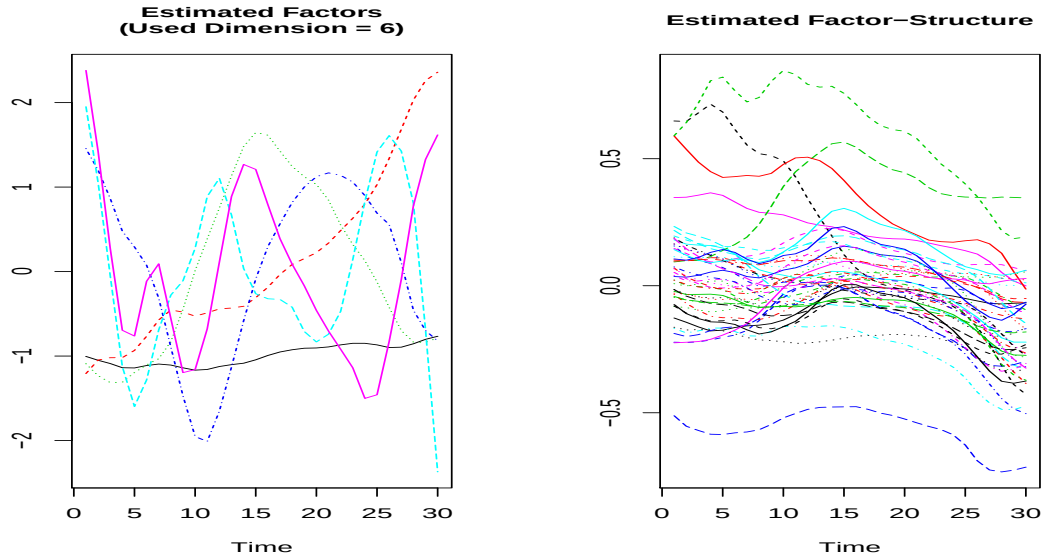

Figure 2: LEFT PANEL: Estimated factors $\hat{f}_1(t), \ldots, \hat{f}_6(t)$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_1(t), \ldots, \hat{v}_n(t)$.

Obviously, there is one nearly time-invariant common factor; this suggests extending the model (20) by additive `individual` effects; see Section 5 for more details.

By setting the logical argument `consult.dim.crit=TRUE`, the user can choose from other dimensionality criteria, which are discussed in Section 3. Note that the consideration of different factor dimensions d would not alter the results for the slope parameters β since the estimation procedure of Kneip *et al.* (2012) for the slope parameters β does not depend on the dimensionality parameter d.

3. Panel criteria for selecting the number of factors

In order to estimate the factor dimension d, Kneip *et al.* (2012) propose a sequential testing procedure based on the following test statistic:

$$KSS(d) = \frac{n \sum_{r=d+1}^{T} \hat{\rho}_r - (n-1)\hat{\sigma}^2 tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2N \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)^2)}} \overset{a}{\sim} N(0,1), \tag{21}$$

where $\hat{\mathcal{P}}_d = I - \frac{1}{T}\sum_{l=1}^{d} f_l f_l^\top$ with $f_l = (f_l(1), \ldots, f_l(T))^\top$, and

$$\hat{\sigma}^2 = \frac{1}{(n-1)tr((I-\mathcal{Z}_\kappa)^2)} \sum_{i=1}^{n} ||(I-\mathcal{Z}_\kappa)(Y_i - X_i\hat{\beta})||^2. \tag{22}$$

The selection method can be described as follows: choose a significance level α (e.g., $\alpha = 1\%$) and begin with $H_0 : d = 0$. Test if $KSS(0) \leq z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution. If the null hypothesis can be rejected, go on with $d = 1, 2, 3, \ldots$ until H_0 cannot be rejected.

The dimensionality criterion of Kneip *et al.* (2012) can be used for stationary as well as non-stationary factors. However, this selection procedure has a tendency to ignore factors that are weakly auto-correlated. As a result, the number of factors can be underestimated.

More robust against this kind of underestimation are the criteria of Bai and Ng (2002). The basic idea of their approach consists simply of finding a suitable penalty term g_{nT}, which countersteers the undesired variance reduction caused by an increasing number of factors $\hat{d}$. Formally, $\hat{d}$ can be obtained by minimizing the following criterion:

$$PC(l) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(l))^2 + l g_{nT} \tag{23}$$

for all $l \in \{1, 2, \ldots\}$, where $\hat{y}_{it}(l)$ is the fitted variable for a given factor dimension l. To estimate consistently the dimension of stationary factors Bai and Ng (2002) propose specifying g_{nT} by one of the following penalty terms:

$$g_{nT}^{(\mathrm{PC1})} = \hat{\sigma}^2 \frac{(n+T)}{nT}\log(\frac{nT}{n+T}) \tag{24}$$

$$g_{nT}^{(\mathrm{PC2})} = \hat{\sigma}^2 \frac{(n+T)}{nT}\log(\min\{n,T\}) \tag{25}$$

$$g_{nT}^{(\mathrm{PC3})} = \hat{\sigma}^2 \frac{\log(\min\{n,T\})}{\min\{n,T\}}, \tag{26}$$

where $\hat{\sigma}^2$ is a consistent estimator of σ^2, the variance of ϵ_{it}. The proposed criteria are denoted by PC1, PC2, and PC3, respectively. In practice, $\hat{\sigma}^2$ can be obtained by

$$\hat{\sigma}^2(d_{max}) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(d_{max}))^2, \tag{27}$$

where d_{max} is an arbitrary maximal dimension that is larger than d. This kind of variance estimation can, however, be inappropriate in some cases, especially when (27) underestimates the true variance σ^2. The latter can be the case, if the error terms are auto-correlated. To overcome this problem, Bai and Ng (2002) propose three additional criteria (IC1, IC2, and IC3):

$$IC(l) = \log\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(l))^2\right) + l g_{nT} \tag{28}$$

with

$$g_{nT}^{(\mathrm{IC1})} = \frac{(n+T)}{nT}\log(\frac{nT}{n+T}) \tag{29}$$

$$g_{nT}^{(\mathrm{IC2})} = \frac{(n+T)}{nT}\log(\min\{n,T\}) \tag{30}$$

$$g_{nT}^{(\mathrm{IC3})} = \frac{\log(\min\{n,T\})}{\min\{n,T\}}. \tag{31}$$

Under similar assumptions, Ahn and Horenstein (2009) propose selecting d by maximizing the ratio of adjacent eigenvalues (or the ratio of their growth rate). The criteria are referred

to as *Eigenvalue Ratio* (ER) and *Growth Ratio* (GR) and defined as following:

$$\begin{aligned} ER &= \frac{\hat{\rho}_l}{\hat{\rho}_{l+1}} \\ GR &= \frac{\log\left(\sum_{r=l}^{T}\hat{\rho}_r / \sum_{r=l+1}^{T}\hat{\rho}_r\right)}{\log\left(\sum_{r=l+1}^{T}\hat{\rho}_r / \sum_{r=l+2}^{T}\hat{\rho}_r\right)}. \end{aligned}$$

Note that the theory of the above dimensionality criteria PC1, PC2, PC3, IC1, IC2, IC3, ER, and GR is developed for stationary factors. In order to estimate the number of unit root factors, Bai (2004) proposes the following panel criteria:

$$IPC(l) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(l))^2 + l g_{nT} \tag{32}$$

where

$$g_{nT}^{(\text{IPC1})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T}\frac{(n+T)}{nT}\log(\frac{nT}{n+T}) \tag{33}$$

$$g_{nT}^{(\text{IPC2})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T}\frac{(n+T)}{nT}\log(\min\{n,T\}) \tag{34}$$

$$g_{nT}^{(\text{IPC3})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T}\frac{\log(\min\{n,T\})}{\min\{n,T\}}. \tag{35}$$

Alternatively, Onatski (2010) has introduced a threshold approach based on the empirical distribution of the largest eigenvalue, which can be used for both stationary and non-stationary factors. The estimated dimension is obtained by

$$\hat{d} = \max\{l \leq d_{max} : \hat{\rho}_l - \hat{\rho}_{l-1} \geq \delta\},$$

where δ is a positive threshold, estimated iteratively from the data.

3.1. Application

The dimensionality criteria introduced above are implemented in the function `OptDim()`, which has the following arguments:

```
R> args(OptDim)

function (Obj, criteria = c("PC1", "PC2", "PC3", "IC1", "IC2",
    "IC3", "IPC1", "IPC2", "IPC3", "KSS.C", "ED", "ER", "GR"),
    standardize = FALSE, d.max, sig2.hat, spar, level = 0.01)
NULL
```

The desired criteria can be selected by one or several of the following character variables: `"KSS.C"`, `"PC1"`, `"PC2"`, `"PC3"`, `"IC1"`, `"IC2"` , `"IC3"`, `"ER"`, `"GR"`, `"IPC1"`, `"IPC2"`, `"IPC3"`, and `"ED"`. The default significance level used for the `"KSS"`-criterion is `level = 0.01`. The values of d_{max} and $\hat{\sigma}^2$ can be specified externally by the arguments `d.max` and `sig2.hat`. By default, `d.max` is computed internally as `d.max`$= \left\lfloor \min\{\sqrt{n}, \sqrt{T}\} \right\rfloor$ and `sig2.hat` as in

(22) and (27). The input variable can be standardized by choosing `standardize = TRUE`. In this case, the calculation of the eigenvalues is based on the correlation matrix instead of the covariance matrix.

As an illustration, imagine that we are interested in the estimation of the factor dimension of the variable $\ln(\texttt{Consumption}_{it})$ with the dimensionality criterion `"PC1"`. The function `OptDim()` requires a $T \times n$ matrix as input variable.

```
R> OptDim(Obj = l.Consumption, criteria = "PC1")

Call: OptDim.default(Obj = l.Consumption, criteria = "PC1")

---------
Criterion of Bai and Ng (2002):

  PC1
    7
```

`OptDim()` offers the possibility of comparing the result of different selection procedures by giving the corresponding criteria to the argument `criteria`. If the argument `criteria` is left unspecified, `OptDim()` automatically compares all 13 dimensionality selection procedures.

```
R> (OptDim.obj <- OptDim(Obj = l.Consumption, criteria = c("PC3",  "ER",  "GR",
+                        "IPC1", "IPC2", "IPC3"), standardize = TRUE))

Call: OptDim.default(Obj = l.Consumption, criteria = c("PC3", "ER",
    "GR", "IPC1", "IPC2", "IPC3"), standardize = TRUE)

---------
Criterion of Bai and Ng (2002):

  PC3
    8

--------
Criteria of Ahn and Horenstein (2008):

  ER GR
   3  3

---------
Criteria of Bai (2004):

  IPC1 IPC2 IPC3
     3    3    2
```

In order help users to choose the most appropriate dimensionality criterion for the data, `OptDim`-objects are provided with a `plot()`-method. This method displays, in descending

order, the magnitude of the eigenvalues in percentage of the total variance and indicates where the given criteria detect the optimal dimension: see Figure 3.

```
R> plot(OptDim.obj)
```

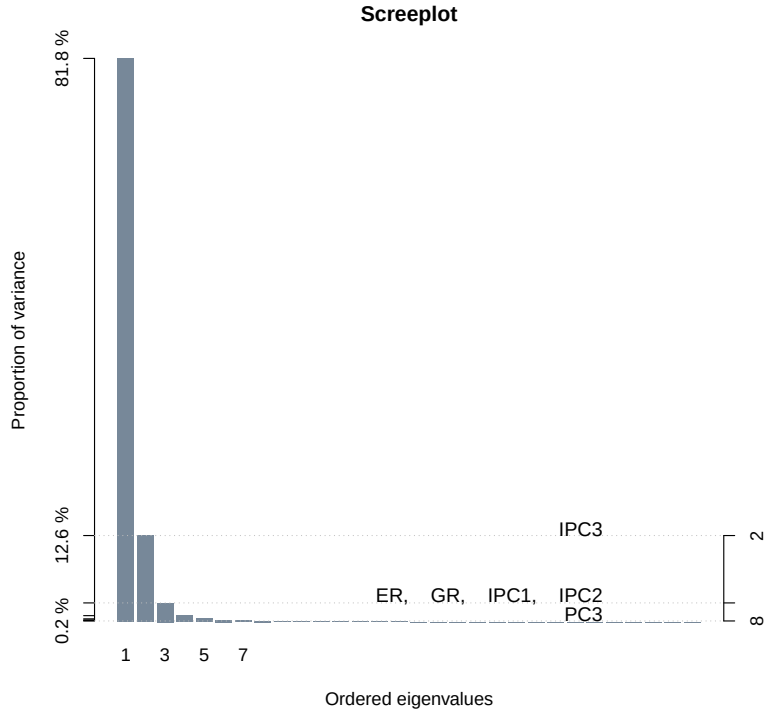

Figure 3: Scree plot produced by the `plot()`-method for `OptDim`-objects. Most of the dimensionality criteria (ER, GR, IPC1 and IPC2) suggest using the dimension $\hat{d} = 3$.

In this regard, the function `KSS()` offers us the ability to compare the results of all dimensionality criteria and to select one of them. If the `KSS()`-argument `consult.dim = TRUE` the results of the dimensionality criteria are printed on the console of R and the user is asked to choose one of the results.

```
R> KSS(formula = l.Consumption ~ -1 + l.Price + l.Income, consult.dim = TRUE)

------------------------------------------------------------
Results of Dimension-Estimations

-Bai:
 PC1  PC2  PC3  IC1  IC2  IC3 IPC3 IPC2 IPC3
   7    6    7    5    5    5    3    3    3

-KSS:
 KSS.C
```

```
     6

-Onatski:
 ED
  3

-RH:
 ER  GR
  3   4

------------------------------------------------------------
Please, choose one of the proposed integers:
```

After entering a number of factors, e.g., 6 we get the following feedback:

```
Used dimension of unobs. factor structure is: 6
------------------------------------------------------------
```

Note that the maximum number of factors that can be given, cannot exceed the maximal number of all proposed factor dimensions (here maximal dimension would be 7). A higher dimension can be chosen using the argument `factor.dim`.

4. Panel models with stationary common factors

The panel model proposed by Bai (2009) can be presented as follows:

$$y_{it} = \sum_{j=1}^{P} x_{itj}\beta_j + v_{it} + \epsilon_{it}, \tag{36}$$

where

$$v_{it} = \sum_{l=1}^{d} \lambda_{il} f_{lt}. \tag{37}$$

Combining (36) with (37) and writing the model in matrix notation, we get

$$Y_i = X_i\beta + F\Lambda_i^\top + \epsilon_i \tag{38}$$

where $Y_i = (y_{i1}, \ldots, y_{iT})^\top$, $X_i = (x_{i1}^\top, \ldots, x_{iT}^\top)^\top$, $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})^\top$, $\Lambda_i = (\lambda_1, \ldots, \lambda_n)^\top$ and $F = (f_1, \ldots, f_T)^\top$ with $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{id})$, $f_t = (f_{1t}, \ldots, f_{dt})$, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})^\top$.
The asymptotic properties of Bai's method rely, among others, on the following assumption:

$$\frac{1}{T}F^\top F \xrightarrow{p} \Sigma_F \quad \text{as} \quad T \to \infty, \tag{39}$$

where Σ_F is a fixed positive definite $d \times d$ matrix. This rules out the large class of non-stationary stochastic processes such as unit root processes.

4.1. Model with known number of factors d

Bai (2009) proposes to estimate the model parameters β, F and Λ_i by minimizing the following least squares objective function:

$$S(\beta, F, \Lambda_i) = \sum_{i}^{n} ||Y_i - X_i\beta - F\Lambda_i^\top||^2 \tag{40}$$

For each given F, the OLS estimator of β can be obtained by

$$\hat{\beta}(F) = \left(\sum_{i=1}^{n} X_i^\top \mathcal{P}_d X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^\top \mathcal{P}_d Y_i\right)$$

where $\mathcal{P}_d = I - F(F^\top F)^{-1}F^\top = I - FF^\top/T$. If β is known, F can be estimated by using the first d eigenvectors $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_d)$ corresponding to the first d eigenvalues of the empirical covariance matrix $\hat{\Sigma} = (nT)^{-1}\sum_{i=1}^{n} w_i w_i^\top$, where $w_i = Y_i - X_i\beta$. That is,

$$\hat{F}(\beta) = \sqrt{T}\hat{\gamma}.$$

The idea of Bai (2009) is to start with initial values for β or F and calculate the estimators iteratively. The method requires, however, the factor dimension d to be known, which is usually not the case in empirical applications.

4.2. Model with unknown number of factors d

Bada and Kneip (2010) propose an algorithmic refinement of the method of Bai (2009) in order to provide a joint estimation of the factor dimension d together with the other parameters β, F, and Λ_i. In this case, the optimization criterion can be defined as a penalized least squares objective function:

$$S(\beta, F, \Lambda_i, l) = \sum_{i}^{N} ||Y_i - X_i\beta - F\Lambda_i^\top||^2 + l g_{nT} \tag{41}$$

The role of the additional term $l g_{nT}$ is to pick up the optimal dimension $\hat{d}$, of the unobserved factor structure. The penalty factor g_{nT} can be chosen according to Bai and Ng (2002). Alternatively, g_{nT} can be replaced by the threshold δ proposed by Onatski (2010). The estimation algorithm is based on the parameter cascading method of Cao and Ramsay (2010) that can be described as follows:

1. Minimizing (41) with respect to Λ_i for each given β, F and d, we get

$$\hat{\Lambda}_i^\top(\beta, F, d) = F^\top (Y_i - X_i\beta)/T. \tag{42}$$

2. Introducing (42) in (41) and minimizing with respect to F for each given β and d, we get

$$\hat{F}(\beta, d) = \sqrt{T}\hat{\gamma}(\beta, d), \tag{43}$$

where $\hat{\gamma}(\beta, d)$ is a $T \times d$ matrix that contains the first d eigenvectors corresponding to the first d eigenvalues $\rho_1, \ldots, \rho_d$ of the covariance matrix $\hat{\Sigma} = (nT)^{-1}\sum_{i=1}^{n} w_i w_i^\top$ with $w_i = Y_i - X_i\beta$.

3. Reintegrating (43) and (42) in (41) and minimizing with respect to β for each given d, we get

$$\hat{\beta}(d) = \left(\sum_{i=1}^{N} X_i^\top X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^\top \left(Y_i - \hat{F}\hat{\Lambda}_i^\top(\hat{\beta}, d)\right)\right). \tag{44}$$

4. Optimizing (41) with respect to l given the results in (42), (43), and (44) allows us to select $\hat{d}$ as

$$\hat{d} = \min_{l} \sum_{i}^{N} ||Y_i - X_i\hat{\beta} - \hat{F}\hat{\Lambda}_i^\top||^2 + l g_{nT}, \qquad \text{for all } l \in \{0, 1, \ldots, d_{max}\}.$$

The final estimators are obtained by alternating between an inner iteration to optimize $\hat{\beta}(d), \hat{F}(d)$, and $\hat{\Lambda}_i(d)$ for each given d and an outer iteration to select the optimal dimension $\hat{d}$. The updating process is repeated in its entirety till the convergence of all the parameters. This is why the estimators are called *entirely updated estimators* (Eup).

It is notable that the objective functions (41) and (40) are not globally convex. There is no guarantee that the iteration algorithm converges to the global optimum. Therefore, it is important to choose reasonable starting values $\hat{d}_{start}$ and $\hat{\beta}_{start}$. We propose to select a large dimension d_{max} and to start the iteration with the following estimate of β:

$$\hat{\beta}_{start} = \left(\sum_{i=1}^{N} X_i^\top (I - GG^\top) X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^\top (I - GG^\top) Y_i\right), \tag{45}$$

where G is the $T \times d_{max}$ matrix of the eigenvectors corresponding to the first d_{max} eigenvalues of the augmented covariance matrix

$$\Gamma^{Aug} = \frac{1}{nT} \sum_{i=1}^{n} (Y_i, X_i)(Y_i^\top, X_i^\top)^\top.$$

The intuition behind these starting estimates relies on the fact that the unobserved factors cannot escape from the space spanned by the eigenvectors G. The orthogonal projection of X_i on G in (45) eliminates the effect of a possible correlation between the observed regressors and unobserved factors, which can heavily distort the value of β^0 if it is neglected. Greenaway-McGrevy, Han, and Sul (2012) give conditions under which (45) is a consistent estimator of β.

According to Bai (2009), the asymptotic distribution of the slope estimator $\hat{\beta}$ for known d and i.i.d. ϵ_{it} is:

$$\sqrt{nT}(\hat{\beta}(d) - \beta) \overset{a}{\sim} N(0, \Sigma_\beta),$$

where $\Sigma_\beta = D_0^{-1}\sigma^2$. Here, $\sigma^2 = \mathrm{Var}(\epsilon_{it})$ and $D_0 = \mathrm{plim}\frac{1}{nT}\sum_{i=1}^{n} Z_i^\top Z_i$ with $Z_i = \mathcal{P}_d X_i - \frac{1}{n}\sum_{k=1}^{n} \mathcal{P}_d X_i a_{ik}$ and $a_{ik} = \Lambda_i(\frac{1}{n}\sum_{i=1}^{n}\Lambda_i^\top\Lambda_i)^{-1}\Lambda_k^\top$. Bada and Kneip (2010) show that if $\lim_{n,T\to\infty} P(\hat{d} = d) = 1$, the entirely updated estimator $\hat{\beta} = \hat{\beta}(\hat{d})$ will have the same asymptotic distribution as $\hat{\beta}(d)$. The asymptotic variance of the estimator $\hat{\beta}$ can be estimated as follows:

$$\hat{\Sigma}_\beta = (\frac{1}{nT}\sum_{i=1}^{n} Z_i^\top Z_i)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \hat{\epsilon}_i^\top \hat{\epsilon}_i,$$

where $\hat{\epsilon}_i = Y_i - X_i\hat{\beta} - \hat{F}\hat{\Lambda}_i^\top$.

4.3. Application

The above described methods are implemented in the function `Eup()`, which takes the following arguments:

```
R> args(Eup)
```

```
function (formula, additive.effects = c("none", "individual",
    "time", "twoways"), dim.criterion = c("PC1", "PC2", "PC3",
    "IC1", "IC2", "IC3", "IPC1", "IPC2", "IPC3", "ED"), d.max = NULL,
    sig2.hat = NULL, factor.dim = NULL, double.iteration = TRUE,
    start.beta = NULL, max.iteration = 500, convergence = 1e-06,
    restrict.mode = c("restrict.factors", "restrict.loadings"),
    ...)
NULL
```

The argument `additive.effects` gives the possibility of extending the model (38) for additional additive effects as discussed in more detail in Section 5. The argument `dim.criterion` specifies the dimensionality criterion to be used if `factor.dim` is left unspecified and defaults to `dim.criterion = "PC1"`. The arguments `d.max` and `sig2.hat` are required for the computation of some dimensionality criteria discussed in Section 3. If their default values are maintained, the function internally computes `d.max` as $\min\{\sqrt{n}, \sqrt{T}\}$ and `sig2.hat` according to (27).

Setting the argument `double.iteration=FALSE` may speed up computations, because the updates of $\hat{d}$ will be done simultaneously with $\hat{F}$ without waiting for their inner convergences. However, in this case, the convergence of the parameters is less stable than in the default setting.

The argument `start.beta` allows us to give a vector of starting values for the slope parameters β_{start}. The maximal number of iteration and the convergence condition can be controlled by `max.iteration` and `convergence`. Finally, by choosing `restrict.mode = c("restrict.loadings")`, the restriction $\frac{1}{T}\sum_t f_{lt}^2 = 1$ will be replaced by the restriction $\frac{1}{n}\sum_i \lambda_{il}^2 = 1$ for all $l \in \{1, \dots d\}$.

In our application, we take first-order differences of the observed time series. This is because some factors show temporal trends, which can violate the stationarity condition (39); see Figure 2. We consider the following modified Cigarettes model:

$$\begin{aligned} \nabla \ln(\texttt{Consumption}_{it}) &= \beta_1 \nabla \ln(\texttt{Price}_{it}) + \beta_2 \nabla \ln(\texttt{Income}_{it}) + e_{it}, \\ \text{with} \quad e_{it} &= \sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it}, \end{aligned}$$

where $\nabla x_t = x_t - x_{t-1}$. In order to avoid notational mess, we use the same notation for the unobserved time-varying individual effects $v_{it} = \sum_{l=1}^{d} \lambda_{il} f_{lt}$ as above in (20). The ∇-transformation can be easily performed in R using the standard `diff()`-function as follows:

```
R> d.l.Consumption  <- diff(l.Consumption)
R> d.l.Price        <- diff(l.Price)
R> d.l.Income       <- diff(l.Income)
```

As previously mentioned for the `KSS()`-function, the `formula` argument of the `Eup()`-function takes balanced panel variables as $T \times n$ dimensional matrices, where the number of rows has to be equal to the temporal dimension T and the number of columns has to be equal to the individual dimension n.

```
R> (Cigar.Eup <- Eup(d.l.Consumption ~  -1 + d.l.Price + d.l.Income,
+                    dim.criterion = "PC3"))

Call:
Eup.default(formula = d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
    dim.criterion = "PC3")

Coeff(s) of the Observed Regressor(s) :

  d.l.Price d.l.Income
 -0.3171044  0.1838808

Additive Effects Type:  none

Dimension of the Unobserved Factors: 7

Number of iterations: 115
```

Inferences about the slope parameters can be obtained by using the method `summary()`.

```
R> summary(Cigar.Eup)

Call:
Eup.default(formula = d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
    dim.criterion = "PC3")

Residuals:
     Min       1Q   Median       3Q      Max
-0.09340 -0.01170  0.00063  0.01260  0.07690


 Slope-Coefficients:
           Estimate Std.Err Z value    Pr(>z)
d.l.Price   -0.3170  0.0237  -13.40 < 2.2e-16 ***
d.l.Income   0.1840  0.0372    4.95  7.48e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Additive Effects Type:  none

Dimension of the Unobserved Factors: 7

Residual standard error: 0.0006995 on 807 degrees of freedom,
R-squared: 0.78
```

The summary output reports that `"PC3"` detects 7 common factors. The effect of the log-real prices for cigarettes on the log-sales is negative and amounts to -0.317104. The estimated effect of the real disposable log-income per capita is 0.183882, which is smaller than the effect estimated by the method of Kneip *et al.* (2012).

The estimated factors $\hat{f}_{tl}$ as well as the individual effects $\hat{v}_{it}$ can be plotted using the `plot()`-method for `summary.Eup`-objects. The corresponding graphics are shown in Figure 4.

```
R> plot(summary(Cigar.Eup))
```

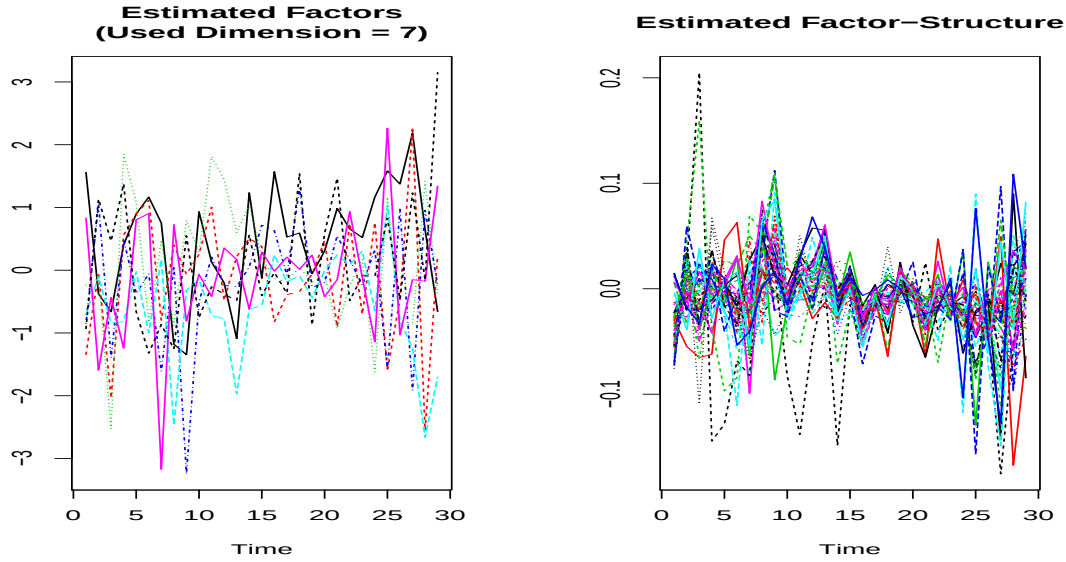

Figure 4: LEFT PANEL: Estimated factors $\hat{f}_{1t}, \ldots, \hat{f}_{7t}$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_{1t}, \ldots, \hat{v}_{nt}$.

5. Models with additive and interactive unobserved effects

Even though the classical additive `"individual"`, `"time"`, and `"twoways"` effects can be absorbed by the factor structure, there are good reasons to model them explicitly. On the one hand, if there are such effects in the true model, then neglecting them will result in non-efficient estimators; see Bai *et al.* (2009). On the other hand, additive effects can be very useful for interpretation purposes.

Consider now the following model:

$$y_{it} = \mu + \alpha_i + \theta_t + x_{it}^{\top}\beta + \nu_{it} + \epsilon_{it} \tag{46}$$

with

$$\nu_{it} = \begin{cases} v_{it} &= \sum_{l=1}^{d} \lambda_{il} f_{lt}, & \text{for the model of Bai (2009),} \\ v_i(t) &= \sum_{l=1}^{d} \lambda_{il} f_l(t), & \text{for the model of Kneip } et\ al. \text{ (2012),} \end{cases}$$

where α_i are time-constant individual effects and θ_t is a common time-varying effect.

In order to ensure identification of the additional additive effects α_i and θ_t, we need the following further restrictions:

(d) $\sum_{i=1}^{n} \lambda_{il} = 0 \quad \text{for all} \quad l \in \{1, \ldots, d\}$

(e) $\sum_{t=1}^{T} f_{lt} = 0 \quad \text{for all} \quad l \in \{1, \ldots, d\}$

(f) $\sum_{i=1}^{n} \alpha_i = 0$

(g) $\sum_{t=1}^{T} \theta_t = 0$

By using the classical within-transformations on the observed variables, we can eliminate the additive effects α_i and θ_t, such that

$$\dot{y}_{it} = \dot{x}_{it}^{\top}\beta + \nu_{it} + \dot{\epsilon}_{it},$$

where $\dot{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} - \frac{1}{n}\sum_{i=1}^{n} y_{it} + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} y_{it}$, $\dot{x}_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it} - \frac{1}{n}\sum_{i=1}^{n} x_{it} + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} x_{it}$, and $\dot{\epsilon}_{it} = \epsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it} - \frac{1}{n}\sum_{i=1}^{n} \epsilon_{it} + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} \epsilon_{it}$. Note that restrictions (d) and (e) insure that the transformation does not affect the time-varying individual effects ν_{it}.

The parameters β and ν_{it} can be estimated by the above introduced estimation procedures. All possible variants of (46) are implemented in the functions `KSS()` and `Eup()`. The appropriate model can be specified by the argument `additive.effects = c("none", "individual", "time", "twoways")`:

`"none"`	$y_{it} = \mu + x_{it}^{\top}\beta + \nu_{it} + \epsilon_{it}$
`"individual"`	$y_{it} = \mu + \alpha_i + x_{it}^{\top}\beta + \nu_{it} + \epsilon_{it}$
`"time"`	$y_{it} = \mu + \theta_t + x_{it}^{\top}\beta + \nu_{it} + \epsilon_{it}$
`"twoways"`	$y_{it} = \mu + \alpha_i + \theta_t + x_{it}^{\top}\beta + \nu_{it} + \epsilon_{it}.$

The presence of μ can be controlled by `-1` in the `formula`-object: a formula with `-1` refers to a model without intercept. However, for identification purposes, if a `twoways` model is specified, the presence `-1` in the `formula` will be ignored.

As an illustration we continue with the application of the `KSS()`-function in Section 2. The left panel of Figure 2 shows that one of the six estimated factors is nearly time-invariant. This motivates us to augment the model (20) for a time-constant additive effects α_i. In this case it is convenient to use an intercept μ, which yields the following model:

$$\begin{aligned} \ln(\texttt{Consumption}_{it}) &= \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + \alpha_i + v_i(t) + \varepsilon_{it}, \qquad (47) \\ \text{where} \quad v_i(t) &= \sum_{l=1}^{d} \lambda_{il}\, f_l(t). \end{aligned}$$

The estimation of the augmented model (47) can be done using the following lines of code.

```
R> Cigar2.KSS <- KSS(formula = l.Consumption ~ l.Price + l.Income,
+                    additive.effects = "individual")
R> (Cigar2.KSS.summary <- summary(Cigar2.KSS))

Call:
KSS.default(formula = l.Consumption ~ l.Price + l.Income, additive.effects = "individual")

Residuals:
   Min     1Q Median     3Q    Max
 -0.11  -0.01   0.00   0.01   0.12


 Slope-Coefficients:
            Estimate  StdErr z.value    Pr(>z)
(Intercept)   4.0500  0.1760   23.10 < 2.2e-16 ***
l.Price      -0.2600  0.0222  -11.70 < 2.2e-16 ***
l.Income      0.1570  0.0381    4.11  3.88e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type:  individual

Used Dimension of the Unobserved Factors: 5

Residual standard error: 0.000734 on 951 degrees of freedom
R-squared: 0.99
```

Again, the `plot()` method provides an useful visualization of the results.

```
R> plot(Cigar2.KSS.summary)
```

The `"individual"`-transformation of the data does not affect the estimation of the slope parameters but reduces the estimated dimension from $\hat{d} = 6$ to $\hat{d} = 5$. The remaining five common factors $\hat{f}_1, \ldots, \hat{f}_5$ correspond to those of model (20); see the middle panel of Figure 5. The estimated time-constant state-specific effects α_i are shown in the left plot of Figure 5. The extraction of the α_i's from the factor structure yields a denser set of time-varying individual effects $\hat{v}_i$ shown in the right panel of Figure 5.

5.1. Specification tests

Model specification is an important step for any empirical analysis. The **phtt** package is equipped with two types of specification tests: the first is a Hausman-type test appropriate for the model of Bai *et al.* (2009); see Section 5.1.1. The second one examines the existence of a factor structure in Bai's model as well as in the model of Kneip *et al.* (2012); see Section 5.1.2.

Testing the sufficiency of classical additive effects

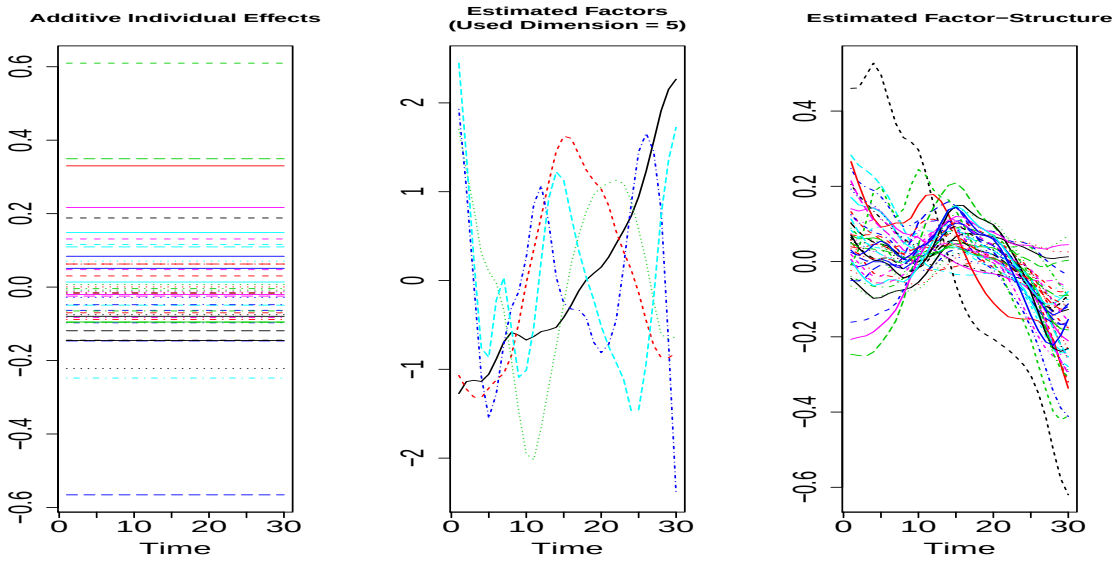

Figure 5: LEFT PANEL: Estimated time-constant state-specific effects $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$. MIDDLE PANEL: Estimated common factors $\hat{f}_1(t), \ldots, \hat{f}_5(t)$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_1(t), \ldots, \hat{v}_n(t)$.

For the case in which there are only one or two factors ($1 \leq d \leq 2$), it is interesting to check whether or not these factors can be interpreted as classical `"individual"`, `"time"`, or `"twoways"` effects. Bai *et al.* (2009) considers the following testing problem:

$$\begin{aligned} H_0&: \quad v_{it} = \alpha_i + \theta_t \\ H_1&: \quad v_{it} = \textstyle\sum_{l=1}^{2} \lambda_{il} f_{lt} \end{aligned}$$

The model with factor structure, as described in Section 4, is consistent under both hypotheses. However, it is less efficient under H_0 than the classical within estimator, while the latter is inconsistent under H_1 if x_{it} and v_{it} are correlated. These conditions are favorable for applying the Hausman test:

$$J_{Bai} = nT \left(\hat{\beta} - \hat{\beta}_{within}\right) \Delta^{-1} \left(\hat{\beta} - \hat{\beta}_{within}\right) \overset{a}{\sim} \chi^2_P, \tag{48}$$

where $\hat{\beta}_{within}$ is the classical within least squares estimator, $\Delta = Var\left(\hat{\beta} - \hat{\beta}_{within}\right)$, P is the vector-dimension of β, and χ^2_P is the χ^2-distribution with P degree of freedom.
The null hypothesis H_0 can be rejected, if $J_{Bai} > \chi^2_{P,1-\alpha}$, where $\chi^2_{P,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2 distribution with P degrees of freedom.
To calculate J_{Bai} we can replace Δ by its consistent estimator

$$\hat{\Delta} = \left(\left(\frac{1}{nT} \sum_{i=1}^{n} Z_i^\top Z_i \right)^{-1} - \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_i^\top \dot{x}_i \right)^{-1} \right) \hat{\sigma}^2, \tag{49}$$

where

$$\hat{\sigma}^2 = \frac{1}{nT - (n+T)\hat{d} - P + 1} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - x_{it}^\top \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{il} \hat{f}_{lt})^2. \tag{50}$$

The test is implemented in the function `checkSpecif()`, which takes the following arguments:

```
R> checkSpecif(obj1, obj2, level = 0.05)
```

The arguments `obj1` and `obj2` take both objects of class `Eup` produced by the function `Eup()`:

`obj1` Takes an `Eup`-object from an estimation with `"individual"`, `"time"`, or `"twoways"` effects and a factor dimension equal to $d = 0$; specified as `factor.dim = 0`.

`obj2` Takes an `Eup`-object from an estimation with `"none"`-effects and a positive factor dimension $1 \leq d \leq 2$:
`factor.dim=1` for testing `"individual"` or `"time"` effects.
`factor.dim=2` for testing `"twoways"` effects.

The argument `level` is used to specify the significance level.

However, the Hausman test of Bai *et al.* (2009) has a clear disadvantage. It is applicable only in situations of one or two factors ($1 \leq d \leq 2$). This is, e.g., not fulfilled in our demonstrations using the `Cigar` dataset, where the estimated factor dimension $\hat{d}$ lies between six and seven; see Figures 2 and 4. The problem is that in such cases the matrix $\hat{\Delta}$ in (49) can become negative definite, which would yield a negative test statistic J_{Bai} in (48). If the test is applied in such situations, an error message is printed:

```
R> twoways.obj     <- Eup(d.l.Consumption ~  -1 + d.l.Price + d.l.Income,
+                         factor.dim = 0, additive.effects = "twoways")
R> not.twoways.obj <- Eup(d.l.Consumption ~  -1 + d.l.Price + d.l.Income,
+                         factor.dim = 2, additive.effects = "none")
R> checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj, level = 0.01)

Error in checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj, level = 0.01) :
The assumptions of the test are not fulfilled.
The (unobserved) true number of factors is probably greater than 2.
```

An alternative test for the sufficiency of a classical additive effects model is given by the following test proposed by Kneip *et al.* (2012). This test can be applied for arbitrary factor dimensions d.

Testing the existence of common factors

This section is concerned with testing the existence of common factors. In contrast to the Hausman type statistic discussed above, the goal of this test is not merely to decide which model specification is more appropriate for the data, but rather to test in general the existence of common factors beyond the eventual presence of additional classical `"individual"`, `"time"` , or `"twoways"` effects in the model.

This test relies on using the dimensionality criterion proposed by Kneip *et al.* (2012) to test the following hypothesis after eliminating eventual additive `"individual"`, `"time"`, or `"twoways"` effects:

$$\begin{aligned} H_0&: \quad d = 0 \\ H_1&: \quad d > 0 \end{aligned}$$

Under H_0 the slope parameters β can be estimated by the classical within estimation method. In this simple case, the dimensionality test of Kneip *et al.* (2012) can be reduced to the following test statistic:

$$J_{KSS} = \frac{n\, tr(\hat{\Sigma}_w) - (n-1)(T-1)\hat{\sigma}^2}{\sqrt{2n}(T-1)\hat{\sigma}^2} \overset{a}{\sim} N(0,1),$$

where $\hat{\Sigma}_w$ is the covariance matrix of the within residuals. The reason for this simplification is that under H_0 there is no need for smoothing, which allows us to set $\kappa = 0$.

We reject H_0: $d = 0$ at a significance level α if $J_{KSS} > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution. It is important to note that the performance of the test depends heavily on the accuracy of the variance estimator $\hat{\sigma}^2$. We propose to use the variance estimators (16) or (50), which are consistent under both hypotheses as long as $\hat{d}$ is greater than the unknown dimension d. Internally, the test procedure sets $\hat{d}$ =`d.max`.

This test can be performed for `Eup`- as well as for `KSS`-objects by using the function `checkSpecif()` leaving the second argument `obj2` unspecified. In the following we apply the test for both models:

```
R> Eup.obj <- Eup(d.l.Consumption ~  -1 + d.l.Price + d.l.Income,
+                 additive.effects = "twoways")
R> checkSpecif(Eup.obj, level = 0.01)

-----------------------------------------------
Testing the Presence of Interactive Effects
Test of Kneip, Sickles, and Song (2012)
-----------------------------------------------
H0: The factor dimension is equal to 0.

Test-Statistic        p-value    crit.-value     sig.-level
         13.29           0.00           2.33           0.01

R> KSS.obj <- KSS(l.Consumption ~  -1 + l.Price + l.Income,
+                 additive.effects = "twoways")
R> checkSpecif(KSS.obj, level = 0.01)

-----------------------------------------------
Testing the Presence of Interactive Effects
Test of Kneip, Sickles, and Song (2012)
-----------------------------------------------
H0: The factor dimension is equal to 0.

Test-Statistic        p-value    crit.-value     sig.-level
     104229.55           0.00           2.33           0.01
```

The null hypothesis H_0: $d = 0$ can be rejected for both models at a significance level $\alpha = 0.01$.

6. Interpretation

This section is intended to outline an exemplary interpretation of the panel model (47), which is estimated by the function `KSS()` in Section 5. The interpretation of models estimated by the function `Eup()` can be done accordingly. For convenience sake we re-write the model (47) in the following:

$$\begin{aligned} \ln(\texttt{Consumption}_{it}) &= \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + \alpha_i + v_i(t) + \varepsilon_{it}, \\ \text{where} \quad v_i(t) &= \sum_{l=1}^{d} \lambda_{il} f_l(t). \end{aligned}$$

A researcher, who chooses the panel models proposed by Kneip *et al.* (2012) or Bai (2009) will probably find them attractive due to their ability to control for very general forms of unobserved heterogeneity. Beyond this a further great advantage of these models is that the time-varying individual effects $v_i(t)$ provide a valuable source of information about the *differences* between the individuals i. These differences are often of particular interest as, e.g., in the literature on stochastic frontier analysis.

The left panel of Figure 5 shows that the different states i have considerable different time-constant levels $\hat{\alpha}_i$ of cigarette consumption. A classical further econometric analysis could be to regress the additive individual effects $\hat{\alpha}_i$ on other time-constant variables, such as the general populations compositions, the cigarette taxes, etc.

The right panel of Figure 5 shows the five estimated common factors $\hat{f}_1(t), \ldots, \hat{f}_5(t)$. It is a good practice to start the interpretation of the single common factors with an overview about their importance in describing the differences between the $v_i(t)$'s, which is reflected in the variances of the individual loadings parameters $\hat{\lambda}_{il}$. A convenient depiction is the quantity of variance-shares of the individual loadings parameters on the total variance of the loadings parameters

$$\texttt{coef(Cigar2.KSS)\$Var.shares.of.loadings.param[}l\texttt{]} = \mathrm{V}(\hat{\lambda}_{il}) / \sum_{k=1}^{\hat{d}} \mathrm{V}(\lambda_{ik}),$$

which is shown for all common functions $\hat{f}_1(t), \ldots, \hat{f}_5(t)$ in the following table:

Common Factor	Share of total variance of $v_i(t)$
$\hat{f}_1(t)$	`coef(Cigar2.KSS)$Var.shares.of.loadings.param[1]` $= 66.32\%$
$\hat{f}_2(t)$	`coef(Cigar2.KSS)$Var.shares.of.loadings.param[2]` $= 24.28\%$
$\hat{f}_3(t)$	`coef(Cigar2.KSS)$Var.shares.of.loadings.param[3]` $= 5.98\%$
$\hat{f}_4(t)$	`coef(Cigar2.KSS)$Var.shares.of.loadings.param[4]` $= 1.92\%$
$\hat{f}_5(t)$	`coef(Cigar2.KSS)$Var.shares.of.loadings.param[5]` $= 1.50\%$

Table 1: List of the variance shares of the common factors $\hat{f}_1(t), \ldots, \hat{f}_5(t)$.

The values in Table 1 suggest to focus on the first two common factors, which explain together about 90% of the total variance of the time-varying individual effects $\hat{v}_i(t)$.

The first two common factors

$$\texttt{coef(Cigar2.KSS)\$Common.factors[,1]} = \hat{f}_1(t) \quad \text{and}$$
$$\texttt{coef(Cigar2.KSS)\$Common.factors[,2]} = \hat{f}_2(t)$$

are plotted as black solid and red dashed lines in the middle panel of Figure 5. Figure 6 visualizes the differences of the time-varying individual effects $v_i(t)$ in the direction of the first common factor (i.e: $\hat{\lambda}_{i1}\hat{f}_1(t)$) and in the direction of the second common factor (i.e: $\hat{\lambda}_{i2}\hat{f}_2(t)$). As for the time-constant individual effects $\hat{\alpha}_i$ a further econometric analysis could be to regress the individual loadings parameters $\hat{\lambda}_{i1}$ and $\hat{\lambda}_{i2}$ on other explanatory time-constant variables.

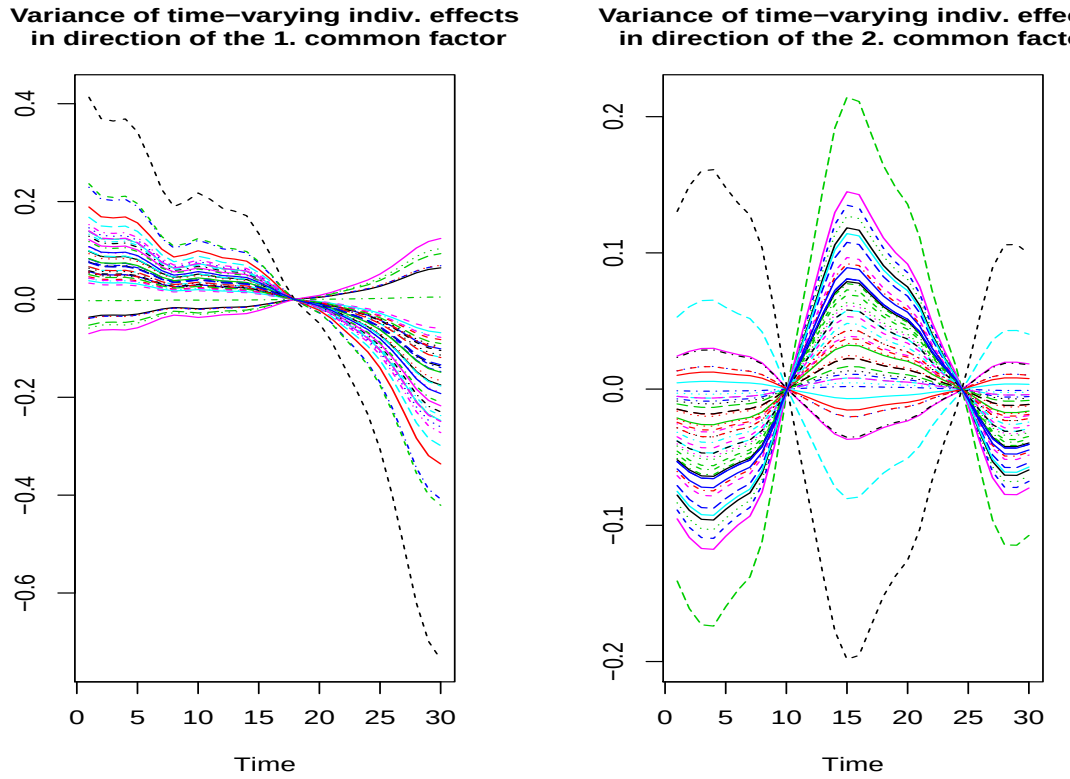

Figure 6: Left Panel: Visualization of the differences of the time-varying individual effects $v_i(t)$ in the direction of the first factor $\hat{f}_1(t)$ (i.e: $\hat{\lambda}_{i1}\hat{f}_1(t)$). Right Panel: Visualization of the differences of the time-varying individual effects $v_i(t)$ in the direction of the second factor $\hat{f}_2(t)$ (i.e: $\hat{\lambda}_{i2}\hat{f}_2(t)$).

Generally, for both models proposed by Kneip *et al.* (2012) and Bai (2009) the time-vaying individual effects

$$\nu_{it} = \sum_{l=1}^{d} \lambda_{il} f_{lt}$$

can be interpreted as it is usually done in the literature on factor models. An important topic that is not covered in this section is the rotation of the common factors. Often, the common factors f_l can be interpreted economically only after the application of an appropriate rotation scheme for the set of factors $\hat{f}_1, \ldots, \hat{f}_{\hat{d}}$. The latter can be done, e.g., using

the function `varimax()` from the **stats** package. Sometimes it is also preferable to standardize the individual loadings parameters instead of the common factors as it is done, e.g., in Ahn, Hoon Lee, and Schmidt (2001). This can be done by choosing `restrict.mode = c("restrict.loadings")` in the functions `KSS()` and `Eup()` respectively.

7. Summary

This paper introduces the R package **phtt** for the new class of panel models proposed by Bai (2009) and Kneip *et al.* (2012). The two main functions of the package are the `Eup()`-function for the estimation procedure proposed in Bai (2009) and the `KSS()`-function for the estimation procedure proposed in Kneip *et al.* (2012). Both of the main functions are supported by several print, summary, and plot methods. While parts of the method of Bai (2009) are available for commercially available software packages, the estimation procedure proposed by Kneip *et al.* (2012) is not available elsewhere. A further remarkable feature of our **phtt** package is the `OptDim()`-function, which provides an ease access to many different dimensionality criteria proposed in the literature on factor models. The usage of the functions is demonstrated in several real data applications.

References

Ahn S, Hoon Lee Y, Schmidt P (2001). "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects." *Journal of Econometrics*, **101**(2), 219–255.

Ahn S, Horenstein AR (2009). "Eigenvalue Ratio Test for the Number of Factors." *Working Paper, Arizona State University*.

Ahn S, Lee Y, Schmidt P (2006). "Panel Data Models with Multiple Time-Varying Individual Effects." *Working Paper 0702, University of Crete, Department of Economics*.

Bada O, Kneip A (2010). "Panel Data Models with Unobserved Multiple Time-Varying Effects to Estimate the Risk Premium of Corporate Bonds." *Working Paper, University of Bonn*.

Bai J (2004). "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data." *Journal of Econometrics*, **122**(1), 137–183.

Bai J (2009). "Panel Data Models with Interactive Fixed Effects." *Econometrica*, **77**(4), 1229–1279.

Bai J, Kao C, Ng S (2009). "Panel Cointegration with Global Stochastic Trends." *Journal of Econometrics*, **149**(1), 82–99.

Bai J, Ng S (2002). "Determining the Number of Factors in Approximate Factor Models." *Econometrica*, **70**(1), 191–221.

Baltagi B (2005). *Econometric Analysis of Panel Data*. Third edition. Wiley.

Baltagi B, Levin D (1986). "Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation and Advertising Reconsidered." *The Review of Economics and Statistics*, pp. 148–155.

Baltagi B, Li D (2004). *Prediction in the Panel Data Model with Spatial Correlation*. First edition. Springer-Verlag.

Bates D, Maechler M, Bolker B (2012). ***lme4****: Linear Mixed-Effects Models Using S4 Classes*. R package version 0.999375-42, URL http://CRAN.R-project.org/package=lme4.

Cao J, Ramsay J (2010). "Linear Mixed-Effects Modeling by Parameter Cascading." *Journal of the American Statistical Association*, **105**(489), 365–374.

Craven P, Wahba G (1978). "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation." *Numerische Mathematik*, **31**(4), 377–403.

Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The **plm** Package." *Journal of Statistical Software*, **27**(2), 1–43.

De Boor C (2001). *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Series*. Revised edition. Springer-Verlag.

Greenaway-McGrevy R, Han C, Sul D (2012). "Asymptotic Distribution of Factor Augmented Estimators for Panel Regression." *Journal of Econometrics (Forthcoming)*.

Kneip A, Sickles RC, Song W (2012). "A New Panel Data Treatment for Heterogeneity in Time Trends." *Econometric Theory*, **28**(3), 590–628.

Millo G, Piras G (2012). "**splm**: Spatial Panel Data Models in R." *Journal of Statistical Software*, **47**(1), 1–38. URL http://www.jstatsoft.org/v47/i01.

Onatski A (2010). "Determining the Number of Factors from Empirical Distribution of Eigenvalues." *The Review of Economics and Statistics*, **92**(4), 1004–1016.

Pesaran HM (2006). "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." *Econometrica*, **74**(4), 967–1012.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core team (2012). ***nlme**: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-103, URL http://CRAN.R-project.org/package=nlme.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL http://www.R-project.org/.

The MathWorks Inc (2012). *MATLAB – The Language of Technical Computing, Version 7.14*. The MathWorks, Inc., Natick, Massachusetts. URL http://www.mathworks.com/products/matlab.

Affiliation:

Oualid Bada and Dominik Liebl
Statistische Abteilung
University of Bonn
Adenauerallee 24-26
53113 Bonn, Germany
E-mail: bada@uni-bonn.de
E-mail: dliebl@uni-bonn.de