# Essays on Large Panel Data Models

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Oualid Bada
aus Monastir (Tunesien)

Bonn 2015

| | |
|---|---|
| Dekan: | Prof. Dr. Rainer Hüttemann |
| Erstreferent: | Prof. Dr. Alois Kneip |
| Zweitreferent: | Prof. Dr. Robin Sickles |

Tag der mündlichen Prüfung: 13.03.2015

UNIVERSITY OF BONN

# *Abstract*

Department of Economics
Institute for Financial Economics and Statistics

Doctoral Thesis

**Essays on Large Panel Data Models**

by Oualid Bada

The standard panel data literature is moving from micro panels, where the cross-section dimension is large and the intertemporal sample size is small, to large panels, where both, the cross-section and the time dimension, are large. This thesis contributes to this new and growing area of panel data treatments called *"large panel data analysis"*. My dissertation consists of three essays: In the first essay, a large panel data model with an omitted factor structure is considered. The role of the factors is to control for the issue of the unobserved time-varying heterogeneity effects. A parameter cascading strategy is proposed to enable efficient estimation of all model parameters when the number of factors is unknown a priori. In the second essay, further models that combine large panel data models with different versions of unobserved latent factors are discussed. Computation-related issues are solved and new specification tests are introduced to check whether or not these factors can be interpreted as classical additive fixed effects. In the third essay, a novel method for estimating panel models with multiple structural changes is proposed. The breaks are allowed to occur at unknown points in time and may affect the multivariate slope parameters individually. Asymptotic results are derived, Monte Carlo experiments are performed, and applications for highlighting these new methods are discussed.

# *Acknowledgements*

I would like to express my heartfelt gratitude to the many people who have supported me in writing this thesis. First, I am deeply indebted and thankful to my supervisor Prof. Dr. Alois Kneip for his continuous guidance and vital advice during the writing of this thesis. His insightful comments helped me to enhance my understanding of advanced and state-of-the-art statistical techniques. He has been an enormous source of inspiration for my research and motivated me to keep improving my papers and the way they might contribute to my research area.

I would like to sincerely thank my co-author Dr. Robin Sickles, Professor of Econometrics at Rice University, for the extraordinary collaboration on the project that constitutes the core of Chapter 3. I am grateful for his time, his enthusiasm, and his generosity during my stay as a visiting researcher in the Department of Econometrics at Rice University. I also want to thank our co-author, James Gualtieri, for investing so much time in collecting data and contributing to the development of the application part.

I would also like to express my gratitude to my co-author and my friend JProf. Dr. Dominik Liebl, who contributed to the joint paper presented in Chapter 2. I benefited much from his numerous helpful remarks and creative ideas. I will miss his extraordinary talent to identify and simplify substantially important and complex statistical problems.

In addition, I would like to give my gratitude to all my colleagues in the department of statistics in the Institute for Financial Economics and Statistics of the University of Bonn: Prof. Dr. Lorens Imhof, Heiko Wagner, Dominik Poß, Prof. Dr. Hans-Joachim Werner, Dr. Klaus Utikal, Fabian Walders, Martin Arndt, and Hildegard Grober. They not only provided me technical help, but also helped me to improve my German language skills and taught me a lot of unconventional idioms.

I would like to express my deepest gratitude to my parents Emna and Romdhan, my uncle Khaled, my aunt Metira, my sister Wihed, and my brother Wissem. I could not have had the opportunity to study in Bonn and to write this dissertation without them. I also would like to thank all my family members, in particular, Naima Moussa, Ali Abbes, Fatma Boughtass, Rami Abbes, Ameur Abbes, Safouane Jguirim, Safa Baccouche Abbes, and Ghada Abada as well as my friends Kacem Kassraoui, Walid Hamdi, and Nadhmi Nefzi. They have played an important moral-support role.

Last but not least, to my wonderful wife Abir, who has supported me in good and bad times during all phases of this thesis: I love you.

# Contents

# List of Figures

# List of Tables

*I would like to dedicate this thesis to my beloved parents, my wonderful wife, and my admirable daughter.*

# Introduction

The standard panel data literature is moving from micro panels, where the cross-section dimension is large and the intertemporal sample size is small, to large panels, where both, the cross-section and the time dimension, are large. This thesis contributes to this new and growing area of panel data treatments called *"large panel data analysis"*. My dissertation consists of three essays: In the first essay, a large panel data model with an omitted factor structure is considered. The role of the factors is to control for the issue of the unobserved time-varying heterogeneity effects. A parameter cascading strategy is proposed to enable efficient estimation of all model parameters when the number of factors is unknown a priori. In the second essay, further models that combine large panel data models with different versions of unobserved latent factors are discussed. Computation-related issues are solved and new specification tests are introduced to check whether or not these factors can be interpreted as classical additive fixed effects. In the third essay, a novel method for estimating panel models with multiple structural changes is proposed. The breaks are allowed to occur at unknown points in time and may affect the multivariate slope parameters individually. Asymptotic results are derived, Monte Carlo experiments are performed, and applications for highlighting these new methods are discussed.

Due to the impressive progress of information technology, econometricians and statisticians, nowadays, have the privilege of working with large dimensional data sets. In the panel data literature, this has been succeeded by the opening of new research perspectives. Recent studies have discussed large panel data models in which the unobserved heterogeneity can be estimated by an *"approximate factor structure"*. The latter, unlike the standard setup of factor models, allows for the presence of weak forms of cross-section and time-series dependence in the idiosyncratic components. The extended regression models combining conventional panel models and factor structures provide a generalization of classical panel models with additive heterogeneity effects. Indeed, it allows for the individual specific effects to be affected by unobserved common time-shocks.

The first chapter is joint work with Alois Kneip. Our paper is published in *Computational Statistics and Data and Analysis*; see Bada and Kneip (2014). In this chapter, we extend the iterated least squares approach of Bai (2009) and Bai et al. (2009) in such a way that we allow the number of factors to be unknown a priori. We propose inserting a penalty term into the objective function to be globally optimized and updating iteratively the estimators of all required parameters in hierarchical order. Our parameter-cascading strategy also includes the update of the penalty term in order to adjust the height of the penalization and to avoid under- or over-parameterization effects. We allow for the presence of stationary and non-stationary factors and discuss the case in which the static factor presentation arises from a small number of dynamic primitive shocks. We show that our extension does not affect the asymptotic properties of the iterated least squares estimator. Our Monte Carlo experiments confirm that, in many configurations of the data, such a refinement provides more efficient estimates in terms of MSE than those that could be achieved if the feasible iterative least squares estimator is calculated with an externally selected factor dimension. In our application, we consider the problem of the so-called *"credit spread puzzle"*–the gap between the observed corporate bond yields and duration-equivalent government bond yields that classical financial models of credit risk fail to explain. Our empirical study detects the presence of two unobserved primitive risk factors affecting the U.S. corporate bond market during the period between September 2006 and March 2008.

The second chapter is written in collaboration with Domink Liebl. Our paper is published in the *Journal of Statistical Software*; see Bada and Liebl (2014b). This chapter focuses on important computational aspects related to the panel methods of Bai (2009) and Kneip et al. (2012). As the estimation procedure of Kneip et al. (2012) involves nonparametric estimation techniques, the choice of an appropriate smoothing parameter can be very crucial in practice. We propose using a slightly modified version of the Generalized Cross Validation (GCV) criterion to determine an upper bound for the optimal smoothing parameter. Using this, we can obtain an enormous gain in terms of computation compared to the cross validation criterion proposed by Kneip et al. (2012). Bai's method relies on an iterated least squares approach and also requires the determination of external parameters such as the number of factors to be considered and appropriate starting values to initiate the iteration process. We propose using the starting estimator and the iteration scheme of Bada and Kneip (2014). Additionally, we review a wide range of recently developed dimensionality criteria that can be applied in this context. To examine the significance of the factors and test whether a model with factor structure is more appropriate than a simple panel model with additive effects, we propose using two testing procedures: a Hausman-type test and a $z$-test based on the method of Kneip et al. (2012). All these methods are implemented in an R package called **phtt** (see Bada

and Liebl (2014a)). To the best of our knowledge, **phtt** is the first software package that provides this large number of methods for analyzing large panel data. To demonstrate the functionality of our package, we re-explore the well known `Cigar` dataset and re-visit the model of Baltagi and Li (2006) to allow for the presence of an omitted factor structure in the idiosyncratic components.

The third chapter is based on joint work with Alois Kneip, James Gualtieri, and Robin Sickles. This chapter proposes a novel method for estimating panel models with multiple structural changes that may occur at unknown points in time. In spite of the growing literature on large panel data analysis, there is an important issue that is scarcely discussed in most of the existing work –the risk of neglecting structural breaks in the data generating process, especially when the observation period is large. Our model generalizes the special model specification in which the slope parameters are time homogeneous. Our theory considers breaks in a two-way panel data model, in which the unobserved heterogeneity is composed of additive individual effects and time specific effects. We show that our method can also be extended to cover the case of panel models with unobserved heterogeneous common factors as proposed by Bai (2009), Kneip et al. (2012), Ahn et al. (2001), Pesaran (2006), and Bada and Kneip (2014). We provide a general setup allowing for endogenous models such as dynamic panel models and/or structural models with simultaneous panel equations. Consistency under weak forms of dependency and heteroscedasticity in the idiosyncratic errors is established and the convergence rate of our slope estimator is derived. To detect the jump locations consistently and test for the statistical significance of the breaks, we propose post-wavelet procedures. Our simulations show that, in many configurations of the data, our method performs very well. Our empirical vehicle for highlighting this new methodology addresses the stability of the relationship between Algorithmic Trading (AT) and Market Quality (MQ). We find evidence that the relationship between AT and MQ was disrupted between 2007 and 2008.

# Chapter 1

# Panel Models with Unknown Number of Unobserved Factors: An Application to the Credit Spread Puzzle

## 1.1 Introduction

In recent years, the use of panel data has attracted increasing attention in many empirical studies. This is motivated by the fact that such data sets allow statisticians to deal with the problem of the *unobserved heterogeneity*. Recent studies have discussed large panel data models in which the unobserved heterogeneity can be modeled by a factor structure; see, e.g., Bai (2009), Bai et al. (2009), Kneip et al. (2012), and Pesaran (2006). While most of the ongoing studies have focused on fitting the model for a given number of factors, the present work considers the problem of estimating the factor dimension jointly with the unknown model parameters. Our estimation algorithm can be applied for models of the form:

$$Y_{it} = X_{it}\beta + \underbrace{F_t\Lambda_i'}_{(1\times d)\times(d\times 1)} + \epsilon_{it} \quad \text{for} \quad i \in \{1, \cdots, N\} \quad \text{and} \; ; t \in \{1, \cdots, T\}, \tag{1.1}$$

where $X_{it}$ is a $(1 \times p)$ vector of observable regressors, $\beta$ is a $(p \times 1)$ vector of unknown parameters, $\Lambda_i$ is a $(1 \times d)$ vector of individual scores (or factor loadings), $F_t$ is a $(1 \times d)$ vector of unobservable common time-varying factors, $\epsilon_{it}$ is the error term, and $d$ is an unknown integer, which has to be estimated jointly with $\beta, \Lambda_i$ and $F_t$.

The difference between (1.1) and the classical panel data models consists in the unobserved factor structure $F_t\Lambda_i^{'}$. It is noted that (1.1) not only provides a generalization of panel data models with additive effects, where $d = 2$, $F_t = (f_{t,1}, 1)$, and $\Lambda_i = (1, \lambda_{2,i})$, but also includes the dynamic factor models in static form as in Stock and Watson (2005). To illustrate this case, consider a static factor model with autocorrelated idiosyncratic errors of order $P$ such that

$$\begin{aligned} Y_{it} &= F_t^*\Lambda_i^{'} + e_{it} \text{ and} \\ e_{it} &= \beta_1 e_{i,t-1} + \ldots + \beta_P e_{i,t-P} + \epsilon_{it}. \end{aligned}$$

It is easily verified that integrating the expansion of $e_{it}$ in the first equation and using $e_{i,t-p} = Y_{i,t-p} - F_{t-p}^*\Lambda_i^{'}$ for each $p = 1, \ldots, P$ results in a panel model of form (1.1), where the regressors are the lags of $Y_{it}$, i.e., $X_{it} = (Y_{i,t-1}, \ldots, Y_{i,t-P})$, and $F_t = F_t^* - \beta_1 F_{t-1}^* - \cdots - \beta_P F_{t-P}^*$.

Moreover, the static presentation of the unobserved factor structure in (1.1) can arise from $q$-dimensional dynamic factors, say $\mathbf{F}_t$, (also called primitive factors or primitive shocks). In this case, $F_t = [\mathbf{F}_t, \ldots, \mathbf{F}_{t-m}]$ and $d = q(m+1) \geq q$.

Stock and Watson (2005) propose to estimate dynamic factor models in static form by the iterated least squares method (also called iterative principal component). Bai (2009) studies (1.1) in the context of panel data models and provides asymptotic theory for the iterative least squares estimators when both $N$ and $T$ are large. However, Stock and Watson (2005) and Bai (2009) assume the factors to be stationary. Bai et al. (2009) extend the theoretical development of Bai (2009) to the case where the cross-sections share unobserved common stochastic trends of unit root processes. They prove that the asymptotic bias arising from the time series in such a case can be consistently estimated and corrected. Ahn et al. (2013) consider the classical case where $T$ is small and $N$ is large and estimate the model by using the generalized method of moments (GMM). They show that, under fixed $T$, the GMM estimator is more efficient than the estimator of Bai (2009).

A second criticism of the conventional iterative least squares method is that the number of unobserved factors has to be known a priori. In this regard, Bai and Ng (2002) and Bai (2004) propose new panel information criteria to assess the number of significant factors in large panels. The performance of these criteria depends, however, on the choice of an appropriate maximal number of factors to be considered in the selection procedure. Hallin and Liška (2007) propose similar criteria in the context of generalized dynamic factor models and provide a calibration strategy to adjust the height of the penalization; however, the calibration requires extensive computations. Alternatively, Kapetanios (2010) proposes a threshold approach based on the empirical distribution properties of

the largest eigenvalue. The method requires i.i.d. errors. Onatski (2010) extends the approach of Kapetanios (2010) by allowing the errors to be either serially correlated or cross-sectionally dependent. Onatski (2009) proposes a test statistic based on the ratios of adjacent eigenvalues in the case of Gaussian errors. Alternative methods for assessing the number of factors in the context of principal component analysis and classical factor analysis can be found in Josse and Husson (2012), Dray (2008), and Chen et al. (2010). But note that all these approaches assume the factors to be extracted directly from observed variables and not estimated with other model parameters.

Kneip et al. (2012) consider the case of observed regressors and unobserved common factors and propose a semi-parametric estimation method and a sequential testing procedure to assess the dimension of the unobserved factor structure. The asymptotic properties of their approach rely on second order differences of the factors and i.i.d. idiosyncratic errors. Pesaran (2006) attempts to control for the hidden factor structure by introducing additional regressors into the model, which are the cross-section averages of the dependent variables and the cross-section averages of the observed explanatory variables. The advantage of this estimation procedure is its invariance to the unknown factor dimension. However, the method does not aim to consistently estimate the factor structure but only deals with the problem of its presence when estimating the remaining model parameters.

In this chapter, we extend the iterative approach of Bai (2009) and Bai et al. (2009) in such a way that we allow for the number of factors to be unknown a priori. We integrate a penalty term into the objective function to be globally optimized and update iteratively the estimators of all required parameters in hierarchical order as described in Cao and Ramsay (2010). Our parameter-cascading strategy also includes the update of the penalty term in order to adjust the height of the penalization and avoid under- and over-parameterization. Monte Carlo experiments show that, in many configurations of the data, such a refinement provides more efficient estimates in terms of MSE than those that could be achieved if the feasible iterative least squares estimator is calculated with an externally selected factor dimension.

There are a lot of examples where the determination of the number of factors in the presence of additional observed regressors is of particular interest. As an example, we consider, in our application section, the problem of the so-called *credit spread puzzle*– the gap between the observed corporate bond yields and duration-equivalent government bond yields that classical financial models of credit risk fail to explain (see, e.g., Huang and Huang (2012), and Elton et al. (2001)). For a long time, the credit spread has been considered a simple compensation for credit default risk. Most empirical studies show, however, that default risk cannot be the unique explanatory variable. Kagraoka

(2010) decomposes the credit spread into credit risk, illiquidity risk, and an unobservable risk component, which he defines as systematic risk premium; however, he assumes the unobserved part to be generated by only one factor. Castagnetti and Rossi (2011) adopt a heterogeneous panel model with multiple factors. Their results suggest that credit spreads are driven by observable as well as unobservable common risk factors.

In our application, we extend the empirical development of Kagraoka (2010) by allowing for the systematic risk premium to be composed of multiple hidden factors. Moreover, we allow for some slope coefficients to be temporally heterogeneous. This differs from the setting of Castagnetti and Rossi (2011), who use a panel model with cross-sectionally heterogeneous slope parameters. Our empirical study relies on daily observations of 111 U.S. corporate bonds over a period of 397 business days. Our results suggest the presence of two unobserved primitive risk factors affecting U.S. corporate bonds during the period between September 2006 and March 2008, while one single factor is sufficient to describe the data for the time periods prior to the beginning of the subprime crisis in 2007.

The remainder of this chapter is organized as follows: Section 2 proposes an algorithmic refinement of the conventional iterative least squares estimation method. In Section 3, we extend the model with additional nominal variables, discuss the model assumptions and present some asymptotic results. Section 4 presents the results of our Monte Carlo simulations. Section 5 describes the empirical application and interprets the results. Conclusions and remarks are provided in Section 6.

## 1.2 Model and Estimation Algorithm

### 1.2.1 Model Identification and Estimation

In vector and matrix notation the model is written as

$$Y_i = X_i\beta + F\Lambda_i' + \epsilon_i, \tag{1.2}$$

where $Y_i = (Y_{i1}, \cdots, Y_{iT})'$, $X_i' = (X_{i1}', \cdots, X_{iT}')'$, $F = (F_1', \ldots, F_T')'$, $\Lambda_i$ is a $(N \times d)$ matrix of loading parameters, and $\epsilon_i$ is a $(T \times 1)$ vector of idiosyncratic errors.

The basic idea of our extension is to treat the conventionally iterated least squares estimators as functions dependent on a run parameter $\mathbf{d}$. The latter is fitted by means of a penalty term that is directly integrated into the global objective function to be optimized. The final solution is obtained by alternating between an inner iteration to

optimize $\hat{\beta}(\mathbf{d}), \hat{F}(\hat{\beta}, \mathbf{d})$, and $\hat{\Lambda}_i(\hat{\beta}, \hat{F}, \mathbf{d})$ for each given $\mathbf{d}$ and an outer iteration to select the optimal dimension $\hat{d}$.

Our optimization criterion can be defined as a penalized least squares objective function of the form:

$$S(\Lambda_i, F, \beta, \mathbf{d}) = \frac{1}{NT} \sum_{i}^{N} ||Y_i - X_i\beta - F\Lambda_i^{'}||^2 + \mathbf{d}g, \tag{1.3}$$

where $g$ is a penalty term that depends on the sample size $N$ and $T$.

Before beginning with the details of the estimation procedure, it is important to mention that the intrinsic problem of factor models consists in the identification of the true factors and the true loading parameters. This is because $F\Lambda_i^{'}$ can be replaced with $F^*\Lambda_i^{*'}$, where $F^* = FH$, $\Lambda_i^{*'} = H^{-1}\Lambda_i^{'}$, and $H$ is an arbitrary $(d \times d)$ full rank matrix. In order ensure the uniqueness of $F$ and $\Lambda_i$ (up to sign change), the following normalization conditions ($d \times d$ restrictions) are required:

(R.1): $\sum_{i=1}^{N} \Lambda_i^{'}\Lambda_i/N$ is a $(d \times d)$ diagonal matrix, where the diagonal elements are ordered decreasingly and

(R.2): $F^{'}F/T^{\delta} = I_d$, where $I_d$ is the $(d \times d)$ identity matrix.

The rate $T^{\delta}$ can be chosen according to the stochastic character of $F_t$. Generally, $\delta$ is set to 1 if the factors are stationary and to 2 if they are integrated of order 1; see, e.g., Bai and Ng (2002), and Bai (2004). To be sparing with notation throughout the chapter, from now on we set $\delta$ to 2.

Our computational algorithm is based on a parameter cascading strategy, which is proposed by Cao and Ramsay (2010) to estimate models with multi-level parameters. The algorithm is relatively easy to program and can be described and implemented by the following logic:

**Step 1 (the individual parameters $\Lambda_i$)**

We estimate the individual parameters by minimizing the objective function $S(\beta, F, \Lambda_i, \mathbf{d})$ with respect to $\Lambda_i$ for each given $F, \beta$, and $\mathbf{d}$. Because the penalty term does not depend on $\Lambda_i$, the optimization criterion at this stage can be expressed as:

$$S_1(\Lambda_i|\beta, F, \mathbf{d}) = \frac{1}{NT} \sum_{i}^{N} ||Y_i - X_i\beta - F\Lambda_i^{'}||^2.$$

Minimizing for $\Lambda_i$ and using restriction (R.2), we get

$$\hat{\Lambda}_i'(F,\beta,\mathbf{d}) = \left[F'F\right]^{-1} F'\left[Y_i - X_i\beta\right] = F'\left[Y_i - X_i\beta\right]/T^2. \tag{1.4}$$

**Step 2 (the time trend effects $F$)**

We make use of result (1.4) and minimize the objective function $S_2(F|\beta,\mathbf{d})$, which depends only on $\beta$ and $\mathbf{d}$:

$$\begin{aligned} S_2(F|\beta,\mathbf{d}) &= \tfrac{1}{NT}\textstyle\sum_i^N ||Y_i - X_i\beta - F\hat{\Lambda}_i'||^2 \\ &= \tfrac{1}{NT}\textstyle\sum_i^N ||\left[Y_i - X_i\beta\right] - \frac{FF'}{T^2}\left[Y_i - X_i\beta\right]||^2. \end{aligned}$$

After rearranging, we can see that minimizing $S_2(F|\beta,\mathbf{d})$ with respect to $F$ is equivalent to maximizing the term $\frac{1}{NT}\sum_i^N ||\frac{FF'}{T^2}(Y_i - X_i\beta)||^2$.

Solving for $F(\beta,\mathbf{d})$ subject to (R.2), we obtain the following result:

$$\hat{F}(\beta,\mathbf{d}) = T\hat{P}(\beta,\mathbf{d}), \tag{1.5}$$

where $\hat{P}(\beta,\mathbf{d})$ is a $(T\times\mathbf{d})$ matrix containing the first $\mathbf{d}$ eigenvectors $[\hat{P}_1,\hat{P}_2,\cdots,\hat{P}_{\mathbf{d}}]$, which correspond to the first $\mathbf{d}$ eigenvalues, $\hat{\rho}_1(\beta,\mathbf{d}) \geq \hat{\rho}_2(\beta,\mathbf{d}),\cdots,\geq \hat{\rho}_{\mathbf{d}}(\beta,\mathbf{d})$, of the matrix

$$\hat{\Sigma}(\beta,\mathbf{d}) = \frac{1}{NT}\sum_{i=1}^{N}\left[Y_i - X_i\beta(\mathbf{d})\right]\left[Y_i - X_i\beta(\mathbf{d})\right]'. \tag{1.6}$$

**Step 3 (the common slope parameter $\beta$)**

To estimate the slope parameter, we reintegrate (1.4) and (1.5) into (1.3) and optimize the new intermediate objective function

$$S_3(\beta|\mathbf{d}) = \frac{1}{NT}\sum_{i}^{N}||Y_i - X_i\beta - \hat{F}(\beta,\mathbf{d})\hat{\Lambda}_i'(\beta,\mathbf{d})||^2. \tag{1.7}$$

Because $\hat{F}(\beta,\mathbf{d})$ depends nonlinearly on $\beta$, the minimization of (1.7) is conventionally done iteratively. For a given $\mathbf{d}$, the estimators of $\beta, F$, and $\Lambda_i$ should satisfy the following equation:

$$\hat{\beta}(\mathbf{d}) = \left[\sum_{i=1}^{N} X_i'X_i\right]^{-1}\left[\sum_{i=1}^{N} X_i'\left[Y_i - \hat{F}(\hat{\beta}(\mathbf{d}))\hat{\Lambda}_i'(\hat{\beta}(\mathbf{d}))\right]\right]. \tag{1.8}$$

We want to emphasize that our setting differs slightly from the development of Stock and Watson (2005), Bai (2009) and Bai et al. (2009) because $\hat{\beta}(\mathbf{d})$, in (1.8), depends on the unknown parameter $\mathbf{d}$, which has to be jointly estimated.

**Step 4 (the dimension $d$)**

The basic idea of constructing consistent panel information criteria consists of finding appropriate penalty functions that reestablish asymptotically the variance minimization when the considered number of factors increases. Explicitly, the optimal dimension $\hat{d}$ can be obtained by minimizing a criterion of the form

$$S_4(\mathbf{d}) = \frac{1}{NT}\sum_{i=1}^{N} ||Y_i - \hat{Y}_i(\mathbf{d})||^2 + \mathbf{d}g, \tag{1.9}$$

where $\hat{Y}_i(\mathbf{d})$ is the fitted response variable based on a given dimension $\mathbf{d}$. The penalty terms proposed by Bai and Ng (2002) and Bai (2004) basically depend on the orders of magnitude in probability of the sequences

$$\frac{1}{NT}\sum_{i=1}^{N} ||Y_i - \hat{Y}_i(\underline{\mathbf{d}})||^2 - \frac{1}{NT}\sum_{i=1}^{N} ||Y_i - \hat{Y}_i(d)||^2 = \Omega_p(\kappa_u) \quad \text{and} \tag{1.10}$$

$$\frac{1}{NT}\sum_{i=1}^{N} ||Y_i - \hat{Y}_i(\overline{\mathbf{d}})||^2 - \frac{1}{NT}\sum_{i=1}^{N} ||Y_i - \hat{Y}_i(d)||^2 = O_p(\kappa_o), \tag{1.11}$$

depending on $N, T \in \mathbb{N}$. Here, $\underline{\mathbf{d}} \in \{0, \ldots, d-1\}$, $\overline{\mathbf{d}} \in \{d+1, \ldots, d_{max}\}$, where $d_{max}$ is an arbitrary large dimension greater than $d$.

To ensure that $\lim_{N,T\to\infty} P\left[S_4(\mathbf{d}) > S_4(d)\right] = 1$ for all $\mathbf{d} = \{0, \ldots, d_{max} | \mathbf{d} \neq d\}$, it is sufficient to choose $g$, such that

$$\begin{aligned} &(i) \quad \lim_{N,T\to\infty} \kappa_u^{-1} g \longrightarrow 0 \quad \text{and} \\ &(ii) \quad \lim_{N,T\to\infty} \kappa_o^{-1} g \longrightarrow \infty. \end{aligned}$$

The existence of a function $g$ satisfying $(i)$ and $(ii)$ requires, of course, $\kappa_u/\kappa_o \to \infty$, which is often fulfilled because the common information, in the presence of a factor structure, is accumulated stochastically faster than the unit specific information in the errors, as $N, T \to \infty$. Intuitively, $g$ can be of the form $\sqrt{\kappa_u \kappa_o}$ or $\log(\kappa_u/\kappa_o)\kappa_o$. For more explicit examples, we refer the reader to the papers of Bai and Ng (2002), Bai (2004), and Hallin and Liška (2007).

The problem with this method is that the degree of freedom in the choice of $g$ is too large, since $g$ is not unique and multiplying it with any finite positive value will not

hurt $(i)$ and $(ii)$. Bai and Ng (2002) and Bai (2004) propose to scale $g$ with $\hat{\sigma}^2_{d_{max}} = \frac{1}{NT}\sum_{i=1}^{N} ||Y_i - \hat{Y}_i(d_{max})||^2$. However, such a penalization can lead to a result that is too sensitive to the choice of $d_{max}$.

In our algorithm, we utilize $\hat{\sigma}^2_{d_{max}}$ only as a starting scaler and make further use of the parameter cascading strategy to calibrate $g$ recursively by updating the sample variance after updating the factor dimension. At the optimum, the obtained estimators, referred to hereafter as *entirely updated estimators* and denoted by $Eup$, satisfy the following system of equations:

$$\begin{aligned}
\hat{\sigma}^2_{Eup} &= \tfrac{1}{NT}\textstyle\sum_i^N ||Y_i - X_i\hat{\beta}_{Eup} - \hat{F}_{Eup}\hat{\Lambda}'_{Eup,i}||^2, \\
\hat{d}_{Eup} &= \arg\min_{\mathbf{d}} \tfrac{1}{NT}\textstyle\sum_i^N ||Y_i - X_i\hat{\beta}_{Eup} - \hat{F}_{Eup}\hat{\Lambda}'_{Eup,i}||^2 + g(\hat{\sigma}^2_{Eup}), \\
\hat{\beta}_{Eup} &= \left[\sum_{i=1}^{N} X_i'X_i\right]^{-1}\left[\sum_{i=1}^{N} X_i'\left[Y_i - \hat{F}_{Eup}\hat{\Lambda}'_{Eup,i}\right]\right], \\
\hat{F}_{Eup} &= \sqrt{T}\hat{P}(\hat{\beta}_{Eup}, \hat{d}_{Eup}), \text{ and} \\
\hat{\Lambda}'_{Eup,i} &= \hat{F}'_{Eup}\left[Y_i - X_i\hat{\beta}_{Eup}\right]/T^2,
\end{aligned} \tag{1.12}$$

where $g(\hat{\sigma}^2_{Eup}) = \hat{\sigma}^2_{Eup} a \frac{\log(b)}{b}$, with $a = 1$ and $b = NT/(N+T)$ if the factor structure is composed of mixed stationary and non-stationary factors or $a = T/(4\log\log(T))$ if all factors are non-stationary stochastic trends.

*Remark* 1.1. As outlined in the introduction, the $d$ static factors can be composed of the leads and the lags of only $q < d$ primitive factors. The intuition behind estimating $\beta$ with a consistent estimator of $d$ and not $q$ is to avoid possible omitted-variable problems. However, assessing the number of the really existing primitive shocks in the data can be very useful for interpretation. For this purpose, a VAR regression with order $m \leq (\hat{d}_{Eup} - 1)$ can be applied to $\hat{F}'_{Eup}$. If the static factors are effectively driven by $q$ dynamic factors, then the rank of the VAR residual covariance matrix is equal to $q$ (at least asymptotically). Bai and Ng (2007) and Stock and Watson (2005) make use of this feature to provide selection procedures based on the principal components of the VAR residuals. Alternatively, one can apply the sequential testing procedure of Onatski (2009) or use the information criterion of Hallin and Liška (2007) on the spectral density matrix of $\hat{F}_{Eup}\hat{\Lambda}'_{Eup,(N)} = (\hat{F}_{Eup}\hat{\Lambda}'_{Eup,1}, \ldots, \hat{F}_{Eup}\hat{\Lambda}'_{Eup,N})$. If the factors are I(1) processes, we replace $\hat{F}_{Eup,t}$ with $\Delta\hat{F}_{Eup,t} = \hat{F}_{Eup,t} - \hat{F}_{Eup,t-1}$ before estimating $q$.

### 1.2.2 Starting Values and Iteration Scheme

Note that the multidimensional objective function $S(\beta, F, \Lambda_i, \mathbf{d})$ in (1.3) is not globally convex. To insure that the iteration algorithm converges to the global optimum, it is important to choose reasonable starting values for $d$ and $\beta$. Because, in practice, only

samples with finite dimensions can be observed, we propose starting with the following integer: $d_{max} = \left\lfloor \min\{\sqrt{N}, \sqrt{T}\} \right\rfloor$, where $\lfloor z \rfloor$ indicates the integer part of $z$. This technical setting provides only a systematic approach to initiate the iteration. For huge data sets, the square root function can be replaced by the natural logarithm or by any truncation method. But this is not necessary, since $d_{max}$ will be replaced directly after the first iteration with the estimated factor dimension and refined iteratively until convergence. More important is the starting value of the slope parameter. We propose using the following estimate:

$$\hat{\beta}_{start} = \left[ \sum_{i=1}^{N} X_i^{'} [I - GG^{'}] X_i \right]^{-1} \left[ \sum_{i=1}^{N} X_i^{'} [I - GG^{'}] Y_i \right], \tag{1.13}$$

where $G$ is the $(T \times d_{max})$ matrix of the eigenvectors $G_1, \ldots, G_{d_{max}}$, corresponding to the first $d_{max}$ eigenvalues of the augmented matrix

$$\Gamma^{Aug} = \frac{1}{NT} \sum_{i=1}^{n} [Y_i, X_i][Y_i, X_i]^{'}.$$

The idea behind these starting estimates relies on the fact that the true unobserved factors $F$ cannot escape from the space spanned by the eigenvectors of the augmented matrix $\Gamma^{Aug}$. The orthogonal projection of $X_i$ on $G$ in (1.13) eliminates the effect of a possible correlation between the observed regressors $X_i$ and the unobserved factors $F$, which can heavily distort the value of $\hat{\beta}_{start}$ from the true $\beta$ if it is neglected. However, the problem of this starting estimate is that $\left[\sum_{i=1}^{N} X_i^{'} [I - GG^{'}] X_i\right]$ and $\left[\sum_{i=1}^{N} X_i^{'} [I - GG^{'}] Y_i\right]$ in (1.13) will be close to zero if the observed regressors are the underlying factors, i.e., $X_i \approx F\Lambda_i^{'}$ . In this case, the estimation algorithm can misleadingly converge to an insignificant estimate of $\beta$ and compensate for that by estimating factors that originally exist only in the observed regressors. To overcome such limitation, we recommend under-scaling the factors $G_l$, $l = 1, \ldots, d_{max}$, with $(1 - \max_{p \in \{1,\ldots,P\}} r^2_{G_l, X_p})^{0.5}$, where $r_{G_l, X_p}$ is the sample correlation coefficient between $G_l$ and the $p$-th element of $X_i$ (or the $p$-th eigenvector of the $(T \times T)$ covariance matrix of $X_i$, for $p \in \{1, \ldots, d_{max}\}$). This automatically eliminates the factors that are perfectly correlated with the observed regressors, and gives the remaining factors appropriate weights.

In spite of the complex form of our estimates, implementing the algorithm that optimizes $S(\Lambda_i, \beta, F, \mathbf{d})$ through optimizing $S_1(\Lambda_i|\beta, F, \mathbf{d})$, $S_2(F|\beta, \mathbf{d})$, $S_3(\beta|\mathbf{d})$ and $S_4(\mathbf{d})$ is relatively easy. The final estimators can be obtained by alternating between

- $r_m$ inner iterations until convergence of the following composite function:

$$\hat{\beta}^{(r_m)} \circ \hat{F}^{(r_m)} \circ \hat{\Lambda}_i^{(r_m)} \left( \hat{d}^{(m)} \right) \approx \hat{\beta}^{(r_m - 1)} \circ \hat{F}^{(r_m - 1)} \circ \hat{\Lambda}_i^{(r_m - 1)} \left( \hat{d}^{(m)} \right),$$

for each given $\hat{d}^{(m)}$, and

- outer iterations until satisfying the following convergence condition:

$$\hat{d}^{(m+1)} \circ \hat{\sigma}^{2(m)} = \hat{d}^{(m)} \circ \hat{\sigma}^{2(m-1)}.$$

Here, the composite notation $c \circ b(z)$ is defined by $c(b(z))$ for each $z$ and used to indicate the application of one estimate on the result of another.

Note that this iteration scheme simplifies the minimization of the dimensionality criterion (1.9), since $\hat{d}^{(m+1)}$ can be obtained by

$$\begin{aligned} \hat{d}^{(m+1)} &= \arg\min_{\mathbf{d}} \tfrac{1}{NT} \textstyle\sum_i^N ||Y_i - X_i\hat{\beta}^{(r_m)} - \hat{F}^{(r_m)}\hat{\Lambda}_i^{(r_m)}||^2 + \mathbf{d}g(\hat{\sigma}^{2(m)}) \\ &= \arg\min_{\mathbf{d}} \textstyle\sum_{l=\mathbf{d}+1}^T \hat{\rho}_l^{(r_m)} + \mathbf{d}g(\hat{\sigma}^{2(m)}), \end{aligned}$$

where $\hat{\rho}_l^{(r_m)}$ are the ordered eigenvalues of the covariance matrix (1.6) required to compute $\hat{F}^{(r_m)}$ at the iteration stage $r_m$ and

$$\hat{\sigma}^{2(m)} = \frac{1}{NT}\sum_i^N ||Y_i - X_i\hat{\beta}^{(r_m)} - \hat{F}^{(r_m)}\hat{\Lambda}_i^{(r_m)}||^2 = \sum_{l=\hat{d}^{(m)}+1}^T \hat{\rho}_l^{(r_m)}. \tag{1.14}$$

Selecting $\hat{d}^{(m+1)}$ reverts therefore to finding the order of the smallest element in the following set:

$$\mathcal{A}^{(m)} = \left\{ \sum_{l=\mathbf{d}+1}^T \hat{\rho}_l^{(r_m)} + \mathbf{d}a\frac{\log(b)}{b} \sum_{l=\hat{d}^{(m)}+1}^T \hat{\rho}_l^{(r_m)} \,\middle|\, \mathbf{d} = 0, 1, \ldots, \hat{d}^{(m)} \right\}. \tag{1.15}$$

A simple pseudo code that optimizes the entirely updated estimators presented in (1.12), can be described as follows:

1. Set $\hat{d}^{(m)} = \begin{cases} d_{max} & \text{if} \quad m = 0 \\ \hat{d}^{(m-1)} & \text{if} \quad m > 0 \end{cases}$
2. Set $\hat{\beta}^{(r_m)} = \begin{cases} \hat{\beta}_{start} & \text{if} \quad r_m = 0 \\ \hat{\beta}^{(r_m-1)} & \text{if} \quad r_m > 0 \end{cases}$
3. Use (1.5) to calculate $\hat{F}^{(r_m)} = \hat{F}(\hat{\beta}^{(r_m)}, \hat{d}^{(m)})$
4. Use (1.4) to calculate $\hat{\Lambda}_i^{(r_m)} = \hat{\Lambda}_i(\hat{F}^{(r_m)}, \hat{\beta}^{(r_m)}, \hat{d}^{(m)})$
5. Use (1.8) to update $\beta^{(r_m+1)} = \hat{\beta}(\hat{d}^{(m)})$ by using $\hat{F}^{(r_m)}$ and $\hat{\Lambda}_i^{(r_m)}$
6. If $\hat{\beta}^{(r_m+1)} \approx \hat{\beta}^{(r_m)}$, go to 7; else, replace the value of $\hat{\beta}^{(r_m)}$ with $\hat{\beta}^{(r_m+1)}$ and repeat $2 - 6$ with $(r_m + 1)$ instead of $(r_m)$

7. Use (1.14) to calculate $\hat{\sigma}^{2(m)}$

8. Select $\hat{d}^{(m+1)}$ that corresponds to the order of the smallest element in the set $\mathcal{A}^{(m)}$ in (1.15)

9. If $\hat{d}^{(m+1)} = \hat{d}^{(m)}$, exit; else, replace the value of $\hat{d}^{(m)}$ with $\hat{d}^{(m+1)}$ and $\hat{\beta}^{(r_m)}$ with $\hat{\beta}^{(r_{m+1})}$ and go to 1 with $(m+1)$ instead of $(m)$ and $(r_{m+1}+1)$ instead of $(r_m)$.

*Remark* 1.2. We can, of course, use the analytic expression of $\hat{\Lambda}'_{Eup,i}$ to write the estimator of $\beta$ in (1.12) conventionally as

$$\hat{\beta}_{Eup} = \left[\sum_{i=1}^{N} X_i' M_{\hat{F}_{Eup}} X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i' M_{\hat{F}_{Eup}} Y_i\right], \tag{1.16}$$

where $M_{\hat{F}_{Eup}} = I_T - \hat{F}_{Eup}\hat{F}'_{Eup}/T^2$. However, implementing the estimation algorithm with (1.16) may destabilize the convergence of the iteration process, since the update of the slope estimator, in this case, requires the inversion of the matrix $\sum_{i=1}^{N} X_i' M_{\hat{F}_{Eup}} X_i$ in each iteration stage and not only at the optimum.

*Remark* 1.3. In order to speed up the computation when $N < T$, we can reconstruct the estimation algorithm with $S_1(F|\Lambda_i, \beta, \mathbf{d})$ and $S_2(\Lambda_i|\beta, \mathbf{d})$ instead of $S_1(\Lambda_i|F, \beta, \mathbf{d})$ and $S_2(F|\beta, \mathbf{d})$. The benefit of such modification is to calculate the eigenvectors of a smaller covariance matrix with a dimension $(N \times N)$ instead of $(T \times T)$. Both computations ultimately lead to the same result.

The routines of this method are provided in an R-Package called `phtt`. For more details about this package, we refer the reader to the paper of Bada and Liebl (2014b).

## 1.3 Model Extension and Theoretical Results

### 1.3.1 Presence of Additional Categorical Variables

Our model assumptions will closely follow the the setup of Bai et al. (2009), who allow for mixed stationary and unit root regressors (I(0)/I(1) regressors) as well as mixed I(0)/I(1) unobserved factors. However, our analysis encounters an additional complication allowing for model (1.2) to be obtained from transforming an underlying model of the form:

$$Y_{it}^{\circ} = X_{it}^{\circ}\beta + \sum_{k=1}^{K} \alpha_{kt}\delta_{ik} + F_t\Lambda_i' + \mu_t + \varepsilon_{it}. \tag{1.17}$$

Here, $Y_{it}^{\circ}$ and $X_{it}^{\circ}$ are the underlying observed variables that change over $i$ and $t$, $\delta_{ik}$ is a nominal variable defined as $\delta_{ik} = 1$ if individual $i$ belongs to category $k$, $k = 1, \ldots, K < N$, and 0 otherwise, $\alpha_{kt}$ is a time heterogeneous parameter describing the effect of the categories on the dependent variable, $\mu_t$ is the time varying general average of $Y_{it}^{\circ}$, $F_t\Lambda_i^{'}$ is, as above, the unobserved factor structure with unknown dimension, and $\varepsilon_{it}$ is the idiosyncratic error.

An application of such a model specification and examples of dependent and independent variables are presented in Section 1.5.1. Note that the identification of the additional parameters $\alpha_{kt}$ and $\mu_t$ in (1.17) requires additional restrictions:

$$\text{(R.3):} \quad \textstyle\sum_k^K \alpha_{kt} = 0, \sum_{i=1}^N \Lambda_i = 0 \text{ and } \sum_{i=1}^N \delta_{ik}\Lambda_i = 0 \text{ for } k = 1, \ldots, K.$$

This condition does not impose any unreasonable limitation but only identifies the model parameters by imposing on the categorical variable $\delta_{ik}$ and the parameters $\mu_t$ and $\alpha_{kt}$ to be unconnected with the parameters of the factor structure.

In order to avoid reverting to constrained optimization techniques that rely explicitly on (R.3), we use a within-group transformation first to eliminate $\alpha_{kt}$ and $\mu_t$ from (1.17) and then to estimate the transformed model as described in Section 1.2. The parameters $\alpha_{kt}$ and $\mu_t$ can be easily estimated in a second step once $\hat{d}_{Eup}, \hat{\beta}_{Eup}, \hat{F}_{Eup}$, and $\hat{\Lambda}_{Eup,i}^{'}$ are obtained. To this end, we define the linear transformation operator $\mathcal{T}(.)$ as

$$\mathcal{T}(Z_{it}) = Z_{it} - \sum_{k=1}^{K} \frac{1}{m_k} \sum_{j=1}^{N} Z_{it}\delta_{jk}, \tag{1.18}$$

where $m_k = \sharp\{j|\delta_{jk} = 1\}$ for $k \in \{1, \ldots, K\}$.

Let $Y_{it} = \mathcal{T}(Y_{it}^{\circ})$, $X_{it} = \mathcal{T}(X_{it}^{\circ})$ and $\epsilon_{it} = \mathcal{T}(\varepsilon_{it})$. By using (R.3), we can easily verify that

$$\mathcal{T}(Y_{it}^{\circ}) = Y_{it} = X_{it}\beta + F_t\Lambda_i^{'} + \epsilon_{it}. \tag{1.19}$$

The transformed Model (1.19) has the same form as Model (1.2) and can be fitted by the entirely updated estimators. In order to estimate the pre-eliminated parameters $\alpha_{kt}$ and $\mu_t$, we propose to use a dummy variable regression once $\hat{\beta}_{Eup}$, $\hat{F}_{Eup}, \hat{\Lambda}_{Eup,i}^{'}$ and $\hat{d}_{Eup}$ are obtained. In fact, estimating $\alpha_{kt}$ and $\mu_t$ does not require any iteration since restriction (R.3) arranges for the orthogonality between $\delta_{ik}$ and $\Lambda_i$. The solution has consequently the same form as the classical fixed effects estimators:

$$\begin{aligned} \hat{\mu}_t &= \overline{Y}_{.t} - \overline{X}_{.t}\hat{\beta}_{Eup} \text{ and} \\ \hat{\alpha}_{kt} &= \overline{Y}_{kt} - \overline{X}_{kt}\hat{\beta}_{Eup} - \hat{\mu}_t, \end{aligned}$$

where $\overline{Y}_{.t} = \frac{1}{N}\sum_i^N Y_{it}$, $\overline{X}_{.t} = \frac{1}{N}\sum_i^N X_{it}$, $\overline{Y}_{kt} = \frac{1}{m_k}\sum_i^N \delta_{ik}Y_{it}$, and $\overline{X}_{kt} = \frac{1}{m_k}\sum_i^N \delta_{ik}X_{it}$.

### 1.3.2 Assumptions

We now consider inference of (1.19) as $(N,T) \to \infty$. Here, $(N,T) \to \infty$ has to be interpreted as a sequential limit: first $T \to \infty$ and then $N \to \infty$. Throughout, we denote by $M$ a finite positive constant, not depending on $N$ and $T$. We use $B(.)$ to denote a Brownian motion process defined on $[0,1]$ and $\lfloor \tau \rfloor$ to denote the largest integer $\leq \tau$. We will use $\beta^\circ, F_t^\circ$ and $\Lambda_i^\circ$ to respectively denote the true slope parameters, the true factors (only identifiable up to rotation), and the true loadings parameters. $E_C(.)$ is used to denote conditional expectation given $F^\circ$. For all $N$, we assume an i.i.d. random sample of individuals.

Our theoretical setup relies on the following assumptions.

**Assumption 1. The observed regressors:**

(a) $\frac{m_k}{N}$ converges a.s. to $E(\delta_{ik})$ as $N \to \infty$, where $\inf_{k=1,\ldots,K} E(\delta_{ik}) > 0$.

(b) Let $X_{it}\beta^\circ = X_{it,1}\beta_1^\circ + X_{it,2}\beta_2^\circ$, where $X_{it,1}$ is $(1 \times P_1)$ vector of a I(1) multivariate process, such that $X_{it,1}^{'} = X_{i,t-1,1}^{'} + \zeta_{it}^{'} - \sum_{k=1}^{K} \delta_{ik}\bar{\zeta}_{kt}^{'}$ where $\zeta_{it}^{'}$ is a zero mean $(P_1 \times 1)$ stationary vector and $\bar{\zeta}_{kt}^{'} = \frac{1}{m_k}\sum_{j=1}^{N} \delta_{jk}\zeta_{jt}^{'}$. $X_{it,2}$ is $(1 \times P_2)$ vector of stationary regressors, such that $X_{it,2} = X_{it,2}^{\circ\prime} - \sum_{k=1}^{K} \delta_{ik}\bar{X}_{kt,2}^{\circ\prime}$ with $\bar{X}_{kt,2}^{\circ\prime} = \frac{1}{m_k}\sum_{j=1}^{N} \delta_{jk}X_{jt}^{\circ\prime}$ and $E_C(X_{it,2}^\circ \zeta_{js}) = 0$ for all $i, j, t$ and $s$.

**Assumption 2. The unobserved factor structure:**

(a) $E||\Lambda_i^\circ||^4 \leq M$; As $N \to \infty$, $E(\Lambda_i^\circ \delta_{ik}) = 0$ for all $k = 1, \ldots, K$, and $\frac{1}{N}\sum_i \Lambda_i^{\circ\prime}\Lambda_i^\circ \xrightarrow{p} \Sigma_\Lambda$, a $(d \times d)$ positive definite matrix.

(b) $F_t^{\circ\prime} = F_{t-1}^{\circ\prime} + \eta_t^{'}$, where $\eta_t^{'}$ is a zero mean random vector with $E||\eta_t^{'}||^{4+\gamma} \leq M$ for some $\gamma > 0$ and for all $t$; As $T \to \infty$, $\frac{1}{T^2}\sum_t F_t^{\circ\prime}F_t^\circ \xrightarrow{d} \int B_\eta^{'} B_\eta$, a $d \times d$ random matrix, where $B_\eta$ is a vector of Brownian motion with a positive definite covariance matrix $\Omega_\eta$. $\eta_t$ is independent of $X_{it,2}^\circ$ for all $i, t, k$.

(c) $\liminf_{T\to\infty} \log\log(T)/T^2 \sum_{t=1}^{T} F_t^{'\circ} F_t^\circ = C$, where $C$ is a nonrandom positive definite matrix.

(d) $\{F_t, X_{it,1}^*\}$ are not cointegrated, where $X_{it}^{*'} = X_{i,t-1}^{*'} + \zeta_{it}^{*'}$, $t = 2, \ldots, T$ with $X_{i1}^{*'} = X_{i1}^{'}$, $\zeta_{it}^{*'} = \zeta_{it}^{'} - \sum_{k=1}^{K} \delta_{ik}\zeta_{kt}^{0'}$, and $\zeta_{kt}^{0'} = E_C(\zeta_{kt}^{'}|\delta_{ik} = 1)$, $k = 1, \ldots, K$.

**Assumption 3. The error terms:**

(a) Let $\epsilon_{it} = \varepsilon_{it} - \sum_{k=1}^{k} \delta_{ik}\bar{\varepsilon}_{kt}$ with $\bar{\varepsilon}_{kt} = \sum_{j=1}^{n} \frac{1}{m_k} \sum_{j=1}^{n} \varepsilon_{jt}\delta_{jk}$. Here, $\epsilon_{it}$ are zero mean error terms and $E_C(\bar{\varepsilon}_{kt}|\delta_{ik} = 1) = 0$ for all $k$. Conditional on $\eta_t$ the error terms $\varepsilon_{it}$ are cross-sectionally independent of each other as well as of $X_{it}^{\circ}$.

(b) The multivariate processes $w_{it} = (\varepsilon_{it}, \zeta_{it}^{*}, \eta_t)$ are stationary. For each $i$, $w_{it} = \sum_{j=0}^{\infty} \Pi_{ij} v_{i,t-j}$, where $v_{it} = (v_{it}^{\varepsilon}, v_{it}^{\zeta}, v_t^{\eta})$ are mutually independent over $i, t$ as well as identically distributed over $t$. Furthermore, $E(v_{it}) = 0$, $E(v_{it}v_{it}^{'}) > 0$, and $E(\|v_{it}\|^8) \leq M$, where $M < \infty$ is independent of $i, t$. In addition, all further conditions of Assumptions 2. and 3 of Bai et al. (2009) are satisfied.

The additional terms $\bar{\zeta}_{kt}$ and $\bar{\varepsilon}_{kt}$ in Assumptions 2 and 3 reflect our subtraction of group means. Assumption 1.a guarantees that the $K$ categories (groups) do not vanish as $N \to \infty$. Assumption 1.b allows for mixed I(1)/I(0) regressors. As in Bai et al. (2009), the I(0) regressors are assumed to be exogenous and linearly independent of the I(1) regressors and the I(1) factors. This is only given for the purpose of simplifying the analysis and avoiding further complications.

The requirement $E(\Lambda_i^{\circ}\delta_{ik}) = 0$ for all $k = 1, \ldots, K$, in Assumption 2, is the population version for our condition (R.3) introduced for identifying $\alpha_{kt}$. We want to emphasize that the transformation $\mathcal{T}(.)$ only influences the structure of the error terms and the explanatory variables, but not the factor structure $F_t^{\circ}\Lambda_i^{'\circ}$. Assumptions 2.b and 2.c are commonly used in the literature of non-stationary factor models with unit roots; see, e.g., Bai (2004) and Bai et al. (2009). Assumption 2.d is a technical assumption used to ensure the non-singularity of the long run covariance matrix $\Omega_{b,i}$ of $(\zeta_{it}^{*'}, \eta_t^{'})^{'}$. This allows for estimating the asymptotic bias of the slope estimator.

Assumption 3 excludes cross-section dependencies of $\epsilon_{it}$ and $\zeta_{it}^{*}$ conditional on $\{\eta_t\}$. But unconditionally, weak cross-section correlations are allowed under Assumptions 3.b of Bai et al. (2009).

*Remark* 1.4. Assumption 2 considers the presence of only I(1) factors. But note that the method is also robust to mixed I(1)/I(0) factors. Bai et al. (2009) argue that, for known $d$, the limiting distribution of the slope estimator, in this case, is the same as when all factors are I(1) (except for small modifications in the expression of the asymptotic variance). Their arguments should also hold for our extended model.

*Remark* 1.5. The last part of Assumption 1.b considerably simplifies the analysis of the asymptotic distributions of the slope parameters. This is because the I(0) and I(1) regressors are asymptotically orthogonal and their asymptotic distributions can be separately analyzed: while the estimator of $\beta_2^{\circ}$ needs no correction and is asymptotically

normal distributed (see Bai (2009) and Bai et al. (2009)), the estimator of $\beta_1^\circ$ has a distribution as if there is no I(0) regressors. Note that the aim of separating I(0) and I(1) variables is to correctly derive the rates of convergence. Bai et al. (2009) argue that if the ultimate purpose is to perform hypothesis testing, one can proceed as if all regressors are I(1) since the scaling factor will be canceled out in the end.

Because of Remark 1.5, we can drop from now on the indexes 1 and 2 respectively from $\beta_1^\circ$ and $\beta_2^\circ$ and focus only on the complicated case, i.e., panel cointegration model with I(1) regressors and I(1) unobserved factors. This allows us to avoid notational mess in the remainder of the chapter.

### 1.3.3 Asymptotic Distribution and Bias Correction

Under Assumptions 1-3, it can be shown that the effects of the model transformation due to $\mathcal{T}(.)$ are asymptotically negligible and that the results of Bai et al. (2009) generalize to our situation. In particular, the slope estimator $\hat{\beta}(d)$ to be obtained for the true factor dimension $d$ is at least $T$ consistent and has following properties.

**Proposition 1.6.** *Under the above assumptions, we have, as* $(N,T) \to \infty$*,*

$$\Sigma_c^{1/2}\left(\sqrt{N}T(\hat{\beta}(d)-\beta^\circ)-\sqrt{N}\phi\right) \xrightarrow{d} N(0,I_p),$$

*for some* $\phi$ *and* $\Sigma_c$*, where* $\hat{\beta}(d)$ *is obtained after transforming Model* (1.17) *with* $\mathcal{T}(.)$*.*

The exact expression of $\phi$ and $\Sigma_c$ is given in the appendix. Proposition 1.6 shows that the limiting distribution of $\sqrt{N}T(\hat{\beta}(d)-\beta^\circ)$ is not centered at zero. Bai et al. (2009) prove that it is possible, in such a case, to construct a consistent estimator $\hat{\phi}_{NT}$ of the bias term $\phi$. Following their suggestion, we define our *entirely updated and bias corrected* (*EupBC*) estimator by

$$\hat{\beta}_{EupBC} = \hat{\beta}_{Eup} - \frac{1}{T}\hat{\phi}_{NT}.$$

This procedure does require extra work (non-parametrical kernel estimation techniques) to estimate the long-run and one-sided long-run covariances of $w_{it} = (\varepsilon_{it}, \zeta_{it}^*, \eta_t)$. The necessary assumptions and precise formulas for constructing $\hat{\phi}_{NT}$ are given in the appendix. Once $\hat{\beta}_{EupBC}$ is obtained, the final bias-corrected estimators of $F^\circ$ and $\Lambda_i^\circ$ are given by

$$\begin{aligned}\hat{F}_{EupBC} &= \sqrt{T}\hat{P}(\hat{\beta}_{EupBC}, \hat{d}_{Eup}), \text{ and}\\ \hat{\Lambda}'_{EupBC,i} &= \hat{F}'_{EupBC}\left[Y_i - X_i\hat{\beta}_{EupBC}\right]/T^2,\end{aligned}$$

respectively.

Note that the main difference between our approach and the methodology of Bai et al. (2009) consists in the fact that our estimation procedure directly incorporates a dimension estimate. We show in Theorem 1 that our final estimator $\hat{\beta}_{EupBC}$ has an asymptotic distribution centered around zero and that $\hat{d}_{Eup}$ and $\hat{F}_{EupBC}$ provide respectively consistent estimators of the true dimension $d$ and the true factors $F^\circ$ (up to rotation).

**Theorem 1.7.** *Under assumptions 1-3, we have, as* $(N,T) \to \infty$*,*

*a)* $P(\hat{d}_{Eup} = d) \to 1$*, if the starting estimate* $d_{max} \geq d$ *and* $g$ *is of the form* $g = cp_{NT}$ *such that* $(i)$ $c = O_P(1)$ *and strictly positive,* $(ii)$ $p_{NT} \to \infty$*, and* $(iii)$ $\frac{\log\log(T)}{T} p_{NT} \to 0$*,*

*b) with the additional Assumption 4,*

$$\Sigma_c^{1/2} \sqrt{NT}(\hat{\beta}_{EupBC} - \beta^\circ) \xrightarrow{d} N(0, I_p),$$

*c) and for some* $(d \times d)$ *invertible matrix* $H$*,*

$$\frac{1}{T}\sum_{t=1}^{T} \|\widehat{F}_{EupBC,t} - F_t^\circ H\|^2 = O_P(\frac{1}{N}) + O_P(\frac{1}{T}).$$

Assumption 4 is given in the appendix and required only for consistency of $\phi$. Examples of $p_{NT}$ can be found in Bai (2004).

*Remark* 1.8. Note that conditions $(ii)$ and $(iii)$ in Theorem 1.7 are sufficient only to consistently estimate the number of I(1) factors. If we assume the presence of additional I(0) factors, we have to choose $g$ such that $(ii)$ $\min\{N,T\}p_{NT} \to \infty$ and $(iii)$ $p_{NT} \to 0$; however, some suitable regularity assumptions are required to forbid strong forms of dependency and heteroscedasticity in the errors. For more details, we refer to the argumentation of Bai and Ng (2002) and Bai et al. (2009) in Section 3.3.

A consistent estimator $\hat{\Sigma}_c$ of $\Sigma_c$ is also defined in the appendix. This allows us to test for the significance of $\hat{\beta}_{EupBC}$.

The bias-corrected estimators of the pre-eliminated effects $\mu_t$ and $\alpha_{kt}$ can be respectively obtained by:

$$\begin{aligned} \hat{\mu}_{BC,t} &= \overline{Y}_{.t} - \overline{X}_{.t}\hat{\beta}_{EupBC} \text{ and} \\ \hat{\alpha}_{BC,kt} &= \overline{Y}_{kt} - \overline{X}_{kt}\hat{\beta}_{EupBC} - \hat{\mu}_{BC,t}. \end{aligned} \tag{1.20}$$

Under our assumptions, it is easy to show that $\hat{\alpha}_{kt}$ is $\sqrt{m_k}$ consistent and has an asymptotic normal distribution, such that

$$\sqrt{m_k}(\hat{\alpha}_{BC,kt} - \alpha_{kt}) \xrightarrow{d} N(0, \sigma_{kt}^2), \tag{1.21}$$

where $\sigma_{kt}^2 = Var(\bar{\varepsilon}_{kt})$, with $\bar{\varepsilon}_{kt} = \frac{1}{m_k}\sum_i^N \delta_{ik}\varepsilon_{it}$.

## 1.4 Monte Carlo Simulations

The goal of this section is to compare, through Monte Carlo experiments, the performance of our algorithmic extension with the performance of the iterative least squares estimators of Bai (2009) and Bai et al. (2009) based on an externally estimated dimension. In a first step, the feasible slope estimator, $\hat{\beta}(d)$, is naively calculated with a high number of factors ($d_{max} = 8$). In a second step, we calculate $\hat{W}_{it} = Y_{it} - X_{it}\hat{\beta}(d_{max})$ and externally estimate the factor dimension by using 5 different criteria: the panel criteria PC1 and IC1 of Bai and Ng (2002), the panel cointegration criterion IPC1 proposed by Bai (2004), the threshold criterion ED of Onatski (2010), and the information criterion $IC_{1;n}^T$ of Hallin and Liška (2007). The maximal number of factors used in PC1, IC1, IPC1, ED, and $IC_{1;n}^T$ is also set to 8. The calibration strategy of Hallin and Liška (2007) (second "'stability interval"' procedure) is applied on a grid interval of length 128 with the borders 0.01 and 3 as they have suggested. Finally, we re-calculate the two-step iterative least squares estimator with the optimizers of these panel criteria. The estimated dimensions are denoted by $\hat{d}_{PC1}, \hat{d}_{IC1}, \hat{d}_{IPC1}, \hat{d}_{ED}$, and $\hat{d}_{IC_{1,n}^T}$, respectively.

Our entirely updated estimator is calculated with the penalty of PC1 as described in Section 1.2, i.e., $g(\hat{\sigma}_{Eup}) = \hat{\sigma}_{Eup}^2 \log(NT/(N+T))(N+T)/NT$. The iteration process is initiated by the starting values described in Section 1.2.2. The bias correction used for estimating panel cointegration models is based on the linearly decreasing weights of Newey and West (1987) with a truncation at $\left\lfloor \min\{\sqrt{N}, \sqrt{T}\} \right\rfloor$. The maximal number of iterations allowed for the feasible iterated least squares estimators and the two-step estimators is 100. The inner iteration of the entirely updated estimator is also limited to 100.

We consider panel models of the form

$$Y_{it} = \sum_{p=1}^{P} X_{pit}\beta_p + c\sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it},$$

for all $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, where $X_{pit}$ are the observed regressors, $f_{lt}$ are the factors to be estimated, $\lambda_{il}$ are the corresponding loading parameters, $c$ controls for the weight of the factor structure in the model, and $\epsilon_{it}$ is the idiosyncratic error term.

The examined panel sets are the outcomes of 6 different DGPs:

DGP1 (panel cointegration model with I(1) factors and endogenous explanatory variables). $d = 2,\ P = 1,\ \beta_1 = 1.5,\ c = 1,\ f_{lt} = f_{l,t-1} + \eta_{lt},\ \lambda_{il} \sim N(1,1),\ l = 1,2,\ X_{1it} = X_{1i,t-1} + \zeta_{1it}$, with

$$\begin{bmatrix} \zeta_{1it} \\ \eta_{1t} \\ \eta_{2t} \\ \epsilon_{it} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.5 & 0.7 & 0.7 \\ -0.5 & 1 & 0 & -0.5 \\ 0.7 & 0 & 1 & 0.7 \\ 0.7 & -0.5 & 0.7 & 1 \end{bmatrix} \right);$$

DGP2 (panel cointegration model with mixed I(1)/I(0) factors). $d = 3,\ P = 1,\ \beta_1 = 1.5,\ c = 1,\ f_{lt} = f_{l,t-1} + \eta_{lt}, l = 1,2,\ f_{3t} = \sqrt{0.5} f_{3,t-1} + \eta_{3t},\ \lambda_{il} \sim N(1,1),\ l = 1,2,3,\ X_{1it} = X_{1i,t-1} + \zeta_{1it}$ with $\eta_{lt}, \zeta_{1it}, \epsilon_{it} \sim N(0,1)$, for $l = 1,2,3$;

DGP3 (the observed regressors are the underlying factors). $c = 0,\ P = 1,\ \beta_1 = 1.5,\ d = 1,\ f_{1t} = f_{1,t-1} + \eta_{1t},\ \lambda_{i1} \sim N(1,1),\ X_{1it} = \lambda_{i1} f_{1t} + \zeta_{1it}$, and $\eta_{1t}, \zeta_{1it}, \epsilon_{it} \sim N(0,1)$;

DGP4 (stationary factors and weakly autocorrelated idiosyncratic errors). $d = 1,\ P = 1,\ \beta_1 = -0.75,\ c = 1,\ f_{1t} = \sqrt{0.5} f_{1,t-1} + \eta_{1t},\ \lambda_{i1} \sim N(1,1),\ X_{1it} = 0.8\lambda_{i1} f_{1t} + \zeta_{1it},\ \epsilon_{it} = \theta_i \epsilon_{i,t-1} + \varepsilon_{it},\ \eta_{1t}, \zeta_{1it}, \varepsilon_{it} \sim N(0,1)$, and $\theta_i \sim U(-0.3, 0.3)$;

DGP5 (stationary factors and strongly autocorrelated idiosyncratic errors). $d = 1,\ P = 1,\ \beta_1 = -0.75,\ c = 1,\ f_{1t} = \sqrt{0.5} f_{1,t-1} + \eta_{1t},\ \lambda_{i1} \sim N(1,1),\ X_{1it} = 0.8\lambda_{i1} f_{1t} + \zeta_{1it},\ \epsilon_{it} = \theta_i \epsilon_{i,t-1} + \varepsilon_{it},\ \eta_{1t}, \zeta_{1it}, \varepsilon_{it} \sim N(0,1)$, and $\theta_i \sim U(0.6, 0.8)$; and

DGP6 (no factor structure and strongly autocorrelated idiosyncratic errors). $d = 0,\ P = 1,\ \beta_1 = -0.75,\ c = 0,\ \epsilon_{it} = \theta_i \epsilon_{i,t-1} + \varepsilon_{it},\ X_{1it}, \varepsilon_{it} \sim N(0,1)$, and $\theta_i \sim U(0.6, 0.8)$.

To see how the properties of the estimators vary with $N$ and $T$, we consider 9 different combinations with the sizes $N = 30, 60, 120$ and $T = 30, 60, 120$. For DGP2, we consider the extra combination $(N, T) = (120, 300)$ to see how the criteria behave with the problem of unproportional factors, which occurs in mixed I(1)/I(0) factor models because the variance of I(1) factors diverge, as $T \to \infty$, whereas the variance of the I(0) factors is bounded. This attitude of mixed I(1)/I(0) factors is closely related to the concept of weak/strong factors discussed in Onatski (2012) and Onatski (2009).

Tables 1.1 and 1.2 report the averages of the estimated dimensions and the mean squared errors (MSE) of the slope estimators obtained from 1000 replications. It is clear that the Eup estimator outperforms the feasible estimator $\hat{\beta}(d_{max})$ in all cases. This result is not surprising, since $\hat{\beta}(d_{max})$ requires the computation of $(d_{max} - \hat{d}_{Eup})NT$ additional parameters comparing to $\hat{\beta}_{Eup}$. Moreover, the number of times the Eup algorithm did not converge within the total of 54000 repetitions is remarkably smaller than the number of times the conventional feasible estimator $\hat{\beta}(d_{max})$ did not converge (5.41% vs.

42.22%). The reason for this outcome is that the naive over-specification of the factor dimension downgrades the degree of freedom available to estimate the slope parameters. The alternation between inner and outer iterations, in our algorithm, seems, hence, to provide a way to stabilize the numerical optimization of the objective function if $d$ is not well-specified.

Tables 1.1 and 1.2 reveal that PC1 has a tendency to overestimate the true dimension if $N$ and/or $T$ are not large enough (30, 60). This is also the case for IPC1 when the factors to be estimated are I(1). The IC1 criterion seems to be more robust than PC1. This is because the penalty of IC1 is less sensitive to the scaling effect; see Bai and Ng (2002). The results of our Monte Carlo experiments show that integrating the penalty term of $PC1$ in the objective function and calibrating the hight of the penalization as described in Section 1.2 provides a gain over the original PC1, IC1, and IPC1.

**DGP1 (I(1) factors and endogenous explanatory variables)** The simulation results for DGP1 (reported in the first part of Table 1.1) show that $\hat{d}_{EupPC1}$ gives a very precise estimation of the factor dimension and outperforms PC1, IPC1, IC1, ED, and $IC^T_{1,n}$. In contrast to all other criteria, ED gets worse as $N$ and $T$ increase. This weakness can be related to the strong endogeneity of the explanatory variables. For $(N,T) = (120,120)$, the MSEs of $\hat{\beta}_{EupBC}, \hat{\beta}(\hat{d}_{IPC1})$, and $\hat{\beta}(\hat{d}_{IC1^T_{1,n}})$ converge to 0. This is not surprising since our estimation strategy and the original method of Bai et al. (2009) used with $d_{max} \geq d$ (or with a consistent external dimensionality criterion) will produce very close outcomes in terms of MSE when both $N$ and $T$ are large enough. The problem, of course, is that the required sizes of $N$ and $T$ ensuring such evidence are unknown in practice. Note also that the outcomes of $\hat{\beta}(\hat{d}_{IPC1})$ are conditional on $d_{max} = 8$ and $\hat{\beta}(\hat{d}_{IPC1})$ can be used only when the $d$ static factors are not driven by the lags of a smaller number of dynamic factors. Such limitations are overcome by our method.

**DGP2 (mixed I(1)/I(0) factors)** The results of this experiment are reported in the second part of Table 1.1. The best estimation is obtained with $\hat{d}_{EupPC1}$. IPC1 has a tendency to overestimate the true number of I(1) factors, especially when $N$ increases and $T$ is fixed. In contrast, ED and $IC^T_{1,n}$ behave properly in such a case. Otherwise, both ED and $IC^T_{1,n}$ get slowly worse as $T$ increases and $N$ is fixed $((\hat{d}_{ED}, \hat{d}_{IC^T_{1,n}}) = (2.99,3),(2.98,2.98),(2.85,82)$ for $(N,T) = (120,60),(120,120)$, and $(120,300)$ respectively). PC1 and IC1 also show a tendency to underestimate the factor dimension. The reason these criteria estimate, on average, a smaller dimension than $\hat{d}_{EupPC1}$ (although $\hat{d}_{EupPC1}$ is obtained by strengthening iteratively the same penalty) is that the starting estimate of $d$ in our algorithm depends on the sample size and is larger than 8 for

$(N,T) = (120,120)$ and $(120,300)$. Unexpectedly, when $(N,T) = (120,300)$, the MSEs of all $\beta$ estimators are larger than those obtained with smaller sample sizes. This result can be explained by the occurrence of two effects when $N$ is fixed and $T$ is relatively large: the first effect is related to the problem of mixed strong/weak factors, which can lead to underestimate the number of I(0) factors since the proportion of the variance explained by the I(1) factors gets much larger than the proportion of the variance explained by the I(0) factors when $T$ grows faster than $N$. Indeed, the worst MSE affiliates with IPC1, which has, on average, the smallest estimate for $d$ when $N = 120$ and $T = 300$; the second effect is related to the inefficiency of estimating a bias that does not exist since the factors and the regressors are exogenous in DGP2. Recall that the bias estimator is the average of $N$ individual bias estimates and converges proportionally to $N$.

**DGP3 (the observed regressor is the underlying factor)** The last two parts of Table 1.1 present the estimation results of DGP3 obtained by initiating the iterations with two different starting estimates of $\beta$: the first estimate is $\hat{\beta}_{start}$ expressed in (1.13); the second estimate is obtained by scaling the factors $G_l, l = 1,\ldots,d_{max}$ in (1.13) with $(1 - r^2_{G_l,X_1})^{0.5}$, where $r_{G_l,X_1}$ is the sample correlation coefficient between $G_l$ and $X_{1i}$, as described Section 1.2.2. The goal of examining DGP3 is only to test whether the calibration of the starting factors $G_l$ in (1.13) will enhance the effectiveness of the estimation algorithm to correctly specify the model. The answer that can be deciphered from the table is: Yes!

**DGP4-6 (stationary panels with weak/strong autocorrelated errors)** Table 1.2 reveals that there is at least one case, where ED and $IC^T_{1,n}$ estimators outperform $\hat{d}_{EupPC1}$. In fact, strong autocorrelations in the idiosyncratic errors (DGP4 and DGP6) seem to inflate the number of factors obtained by using the penalty of Bai and Ng (2002). IPC1 seems to work well for DGP4 and DGP6 (except for $N = T = 30$), although theoretically this criterion is only appropriate for detecting integrated factors. The explanation of this outcome could lie in the penalty term of IPC1, which is, by construction, higher than the penalty of PC1 and IC1. The ED criterion also seems to be robust against the problem of high autocorrelated idiosyncratic errors, even if the autocorrelation coefficients are individual specific. $IC^T_{1,n}$ is based on the spectral density decomposition of the (disturbed) factor structure and is expected to be the most appropriate criterion for these cases. Our estimator $\hat{d}_{EupPC1}$ behaves similarly to $\hat{d}_{PC1}$ for DGP5 and DGP6. Both estimators have a tendency to select the largest possible dimension in the interval of the run parameter, i.e., $\min\{\sqrt{N},\sqrt{T}\}$ for $\hat{d}_{EupPC1}$ and 8 for $\hat{d}_{PC1}$. This result is in line with the simulation results of Onatski (2010) and Greenaway-McGrevy et al. (2012) for IC1 and PC1 when the autocorrelation coefficient in the

idiosyncratic errors is large (e.g., $\geq 0.7$). In fact, Assumption C in Bai and Ng (2002) forbid strong forms of correlation and heteroskedasticity in the error term. The Monte Carlo experiments of Bai and Ng (2002) consider only cases in which the correlation coefficient is smaller or equal to 0.5. Note that such a limitation is not necessary if the factors are I(1). This is because we can replace $a = 1$ with $a = T/(4 \log \log T)$ in the penalty term. The latter will diverge with $N$ and $T$ and dominate asymptotically any $O_p(1)$ structure in the idiosyncratic errors.

Finally, we want to emphasize that the goal of estimating panel models with unobserved common factors is not only to assess the dimension of the factor structure but also to efficiently estimate the slope parameters. Inspection of the MSE values reported in the second and third part of Table 1.2 shows that $\hat{\beta}_{Eup}$ does not suffer from an over-parameterization effect when the errors are strongly autocorrelated and the factors are over estimated. The additional factors seem to compensate for the unparameterized linear dependency in the idiosyncratic term. From the second and third part of Table 1.2, we can see that, for $(N, T) = (60, 60)$, the MSE of $\hat{\beta}_{Eup}$ is smaller than the MSEs of $\hat{\beta}(\hat{d}_{IPC1})$, $\hat{\beta}(\hat{d}_{ED})$, and $\hat{\beta}(\hat{d}_{IC^T_{1,n}})$. The first part of Table 1.2 shows that all criteria behave very well when the autocorrelation in the idiosyncratic errors is weak, in particular, $\hat{d}_{EupPC1}$, $\hat{d}_{IC1}$, and $\hat{d}_{IC^T_{1,n}}$.

The Monte Carlo experiments show that, in many configurations of the data, our algorithmic refinement provides more efficient estimates in terms of MSE for the estimator of $\beta$ than those that can be achieved if the feasible iterative least squares estimator is calculated with an externally selected factor dimension. Moreover, our results show that the iterative calibration of the penalty term makes the criteria of Bai and Ng (2002) more robust in a practical context, especially when $N$ and/or $T$ are small. If the idiosyncratic errors are strongly autocorrelated, the number of stationary factors will be overestimated but without affecting the efficiency of the slope estimator.

| N | T | MEAN $\hat{d}_{EupPC1}$ | MEAN $\hat{d}_{PC1}$ | MEAN $\hat{d}_{IC1}$ | MEAN $\hat{d}_{IPC1}$ | MEAN $\hat{d}_{ED}$ | MEAN $\hat{d}_{IC^T_{1,n}}$ | MSE $\hat{\beta}_{EupBC}$ | MSE $\hat{\beta}(d_{max})$ | MSE $\hat{\beta}(\hat{d}_{PC1})$ | MSE $\hat{\beta}(\hat{d}_{IC1})$ | MSE $\hat{\beta}(\hat{d}_{IPC1})$ | MSE $\hat{\beta}(\hat{d}_{ED})$ | MSE $\hat{\beta}(\hat{d}_{IC^T_{1,n}})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DGP1: panel cointegration model with I(1) factors and endogenous explanatory variables ($d = 2$) | | | | | | | | | | | | | | |
| 30 | 30 | 2.05 | 6.53 | 2.62 | 2.15 | 2.49 | 3.35 | 0.001 | 0.023 | 0.021 | 0.007 | 0.003 | 0.006 | 0.006 |
| 30 | 60 | 2.00 | 3.87 | 2.11 | 2.01 | 2.10 | 2.00 | 0.000 | 0.003 | 0.002 | 0.001 | 0.000 | 0.001 | 0.000 |
| 30 | 120 | 2.00 | 2.07 | 2.02 | 2.00 | 2.03 | 2.00 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 60 | 30 | 2.02 | 4.94 | 3.37 | 2.48 | 3.24 | 2.18 | 0.001 | 0.068 | 0.052 | 0.023 | 0.006 | 0.020 | 0.001 |
| 60 | 60 | 2.00 | 3.80 | 3.30 | 2.20 | 3.28 | 2.03 | 0.000 | 0.028 | 0.009 | 0.005 | 0.001 | 0.005 | 0.000 |
| 60 | 120 | 2.00 | 2.55 | 2.38 | 2.00 | 2.43 | 2.00 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 120 | 30 | 2.02 | 4.61 | 3.77 | 2.74 | 3.65 | 2.46 | 0.001 | 0.104 | 0.070 | 0.040 | 0.009 | 0.036 | 0.004 |
| 120 | 60 | 2.00 | 4.60 | 4.14 | 2.65 | 4.02 | 2.15 | 0.000 | 0.063 | 0.025 | 0.017 | 0.002 | 0.016 | 0.000 |
| 120 | 120 | 2.00 | 4.29 | 4.02 | 2.25 | 4.03 | 2.10 | 0.000 | 0.023 | 0.005 | 0.004 | 0.000 | 0.004 | 0.000 |
| DGP2: panel cointegration with mixed I(1)/I(0) factors: 1 I(1) factor and 2 I(0) factors ($d = 3$) | | | | | | | | | | | | | | |
| 30 | 30 | 3.08 | 6.45 | 3.00 | 2.92 | 2.97 | 4.52 | 0.002 | 0.022 | 0.012 | 0.002 | 0.006 | 0.006 | 0.010 |
| 30 | 60 | 3.00 | 4.14 | 2.99 | 2.46 | 2.97 | 2.98 | 0.001 | 0.052 | 0.013 | 0.006 | 0.040 | 0.009 | 0.008 |
| 30 | 120 | 3.00 | 2.98 | 2.94 | 1.68 | 2.92 | 2.94 | 0.014 | 0.260 | 0.059 | 0.076 | 0.263 | 0.080 | 0.075 |
| 60 | 30 | 3.00 | 4.15 | 3.00 | 2.95 | 2.98 | 3.00 | 0.001 | 0.004 | 0.001 | 0.001 | 0.002 | 0.003 | 0.001 |
| 60 | 60 | 3.00 | 3.00 | 3.00 | 2.61 | 2.98 | 2.99 | 0.001 | 0.029 | 0.003 | 0.004 | 0.029 | 0.006 | 0.005 |
| 60 | 120 | 3.00 | 2.98 | 2.96 | 1.90 | 2.95 | 2.95 | 0.011 | 0.177 | 0.042 | 0.058 | 0.190 | 0.056 | 0.061 |
| 120 | 30 | 3.00 | 3.00 | 3.00 | 2.96 | 2.98 | 3.00 | 0.000 | 0.003 | 0.000 | 0.000 | 0.002 | 0.002 | 0.000 |
| 120 | 60 | 3.00 | 3.00 | 3.00 | 2.80 | 2.99 | 3.00 | 0.000 | 0.020 | 0.000 | 0.000 | 0.019 | 0.003 | 0.001 |
| 120 | 120 | 3.00 | 2.99 | 2.99 | 2.16 | 2.98 | 2.98 | 0.007 | 0.099 | 0.020 | 0.023 | 0.127 | 0.023 | 0.025 |
| 120 | 300 | 3.00 | 2.87 | 2.84 | 1.33 | 2.85 | 2.82 | 0.160 | 0.941 | 0.446 | 0.461 | 0.978 | 0.441 | 0.444 |
| DGP3: the observed regressors are the underlying factors (naive starting slope estimate) | | | | | | | | | | | | | | |
| 30 | 30 | 1.01 | 6.18 | 1.00 | 1.00 | 1.00 | 1.84 | 4.679 | 4.681 | 4.681 | 4.679 | 4.679 | 4.679 | 4.680 |
| 60 | 60 | 1.00 | 1.02 | 1.00 | 1.00 | 1.00 | 1.00 | 4.693 | 4.693 | 4.693 | 4.693 | 4.693 | 4.693 | 4.693 |
| 120 | 120 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.693 | 4.693 | 4.693 | 4.693 | 4.693 | 4.693 | 4.693 |
| DGP3: the observed regressors are the underlying factors (calibrated starting slope estimate) | | | | | | | | | | | | | | |
| 30 | 30 | 0.01 | 6.12 | 0.01 | 0.00 | 0.01 | 7.74 | 0.00 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 30 | 120 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 120 | 120 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

TABLE 1.1: Simulation results for DGP1 - DGP3. The entries are the averages of the estimated dimensions and the MSEs of the slope estimator over 1000 replications.

| N | T | MEAN $\hat{d}_{EupPC1}$ | MEAN $\hat{d}_{PC1}$ | MEAN $\hat{d}_{IC1}$ | MEAN $\hat{d}_{IPC1}$ | MEAN $\hat{d}_{ED}$ | MEAN $\hat{d}_{IC^T_{1,n}}$ | MSE $\hat{\beta}_{Eup}$ | MSE $\hat{\beta}(d_{max})$ | MSE $\hat{\beta}(\hat{d}_{PC1})$ | MSE $\hat{\beta}(\hat{d}_{IC1})$ | MSE $\hat{\beta}(\hat{d}_{IPC1})$ | MSE $\hat{\beta}(\hat{d}_{ED})$ | MSE $\hat{\beta}(\hat{d}_{IC^T_{1,n}})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DGP4: 3 stationary factors and weak autocorrelations in the errors ($d = 3$) | | | | | | | | | | | | | | |
| 30 | 30 | 3.07 | 6.53 | 3.01 | 2.94 | 2.98 | 4.81 | 0.001 | 0.004 | 0.003 | 0.001 | 0.006 | 0.007 | 0.002 |
| 30 | 60 | 3.00 | 4.18 | 3.00 | 2.80 | 2.97 | 3.00 | 0.001 | 0.001 | 0.001 | 0.001 | 0.036 | 0.011 | 0.001 |
| 30 | 120 | 3.00 | 3.01 | 3.00 | 2.20 | 2.97 | 3.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.242 | 0.014 | 0.000 |
| 60 | 30 | 3.00 | 4.21 | 3.00 | 2.97 | 2.98 | 3.00 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.008 | 0.001 |
| 60 | 60 | 3.00 | 3.00 | 3.00 | 2.93 | 2.98 | 3.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.011 | 0.008 | 0.000 |
| 60 | 120 | 3.00 | 3.00 | 3.00 | 2.57 | 2.96 | 3.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.121 | 0.019 | 0.000 |
| 120 | 30 | 3.00 | 3.01 | 3.00 | 2.98 | 2.97 | 3.00 | 0.000 | 0.001 | 0.000 | 0.000 | 0.002 | 0.010 | 0.000 |
| 120 | 60 | 3.00 | 3.00 | 3.00 | 2.97 | 2.98 | 3.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 | 0.010 | 0.000 |
| 120 | 120 | 3.00 | 3.00 | 3.00 | 2.91 | 2.95 | 3.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.021 | 0.023 | 0.000 |
| DGP5: 1 stationary factor and strong autocorrelations in the errors ($d = 1$) | | | | | | | | | | | | | | |
| 30 | 30 | 5.00 | 7.97 | 7.71 | 1.97 | 1.94 | 6.56 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 30 | 60 | 4.93 | 7.80 | 6.47 | 1.00 | 1.04 | 1.39 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 30 | 120 | 2.83 | 6.64 | 1.68 | 0.97 | 1.00 | 1.08 | 0.001 | 0.001 | 0.001 | 0.001 | 0.012 | 0.001 | 0.001 |
| 60 | 30 | 5.00 | 7.97 | 7.81 | 1.68 | 1.57 | 2.00 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 60 | 60 | 7.88 | 7.90 | 7.49 | 1.00 | 1.00 | 1.09 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 |
| 60 | 120 | 4.46 | 6.59 | 2.67 | 1.00 | 1.00 | 1.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 |
| 120 | 30 | 5.00 | 7.99 | 7.94 | 1.53 | 1.25 | 2.75 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.000 |
| 120 | 60 | 8.00 | 8.00 | 7.98 | 1.00 | 1.00 | 1.06 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 120 | 120 | 8.82 | 7.62 | 6.24 | 1.00 | 1.00 | 1.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| DGP6: no factors and strong autocorrelations in the errors ($d = 0$) | | | | | | | | | | | | | | |
| 30 | 30 | 4.99 | 7.94 | 7.54 | 1.48 | 0.71 | 7.51 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| 30 | 60 | 4.83 | 7.72 | 5.62 | 0.00 | 0.03 | 1.09 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 30 | 120 | 1.79 | 6.42 | 0.57 | 0.00 | 0.00 | 0.17 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 60 | 30 | 5.00 | 7.93 | 7.65 | 1.17 | 0.27 | 2.14 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| 60 | 60 | 7.77 | 7.83 | 7.16 | 0.00 | 0.01 | 0.95 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 |
| 60 | 120 | 3.23 | 6.34 | 1.51 | 0.00 | 0.00 | 0.01 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 120 | 30 | 5.00 | 7.98 | 7.87 | 1.07 | 0.07 | 2.64 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| 120 | 60 | 7.99 | 7.99 | 7.94 | 0.00 | 0.00 | 0.62 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 120 | 120 | 8.11 | 7.44 | 5.47 | 0.00 | 0.00 | 0.00 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

TABLE 1.2: Simulation results for DGP4 - DGP6. The entries are the averages of the estimated dimensions and the MSEs of the slope estimator over 1000 replications.

## 1.5 Application: The Unobserved Risk Premia of Corporate Bonds

### 1.5.1 The Empirical Model

The empirical evidence shows that there is a discrepancy between the observed credit spreads and the theoretical spreads implied by the financial models of credit risk. Elton et al. (2001) assert that default risk cannot explain more than 25% of the credit spread variation. Longstaff et al. (2005) argue that non-default components such as bond-specific illiquidity and overall illiquidity risk do exist. Collin-Dufresne et al. (2001) examine the effect of a large number of risk proxies such as changes in the spot rate, changes in the slope of the yield curve, changes in Leverage, changes in the probability, and changes in the business climate. They detect high cross-correlations in the residuals of the regressed time series and conjecture that undefined missing factors generate these dependencies. The authors also examine the effects of several macroeconomic and financial determinants and argue that such variables can not solve the mystery.

In this chapter, we decompose the credit spread into individual specific components and unobserved common components generated by common risk factors. Because our focus is on estimating the dimension of the puzzling part of the credit spread, we restrict the observed individual specific components to the two most frequently used determinants in the literature on corporate finance, namely the credit default risk and the illiquidity risk; see, e.g., Huang and Huang (2012), Elton et al. (2001), and Longstaff et al. (2005). While the default risk is the basic component of structural models, the illiquidity risk is commonly used in reduced form models. We avoid introducing variables that are controversial in the literature and let the data inform us about the dimension of the missing risk factors.

Our empirical model is expressed as follows:

$$\mathrm{CS}_{it} \approx \mu_t + \sum_{k=1}^{K} \alpha_{kt}\delta_{ik} + \beta \mathrm{LR}_{it} + \sum_{l=1}^{d} \lambda_{il} f_{lt}. \tag{1.22}$$

The index $i \in \{1, \ldots, N\}$ denotes the single bonds, the index $t \in \{1, \ldots, T\}$ denotes the date. The explained variable $\mathrm{CS}_{it}$ is the corporate-government credit spread defined as

$$\mathrm{CS}_{it} = \mathrm{R}_{it} - \mathrm{R}_{G,it},$$

where $\mathrm{R}_{it}$ is the yield of the corporate bond $i$ at time $t$ and $\mathrm{R}_{G,it}$ is its duration-equivalent government bond. $\mu_t$ denotes the general average time-process of the credit spread.

Following many previous studies, we consider the rating class (or rating level) to be a measure for assessing the credit default risk; see, e.g., Gebhardt et al. (2005), Houweling et al. (2005), and Kagraoka (2010). In fact, the rating class constitutes the synthetic evaluation of the rating agencies, that takes into account the default probability as well as the recovery rate. In our model, this proxy is presented by the delta function $\delta_{ik}$, which we define as:

$$\delta_{ik} = \begin{cases} 1 & \text{if bond } i \text{ has the rating class } k \text{ and} \\ 0 & \text{else.} \end{cases}$$

In order to focus on the unobserved systematic risk premia, we follow Kagraoka (2010) and consider only bonds that did not experience a rating migration during the observation period. Unlike most existing work, we allow for time-varying rating effects $\alpha_{kt}$. This establishes a general framework that enables us to assess the response of the investors over time to the credit evaluation of the rating agency.

The explanatory variable $\mathrm{LR}_{it}$ measures the illiquidity risk of bond $i$ at time $t$. Because the trading frequency of corporate bonds is generally low, arbitrage theory implies that the price of illiquidity will be reflected in the bond yield. In the literature, several proxies of illiquidity have been considered; see, e.g., Chen et al. (2002), Bessembinder et al. (2006), Houweling et al. (2005), and Lesmond (1999). Following Bessembinder et al. (2006), we construct a measure based on the following quoted bid-ask spread:

$$\mathrm{LR}_{it} = \left| \frac{\mathrm{R}_{it}^{A} - \mathrm{R}_{it}^{B}}{\mathrm{R}_{it}^{B}} \right| \times 100,$$

where $\mathrm{R}_{it}^{A}$ and $\mathrm{R}_{it}^{B}$ are the ask yield and the bid yield of bond $i$ at time $t$. It is easy to realize that the larger the spread is, the more problematic the immediate trading becomes and vice versa. We expect the credit spread to be larger for less liquid bonds. Because this relationship can be generalized across all corporate bonds, we assume the slope parameter $\beta$ to be cross-sectionally homogeneous. Moreover, assuming $\beta$ to be constant over $i$ and $t$ not only simplifies the economic interpretation, but also greatly improves the asymptotic property of its estimator.

The term $f_{lt}$ represents the stochastic process describing the time pattern of the underlying common risk factors. We may interpret $f_{lt}$ as systematic risks, since they do not depend on $i$. The scores $\lambda_{il}$ are the corresponding individual loading parameters describing the effect of $f_{lt}$ on each bond $i$. Like Kagraoka (2010), we interpret the interaction between $\lambda_{il}$ and $f_{lt}$ as the systematic risk premium imposed by the investor on bond $i$ at time $t$.

The role of $d$ is intended to determine the number of missing factors in the credit spread puzzle. We interpret $d$ as a measure reflecting the degree of difficulty in diversifying the individual bond risks in the market. The higher the dimension of the existing common risk factors, the more difficult it is to diversify the risk.

The parameters $\beta, \lambda_{il}, f_{lt}$ and $d$ are estimated under Conditions (R.1)-(R.3) after eliminating the rating effects as described in Sections 1.2 and 1.3. We denote the transformations of $\mathrm{CS}_{it}$ and $\mathrm{LR}_{it}$ by $Y_{it}$ and $X_{it}$ respectively. .

For more convincing evidence, we compare our result with the results of the following model specifications:

$$\begin{aligned}
(M.1): \quad Y_{it} &= \textstyle\sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it}, \\
(M.2): \quad Y_{it} &= X_{it}\beta + \textstyle\sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it}, \\
(M.3): \quad Y_{it} &= D(L)Y_{i,t-1} + \textstyle\sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it}, \text{ and} \\
(M.4): \quad Y_{it} &= D(L)Y_{i,t-1} + X_{it}\beta + \textstyle\sum_{l=1}^{d} \lambda_{il} f_{lt} + \epsilon_{it},
\end{aligned}$$

where $D(L)$ is an $m$ polynomial lag-operator. Model $(M.1)$ is a standard static factor model that can be estimated by principal component (PC) analysis; see, e.g., Stock and Watson (2002) and Bai and Ng (2002). In Model $(M.2)$, we assume that the regressors are exogenous and estimate the parameters by using the two-step PC estimator of Coakley et al. (2002). In the first step, the OLS estimator of $\beta$ is obtained by ignoring the factor structure. In a second step, we use the OLS residuals and estimate the factor structure by the PC method. Models $(M.3)$ and $(M.4)$ are estimated by the feasible iterative least squares method.

### 1.5.2 Data Description

Our data are extracted from Datastream, which is an online database containing a broad range of financial entities and instruments. Our explained variable is the credit spread. Because the maturities of many bonds do not exactly match the maturity of the available government benchmark bonds, Datastream uses a linear interpolation to approximate the yield of the duration-equivalent government benchmark. The spread is expressed as yield differences in basis points. The explanatory variables are the credit rating levels and the quoted bid-ask yield spread of the corresponding bonds. We choose the U.S. corporate bonds rated by S&P. Our observation period extends from September 18, 2006, to May 25, 2008. This allows us to compare the dimension of the unobserved factor structure before and after the subprime crisis emerged in the middle of this time interval. Moreover, we choose fixed rate bonds with long remaining time to maturity. This is to marginalize the possible term structure effects. Finally, we ignore securities

that have missing prices. We then obtain an equidistant balanced panel data set based on 111 U.S. corporate bonds over a period of 397 business days. The retained rating classes are AAA, AA, A, and BBB; see Table 1.3.

| Rating class | AAA | AA | A | BBB | Total |
|---|---|---|---|---|---|
| Number | 24 | 29 | 21 | 31 | 111 |

TABLE 1.3: Number of corporate bonds by rating class.

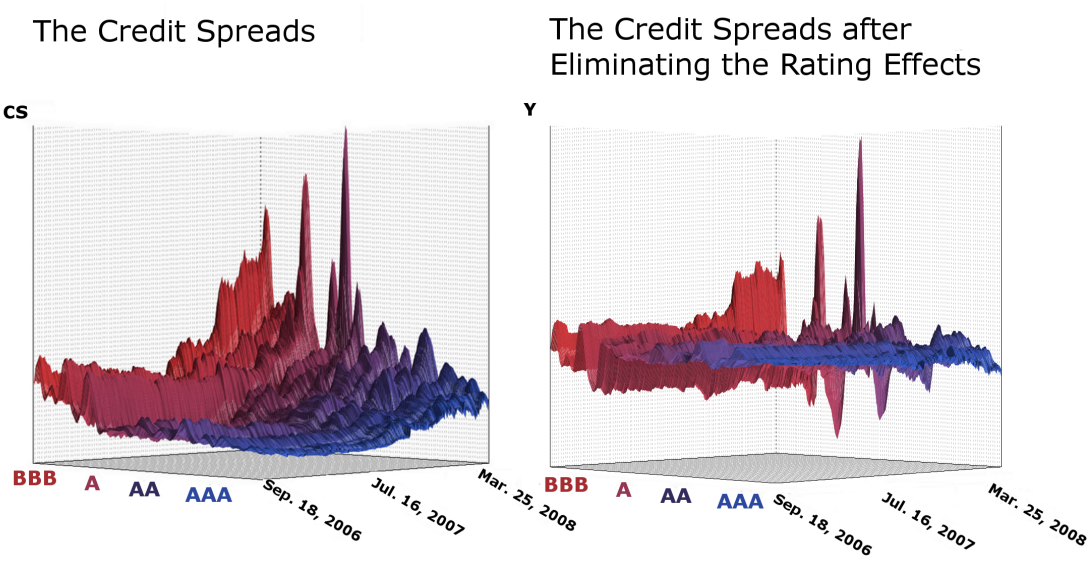


FIGURE 1.1: Three-dimensional plot of the credit spread curves before and after performing the within-group transformation to eliminate the rating effects. The credit spread values are on the y-axis, time is on the x-axis, and the index of bonds ordered by rating class is on the z-axis.

Figure 1.1 displays a three-dimensional plot of the credit spread curves before and after performing the within-group transformation discussed in Section 1.3.1.

To examine whether the data sets are affected by unit root common factors, we use the $MQ$ test of the *Panel Analysis of Non-stationarity in Idiosyncratic and Common Components* (PANIC) proposed by Bai (2004). PANIC is developed to detect the source and the nature of the non-stationarity in the data. In our application, we restrict our test to the following simplified test problem:

$$\begin{aligned} H_0 &: \quad k \geq 1 \text{ and} \\ H_1 &: \quad k < 1, \end{aligned}$$

where $k$ is the number of independent common unit root processes. The intuition behind this test is that if the first factor (corresponding to the largest eigenvalue) is integrated, then $MQ$ cannot reject the null hypothesis. The test is applied to the credit spread

variable $Y_i$ (after eliminating the rating effects) and to the remainder term $\hat{W}_i = Y_i - X_i\hat{\beta}_{EupBC}$ obtained from the entirely updated regression. For $k = 1$, we obtain the following result:

| Panel | $MQ_f^c$-value | Lag order | Critical value at significance level 0.05 |
|---|---|---|---|
| $Y_i$ | $-2.2774$ | 7 | $-11.022$ |
| $\hat{W}_i$ | $-1.6019$ | 7 | $-11.022$ |

The null hypothesis cannot, hence, be rejected for both $Y_i$ and $\hat{W}_i$. Alternatively, one can use the panel data unit root tests of Karavias and Tzavalis (2012), who allow for a common structural break in the individual effects. The estimation results of Models (M.3) and (M.4) considered with $D(L)Y_{i,t-1} = D_0Y_{i,t-1} + D_1Y_{i,t-2} + D_2Y_{i,t-3}$, support qualitatively the hypothesis of the unit root because $1 - \hat{D}_0 - \hat{D}_1L - \hat{D}_2L^2 \approx 0$ for $L = 1$ (see Table 1.4).

### 1.5.3 Empirical Results and Interpretations

The estimation results are reported in Table 1.4. The effect of the illiquidity risk is positive and significant in our Model as well as in (M.2), and (M.4). $\hat{\beta}_{EupBC}$ amounts to 0.0217 (with a standard deviation of 0.004). These results are in line with the previous findings of Chen et al. (2002), Elton et al. (2001), and Kagraoka (2010): The more illiquid the bond, the higher the expected credit spread.

Figure 1.2 shows the percentages of the ordered eigenvalues related to the unobserved factor structure in its static form as well as the eigenvalues obtained after applying on the estimated factors a VAR-regression. By using the penalty term of PC1 in our algorithm, we estimate $\hat{d}_{Eup} = 11$. The test $\mathcal{K}_3$ of Bai and Ng (2007), however, detects the presence of only 2 primitive shocks. This result is confirmed by the information criterion $IC_{1,n}^T$ of Hallin and Liška (2007), which suggests the presence of 2 dynamic factors in all models. The ED criterion of Onatski (2010) is optimized at $7, 6, 2$, and 2 for Models (M.1), (M.2), (M.3), and (M.4) respectively. The criteria of Bai (2004) indicate the presence of at least 2 unit root factors in all models except for (M.3) and (M.4), where the unit root source seems to be automatically integrated by the lags of $Y_{it}$. 1.3 (a).

Our time-varying estimates $\hat{\alpha}_{kt}$ and their corresponding 95% confidence intervals are depicted in Figure 1.3(a). The confidence intervals of the default risk parameters indicate that the rating effects are statistically significant, except for class A during the time between January and February in 2008. The part of the variance explained by the default risk accounts for 24.06%. This result agrees with the results of most research on

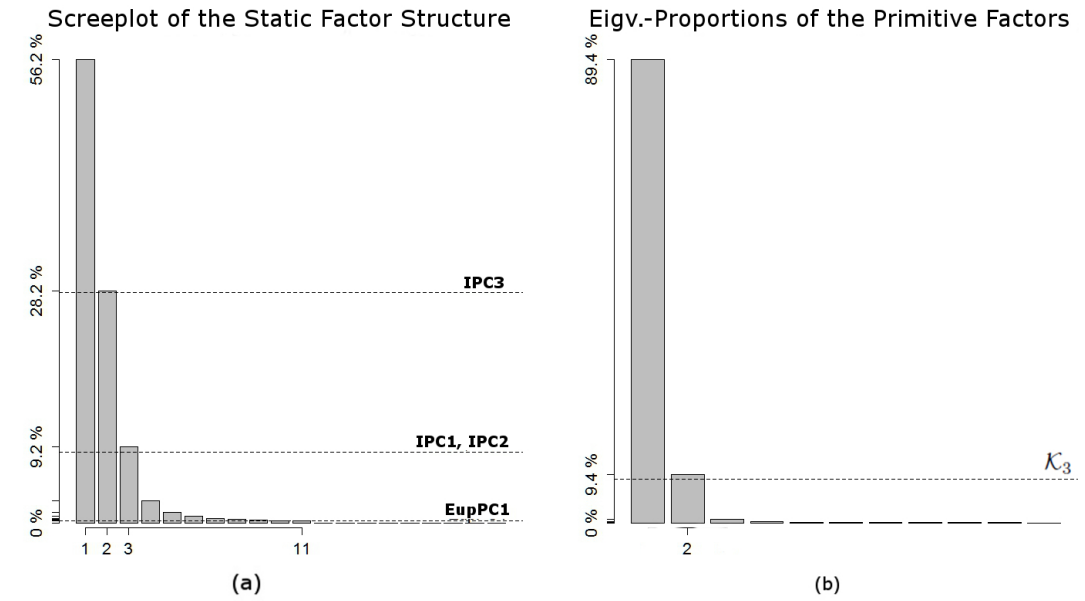


FIGURE 1.2: (a) The screeplot of the eigenvalues obtained from the matrix of $\hat{\Sigma}(\hat{\beta}_{CupBC}, \hat{d}_{Eup})$; (b) Proportions of the (squared) eigenvalues obtained from the residuals of the VAR-regression (with $p = 1$) applied on the estimated factors (after integrating the first 3 I(1) factors).

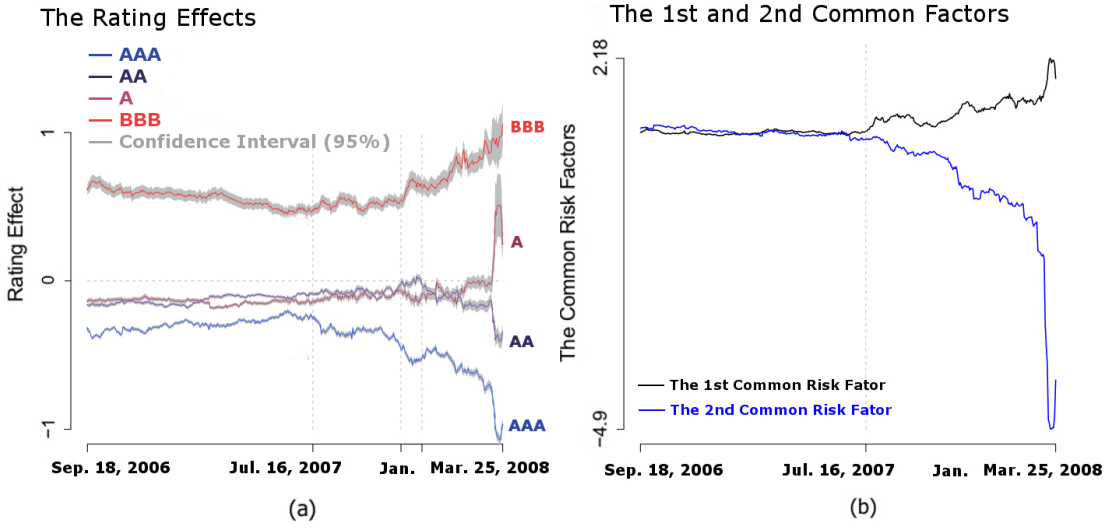


FIGURE 1.3: (a) the time series of the estimated rating effects; (b) the EupBC estimated first and second common factors.

| Regressor | EupBC | M.1 | M.2 | M.3 | M.4 |
|---|---|---|---|---|---|
| $X_t$ | 0.0217*** | - | 0.3013*** | - | 0.0047*** |
| | (0.004) | - | (0.006) | - | (0.001) |
| $Y_{t-1}$ | - | - | - | 0.8029*** | 0.8021*** |
| | - | - | - | (0.005) | (0.005) |
| $Y_{t-2}$ | - | - | - | 0.1061*** | 0.1055*** |
| | - | - | - | (0.007) | (0.007) |
| $Y_{t-3}$ | - | - | - | 0.0865*** | 0.0869*** |
| | - | - | - | (0.005) | (0.005) |
| Number of Factors | | | | | |
| Static I(1)/I(0) | | | | | |
| $g_{NT,Eup}$ | 11 | - | - | - | - |
| PC1 ($d_{max} = 11$) | - | 11 | 11 | 11 | 11 |
| ED | - | 7 | 6 | 2 | 2 |
| Static I(1) | | | | | |
| IPC1 ($d_{max} = 11$) | 3 | 3 | 3 | 0 | 0 |
| IPC2 ($d_{max} = 11$) | 3 | 3 | 3 | 0 | 0 |
| IPC3 ($d_{max} = 11$) | 2 | 2 | 2 | 0 | 0 |
| Primitive (dynamic) | | | | | |
| $\mathcal{K}_3(m = 1, \delta = 1/4)$ | 2 | 3 | 2 | 2 | 2 |
| $IC^T_{1,n}$ | 2 | 2 | 2 | 2 | 2 |

TABLE 1.4: Estimation results for Models (M.1)-(M.4).
The columns labeled with M.1-M.4 respectively present the estimation results of Models (M.1)- (M.4). PC1 is the panel criterion of Bai and Ng (2002). ED is the threshold criterion of Onatski (2010). IPC1-IPC3 are from Bai (2004). $\mathcal{K}_3$ is the selection criterion of Bai and Ng (2007). $IC^T_{1,n}$ is the information criterion of Hallin and Liška (2007). High significant coefficients (p-value $< 1\%$) are indexed by "***". The values between parentheses are the corresponding estimated standard deviations.

the credit spread; see, e.g., Collin-Dufresne et al. (2001), Huang and Huang (2012), and Kagraoka (2010).

From Figure 1.3(a), we can see that the time patterns of $\hat{\alpha}_{kt}$ exhibit some structural changes after July 16, 2007, in particular, the volatility of $\hat{\alpha}_{kt}$ for AAA, A, and BBB. The negative effect of the rating class A registered during the periods prior to mid-July became unstable and positive in 2008. These structural changes coincide with the beginning of the subprime crisis in the U.S. market. The market perception of the credit risk assessment performed by an external rating agency seems to depend on the market situation and is not constant over time even if bonds remain in the same rating class.

The estimated factors, $\hat{f}_{1t}$ and $\hat{f}_{2t}$, are displayed in Figure 1.3(b). These factors explain about 84.5% of the variance in the factor structure. The forms of $\hat{f}_{1t}$ and $\hat{f}_{2t}$ over time support the non-stationarity hypothesis. But note that these factors do not necessary affect the totality of bonds.

The first and second risk component, defined respectively as $\hat{C}_{it1} = \hat{\lambda}_{i1}\hat{f}_{1t}$ and $\hat{C}_{it2} = \hat{\lambda}_{i2}\hat{f}_{2t}$, are displayed in Figure 1.4. $\hat{C}_{it1}$ explains about 17.27% of the total variance

of the credit spread. Bonds, which had positive $\hat{C}_{it1}$ values during the period between

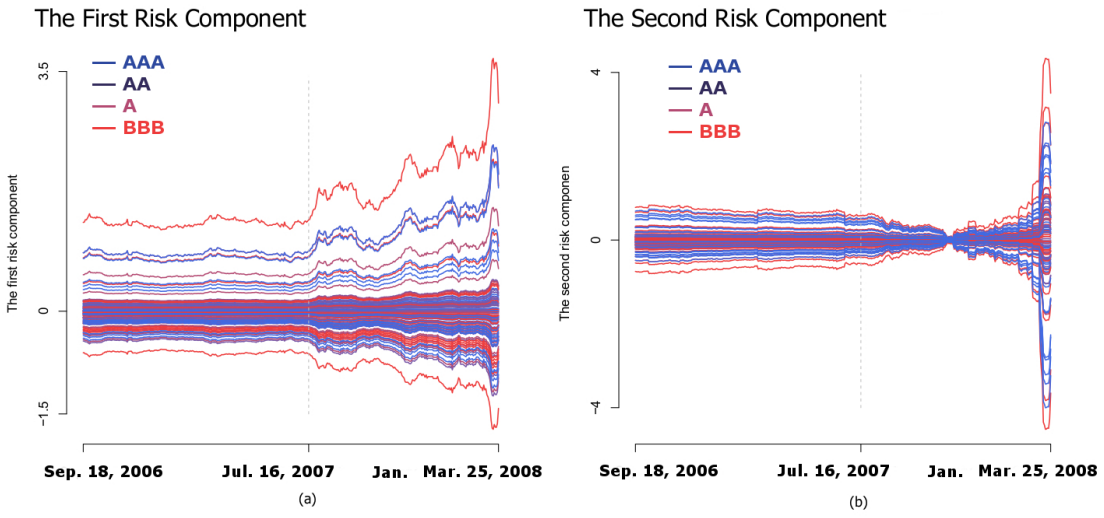


FIGURE 1.4: (a) the first risk component $\hat{C}_{it1} = \hat{\lambda}_{i1}\hat{f}_{1t}$; (b) the second risk component $\hat{C}_{it2} = \hat{\lambda}_{i2}\hat{f}_{2t}$.

September 18, 2006 and July 16, 2007, experienced an important rise in the next period, while bonds with negative $\hat{C}_{it1}$ experienced further decreases after July 16, 2007. This result confirms the hypothesis of Jegadeesh and Titman (1993) and Chan et al. (1996), who assert that security returns are affected by a so-called *momentum effect*, because investors typically buy stocks that have performed well in the past, and sell stocks that have performed poorly. Our analysis thus sheds some light on an ongoing discussion in the literature on stock market prices. The part of variance explained by the second risk component amounts to 12.94%. The individual patterns of $\hat{C}_{it2}$ seem to reflect the complexity of the market behavior in the subprime period.

When re-estimating our panel model for the period spanning only the time before July 16, 2007, we detect the presence of only one primitive factor. The number of detected common factors can therefore be interpreted as an index for assessing the complexity of the market and the difficulty of diversification, as mentioned in Elton et al. (2001) and Amato and Remolona (2003). The higher the number of common risk factors, the more complex the market is.

## 1.6 Conclusion

In this chapter, we extend the iterative least squares approach developed to estimate panel data models with unobserved factor structure in such a way that we allow for the number of factors to be unknown a priori. The basic idea of our extension is to treat

the iterative least squares estimator of Bai (2009) and Bai et al. (2009) as functions depending on a run parameter $\mathbf{d}$. The latter is fitted by means of a penalty term that is directly integrated into the global objective function to be optimized. The final solution is obtained by alternating between an inner iteration to optimize $\hat{\beta}(\mathbf{d}), \hat{F}(\hat{\beta}, \mathbf{d})$, and $\hat{\Lambda}_i(\hat{\beta}, \hat{F}, \mathbf{d})$ for each given $\mathbf{d}$ and an outer iteration to select the optimal dimension $\hat{d}$. Monte Carlo experiments show that our algorithm provides more efficient estimates in terms of MSE than could be achieved if the estimator of Bai et al. (2009) is calculated with an externally selected factor dimension. We consider, in our application, the problem of the credit spread puzzle and estimate the number of the hidden risk factors jointly with the effect of the observed risk components. Our result proves the presence of two unobserved common risk factors affecting the U.S. corporate bonds during the period between September 2006 and March 2008, while one single risk factor is sufficient to describe the data for all time periods prior to the beginning of the subprime crisis in 2007. Our analysis neglects, however, the possible effect of taxes. This component can be introduced in the regression function by means of a reasonable determinant. There is also a large potential for expanding panel data models with structural breaks on the individual effects, as proposed by De Wachter and Tzavalis (2012), to panel models with structural breaks in the unobserved factor structure.

# Chapter 2

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

## 2.1 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 the 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., H. (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 very often 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 et al. (2013), Pesaran (2006), Bai (2009), Bai et al. (2009), and Kneip et al. (2012) have focused on advanced panel models for which the unobservable individual effects are allowed to have heterogeneous (i.e., individual specific) 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:

$$y_{it} = \sum_{j=1}^{P} x_{itj}\beta_j + \nu_{it} + \epsilon_{it} \quad \text{for } i \in \{1, \ldots, n\} \quad \text{and } t \in \{1, \ldots, T\}, \tag{2.1}$$

where $y_{it}$ is the dependent variable for each individual $i$ at time $t$, $x_{itj}$ is the $j$th element of the vector of explanatory variables $x_{it} \in \mathbb{R}^P$, and $\epsilon_{it}$ 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 be generated by $d$ common time-varying factors. The following two specifications of the time-varying individual effects $\nu_{it}$ are implemented in our R package **phtt**:

$$\nu_{it} = \begin{cases} v_{it} &= \sum_{l=1}^{d} \lambda_{il} f_{lt}, & \text{for the model of Bai (2009),} \\ v_i(t) &= \sum_{l=1}^{d} \lambda_{il} f_l(t), & \text{for the model of Kneip et al. (2012).} \end{cases} \tag{2.2}$$

Here, $\lambda_{il}$ 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.

Note that the explicit consideration of an intercept in model (2.1) is not necessary but may facilitate interpretation. If $x_{it}$ includes an intercept, the time-varying individual effects $\nu_{it}$ are centered around zero. If $x_{it}$ does not include an intercept, the time-varying individual effects $\nu_{it}$ are centered around the overall mean.

Model (2.1) includes the classical panel data models with additive time-invariant individual effects and common time-specific effects. This model is obtained by choosing $d = 2$ with a first common factor $f_{1t} = 1$ for all $t \in \{1, \ldots, T\}$ that has individual loadings parameters $\lambda_{i1}$, 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

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

see, e.g., Bai (2009) and Kneip et al. (2012). For the model of Kneip et al. (2012), $f_{lt}$ in conditions (a) and (b) has to be replaced by $f_l(t)$. As usual in factor models, a certain degree of indeterminacy remains, because the factors can only be determined up to sign changes and different ordering schemes.

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 nonparametric functions, say, $\vartheta_i(t)$. In this way (2.1) becomes a semi-parametric model and its estimation is done using a two-step estimation procedure, which we explain in more detail in Section 2.2. The asymptotic properties of this method rely, however, on independent and identically distributed errors.

Alternatively, Bai (2009) allows for weak forms of heteroskedasticity and dependency in both time and cross-section dimensions and proposes an iterated least squares approach to estimate (2.1) for stationary time-varying individual effects $v_{it}$ such as ARMA processes or non-stationary deterministic trends. However, Bai (2009) rules out a large class of non-stationary processes such as stochastic processes with integration.

Moreover, 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 (2014) in order to estimate the number of unobserved common factors $d$ jointly with the remaining model parameters; see Section 2.4 for more details.

Besides the implementations of the methods proposed by Kneip et al. (2012), Bai (2009), and Bada and Kneip (2014) the R package **phtt** comes with a wide range of criteria (16 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 2.2.
- `Eup()`: Computes the estimators of the model parameters according to the method of Bai (2009) and Bada and Kneip (2014); see Section 2.4.
- `OptDim()`: Allows for a comparison of the estimated factor dimensions $\hat{d}$ obtained from many different (in total 16) criteria; see Section 2.3.
- `checkSpecif()`: Tests whether to use a classical fixed effects panel model or a panel model with individual effects $\nu_{it}$; see Section 2.5.1.

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

Standard methods for estimating models for panel and longitudinal data are also implemented in the R packages **plm** (Croissant and Millo, 2008), **nlme** (Pinheiro et al., 2014), and **lme4** (Bates et al., 2014); 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 with

large dimensions $n$ and $T$ and considers in the case when the unobserved heterogeneity effects are time-varying.

To the best of our knowledge, our **phtt** package Bada and Liebl (2014a) is the first software package that offers the estimation methods of Bai (2009) and Kneip et al. (2012). Regarding the different dimensionality criteria that can by accessed via the function `OptDim()` only those of Bai and Ng (2002) are publicly available as MATLAB codes 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 per capita cigarette consumptions of $n = 46$ American states from 1963 to 1992 ($T = 30$) as well as data about the income per capita and cigarette prices (see, e.g., Baltagi and Levin (1986) for more details on the dataset).

We follow Baltagi and Li (2006), who estimate the following panel model:

$$\ln(\texttt{Consumption}_{it}) \quad = \quad \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + e_{it}. \tag{2.3}$$

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.

We revisit this model, but allow for a multidimensional factor structure such that

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

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 2.1.

Section 2.2 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 2.3 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 2.4 deals with the explanation as well as application of the panel method proposed by Bai (2009), which is basically appropriate for stationary and relatively unstructured common factors $f_{lt}$.

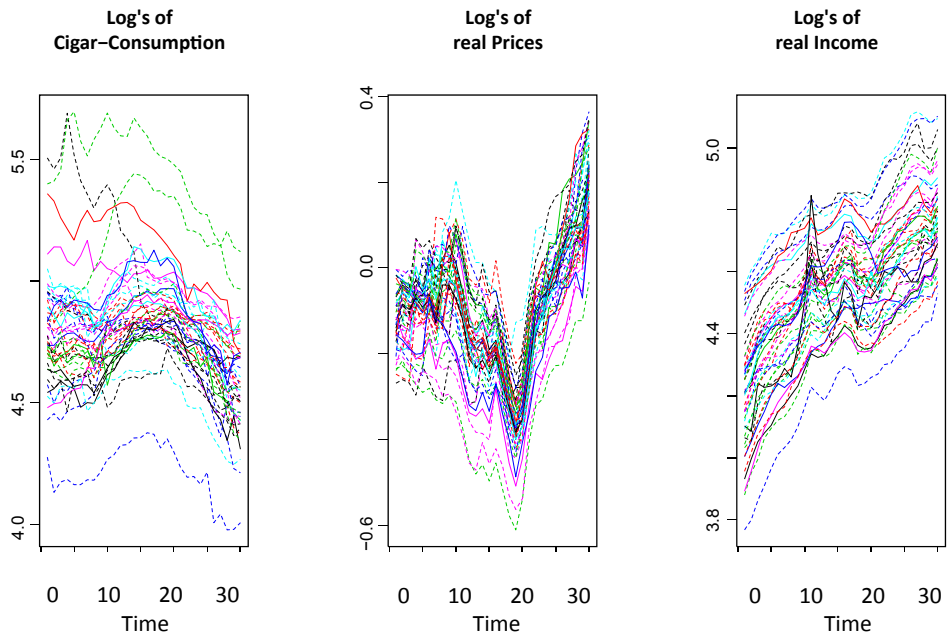


FIGURE 2.1: Time series of the dependent variable $\ln(\texttt{Consumption}_{it})$ and regressor variables $\ln(\texttt{Price}_{it})$ and $\ln(\texttt{Income}_{it})$.

## 2.2 Panel Models for Heterogeneity in Time Trends

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

$$y_{it} = \sum_{j=1}^{P} x_{itj}\beta_j + v_i(t) + \epsilon_{it}, \tag{2.4}$$

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

$$v_i(t) = \sum_{l=1}^{d} \lambda_{il} f_l(t). \tag{2.5}$$

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 discrete 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 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 $\beta_j$ 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 $\beta_j$ and $v_i(t)$ are estimated by the minimization of

$$\sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} \left( y_{it} - \sum_{j=1}^{P} x_{itj}\beta_j - \vartheta_i(t) \right)^2 + \sum_{i=1}^{n} \kappa \int_1^T \frac{1}{T} \left( \vartheta_i^{(m)}(s) \right)^2 ds, \tag{2.6}$$

over all $\beta_j \in \mathbb{R}$ and all $m$-times continuously differentiable functions $\vartheta_i(t)$, where $\vartheta_i^{(m)}(t)$ denotes the $m$th derivative of the function $\vartheta_i(t)$. 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 (2.6) to formalize the following objective function:

$$S(\beta, \zeta) = \sum_{i=1}^{n} \left( ||Y_i - X_i\beta - Z\zeta_i||^2 + \kappa \zeta_i^\top R \zeta_i \right), \tag{2.7}$$

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 $R$ 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. $\kappa$ 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.

In contrast to Kneip et al. (2012), we do not specify a common time effect in model (2.4), but the vector of explanatory variables is allowed to contain an intercept. This means that the time-varying individual effects $v_i(t)$ are not centered around zero for each specific time point $t$, but around a common intercept term. The separate estimation of the common time effect, say $\theta_t$, is also possible with our **phtt** package; we discuss this in detail in Section 2.5.

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

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa) X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa) Y_i\right), \tag{2.8}$$

$$\hat{\zeta}_i = (Z^\top Z + \kappa R)^{-1} Z^\top (Y_i - X_i \hat{\beta}), \text{ and} \tag{2.9}$$

$$\tilde{v}_i = \mathcal{Z}_\kappa \left(Y_i - X_i \hat{\beta}\right), \text{ where } \mathcal{Z}_\kappa = Z \left(Z^\top Z + \kappa R\right)^{-1} Z^\top. \tag{2.10}$$

**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

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \tilde{v}_i \tilde{v}_i^\top. \tag{2.11}$$

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

$$\hat{f}_l(t) = \sqrt{T} \hat{\gamma}_{lt} \text{ for all } l \in \{1, \ldots, d\}, \tag{2.12}$$

where $\hat{\gamma}_{lt}$ is the $t$th 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 2.1. The estimates of the individual loadings parameters $\lambda_{il}$ are obtained by ordinary least squares regressions of $\left(Y_i - X_i \hat{\beta}\right)$ on $\hat{f}_l$, where $\hat{f}_l = (\hat{f}_l(1), \ldots, \hat{f}_l(T))^\top$. Recall from conditions (a) and (b) that $\hat{\lambda}_{il}$ can be calculated as follows:

$$\hat{\lambda}_{il} = \frac{1}{T} \hat{f}_l^\top \left(Y_i - X_i \hat{\beta}\right). \tag{2.13}$$

A crucial part of the estimation procedure of Kneip et al. (2012) is the re-estimation of the time-varying individual effects $v_i(t)$ in **Step 2** by $\hat{v}_i(t) := \sum_{l=1}^{d} \hat{\lambda}_{il} \hat{f}_l(t)$, 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 criterion; see also Section 2.3. This re-estimation leads to more efficiently estimated time-varying individual effects.

Kneip et al. (2012) derive the consistency of the estimators as $n, T \to \infty$ and show that the asymptotic distribution of common slope estimators is given by $\hat{\Sigma}_\beta^{-1/2}(\hat{\beta} - \mathrm{E}_\epsilon(\hat{\beta})) \xrightarrow{d} \mathbf{N}(0, I)$, where

$$\hat{\Sigma}_\beta = \sigma^2 \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa) X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa)^2 X_i\right) \left(\sum_{i=1}^{n} X_i^\top (I - \mathcal{Z}_\kappa) X_i\right)^{-1}. \tag{2.14}$$

A consistent estimator of $\sigma^2$ can be obtained by

$$\hat{\sigma}^2 = \frac{1}{(n-1)T} \sum_{i=1}^{n} ||Y_i - X_i\hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{i,l}\hat{f}_l||^2. \tag{2.15}$$

To determine the optimal smoothing parameter $\kappa_{opt}$, Kneip et al. (2012) propose the following cross validation (CV) criterion:

$$CV(\kappa) = \sum_{i=1}^{n} ||Y_i - X_i\hat{\beta}_{-i} - \sum_{l=1}^{d} \hat{\lambda}_{-i,l}\hat{f}_{-i,l}||^2, \tag{2.16}$$

where $\hat{\beta}_{-i}$, $\hat{\lambda}_{-i,l}$, and $\hat{f}_{-i,l}$ are estimates of the parameters $\beta$, $\lambda$, and $f_l$ based on the dataset without the $i$th observation. 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 the following Section 2.2.1.

### 2.2.1 Computational Details

Theoretically, it is possible to determine $\kappa$ by the CV criterion in (2.16); however, 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 the estimator $\hat{d}$ is influenced by the choice of $\kappa$.

In order to get a quick and effective solution, we propose to determine the smoothing parameter $\kappa$ by generalized cross validation (GCV). However, we cannot apply the classical GCV formulas as proposed, e.g., in Craven and Wahba (1979) since we do not know the parameters $\beta$ and $v_i(t)$. Our computational algorithm for determining the GCV smoothing parameter $\kappa_{GCV}$ is based on the method of Cao and Ramsay (2010), who propose optimizing objective functions of the form (2.7) by updating the parameters iteratively in a functional hierarchy. Formally, the iteration algorithm can be described as follows:

1. For given $\kappa$ and $\beta$, we optimize (2.7) with respect to $\zeta_i$ to get

$$\hat{\zeta}_i = (Z^\top Z + \kappa R)^{-1} Z^\top (Y_i - X_i\beta). \tag{2.17}$$

2. By using (2.17), we minimize (2.7) with respect to $\beta$ to get

$$\hat{\beta} = \left( \sum_{i=1}^{n} X_i^\top X_i \right)^{-1} \left( \sum_{i=1}^{n} X_i^\top (Y_i - Z\hat{\zeta}_i) \right) \tag{2.18}$$

3. Once (2.17) and (2.18) are obtained, we optimize the following GCV criterion to calculate $\kappa_{GCV}$:

$$\kappa_{GCV} = \arg\min_{\kappa} \frac{1}{\frac{n}{T} tr(I - \mathcal{Z}_\kappa)^2} \sum_{i=1}^{n} ||Y_i - X_i\hat{\beta} - \mathcal{Z}_\kappa(Y_i - X_i\hat{\beta})||^2. \tag{2.19}$$

The program starts with initial estimates of $\beta$ and $\kappa$ 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 (2.50) and the initial value $\kappa_{start}$ 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 (2.18) does not have to be updated during the iteration process. Moreover, the determination of the GCV-minimizer in (2.19) can be easily performed in R using the function `smooth.spline()`, which calls on a rapid C-routine.

But note that the GCV smoothing parameter $\kappa_{GCV}$ in (2.19) does not explicitly account for the factor structure of the time-varying individual effects $v_i(t)$ as formalized in (2.2). In fact, given that the assumption of a factor structure is true, the goal shall not be 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 $\kappa_{opt}$ will be smaller than $\kappa_{GCV}$; see Kneip et al. (2012).

If the goal is to obtain optimal estimates of $f_l(t)$, $\kappa_{opt}$ will be used as an upper bound when minimizing the CV criterion (2.16) (via setting the argument `CV = TRUE`); which, however, can take some time. Note that, this optimal smoothing parameter $\kappa_{opt}$ depends on the unknown factor dimension $d$. Therefore, we propose to, first, estimate the dimension based on the smoothing parameter $\kappa_{GCV}$ and, second, to use the estimated dimension $\hat{d}$ (via explicitly setting the dimension argument `factor.dim`$= \hat{d}$) in order to determine the dimension-specific smoothing parameter $\kappa_{opt}$ (via setting the argument `CV = TRUE`).

### 2.2.2 Application

This section is devoted to the application of the method of Kneip et al. (2012) discussed above. The computation of this method is accessible through the function `KSS()`, which

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 `formula` is compatible with the usual `R`-specific symbolic designation of the model. The unique specificity here is that the variables should be defined as $T \times n$ matrices, where $T$ is the temporal dimension and $n$ is the number of the cross-section unites.[1]

The argument `additive.effects` makes it possible to extend the model (2.4) for additional additive `individual`, `time`, or `twoways` effects as discussed in Section 2.5.

If the logical argument `consult.dim.crit` is set to `TRUE` all dimensionality criteria discussed in Section 2.3 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 2.3. 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 (2.15), where $\lfloor x \rfloor$ indicates the integer part of $x$. The argument `level` allows to adjust the significance level for the dimensionality testing procedure (2.21) of Kneip et al. (2012); see Section 2.3.

`CV` is a logical argument. If it is set to `TRUE` the cross validation criterion (2.16) of Kneip et al. (2012) will be computed. In the default case, the function uses the GCV method discussed above in Section 2.2.1.

The factor dimension $d$ can 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 $\frac{1}{n}\sum_{i=1}^{n} \hat{\lambda}_{il}^2 = 1$, which can be done by setting `restrict.mode = "restrict.loadings"`.

[1] Note that **phtt** is written for balanced panels. Missing values have to be replaced in a pre-processing step by appropriate imputation methods.

As an illustration we estimate the Cigarettes model (2.3) introduced in Section 2.1:

$$\begin{aligned} \ln(\texttt{Consumption}_{it}) &= \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + e_{it} \qquad (2.20) \\ \text{with} \quad e_{it} &= \sum_{l=1}^{d} \lambda_{il} f_l(t) + \epsilon_{it}. \end{aligned}$$

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. The variables are stored 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 $n$.

```
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 $\beta_1$, $\beta_2$, the factors $f_l(t)$, the loadings parameters $\lambda_{il}$, 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 and the log-real incomes on the log-sales of cigarettes 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)
```

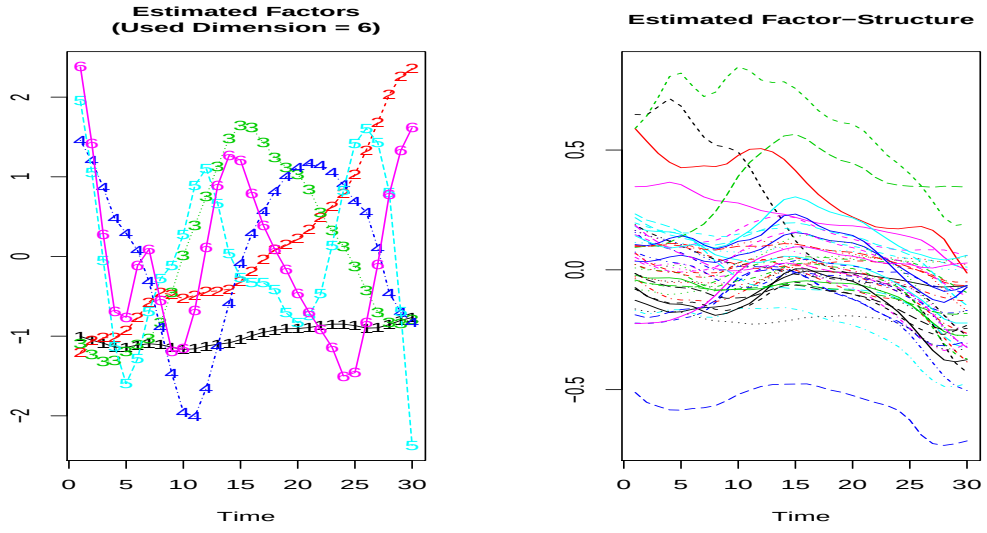


FIGURE 2.2: LEFT PANEL: Estimated factors $\hat{f}_1(t), \ldots, \hat{f}_6(t)$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_1(t), \ldots, \hat{v}_n(t)$.

The left panel of Figure 2.2 shows the six estimated common factors $\hat{f}_1(t), \ldots, \hat{f}_6(t)$ and the right panel of Figure 2.2 shows the $n = 46$ estimated time-varying individual effects

$\hat{v}_1(t), \ldots, \hat{v}_n(t)$. The common factors are ordered correspondingly to the decreasing sequence of their eigenvalues. Obviously, the first common factor is nearly time-invariant; this suggests extending the model (2.20) by additive `individual` (time-invariante) effects; see Section 2.5 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 2.3. Note that the consideration of different factor dimensions $d$ would not alter the results for the slope parameters $\beta$ since the estimation procedure of Kneip et al. (2012) for the slope parameters $\beta$ does not depend on the dimensionality parameter $d$.

## 2.3 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:

$$KSS(d) = \frac{n \sum_{r=d+1}^{T} \hat{\rho}_r - (n-1)\hat{\sigma}^2 tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2n \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)^2)}} \overset{a}{\sim} N(0,1), \tag{2.21}$$

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

$$\hat{\sigma}^2 = \frac{1}{(n-1)tr((I-\mathcal{Z}_\kappa)^2)} \sum_{i=1}^{n} ||(I-\mathcal{Z}_\kappa)(Y_i - X_i\hat{\beta})||^2. \tag{2.22}$$

The selection method can be described as follows: choose a significance level $\alpha$ (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. Finally, the estimated dimension is then given by the smallest dimension $d$, which leads a rejection of $H_0$.

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:

$$PC(l) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(l))^2 + l g_{nT} \tag{2.23}$$

for all $l \in \{1, 2, \ldots\}$, where $\hat{y}_{it}(l)$ is the fitted value 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:

$$g_{nT}^{(\mathrm{PC1})} = \hat{\sigma}^2 \frac{(n+T)}{nT} \log\left(\frac{nT}{n+T}\right), \tag{2.24}$$

$$g_{nT}^{(\mathrm{PC2})} = \hat{\sigma}^2 \frac{(n+T)}{nT} \log(\min\{n,T\}), \tag{2.25}$$

$$g_{nT}^{(\mathrm{PC3})} = \hat{\sigma}^2 \frac{\log(\min\{n,T\})}{\min\{n,T\}}, \text{ and} \tag{2.26}$$

$$g_{nT}^{(\mathrm{BIC3})} = \hat{\sigma}^2 \frac{(n+T-l)}{nT} \log(nT), \tag{2.27}$$

where $\hat{\sigma}^2$ is the sample variance estimator of the residuals $\hat{\epsilon}_{it}$. The proposed criteria are denoted by PC1, PC2, PC3, and BIC3, respectively. Note that only the first three criteria satisfy the requirements of Theorem 2 in Bai and Ng (2002), i.e., $(i)$ $g_{nT} \to 0$ and $(ii)\min\{n,T\}g_{nt} \to \infty$, as $n, T \to \infty$. These conditions ensure consistency of the selection procedure without imposing additional restrictions on the proportional behavior of $n$ and $T$. The requirement $(i)$ is not always fulfilled for BIC3, especially when $n$ is too large relative to $T$ or $T$ is too large relative to $n$ (e.g., $n = \exp(T)$ or $T = \exp(n)$). In practice, BIC3 seems to perform very well, especially when the idiosyncratic errors are cross-correlated.

The variance estimator $\hat{\sigma}^2$ can be obtained by

$$\hat{\sigma}^2(d_{max}) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(d_{max}))^2, \tag{2.28}$$

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 $\hat{\sigma}^2(d_{max})$ underestimates the true variance. To overcome this problem, Bai and Ng (2002) propose three additional criteria (IC1, IC2, and IC3):

$$IC(l) = \log\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(l))^2\right) + l g_{nT} \tag{2.29}$$

with

$$g_{nT}^{(\mathrm{IC1})} = \frac{(n+T)}{nT}\log(\frac{nT}{n+T}), \tag{2.30}$$

$$g_{nT}^{(\mathrm{IC2})} = \frac{(n+T)}{nT}\log(\min\{n,T\}), \text{ and} \tag{2.31}$$

$$g_{nT}^{(\mathrm{IC3})} = \frac{\log(\min\{n,T\})}{\min\{n,T\}}. \tag{2.32}$$

In order to improve the finite sample performance of IC1 and IC2, Alessi et al. (2010) propose to multiply the penalties $g_{nT}^{(\mathrm{IC1})}$ and $g_{nT}^{(\mathrm{IC2})}$ with a positive constant $c$ and apply the calibration strategy of Hallin and Liška (2007). The choice of $c$ is based on the inspection of the criterion behavior through $J$-different tuples of $n$ and $T$, i.e., $(n_1, T_1), \ldots, (n_J, T_J)$, and for different values of $c$ in a pre-specified grid interval. We denote the refined criteria in our package by ABC.IC1 and ABC.IC2 respectively. Note that such a modification does not affect the asymptotic properties of the dimensionality estimator.

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:

$$ER = \frac{\hat{\rho}_l}{\hat{\rho}_{l+1}} \tag{2.33}$$

$$\tag{2.34}$$

$$GR = \frac{\log\left(\sum_{r=l}^{T}\hat{\rho}_r / \sum_{r=l+1}^{T}\hat{\rho}_r\right)}{\log\left(\sum_{r=l+1}^{T}\hat{\rho}_r / \sum_{r=l+2}^{T}\hat{\rho}_r\right)}. \tag{2.35}$$

Note that the theory of the above dimensionality criteria PC1, PC2, PC3, BIC3, IC1, IC2, IC3, IPC1,IPC2, IPC3, ABC.IC1, ABC.IC2, KSS.C, ER, and GR are developed for stochastically bounded factors. In order to estimate the number of unit root factors, Bai (2004) proposes the following panel criteria:

$$IPC(l) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it}(l))^2 + lg_{nT}, \tag{2.36}$$

where

$$g_{nT}^{(\mathrm{IPC1})} = \hat{\sigma}^2\frac{\log(\log(T))}{T}\frac{(n+T)}{nT}\log\left(\frac{nT}{n+T}\right), \tag{2.37}$$

$$g_{nT}^{(\mathrm{IPC2})} = \hat{\sigma}^2\frac{\log(\log(T))}{T}\frac{(n+T)}{nT}\log(\min\{n,T\}), \text{ and} \tag{2.38}$$

$$g_{nT}^{(\mathrm{IPC3})} = \hat{\sigma}^2\frac{\log(\log(T))}{T}\frac{(n+T-l)}{nT}\log(nT). \tag{2.39}$$

Alternatively, Onatski (2010) has introduced a threshold approach based on the empirical distribution of the sample covariance eigenvalues, 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 $\delta$ is a positive threshold, estimated iteratively from the data. We refer to this criterion as ED, which stands for Eigenvalue Differences.

### 2.3.1 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", "BIC3", "IC1",
    "IC2", "IC3", "IPC1", "IPC2", "IPC3", "ABC.IC1", "ABC.IC2",
    "KSS.C", "ED", "ER", "GR"), standardize = FALSE, d.max, sig2.hat,
    spar, level = 0.01, c.grid = seq(0, 5, length.out = 128),
    T.seq, n.seq)
NULL
```

The desired criteria can be selected by one or several of the following character variables: `"KSS.C"`, `"PC1"`, `"PC2"`, `"PC3"`, `"BIC2"`, `"IC1"`, `"IC2"` , `"IC3"`, `"ABC.IC1"`, `"ABC.IC2"`, `"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 (2.22) and (2.28). The arguments `"c.grid"`, `"T.seq"`, and `"n.seq"` are required for computing `"ABC.IC1"` and `"ABC.IC2"`. The grid interval of the calibration parameter can be externally specified with `"c.grid"`. The $J$-Tuples, $(n_1, T_1), \ldots, (n_J, T_J)$, can be specified by using appropriate vectors in `"T.seq"`, and `"n.seq"`. If these two arguments are left unspecified, the function constructs internally the following sequences: $T - C, T - C + 1, \ldots, T$, and $n - C, n - C + 1, \ldots, n$, for $C = min\sqrt{n}, \sqrt{T}, 30$. Alternatively, the user can specify only the length of the sequences by giving appropriate integers to the arguments `"T.seq"`, and `"n.seq"`, to control for $C$.

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 for all criteria.

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
    5
```

`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 16 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
    5

--------
Criteria of Ahn and Horenstein (2013):

  ER GR
   3  3
```

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

  IPC1 IPC2 IPC3
     3    3    2
```

In order to 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 selected criteria detect the dimension; see Figure 2.3.

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

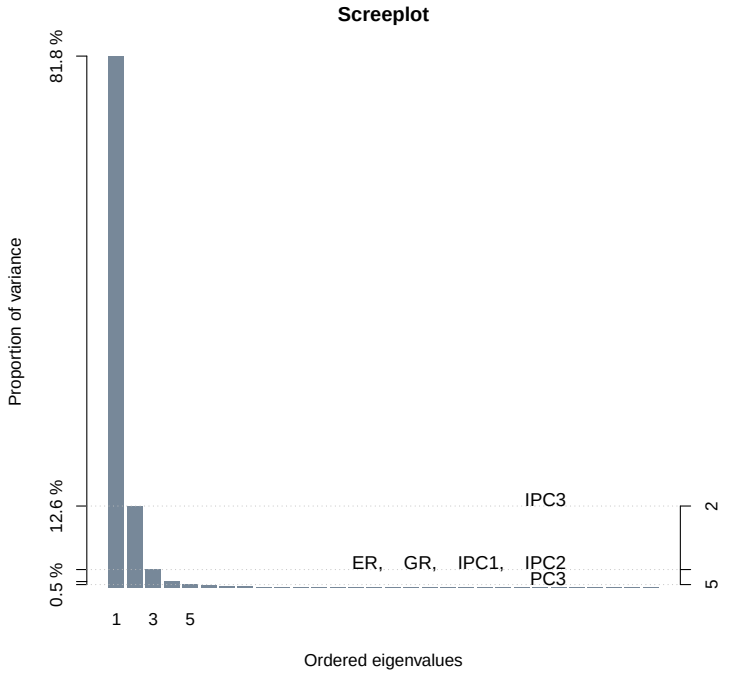


FIGURE 2.3: Scree plot produced by the `plot()`-method for `OptDim`-objects. Most of the dimensionality criteria (ER, GR, IPC1 and IPC2) detect $\hat{d} = 3$.

We, now, come back to the `KSS`- function, which offers an additional way 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 and Ng (2002):
 PC1  PC2  PC3 BIC3  IC1  IC2  IC3
   5    5    5    4    5    5    5


-Bai (2004):
IPC1 IPC2 IPC3
   3    3    2


-Alessi et al. (2010):
ABC.IC1 ABC.IC2
      3       3

-Kneip et al. (2012):
 KSS.C
     6

-Onatski (2009):
 ED
  3

-Ahn and Horenstein (2013):
 ER  GR
  3   6

------------------------------------------------------------
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 highest estimated factor dimension (here maximal dimension would be 6). A higher dimension can be chosen using the argument `factor.dim`.

## 2.4 Panel Models with Stochastically Bounded Factors

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

$$y_{it} = \sum_{j=1}^{P} x_{itj}\beta_j + v_{it} + \epsilon_{it}, \tag{2.40}$$

where

$$v_{it} = \sum_{l=1}^{d} \lambda_{il} f_{lt}. \tag{2.41}$$

Combining (2.40) with (2.41) and writing the model in matrix notation we get

$$Y_i = X_i\beta + F\Lambda_i^\top + \epsilon_i, \tag{2.42}$$

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:

$$\frac{1}{T}F^\top F \xrightarrow{p} \Sigma_F, \quad \text{as} \quad T \to \infty, \tag{2.43}$$

where $\Sigma_F$ is a fixed positive definite $d \times d$ matrix. This allows for the factors to follow a deterministic time trend such as $f_t = t/T$ or to be stationary dynamic processes such that $f_t = \sum_{j=1}^{\infty} C_j e_{t-j}$, where $e_t$ are i.i.d. zero mean stochastic components. It is, however, important to note that such an assumption rules out a large class of non-stationary factors such as I($p$) processes with $p \geq 1$.

### 2.4.1 Model with Known Number of Factors

Bai (2009) proposes to estimate the model parameters $\beta, F$ and $\Lambda_i$ by minimizing the following least squares objective function:

$$S(\beta, F, \Lambda_i) = \sum_{i}^{n} ||Y_i - X_i\beta - F\Lambda_i^\top||^2. \tag{2.44}$$

For each given $F$, the OLS estimator of $\beta$ can be obtained by

$$\hat{\beta}(F) = \left( \sum_{i=1}^{n} X_i^\top \mathcal{P}_d X_i \right)^{-1} \left( \sum_{i=1}^{n} X_i^\top \mathcal{P}_d Y_i \right) \tag{2.45}$$

where $\mathcal{P}_d = I - F(F^\top F)^{-1} F^\top = I - FF^\top/T$. If $\beta$ 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 $\beta$ 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.

A feasible estimator of (2.45) can be obtained by using an arbitrary large dimension $d_{max}$ greater than $d$. The factor dimension can be estimated subsequently by using the criteria of Bai and Ng (2002) to the remainder term $Y_i = X_i\hat{\beta}(\hat{F}(d_{max}))$, as suggested by Bai (2009). This strategy can lead, however, to inefficient estimation and spurious interpretation of $\beta$ due to over-parameterization.

### 2.4.2 Model with Unknown Number of Factors

In order to estimate $d$ jointly with $\beta, F$, and $\Lambda_i$, Bada and Kneip (2014) propose to integrate a penalty term into the objective function to be globally optimized. In this case, the optimization criterion can be defined as a penalized least squares objective function of the form:

$$S(\beta, F, \Lambda_i, l) = \sum_{i}^{n} ||Y_i - X_i\beta - F\Lambda_i^\top||^2 + l g_{nT} \tag{2.46}$$

The role of the additional term $l g_{nT}$ is to pick up the dimension $\hat{d}$, of the unobserved factor structure. The penalty $g_{nT}$ can be chosen according to Bai and Ng (2002). The estimation algorithm is based on the parameter cascading strategy of Cao and Ramsay (2010), which in this case can be described as follows:

1. Minimizing (2.46) with respect to $\Lambda_i$ for each given $\beta, F$ and $d$, we get

$$\hat{\Lambda}_i^\top(\beta, F, d) = F^\top (Y_i - X_i\beta)/T. \tag{2.47}$$

2. Introducing (2.47) in (2.46) and minimizing with respect to $F$ for each given $\beta$ and $d$, we get

$$\hat{F}(\beta, d) = \sqrt{T}\hat{\gamma}(\beta, d), \tag{2.48}$$

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 (2.48) and (2.47) in (2.46) and minimizing with respect to $\beta$ for each given $d$, we get

$$\hat{\beta}(d) = \left( \sum_{i=1}^{n} X_i^\top X_i \right)^{-1} \left( \sum_{i=1}^{n} X_i^\top \left( Y_i - \hat{F}\hat{\Lambda}_i^\top(\hat{\beta}, d) \right) \right). \tag{2.49}$$

4. Optimizing (2.46) with respect to $l$ given the results in (2.47), (2.48), and (2.49) allows us to select $\hat{d}$ as

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

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 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). In order to avoid over-estimation, Bada and Kneip (2014) propose to re-scale $g_{nT}$ in each iteration stage with $\hat{\sigma}^2 = \sum_i^n ||Y_i - X_i\hat{\beta} - \hat{F}\hat{\Lambda}_i^\top||^2$ in stead of $\hat{\sigma}^2(d_{max})$. Simulations show that such a calibration can improve the finite sample properties of the estimation method.

It is notable that the objective functions (2.46) and (2.44) 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 $\beta$:

$$\hat{\beta}_{start} = \left( \sum_{i=1}^{n} X_i^\top (I - GG^\top) X_i \right)^{-1} \left( \sum_{i=1}^{n} X_i^\top (I - GG^\top) Y_i \right), \tag{2.50}$$

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 projection of $X_i$ on the orthogonal complement of $G$ in (2.50) eliminates the effect of a possible correlation between the observed regressors and unobserved factors, which can heavily distort the value of $\beta^0$ if it is neglected. Greenaway-mcgrevy (2012) give conditions under which (2.50) is a consistent estimator of $\beta$. In order to avoid miss-specifying the model through identifying factors that only exist in $X_i$ and not $Y_i$, Bada and Kneip (2014) recommend to under-scale the starting common factors $G_l$ that are highly correlated with $X_i$.

According to Bai (2009), the asymptotic distribution of the slope estimator $\hat{\beta}(d)$ for known $d$ is given by

$$\sqrt{nT}(\hat{\beta}(d) - \beta) \overset{a}{\sim} N(0, D_0^{-1} D_Z D_0^{-1}),$$

where $D_0 = \text{plim} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{it}^\top Z_{it}$ with $Z_i = (Z_{i1}, \ldots, Z_{iT})^\top = \mathcal{P}_d X_i - \frac{1}{n} \sum_{k=1}^{n} \mathcal{P}_d X_i a_{ik}$ and $a_{ik} = \Lambda_i (\frac{1}{n} \sum_{i=1}^{n} \Lambda_i^\top \Lambda_i)^{-1} \Lambda_k^\top$, and

Case 1. $D_Z = D_0^{-1} \sigma^2$ if the errors are i.i.d. with zero mean and variance $\sigma^2$,

Case 2. $D_Z = \text{plim} \frac{1}{nT} \sum_{i=1}^{n} \sigma_i^2 \sum_{t=1}^{T} Z_{it}^\top Z_{it}$, where $\sigma_i^2 = E(\epsilon_{it}^2)$ with $E(\epsilon_{it}) = 0$, if cross-section heteroskedasticity exists and $n/T \to 0$,

Case 3. $D_Z = \text{plim} \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{ij} \sum_{t=1}^{T} Z_{it}^\top Z_{jt}$, where $\omega_{ij} = E(\epsilon_{it} \epsilon_{jt})$ with $E(\epsilon_{it}) = 0$, if cross-section correlation and heteroskedasticity exist and $n/T \to 0$,

Case 4. $D_Z = \text{plim} \frac{1}{nT} \sum_{t=1}^{T} \sigma_t^2 \sum_{i=1}^{n} Z_{it}^\top Z_{it}$, where $\sigma_t^2 = E(\epsilon_{it}^2)$ with $E(\epsilon_{it}) = 0$, if heteroskedasticity in the time dimension exists and $T/n \to 0$,

Case 5. $D_Z = \text{plim} \frac{1}{nT} \sum_{t=1}^{T} \sum_{s=1}^{T} \rho(t,s) \sum_{i=1}^{n} Z_{it}^\top Z_{is}$, where $\rho(t,s) = E(\epsilon_{it} \epsilon_{is})$ with $E(\epsilon_{it}) = 0$ , if correlation and heteroskedasticity in the time dimension exist and $T/n \to 0$, and

Case 6. $D_Z = \text{plim} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \sigma_{it}^2 Z_{it}^\top Z_{is}$, where $\sigma_{it}^2 = E(\epsilon_{it}^2)$ with $E(\epsilon_{it}) = 0$, if heteroskedasticity in both time and cross-section dimensions exists with $T/n^2 \to 0$ and $n/T^2 \to 0$.

In presence of correlation and heteroskedasticity in panels with proportional dimensions $n$ and $T$, i.e., $n/T \to c > 0$, the asymptotic distribution of $\hat{\beta}(d)$ will be not centered at zero. This can lead to false inference when using the usual test statistics such as $t$- and $\chi^2$-statistic. To overcome this problem, Bai (2009) propose to estimate the asymptotic bias and correct the estimator as follows:

$$\hat{\beta}^*(d) = \hat{\beta}(d) - \frac{1}{n}\hat{B} - \frac{1}{T}\hat{C} \tag{2.51}$$

where $\hat{B}$ and $\hat{C}$ are the estimators of

$$\begin{aligned} B &= -\left(\tfrac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}^{\top} Z_{it}\right)^{-1} \tfrac{1}{nT}\sum_{i=1}^{n}\sum_{k=1}^{n}(X_i - V_i)^{\top} F \left(\tfrac{1}{T}F^{\top}F\right)^{-1} W_{ik} \\ C &= -\left(\tfrac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}^{\top} Z_{it}\right)^{-1} \tfrac{1}{nT}\sum_{i=1}^{n} X_i^{\top} M_F \Omega F \left(\tfrac{1}{T}F^{\top}F\right)^{-1} \left(\tfrac{1}{n}\sum_{k=1}^{n}\Lambda_k^{\top}\Lambda_k\right)^{-1}\Lambda_i^{\top} \end{aligned}$$

respectively. Here, $V_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}X_j$, $W_{ik} = \left(\frac{1}{n}\sum_{j=1}^{n}\Lambda_j^{\top}\Lambda_j\right)^{-1}\Lambda_k^{\top}\frac{1}{T}\sum_{t=1}^{T}E(\epsilon_{it}\epsilon_{kt})$, and $\Omega = \frac{1}{n}\sum_{k=1}^{n}\Omega_k$ with

Case 7. $\Omega_k$ is a $T \times T$ diagonal matrix with elements $\omega_{kt} = E(\epsilon_{kt}^2)$ if heteroskedasticity in both time and cross-section dimensions exist and $n/T \to c > 0$ and,

Case 8. $\Omega_k$ is a $T \times T$ matrix with elements $\Omega_{k,ts} = E(\epsilon_{kt}\epsilon_{ks})$ if correlation and heteroskedasticity in both time and cross-section dimensions exist and $n/T \to c > 0$.

In a similar context, Bada and Kneip (2014) prove that estimating $d$ with the remaining model parameters does not affect the asymptotic properties of $\hat{\beta}(d)$. The asymptotic distribution of $\hat{\beta} = \hat{\beta}(\hat{d})$ is given by

$$\sqrt{nT}(\hat{\beta} - \beta) \overset{a}{\sim} N(0, D_0^{-1} D_Z D_0^{-1})$$

under Cases 1-6, and

$$\sqrt{nT}(\hat{\beta}^* - \beta) \overset{a}{\sim} N(0, D_0^{-1} D_Z D_0^{-1})$$

under Cases 7-8, where $\hat{\beta}^* = \hat{\beta}^*(\hat{d})$.

The asymptotic variance of $\hat{\beta}$ and the bias terms $B$ and $C$ can be estimated by replacing $F$, $\Lambda_i$, $Z_{it}$, and $\epsilon_{it}$ with $\hat{F}$, $\hat{\Lambda}_i$, $\hat{Z}_{it}$, and $\hat{\epsilon}_{it}$ respectively.

In presence of serial correlation (cases 5 and 8), consistent estimators for $D_Z$ and $C$ can be obtained by using the usual heteroskedasticity and autocorrelation (HAC) robust limiting covariance. In presence of cross-section correlation (case 3), $D_Z$ is estimated by $\hat{D}_Z = \frac{1}{mT}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{t=1}^{T}\hat{Z}_{it}^{\top}\hat{Z}_{jt}\hat{\epsilon}_{it}\hat{\epsilon}_{jt}$, where $m = \sqrt{n}$. If both cross-section and serial correlation exist (case 8), we estimate the long-run covariance of $\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\hat{Z}_{it}\hat{\epsilon}_{it}$.

### 2.4.3 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",
    "BIC3", "IC1", "IC2", "IC3", "IPC1", "IPC2", "IPC3"), 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 arguments `additive.effects`, `d.max`, `sig2.hat`, and `restrict.mode` have the same roles as in `KSS()`; see Section 2.2.2. The argument `dim.criterion` specifies the dimensionality criterion to be used if `factor.dim` is left unspecified and defaults to `dim.criterion = "PC1"`.

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 $\beta_{start}$. The maximal number of iteration and the convergence condition can be controlled by `max.iteration` and `convergence`.

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 (2.43); see Figure 2.2. We consider the following modified cigarettes model:

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

where $\triangle x_t = x_t - x_{t-1}$. 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 (2.20). The $\triangle$-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
 -0.3140143   0.159392

Additive Effects Type:  none

Dimension of the Unobserved Factors: 5

Number of iterations: 55
```

Inferences about the slope parameters can be obtained by using the method `summary()`. The type of correlation and heteroskedasticity in the idiosyncratic errors can be specified by choosing one of the corresponding Cases 1-8 described above using the argument `error.type = c(1, 2, 3, 4, 5, 6, 7, 8)`.

In presence of serial correlations (cases 5 and 8), the kernel weights required for estimating the long-run covariance can be externally specified by giving a vector of weights in the argument `kernel.weights`. By default, the function uses internally the linearly decreasing weights of Newey and West (1987) and a truncation at $\left\lfloor \min\{\sqrt{n}, \sqrt{T}\} \right\rfloor$. If case 7 or 8 is chosen, the method `summary()` calculates the realization of the bias corrected estimators and gives appropriate inferences. The bias corrected coefficients can be called by using the method `coef()` to the object produced by `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
-0.147000 -0.013700  0.000889  0.014100  0.093300


 Slope-Coefficients:
           Estimate Std.Err Z value    Pr(>z)
d.l.Price   -0.3140  0.0227  -13.90 < 2.2e-16 ***
d.l.Income   0.1590  0.0358    4.45  8.39e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type:  none

Dimension of the Unobserved Factors: 5

Residual standard error: 0.02804 on 957 degrees of freedom,
R-squared: 0.7033
```

The summary output reports that `"PC3"` detects 5 common factors. The effect of the differenced log-real prices for cigarettes on the differenced log-sales is negative and amounts to $-0.31$. The estimated effect of the differenced real disposable log-income per capita is 0.16.

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 2.4.

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

## 2.5 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.

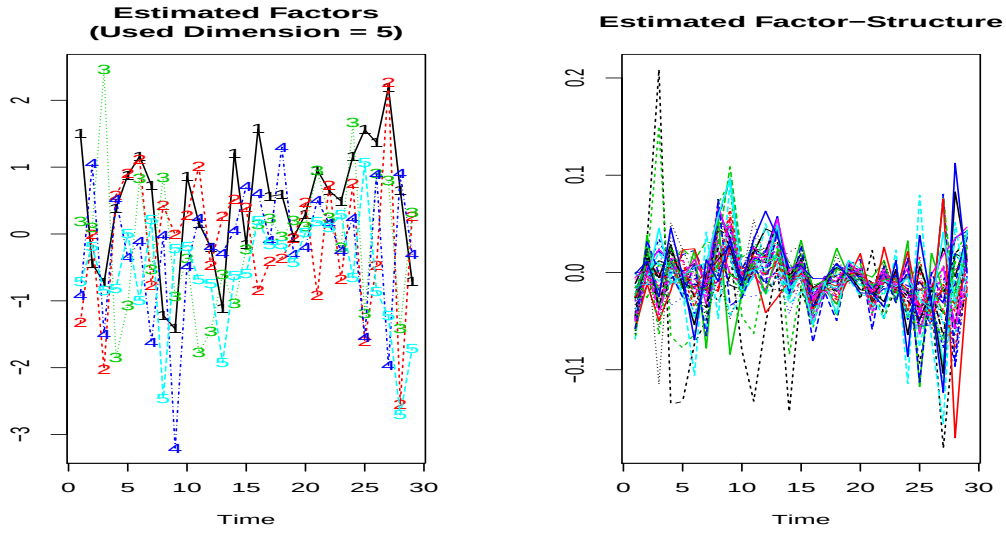


FIGURE 2.4: LEFT PANEL: Estimated factors $\hat{f}_{1t}, \ldots, \hat{f}_{7t}$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_{1t}, \ldots, \hat{v}_{nt}$.

On the one hand, if there are such effects in the true model, then neglecting them will result in non-efficient estimators; see Bai (2009). On the other hand, additive effects can be very useful for interpretation.

Consider now the following model:

$$y_{it} = \mu + \alpha_i + \theta_t + x_{it}^\top \beta + \nu_{it} + \epsilon_{it} \tag{2.52}$$

with

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

where $\alpha_i$ are time-constant individual effects and $\theta_t$ is a common time-varying effect.

In order to ensure identification of the additional additive effects $\alpha_i$ and $\theta_t$, 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 $\alpha_i$ and $\theta_t$, 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) ensure that the transformation does not affect the time-varying individual effects $\nu_{it}$. The parameters $\mu, \alpha_i$ and $\theta_t$ can be easily estimated in a second step once an estimate of $\beta$ is obtained. Because of Restrictions (d) and (e), the solution has the same form as the classical fixed effects model.

The parameters $\beta$ and $\nu_{it}$ can be estimated by the above introduced estimation procedures. All possible variants of model (2.52) 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 $\mu$ 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 2.2. The left panel of Figure 2.2 shows that the first common factor is nearly time-invariant. This motivates us to augment the model (2.20) for a time-constant additive effects $\alpha_i$. In this case, it is convenient to use an intercept $\mu$, which yields the following model:

$$\begin{aligned} \ln(\texttt{Consumption}_{it}) &= \mu + \beta_1 \ln(\texttt{Price}_{it}) + \beta_2 \ln(\texttt{Income}_{it}) + \alpha_i + v_i(t) + \varepsilon_{it} \quad (2.53) \\ \text{where} \quad v_i(t) &= \sum_{l=1}^{d} \lambda_{il}\, f_l(t). \end{aligned}$$

The estimation of the augmented model (2.53) 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 a 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 (2.20); see the middle panel of Figure 2.5. The estimated time-constant state-specific effects $\alpha_i$ are shown in the left plot of Figure 2.5. The extraction of the $\alpha_i$'s from the factor structure yields a denser set of time-varying individual effects $\hat{v}_i$ shown in the right panel of Figure 2.5.

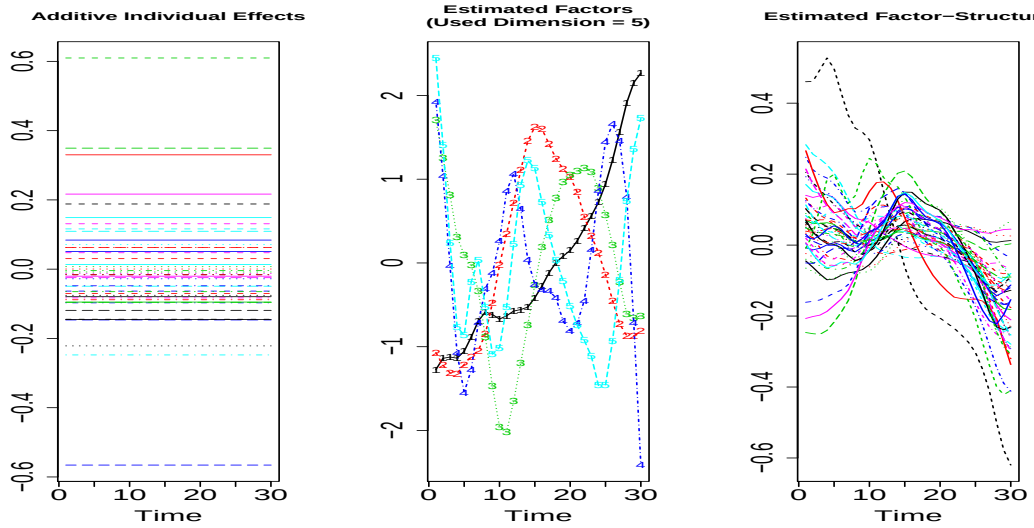


FIGURE 2.5: LEFT PANEL: Estimated time-constant state-specific effects $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$. MIDDLE PANEL: Estimated common factors $\hat{f}_1(t), \ldots, \hat{f}_5(t)$. RIGHT PANEL: Estimated time-varying individual effects $\hat{v}_1(t), \ldots, \hat{v}_n(t)$.

### 2.5.1 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 (2009). 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).

#### 2.5.1.1 Testing the Sufficiency of Classical Additive Effects

For the case in which the estimated number of factors amounts to one or two ($1 \leq \hat{d} \leq 2$), it is interesting to check whether or not these factors can be interpreted as classical `"individual"`, `"time"`, or `"twoways"` effects. Bai (2009) considers the following testing problem:

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

The model with factor structure, as described in Section 2.4, 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:

$$J_{Bai} = nT \left( \hat{\beta} - \hat{\beta}_{within} \right) \Delta^{-1} \left( \hat{\beta} - \hat{\beta}_{within} \right) \overset{a}{\sim} \chi^2_P, \tag{2.54}$$

where $\hat{\beta}_{within}$ is the classical within least squares estimator, $\Delta$ is the asymptotic variance of $\sqrt{nT}\left(\hat{\beta} - \hat{\beta}_{within}\right)$, $P$ is the vector-dimension of $\beta$, and $\chi^2_P$ is the $\chi^2$-distribution with $P$ degrees 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 $\chi^2$ distribution with $P$ degrees of freedom.

Under i.i.d. errors, $J_{Bai}$ can be calculated by replacing $\Delta$ with its consistent estimator

$$\hat{\Delta} = \left( \left( \frac{1}{nT} \sum_{i=1}^{n} Z_i^\top Z_i \right)^{-1} - \left( \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_{it} \dot{x}_{it}^\top \right)^{-1} \right) \hat{\sigma}^2, \tag{2.55}$$

where

$$\hat{\sigma}^2 = \frac{1}{nT - (n+T)\hat{d} - P + 1} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - x_{it}^\top \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{il} \hat{f}_{lt})^2. \tag{2.56}$$

The used residual variance estimator $\hat{\sigma}^2$ is chosen here, since it is supposed to be consistent under the null as well as the alternative hypothesis. The idea behind this trick is to avoid negative definiteness of $\hat{\Delta}$. But notice that even with using this construction, the possibility of getting a negative definite variance estimator cannot be excluded. As an illustration, consider the case in which the true number of factors is greater than the number of factors used under the alternative hypothesis, i.e., the true $d > 2$. In such a case, the favorable conditions for applying the test can be violated, since the iterated least squares estimator $\hat{\beta}$ is computed with $\hat{d} \leq 2$ and can be inconsistent under both hypothesis. To avoid such a scenario, we recommended to the user to calculate $\hat{\beta}$ with a large dimension $d_{max}$ instead of $\hat{d} \leq 2$.

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

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

The argument `level` is used to specify the significance level. 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 large factor dimension $d_{max}$; specified with the argument `factor.dim`.

If the test statistic is negative (due to the negative definiteness of $\hat{\Delta}$), the `checkSpecif()` prints an error message.

```
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.
```

Notice that the Hausman test of Bai (2009) assumes the within estimator to be inconsistent under the alternative hypothesis, which requires $x_{it}$ to be correlated with $v_{it}$. If this assumption is violated, the test can suffer from power to reject the null hypothesis, since the within estimator becomes consistent under both hypothesis.

Bai (2009) discusses in his supplementary material another way to check whether a classical panel data with fixed additive effects is sufficient to describe the data. His idea consists of estimating the factor dimension after eliminating the additive effects as described in Section 2.5. If the obtained estimate of $d$ is zero, the additive model can be considered as a reasonable alternative for the model with factor structure. But note that this procedure can not be considered as a formal testing procedure, since information about the significance level of the decision are not provided.

An alternative test for the sufficiency of a classical additive effects model can be given by manipulating the test proposed by Kneip et al. (2012) as described in the following section.

#### 2.5.1.2 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 possible 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:

$$H_0: \quad d = 0$$
$$H_1: \quad d > 0$$

Under $H_0$ the slope parameters $\beta$ 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 $\alpha$, 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 (2.15) or (2.56), 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` and $\sigma^2$ as in (2.56).

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:

For the model of Bai (2009):

```
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
```

For the model of Kneip et al. (2012):

```
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 $\alpha = 0.01$.

## 2.6 Interpretation

This section is intended to outline an exemplary interpretation of the panel model (2.53), which is estimated by the function `KSS()` in Section 2.5. The interpretation of models estimated by the function `Eup()` can be done accordingly. For convenience sake, we rewrite the model (2.53) 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 2.5 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 2.5 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

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

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

| Common Factor | Command of the share of total variance of $v_i(t)$ | Value |
|---|---|---|
| $\hat{f}_1(t)$ | `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 2.1: List of the variance shares of the estimated common factors

The values in Table 2.1 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

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

are plotted as black and red lines in the middle panel of Figure 2.5. Figure 2.6 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.

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

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

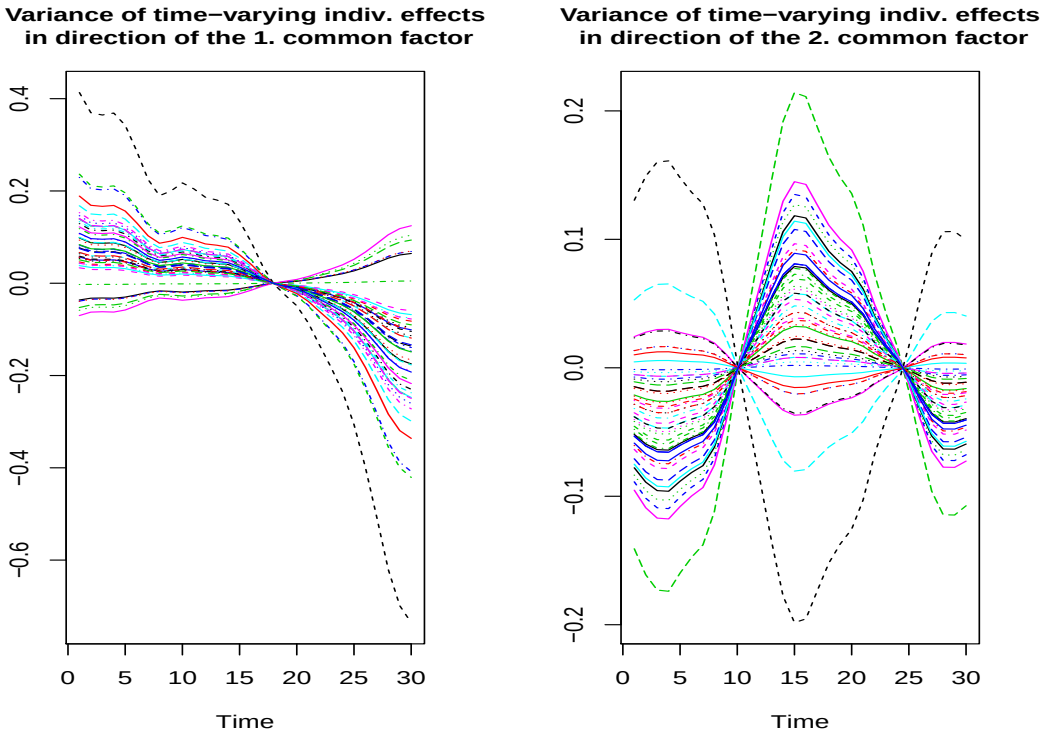


FIGURE 2.6: LEFT PANEL: Visualization of the differences of the time-varying individual effects $v_i(t)$ in the direction of the first factor $\hat{f}_1(t)$ (i.e., $\hat{\lambda}_{i1}\hat{f}_1(t)$). RIGHT PANEL: Visualization of the differences of the time-varying individual effects $v_i(t)$ in the direction of the second factor $\hat{f}_2(t)$ (i.e., $\hat{\lambda}_{i2}\hat{f}_2(t)$).

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 et al. (2001). This can be done by choosing `restrict.mode = c("restrict.loadings")` in the functions `KSS()` and `Eup()` respectively.

## 2.7 Summary

This chapter 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 the usual `print()`-, `summary()`-, `plot()`-, `coef()`- and `residuals()`-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 by a real data application.

# Chapter 3

# Panel Models with Multiple Jumps in the Parameters

## 3.1 Introduction

Panel datasets with large cross-sectional dimensions and large periods of time observations are becoming more and more available due to the impressive progress of information technology. This has been succeeded, in the econometric literature, by the development of new methods and techniques for analyzing large panels. There is, however, an important issue, that is scarcely discussed in most of the existing work –the risk of neglecting structural breaks in the data generating process, especially when the observation period is large. In the field of empirical macroeconomics, this problem is most famously considered by Lucas (1976), who points out the risk of predicting naively the effects of economic policy changes based on historical data, since the emergence of important economic events and shocks may induce changes in the model parameters during the time. Of course, the larger the observation period, the more likely the occurrence of such shocks. While a vast literature on change point analysis exists for univariate time series, little research has been done on panel data models.

In this chapter, we propose a novel method for estimating panel models with multiple structural changes that occur at unknown points in time and may affect each slope parameter individually. We consider, for $i \in \{1, \dots, n\}$, $t \in \{1, \dots, T\}$, models of the form

$$Y_{it} = \sum_{p=1}^{P} \sum_{j=1}^{S_p+1} X_{it,p} \mathbf{I}\big(\tau_{j-1} < t \leq \tau_j\big) \beta_{\tau_j,p} + \alpha_i + \theta_t + \varepsilon_{it}, \tag{3.1}$$

where $\mathbf{I}(.)$ is the indicator function, the set of jump points $\{\tau_{0,p}, \tau_{1,p}, \dots, \tau_{S_p+1,p} | \tau_{0,p} = 1 < \tau_{1,p} < \dots < \tau_{S_p+1,p} = T\} \subseteq \{1, \dots, T\}$ and $S_p$ are unknown, $Y_{it}$ is the dependent

variable for individual $i$ at time $t$, $X_{it,p}$ is the $p$th explaining variable, $\alpha_i$ is an individual specific effect, $\theta_t$ is a common time parameter, and $\varepsilon_{it}$ is an unobserved idiosyncratic term that may be correlated with one or more explaining variables.

In single time series, the available information is often not sufficient to uncover the true dates of the structural breaks. Only the time fractions of the break locations can be consistently estimated and tested; see, e.g., Bai (1997), Bai and Perron (1998, 2003), Inoue and Rossi (2011), Pesaran et al. (2011), Aït-Sahalia and Jacod (2009), and Carr and Wu (2003). In panel data models, such a limitation can be alleviated since the cross-section dimension provides an important source of additional information. Besides the virtue of getting improved statistical efficiency, the determination of the change point locations can be, in many applications, of particular interest. Indeed, estimating the number and locations of the structural breaks from the data alleviates concerns about ad-hoc subsample selection, enables interpretation of historical events that are not explicitly considered in the model, and avoids statistical under- or over-parametrization related issues.

One of the earliest contributions in testing the structural breaks in panel data literature is the work of Han and Park (1989). The authors propose a multivariate version of the consum-test, which can be seen as a direct extension of the univariate time series test proposed by Brown et al. (1975). Qu and Perron (2007) extend the work of Bai and Perron (2003) and consider the problem of estimating, computing, and testing multiple structural changes that occur at unknown dates in linear multivariate regression models. They propose a quasi-maximum likelihood method and a likelihood ratio-type statistics based on Gaussian errors. The method requires, however, the number of equations to be fixed and does not consider the case of large panel models with unobserved effects and possible endogenous regressors. Based on the work of Andrews (1993), De Wachter and Tzavalis (2012) propose a break testing procedure for dynamic panel data models with exogenous or pre-determined regressors when $n$ is large and $T$ is fixed. The method can be used to test for the presence of a structural break in the slope parameters and/or in the unobserved fixed effects. But their assumptions allow only for the presence of a single break. Bai (2010) proposes a framework to estimate the break in means and variance. Bai (2010) also considers the case of one break and establishes consistency for both large and fixed $T$. Kim (2014) extends the work of Bai (2010) to allow for the presence of unobserved common factors in the model. Pauwels et al. (2012) analyze the cases of a known and an unknown break date and propose a Chow-type test allowing for the break to affect some, but not all, cross-section units. Although the method concerns the one-break case, it requires intensive computation to select the most likely individual breaks from all possible sub-intervals when the break date is unknown.

To the best of the authors' knowledge, ours is the first work to deal with the problem of multiple jump discontinuities in the parameters of panel models without imposing restrictive assumptions on the number, the location, and/ or the aspect of the breaks. The method can be applied to panel data with large time span $T$ and large cross-section dimension $n$ and allows for $T$ to be very long compared to $n$. We also consider the classic case of panel data, in which $T$ is fixed and only $n$ is large. Our model generalizes the special model specifications in which the slope parameters are either constant over time, so that $S_p = 0$, or extremely time heterogeneous so that, for all $p$, $\tau_{0,p} = 1, \tau_{1,p} = 2, \ldots, \tau_{S_p+1,p} = T$ when $T$ is fixed. Our theory considers breaks in a two-way panel data model, in which the unobserved heterogeneity is composed of additive individual effects and time specific effects. We show that our method can also be extended to cover the case of panel models with unobserved heterogeneous common factors as proposed by Ahn et al. (2001), Pesaran (2006), Bai (2009), Kneip et al. (2012), and Bada and Kneip (2014). Our estimation procedure is related to the Haar wavelet technique, which we transform and adapt to the structure of the observed variables in order to detect the location of the break points consistently. We propose a general setup allowing for endogenous models such as dynamic panel models and/or structural models with simultaneous panel equations. Consistency under weak forms of dependency and heteroscedasticity in the idiosyncratic errors is established and the convergence rate of our slope estimator is derived. To detect consistently the jump locations and test for the statistical significance of the breaks, we propose post-wavelet procedures. Our simulations show that, in many configurations of the data, our method performs very well even when the idiosyncratic errors are affected by weak forms of serial-autocorrelation and/or heteroskedasticity.

Our empirical vehicle for highlighting this new methodology addresses the stability of the relationship between Algorithmic Trading (AT) and Market Quality (MQ). We propose to automatically detect jumps in regression slope parameters to examine the effect of algorithmic trading on market quality in different market situations. We find evidence that the relationship between AT and MQ was disrupted between 2007 and 2008. This period coincides with the beginning of the subprime crisis in the US market and the bankruptcy of the big financial services firm Lehman Brothers.

The remainder of the chapter is organized as follows. Section 3.2 explains the basic idea of our estimation procedure by using a relatively straightforward centered univariate panel model. In Section 3.3, we consider panel models with unobserved effects and multiple jumping slope parameters, present our model assumptions, and derive the main asymptotic results. Section 3.4 proposes a post-wavelet procedure to estimate the jump locations, derives the asymptotic distribution of the final estimator, and describes selective testing procedures. In Section 3.5, we discuss models with an issue of omitted

common factors and endogenous models arising from structural simultaneous equation systems. Section 3.6 presents the simulation results of our Monte Carlo experiments. Section 3.7 is concerned with the empirical application. The conclusion follows in Section 3.8. The mathematical proofs are collected in Appendix B.

## 3.2 Preliminaries

### A Simple Panel Model with one Jumping Parameter

To simplify the exposition, we begin with a relatively straightforward version of (3.1). We consider a centered univariate panel data model of the form

$$Y_{it} = X_{it}\beta_t + e_{it} \quad \text{for} \quad i \in \{1, \ldots, n\} \quad \text{and} \quad t \in \{1, \ldots, T\}, \tag{3.2}$$

where $X_{it}$ is an univariate regressor, $\beta_t$ is a scalar, and $E(e_{it}) = 0$.

We allow for the slope parameter $\beta_t$ to change at unknown time points, say $\tau_1, \ldots, \tau_S$, such that

$$\beta_t = \begin{cases} \beta_{\tau_1} & \text{for} \quad t \in \{1, \ldots, \tau_1\}, \\ \beta_{\tau_2} & \text{for} \quad t \in \{\tau_1 + 1, \ldots, \tau_2\}, \\ \vdots & \\ \beta_{\tau_S} & \text{for} \quad t \in \{\tau_{S-1} + 1, \ldots, \tau_S\}, \text{ and} \\ \beta_{\tau_S+1} & \text{for} \quad t \in \{\tau_S + 1, \ldots, T\}. \end{cases} \tag{3.3}$$

### Some Fundamental Concepts of Wavelet Transform

The idea behind our approach consists basically of using the Haar wavelet expansion of $\beta_t$ to control for its piecewise changing character. Before continuing with the estimation method, we introduce some important concepts and notations that are necessary for our analysis.

We assume that the intertemporal sample size $T$ is *dyadic*, i.e., $T = 2^{L-1}$ for some positive integer $L \geq 2$. This is because wavelet functions are constructed via dyadic dilations of order $2^l$, for $l \in \{1, \ldots, L\}$. The case of a non-dyadic time dimension will be discussed later. Technically, the discrete wavelet transformation is much like the Fourier transformation, except that the wavelet expansion is constructed with a two parameter system: a dilation level $l \in \{1, \ldots, L\}$ and a translation index $k \leq 2^{l-2}$.

Let $\{\varphi_{l_0,k}, k = 1, \ldots, K_{l_0}\}$, and $\{\psi_{l,k}, l = l_0 + 1, \ldots, L; k = 1, \ldots, 2^{l-2}\}$, respectively, represent collections of discrete scaling and wavelet functions defined on the discrete

interval $\{1,\ldots,2^{L-1}\}$ such that

$$\psi_{l,k}(t) = a_l^{\psi} I_{l,2k-1}(t) - a_l^{\psi} I_{l,2k}(t) \text{ and} \tag{3.4}$$

$$\varphi_{l_0,k}(t) = a_{l_0}^{\varphi} I_{l+1,2k-1}(t) + a_{l_0}^{\varphi} I_{l+1,2k}(t), \tag{3.5}$$

where $a_{l_0}^{\varphi} = \sqrt{2^{l_0-1}}$, $a_l^{\psi} = \sqrt{2^{l-2}}$, and $I_{l,m}(t)$ is the indicator function that carries the value one if $t \in \{2^{L-l}(m-1)+1,\ldots,2^{L-l}m\}$ and zero otherwise.

The multiscale discrete Haar wavelet expansion of $\beta_t$ can be presented as follows:

$$\beta_t = \sum_{k=1}^{K_{l_0}} \varphi_{l_0,k}(t) d_{l_0,k} + \sum_{l=l_0+1}^{L} \sum_{k=1}^{K_l} \psi_{l,k}(t) c_{l,k}, \qquad \text{for } t \in \{1,\ldots,T\}, \tag{3.6}$$

where $K_l = 2^{l-2}$, for $l > 1$, and $K_1 = 1$. The coefficients $d_{l,k}$ and $c_{l,k}$ are called scaling and wavelet coefficients, respectively. Because $\varphi_{l_0,k}(t)$ and $\psi_{l,k}(t)$ are orthonormal, $d_{l,k}$ and $c_{l,k}$ are unique and can be interpreted as the projection of $\beta_t$ on their corresponding bases, i.e., $d_{l_0,k} = \frac{1}{2^{L-1}} \sum_{t=1}^{2^{L-1}} \varphi_{l_0,k}(t)\beta_t$ and $c_{l,k} = \frac{1}{2^{L-1}} \sum_{t=1}^{2^{L-1}} \psi_{l,k}(t)\beta_t$.

Although the Haar wavelet basis functions are the simplest basis within the family of wavelet transforms, they exhibit an interesting property allowing for analyzing functions with piecewise sudden changes.

### Orthonormalization and Estimation

Note that the collection of functions, in (3.6), is not unique. Here, we set $l_0 = 1$, to fix the primary scale to be the coarsest possible with only one parameter that reflects the general mean of $\beta_t$. In addition, we propose a slightly modified version of wavelet expansion to adapt the orthonormalization conditions to the requirements of our panel data method.

We consider the following expansion:

$$\beta_t = \sum_{l=1}^{L} \sum_{k=1}^{K_l} w_{l,k}(t) b_{l,k} \qquad \text{for } t \in \{1,\ldots,T\}, \tag{3.7}$$

where

$$w_{l,k}(t) = \begin{cases} a_{1,1} = a_{2,1}h_{2,1}(t) + a_{2,2}h_{2,2}(t) & \text{if} \quad l = 1, \text{ and} \\ a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t) & \text{if} \quad l > 1, \end{cases} \tag{3.8}$$

for some positive standardizing scales $a_{l,2k-1}$ and $a_{l,2k}$ that, unlike the conventional wavelets, do not only depend on the dilation level $l$ but also on the translation index $k$. Their exact form will be discussed in detail below. We define the function $h_{l,m}(t)$ as

follows:

$$h_{l,m}(t) = \sqrt{2^{l-2}} I_{l,m}(t). \tag{3.9}$$

The most appealing feature of expansion (3.7) (and (3.6) with $l_0 = 1$) is that the set of the wavelet coefficients $\{b_{l,k}\}$ contains at most $(S+1)L$ non-zero-wavelet coefficients. This important property results from the fact that each jump in $\beta_t$ can be sensed at each dilation level by at most one translation function. Proposition 3.1 states the existence of (3.7) for any arbitrary positive real scales $a_{l,2k}$ and $a_{l,2k-1}$.

**Proposition 3.1.** *Suppose $T = 2^{L-1}$, for some integer $L \geq 2$, and $\beta = (\beta_1, \ldots, \beta_T)' \in \mathbb{R}^T$ a vector that possesses exactly $S \geq 1$ jumps at $\{\tau_1, \ldots, \tau_S | \tau_1 < \ldots < \tau_S\} \subseteq \{1, \ldots, T\}$ as in (3.3). Let $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ be arbitrary positive real values for all $l \in \{1, \ldots, L\}$, and $k \in \{1, \ldots, K_l\}$. Thus, Expansion (3.7) exists and the set of the wavelet coefficients $\{b_{lk} | l = 1, \ldots, L; k = 1, \ldots, K_l\}$ contains at most $(S+1)L$ non-zero coefficients.*

Using (3.7), we can rewrite Model (3.2) as

$$Y_{it} = \sum_{l=1}^{L} \sum_{k=1}^{K_l} \mathcal{X}_{l,k,it} b_{l,k} + e_{it}, \tag{3.10}$$

where

$$\mathcal{X}_{l,k,it} = X_{it} w_{l,k}(t).$$

In vector notation,

$$Y_{it} = \mathcal{X}_{it}' b + e_{it}, \tag{3.11}$$

where $\mathcal{X}_{it} = (\mathcal{X}_{1,1,it}, \ldots, \mathcal{X}_{L,K_L,it})'$ and $b = (b_{1,1}, \ldots, b_{L,K_L})'$.

Throughout, we assume the existence of an instrument $Z_{it}$ that is correlated with $X_{it}$ and fulfills $E(Z_{it} e_{it}) = 0$ for all $i$ and $t$. The idea behind this assumption is to provide a general treatment that allows for estimating models with endogenous regressors such as dynamic models or structural models with simultaneous equations. Let $\mathcal{Z}_{l,k,it} = Z_{it} w_{l,k}(t)$ and $\mathcal{Z}_{it} = (\mathcal{Z}_{1,1,it}, \ldots, \mathcal{Z}_{L,K_L,it})'$. Because $E(Z_{it} e_{it}) = 0$ for all $i$ and $t$, we can infer that $E(\mathcal{Z}_{l,k,it} e_{it}) = 0$, for all $l$ and $k$. The required theoretical moment condition for estimating $b$ is

$$E\big(\mathcal{Z}_{it}(Y_{it} - \mathcal{X}_{it}' b)\big) = 0. \tag{3.12}$$

The IV estimator of $b$ (hereafter, denoted by $\tilde{b}$) is obtained by solving the empirical counterpart of (3.12), i.e.,

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathcal{Z}_{it}(Y_{it}-\mathcal{X}_{it}^{'}\tilde{b})\right)=0. \tag{3.13}$$

*Remark* 3.2. We know from the Generalized Method of Moments (GMM) that the IV estimator is equivalent to the *just-identified* GMM estimator, in which the number of instruments is equal to the number of parameters to be estimated. Our estimator of $b$ can be, hence, seen as a GMM estimator:

$$\tilde{b}=\arg\min_{b}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\mathcal{X}_{it}^{'}b)\mathcal{Z}_{it}^{'}\mathcal{W}_{T}\mathcal{Z}_{it}(Y_{it}-\mathcal{X}_{it}^{'}b), \tag{3.14}$$

where $\mathcal{W}_T$ is an arbitrary symmetric $(T\times T)$ full rank matrix. Since there is no matter how to choose $\mathcal{W}_T$ in the just-identified case, we can use the identity matrix to solve (3.14).

Under general assumptions, we can state the consistency of $\tilde{b}$ for any arbitrary collection of wavelet functions. But the problem with naively using the conventional basis functions is that the identification of the zero- and non-zero coefficients will be complicated. Not only will the presence of the error term in (3.10) affect the estimates of $b_{l,k}$ but also the non-orthogonality of $\mathcal{Z}_{l^{'},k^{'},it}$ to $\mathcal{X}_{l,k,it}$ across different dilation and translation levels in the objective function (the IV moment condition) will move the problem from a classical wavelets shrinkage scheme to a complex model selection problem.

Our idea consists of adjusting the scales $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ in (3.8) to the structure of $X_{it}$ and $Z_{it}$ so that following normalization conditions are satisfied.

*(a):* $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it}=1$ if $(l,k)=(l^{'},k^{'})$ and

*(b):* $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it}=0$ for all $(l,k)\neq(l^{'},k^{'})$.

Proposition B.2, in Appendix B.1, gives the mathematical conditions for $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ to ensure (a) and (b). The solution is

$$\begin{aligned} a_{1,1} &= Q_{1,1}^{-\frac{1}{2}}, \\ a_{l,2k-1} &= Q_{l,2k-1}^{-1}\left(Q_{l,2k-1}^{-1}+Q_{l,2k}^{-1}\right)^{-\frac{1}{2}}, \text{ and} \\ a_{l,2k} &= Q_{l,2k}^{-1}\left(Q_{l,2k-1}^{-1}+Q_{l,2k}^{-1}\right)^{-\frac{1}{2}}, \end{aligned}$$

where $Q_{1,1}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}Z_{it}$, $Q_{l,2k-1}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}Z_{it}h_{l,2k-1}^{2}(t)$, and $Q_{l,2k}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}Z_{it}h_{l,2k}^{2}(t)$.

Solving (3.13) (or (3.14)) with respect to $b_{l,k}$ under Restrictions (a) and (b), we obtain

$$\tilde{b}_{l,k} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathcal{Z}_{l,k,it} Y_{it}. \tag{3.15}$$

Making use of orthonormality, we can directly perform the universal thresholding scheme of Donoho and Johnstone (1994). Our structure adapted wavelet estimator of $\beta_t$ (hereafter, the SAW estimator) can be obtained by

$$\tilde{\beta}_t = \sum_{l=1}^{L} \sum_{k=1}^{K_l} w_{l,k}(t) \hat{b}_{l,k}, \tag{3.16}$$

where

$$\hat{b}_{l,k} = \begin{cases} \tilde{b}_{l,k} & \text{if} \quad |\tilde{b}_{l,k}| > \lambda_{n,T} \text{ and} \\ 0 & \text{else,} \end{cases} \tag{3.17}$$

for some threshold $\lambda_{n,T}$ that depends on $n$ and $T$. Theorems 3.5, 3.7 give the necessary conditions for $\lambda_{n,T}$ to ensure consistency under Assumptions A-C presented in Section 3.3.

*Remark* 3.3. If the explaining variable $X_{it}$ is exogenous, we can choose $Z_{it} = X_{it}$ to instrument all elements in $\mathcal{X}_{l,k,it}$ with themselves. In this case, our shrinkage estimator $\hat{b}_{l,k}$ can be interpreted as a Lasso estimator with the advantage of perfect orthogonal regressors; see, e.g., Tibshirani (1996). More generally, if $X_{it}$ is allowed to be endogenous and $\mathcal{Z}_{l,k,it} \neq \mathcal{X}_{l,k,it}$, $\hat{b}_{l,k}$ can be obtained by minimizing a Lasso-penalized just-identified GMM objective function. That is,

$$\hat{b} = \arg\min_{b} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \mathcal{X}_{it}' b) \mathcal{Z}_{it}' \mathcal{W}_T \mathcal{Z}_{it} (Y_{it} - \mathcal{X}_{it}' b) + \lambda_{n,T} |b|, \tag{3.18}$$

where $|b| = \sum_{l=1}^{L} \sum_{k=1}^{K_l} |b_{l,k}|$ and $\mathcal{W}_T$ is an arbitrary symmetric $(T \times T)$ full rank matrix. Note that (3.18) and (3.17) lead to the same result independently of the choice of $\mathcal{W}_T$.

First step SAW estimation for a straightforward centered panel model is done. Generalization for multivariate models with unobserved heterogeneity effects and post-SAW procedures follow.

## 3.3 Two-way Panel Models with Multiple Jumps

### Model

One of the main advantages of using panel datasets is the possibility of dealing with problems related to the potential effect of unobserved heterogeneity in time- and cross-section dimensions. In this section, we generalize the SAW method to models with unobserved individual and time effects and allow for multiple jumping parameters. Collecting the slope parameters in a $(P \times 1)$ time-varying vector, we can write Model (3.1) as

$$Y_{it} = \mu + X_{it}^{'}\beta_t + \alpha_i + \theta_t + e_{it}, \tag{3.19}$$

where $X_{it} = (X_{1,it}, \ldots, X_{P,it})^{'}$ is a $(P \times 1)$ vector of regressors, $\beta_t = (\beta_{t,1} \ldots, \beta_{t,P})^{'}$ is a $(P \times 1)$ vector of slope parameters, $\alpha_i$ is a time-constant individual effect and $\theta_t$ is a common time-varying effect. We allow for each $\beta_{t,p}$, $p \in \{1, \ldots, P\}$, to jump at unknown $S_p$ break points, say $\tau_{1,p}, \ldots, \tau_{S_p,p}$, such that

$$\beta_{t,p} = \begin{cases} \beta_{\tau_{1,p}} & \text{for} \quad t \in \{1, 2, \ldots, \tau_{1,p}\}, \\ \beta_{\tau_{2,p}} & \text{for} \quad t \in \{\tau_{1,p}+1, \ldots, \tau_{2,p}\}, \\ \vdots & \\ \beta_{\tau_{S_p+1,p}} & \text{for} \quad t \in \{\tau_{S_p,p}+1, \ldots, T\}. \end{cases} \tag{3.20}$$

The estimation procedure for this model is conceptually similar to the univariate method discussed in Section 3.2. However, besides the need to deal with multivariate wavelets, we have to control for the additional unknown parameters $\mu$, $\alpha_i$, and $\theta_t$.

From the literature on panel models, we know that uniqueness of $\mu$, $\alpha_i$, and $\theta_t$ requires the following identification conditions:

$$\begin{array}{lrcl} \text{C.1:} & \sum_{i=1}^{n} \alpha_i & = & 0, \quad \text{and} \\ \text{C.2:} & \sum_{t=1}^{T} \theta_t & = & 0. \end{array} \tag{3.21}$$

### 3.3.1 Estimation

In order to cover the case of dynamic models with both small and large $T$, we conventionally start with differencing the model to eliminate the individual effects and assume the existence of appropriate instruments.

By taking the difference on the left and the right hand side of (3.19), we get

$$(Y_{it} - Y_{it-1}) = X_{it}^{'}\beta_t - X_{it-1}^{'}\beta_{t-1} + (\theta_t - \theta_{t-1}) + (e_{it} - e_{it-1}), \tag{3.22}$$

for $i \in \{1, \ldots, n\}$ and $t \in \{2, \ldots, T\}$.

Because $n$ is usually supposed to be large, $\Delta\theta_t = \theta_t - \theta_{t-1}$, can be eliminated by using the classical within transformation on the model, i.e., $\Delta\dot{Y}_{it} = \Delta Y_{it} - \frac{1}{n}\sum_{i=1}^{n} \Delta Y_{it}$. Alternatively, we can associate $\Delta\theta_t$ with an additional unit regressor in the model and estimate it jointly with $\beta_t$ as a potential jumping parameter. Indeed, allowing for $\Delta\theta_t$ to be piecewise constant over time can be very useful for interpretation, especially when the original time effect $\theta_t$ has approximately a piecewise changing linear trend.

Let $\underline{X}_{it} = (X_{it}^{'}, -X_{it-1}^{'}, 1)^{'}$ and $\gamma_t = (\beta_t^{'}, \beta_{t-1}^{'}, \Delta\theta_t)^{'}$ be $(\underline{P} \times 1)$ extended vectors, where $\underline{P} = 2P + 1$. We can rewrite Model (3.22) as

$$\begin{aligned} \Delta Y_{it} &= (X_{it}^{'}, -X_{it-1}^{'}, 1) \begin{pmatrix} \beta_t \\ \beta_{t-1} \\ \Delta\theta_t \end{pmatrix} + \Delta e_{it}, \\ &= \underline{X}_{it}^{'}\gamma_t + \Delta e_{it}, \end{aligned} \tag{3.23}$$

for $i \in \{1, \ldots, n\}$ and $t \in \{2, \ldots, T\}$.

By using multivariate structure adapted wavelet functions, we can estimate $\gamma_t$ in a way similar to the way discussed in Section 3.2.

The multivariate structure adapted wavelet expansion of $\gamma_t$ can be presented as follows:

$$\gamma_t = \sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{lk}(t)\underline{b}_{l,k} \qquad \text{for } \; t \in \{2, \ldots, T\}, \tag{3.24}$$

where $\underline{b}_{lk} = (b_{l,k,1}, \ldots, b_{l,k,\underline{P}})^{'}$ is a $(\underline{P} \times 1)$ vector of wavelet coefficients and $W_{lk}(t)$ is a $(\underline{P} \times \underline{P})$ multivariate wavelet basis matrix defined as

$$W_{l,k}(t) = \begin{cases} A_{1,1} = A_{2,1}H_{2,1}(t) + A_{2,2}H_{2,2}(t) & \text{if} \quad l = 1, \;\; \text{and} \\ A_{l,2k-1}H_{l,2k-1}(t) - A_{l,2k}H_{l,2k}(t) & \text{if} \quad l > 1, \end{cases} \tag{3.25}$$

with

$$H_{l,m}(t) = \sqrt{2^{l-2}}I_{l,m}(t-1),$$

and $A_{1,1}, A_{l,2k-1}$, and $A_{l,2k}$ are constructed so that the following orthonormality conditions are fulfilled:

*(A):* $\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T} \underline{\mathcal{Z}}_{l,k,it}\underline{\mathcal{X}}_{l',k',it}^{'} = I_{\underline{P}\times\underline{P}}$ if $(l,k) = (l^{'}, k^{'})$ and

*(B):* $\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T} \underline{\mathcal{Z}}_{l,k,it}\underline{\mathcal{X}}_{l',k',it}^{'} = \mathbf{0}_{\underline{P}\times\underline{P}}$ for all $(l,k) \neq (l^{'}, k^{'})$.

Here, $\underline{\mathcal{X}}'_{l,k,it} = \underline{X}'_{it} W_{lk}(t)$, $I_{\underline{P}\times\underline{P}}$ is the $(\underline{P}\times\underline{P})$ identity matrix, $\mathbf{0}_{\underline{P}\times\underline{P}}$ is a $(\underline{P}\times\underline{P})$ matrix of zeros, and $\underline{\mathcal{Z}}'_{l,k,it} = \underline{Z}'_{it} W_{l,k}(t)$, where $\underline{Z}_{it}$ is a $(\underline{P}\times 1)$ vector used to instrument the $\underline{P}$ variables in $\underline{X}_{it}$; the unit regressor associated with $\Delta\theta_t$ and the remaining exogenous regressors (if they exist) can be, of course, instrumented by themselves.

We can easily verify that

$$\begin{aligned} A_{1,1} &= \underline{Q}_{1,1}^{-\frac{1}{2}}, \\ A_{l,2k-1} &= \underline{Q}_{l,2k-1}^{-1}\left(\underline{Q}_{l,2k-1}^{-1} + \underline{Q}_{l,2k}^{-1}\right)^{-\frac{1}{2}}, \text{ and} \\ A_{l,2k} &= \underline{Q}_{l,2k}^{-1}\left(\underline{Q}_{l,2k-1}^{-1} + \underline{Q}_{l,2k}^{-1}\right)^{-\frac{1}{2}}, \end{aligned}$$

with

$$\begin{aligned} \underline{Q}_{1,1} &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\underline{Z}_{it}\underline{X}'_{it}, \\ \underline{Q}_{l,2k-1} &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\underline{Z}_{it}\underline{X}'_{it}h_{l,2k-1}(t)^2, \text{ and} \\ \underline{Q}_{l,2k} &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\underline{Z}_{it}\underline{X}'_{it}h_{l,2k}(t)^2. \end{aligned}$$

ensure conditions $(A)$ and $(B)$.

The IV estimator of $\underline{b}_{l,k}$ is the solution of the empirical moment condition

$$\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left(\underline{\mathcal{Z}}_{l,k,it}(\Delta Y_{it} - \underline{\mathcal{X}}'_{l,kit}\tilde{\underline{b}}_{l,k})\right) = 0. \tag{3.26}$$

Solving (3.26) for $\tilde{\underline{b}}_{l,k}$ under the the normalization Conditions $(A)$ and $(B)$, we obtain

$$\tilde{b}_{l,k,p} = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\mathcal{Z}_{l,k,it,p}\Delta Y_{it},$$

where $\tilde{b}_{l,k,p}$ and $\mathcal{Z}_{l,k,it,p}$ are the $p$th elements of $\tilde{\underline{b}}_{l,k}$ and $\underline{\mathcal{Z}}_{l,k,it}$, respectively.

The SAW estimator of $\beta_{t,p}$ can be obtained by

$$\hat{\gamma}_{t,p} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}} W_{lk,p,q}(t)\hat{b}_{l,k,q}, \tag{3.27}$$

or

$$\hat{\gamma}_{t+1,p+P} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}} W_{lk,p+P,q}(t+1)\hat{b}_{l,k,q}, \tag{3.28}$$

where $W_{lk,p,q}$ is the $(p,q)$- element of the basis matrix $W_{lk}(t)$, and

$$\hat{b}_{l,k,q} = \begin{cases} \tilde{b}_{l,k,q} & \text{if} \quad |\tilde{b}_{l,k,q}| > \lambda_{n,T} \text{ and} \\ 0 & \text{else.} \end{cases} \tag{3.29}$$

### 3.3.2 Assumptions and Main Asymptotic Results

We present a set of assumptions that are necessary for our asymptotic analysis. Throughout, we use $E_c(.)$ to define the conditional expectation given $\{X_{it}\}_{i,t\in\mathbb{N}^{*2}}$ and $\{Z_{it}\}_{i,t\in\mathbb{N}^{*2}}$, where $\mathbb{N}^* = \mathbb{N}\setminus\{0\}$. We denote by $M$ a finite positive constant, not dependent on $n$ and $T$. The operators $\xrightarrow{p}$ and $\xrightarrow{d}$ denote the convergence in probability and distribution. $O_p(.)$ and $o_p(.)$ are the usual Landau-symbols. The Frobenius norm of a $(p\times k)$ matrix $A$ is denoted by $||A|| = [tr(A'A)]^{1/2}$, where $A'$ denotes the transpose of $A$. $\Delta$ denotes the difference operator of first order.

Our theoretical setup relies on the following assumptions.

**Assumption A - Data Dimension and Stability Intervals:**

(i) $T-1 = 2^{L-1}$ for some natural number $L>1$; the number of regressors $P$ is fixed.

(ii) $n\to\infty$; $T$ is either fixed or passes to infinity simultaneously with $n$ such that $\log(T)/n \to 0$.

(iii) $\min_{j,p}|\beta_{\tau_{j,p}} - \beta_{\tau_{j-1,p}}|$ does not vanish when $n$ and $T$ pass to infinity; all stability intervals $(\tau_{j,p} - \tau_{j-1,p}) \to \infty$ uniformly in $n$, as $T\to\infty$.

**Assumption B - Regressors and Instruments:**

(i) for all $i$ and $t$, $E_c(\underline{Z}_{it}e_{it}) = 0$; for all $l\in\{1,\dots,L\}$ and $k\in\{1,\dots,K_l\}$,

$$\underline{Q}_{l,k} = \frac{1}{n\cdot\sharp\{s|h_{l,k}(s)\neq 0\}} \sum_{t\in\{s|h_{l,k}(s)\neq 0\}} \sum_{i=1}^{n} \underline{Z}_{it}\underline{X}'_{it} \xrightarrow{p} \underline{Q}^{\circ}_{l,k},$$

where $\underline{Q}^{\circ}_{l,k}$ is a $(\underline{P}\times\underline{P})$ full rank finite matrix with distinct eigenvectors.

(ii) The moments $E||\underline{Z}_{it}||^4$ and $E||\underline{X}_{it}||^4$ are bounded uniformly in $i$ and $t$; for $A_{l,2k} = \underline{Q}^{-1}_{l,2k}(\underline{Q}^{-1}_{l,2k} + \underline{Q}^{-1}_{l,2k-1})^{-1/2}$ and $A_{l,2k-1} = \underline{Q}^{-1}_{l,2k-1}(\underline{Q}^{-1}_{l,2k} + \underline{Q}^{-1}_{l,2k-1})^{-1/2}$, the moments $E||A_{l,2k}||^4$ and $E||A_{l,2k-1}||^4$ are bounded uniformly in $l$ and $k$.

(iii) the multivariate distribution of $\{\Delta e_{it}\}_{i\in\mathbb{N}^*,t\in\mathbb{N}^*\setminus\{1\}}$ is Sub-Gaussian so that every linear combination

$$\Pi_{nT}(a_{s,s'}) = \sum_{t=s+1}^{s'}\sum_{i=1}^{n}\frac{a_{s,s',it}}{\sqrt{n(s'-s)}}\Delta e_{it},$$

with $E(a_{s,s',it}\Delta e_{it}) = 0$ and $E\big(\Pi^2_{nT}(a_{s,s'})\big) \leq M$, is Sub-Gaussian distributed of order $\Sigma_{nT}(a_{s,s'}) = E\big(\Pi^2_{nT}(a_{s,s'})\big)$, i.e.,

$$P\big(\Sigma^{-\frac{1}{2}}_{nT}(a_{s,s'})|\Pi_{nT}(a_{s,s'})| \geq c\big) \leq \frac{1}{c}\exp(-\frac{c^2}{2}),$$

for any $c > 0$.

**Assumption C - Weak Dependencies and Heteroskedasticity in the Error Term:** $E_c(\Delta e_{it}\Delta e_{jm}) = \sigma_{ij,tm},\ |\sigma_{ij,tm}| \leq \bar{\sigma}$ for all $(i,j,t,m)$ such that

$$\frac{1}{n(s'-s+1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=s+1}^{s'}\sum_{m=s+1}^{s'}|\sigma_{ij,tm}| \leq M.$$

Assumption A.(i) specifies a dyadic condition on the intertemporal data size $T$. This is a technical assumption that is only required for constructing the dyadic wavelet basis functions. In practice, we can replicate the data by reflecting the observations at the boundaries to get the desired dimension. If, for instance, $T-1 = 125$, we can extend the sample $(Y_{i1}, X_{i1}), \ldots, (Y_{iT}, X_{iT})$ with the three last observations $(Y_{iT-1}, X_{iT-1})$, $(Y_{iT-2}, X_{iT-2})$, and $(Y_{iT-3}, X_{iT-3})$ for $T+1, T+2$, and $T+3$, respectively. The asymptotic property of the estimator, will depend, of course, on the original data size and not on the size of the replicated data. Assumption A.(ii) allows for the time dimension $T$ to be very long compared to $n$ but in such a way that $\log(T) = o(n)$. A.(ii) considers also the classical case of panel data, in which $T$ is fixed and only $n \to \infty$. Assumption A.(iii) guarantees that the jumps do not vanish as $n$ and/or $T$ pass to infinity. The second part of Assumption A.(iii) can be alleviated to allow for some stability intervals to stay fixed if $T \to \infty$. Assuming the stability intervals to pass to infinity when $T$ gets large allows for interpreting the $T$-asymptotic as a full-in asymptotic.

Assumption B.(i) requires that the probability limit of $\underline{Q}_{l,k}$ is a full rank finite matrix with distinct eigenvectors. This is to ensure that its eigendecompostion exists. Assumption B.(ii) specifies commonly used moment conditions to allow for some limiting terms

to be $O_p(1)$ when using Chebyshev inequality. The Sub-Gaussian condition in Assumption B.(iii) excludes heavy tailed distributed errors but does not impose any specific exact distribution.

Assumption C allows for a weak form of time series and cross section dependence in the errors as well as heteroskedasticities in both time and cross-section dimension. It implies that the covariances and variances are uniformly bounded and the double summations over all possible time partitions are well behaved. The assumption generalizes the restricted case of independent and identically distributed errors.

The following Lemma establishes the main asymptotic results for the structure adapted wavelet coefficients.

**Lemma 3.4.** *Suppose Assumptions A-C hold, then*

*(i)*

$$\sup_{l,k,q}\left|\tilde{b}_{l,k,q}-b_{l,k,q}\right| = O_p\big(\sqrt{\log(T-1)/n(T-1)}\big),$$

*(ii) for some finite* $M>\sqrt{2}$,

$$\sup_{l,k,q}\left|\tilde{b}_{l,k,q}-b_{l,k,q}\right| \leq M\sqrt{\log((T-1)\underline{P})/n(T-1)}$$

*holds with a probability that converges to* 1 *independently of* $n$, *as* $T\to\infty$.

Theorem 3.5 establishes the uniform and the mean square consistency of $\tilde{\gamma}_{t,p}$.

**Theorem 3.5.** *Assume Assumptions A-C, then the following statements hold:*

*(i)* $\sup_t|\hat{\gamma}_{t,p}-\gamma_{t,p}| = o_p(1)$ *for all* $p\in\{1,\dots,\underline{P}\}$, *if* $\sqrt{T-1}\lambda_{n,T}\to 0$, *as* $n,T\to\infty$ *or* $n\to\infty$ *and* $T$ *is fixed, and*

*(ii)* $\frac{1}{T-1}\sum_{t=2}^{T}||\hat{\gamma}_t-\gamma_t||^2 = O_p\big(\frac{J^*}{(T-1)}(\log(T-1)/n)^{\kappa}\big)$, *where* $J^*=\min\{(\sum_{p=1}^{\underline{P}}S_p+1)\log(T-1),(T-1)\}$ , *if* $\sqrt{T-1}\lambda_{n,T}\sim(\log(T-1)/n)^{\kappa/2}$, *for any* $\kappa\in]0,1[$.

Uniform consistency is obtained when $n\to\infty$ and $T$ is fixed or $n,T\to\infty$ with $\log(T)/n\to 0$. If the maximum number of jumps is fixed, the mean square consistency is obtained even when $n$ is fixed and only $T\to\infty$.

A threshold that satisfies Conditions $(i)$ and $(ii)$ in theorem 3.5, can be constructed as follows:

$$\lambda_{nT}=\hat{V}_{nT}^{\frac{1}{2}}\left(\frac{2\log((T-1)\underline{P})}{n(T-1)^{1/\kappa}}\right)^{\kappa/2},\ \text{for some } \kappa\in]0,1[, \tag{3.30}$$

where $\hat{V}_{nT}$ is the empirical variance estimator corresponding to the largest variance of $\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathcal{Z}_{it,l,k,p}\Delta e_{it}$ over $l, k$, and $p$. Such an estimator can be obtained by using the residuals $\tilde{e}_{it}$ of a pre-intermediate SAW regression performed with a plug-in threshold $\lambda^{*}_{nT} = 0$. We want to emphasize that asymptotically all we need is that $\hat{V}_{nT}$ be strictly positive and bounded. The role of $\hat{V}_{nT}^{\frac{1}{2}}$ is only to give the threshold a convenient amplitude. The role of $\kappa < 1$ is to trade off the under-estimation effect that can arise from the plug-in threshold $\lambda^{*}_{n,T} = 0$. An ad-hoc choice of $\kappa$ is $1 - \log\log(nT)/\log(nT)$. For more accurate choices, we refer to the calibration strategies proposed by Hallin and Liška (2007) and Alessi et al. (2010).

## 3.4 Post-SAW Procedures

### 3.4.1 Tree-Structured Representation

The intrinsic problem of wavelets is that wavelet functions are constructed via dyadic dilations. Error may make this feature spuriously generate some additional mini jumps to stimulate the big (true) jump when it is located at a non-dyadic position. To construct a selective inference for testing the systematic jumps it is important to encode the coefficients that may generate such effects. One possible approach is to examine the so-called *tree-structured* representation, which is based on the hierarchical interpretation of the wavelet coefficients. Recall that the wavelet basis functions are nested over a binary multiscale structure so that the support of an $(l, k)$- basis (the time interval in which the basis function is not zero) contains the supports of the basis $(l + 1, 2k - 1)$ and $(l + 1, 2k)$. We say that the wavelet coefficient $b_{l,k}$ is the *parent* of the two *children* $b_{l+1,2k-1}$ and $b_{l+1,2k}$. This induces a dyadic tree structure rooted to the primary parent $b_{1,1}$. To encode the possible systematic jumps, we have to traverse the tree up to the root parent in a recursive trajectory starting from the non-zero coefficients at the finest resolution (highest dilation level). While the presence of a non-zero coefficient, at the highest level, indicates the presence of a jump, the parent may have a non-zero coefficient only to indicate that the stability interval around this jump is larger than its support.

As an illustration, consider the tree-structured representation in Figure 3.1. The coefficients at the not-ringed nodes fall in the interval $[-\lambda_{n,T}, \lambda_{n,T}]$ and carry the value zero. Starting from the non-zero coefficient $\tilde{b}_{5,6}$ at the finest resolution and traversing the tree up to the root parent, we can identify $\tilde{b}_{4,3}, \tilde{b}_{3,2}$, and $\tilde{b}_{2,1}$ as candidates for generating potential visual artifacts at points $8, 10$, and $12$ if a jump exists only at $11$. These selected jump points can be tested by using, e.g., the equality test of Chow (1960).

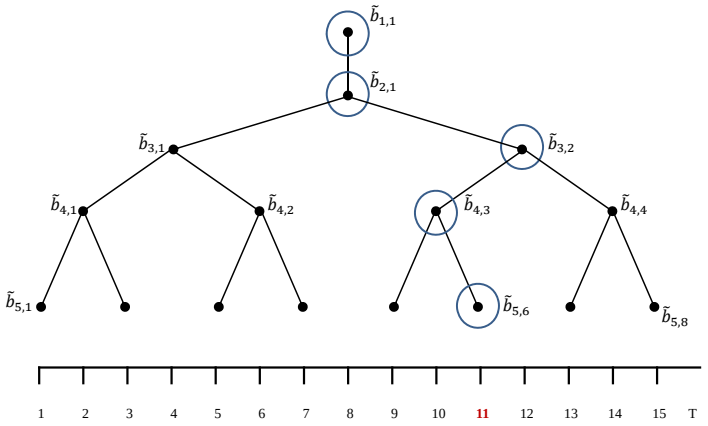


FIGURE 3.1: An illustrating example of a tree-structured representation for the wavelet coefficients.

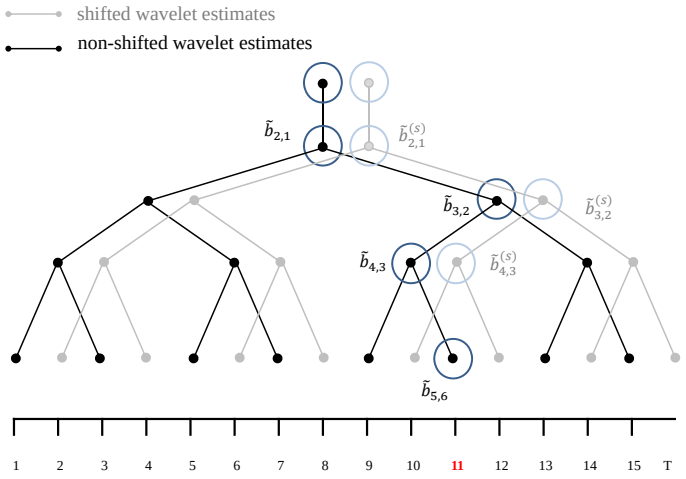


FIGURE 3.2: An illustrating example of a tree-structured representation for the shifted and non-shifted coefficients.

If we have an additional observation, we can construct a shifted wavelet expansion on a second (shifted) dyadic interval. The tree-structured representation of the new coefficients can provide important information about the significance of the potential jumps detected in the first tree. Continuing with the same example of Figure 3.1, we can see that the tree-structured representations of the shifted and non-shifted coefficients presented in Figure 3.2 support the hypothesis of only one jump at 11.

In the multivariate case, the interpretation of the tree-structured representation can

be complicated since the nodes represent vectors that contain simultaneous information about multiple regressors. In order to construct an individual tree for each parameter, we can re-transform each element of the $(\underline{P} \times 1)$ vector $\gamma_t$ with the conventional univariate wavelet basis functions defined in (3.4). Recall that, in our differenced model, $\gamma_{t,p} = \beta_{t,p}$ and $\gamma_{t,p+P} = \beta_{t-1,p}$. This allows us to obtain for each slope parameter, $\beta_p$, two sets of univariate wavelet coefficients:

$$c_{l,k,p}^{(s)} = \frac{1}{T-1}\sum_{t=2}^{T}\psi_{l,k}(t-1)\gamma_{t,p}, \tag{3.31}$$

and

$$c_{l,k,p}^{(u)} = \frac{1}{T-1}\sum_{t=1}^{T-1}\psi_{l,k}(t)\gamma_{t+1,p+P}. \tag{3.32}$$

We use the superscripts $(s)$ and $(u)$ in (3.31) and (3.32) to denote the shifted and non-shifted coefficients, respectively.

Replacing $\gamma_{t,p}$ with $\tilde{\gamma}_{t,p} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}} W_{lk,p,q}(t)\tilde{b}_{l,k,q}$ and $\gamma_{t+1,p+P}$ with $\tilde{\gamma}_{t+1,p+P} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}} W_{lk,p+P,q}(t+1)\tilde{b}_{l,k,q}$, we obtain

$$\tilde{c}_{l,k,p}^{(s)} = \frac{1}{T-1}\sum_{t=2}^{T}\psi_{l,k}(t-1)\tilde{\gamma}_{t,p}, \tag{3.33}$$

and

$$\tilde{c}_{l,k,p}^{(u)} = \frac{1}{T-1}\sum_{t=1}^{T-1}\psi_{l,k}(t)\tilde{\gamma}_{t+1,p+P}. \tag{3.34}$$

Having an appropriate threshold for $\tilde{c}_{l,k,p}^{(u)}$, we can construct the shifted and non-shifted tree-structured representation for each parameter, as before. This can provide important information about the potential spurious jumps since all low level parameters in the shifted tree fall in the highest level of the non-shifted tree and vice versa. Based on this predicate, we propose a selection method for consistently detecting the jump locations. All we need is an appropriate threshold for the highest coefficients.

The following Lemma establishes the uniform consistency in $k$ and $p$ of both $\tilde{c}_{L,k,p}^{(s)}$ and $\tilde{c}_{L,k,p}^{(u)}$ and states their order of magnitude in probability.

**Lemma 3.6.** *Suppose Assumptions A-C hold, then, for all* $p \in \{1,\dots,P\}$ *and* $m \in \{m,s\}$

$$\sup_k \left|\tilde{c}_{L,k,p}^{(m)} - c_{L,k,p}^{(m)}\right| = O_p\big(\sqrt{\log(T-1)/n(T-1)}\big).$$

From Lemma 3.6, we can intuitively see that asymptotically both $\tilde{c}^{(m)}_{L,k,p}$ and $\tilde{b}_{l,k,p}$ can be shrunk by the same threshold $\lambda_{n,T}$. Theorem 3.7 gives the necessary asymptotic conditions to ensure consistency of the jump selection method.

### 3.4.2 Detecting the Jump Locations

As mentioned earlier, interpreting all jumps of the SAW estimator as structural breaks may lead to an over-specification of the break points. In this Section, we exploit the information existing in the shifted and unshifted univariate wavelet coefficients (3.33) and (3.34) to construct a consistent selection method for detecting the jump locations.

We use (3.33) and (3.34) to obtain the following two estimators of $\Delta\beta_t$:

$$\Delta\tilde{\beta}^{(u)}_{t,p} = \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t)\hat{c}^{(u)}_{L,k,p}, \quad \text{for } t \in \mathcal{E}, \tag{3.35}$$

and

$$\Delta\tilde{\beta}^{(s)}_{t,p} = \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t-1)\hat{c}^{(s)}_{l,k,p}, \quad \text{for } t \in \mathcal{E}^c, \tag{3.36}$$

where

$$\hat{c}^{(.)}_{l,k,p} = \mathbf{I}(|\tilde{c}^{(.)}_{l,k,p}| > \lambda_{n,T}),$$

$\mathcal{E}$ is the set of the even time locations $\{2, 4, \ldots, T-1\}$, $\mathcal{E}^c$ is the complement set composed of the odd time locations $\{2, 3, \ldots, T\} \setminus \mathcal{E}$, and $\mathbf{I}(.)$ is the indicator function.

The number of jumps of each parameter can be estimated by

$$\tilde{S}_p = \sum_{t\in\mathcal{E}} \mathbf{I}(\Delta\tilde{\beta}^{(u)}_{t,p} \neq 0) + \sum_{t\in\mathcal{E}^c} \mathbf{I}(\Delta\tilde{\beta}^{(s)}_{t,p} \neq 0). \tag{3.37}$$

The jump locations $\tilde{\tau}_{1,p}, \ldots, \tilde{\tau}_{\tilde{S}_p,p}$ can be identified as follows:

$$\tilde{\tau}_{j,p} = \min\left\{ s \,\middle|\, j = \sum_{t=2}^{s} \mathbf{I}\left(\Delta\tilde{\beta}^{(u)}_{t,p} \neq 0,\, t \in \mathcal{E}\right) + \sum_{t=3}^{s} \mathbf{I}\left(\Delta\tilde{\beta}^{(s)}_{t,p} \neq 0,\, t \in \mathcal{E}^c\right) \right\}, \tag{3.38}$$

for $j \in \{1, \ldots, \tilde{S}_p\}$. The maximal number of breaks $S = \sum_{p=1}^{P} = S_p$ can be estimated by $\tilde{S} = \sum_{p=1}^{P} \tilde{S}_p$.

**Theorem 3.7.** *Under Assumptions A-C, if* $(c.1)$ : $\sqrt{\frac{n(T-1)}{\log((T-1))}}\lambda_{n,T} \to \infty$ *and* $(c.2)$ : $\sqrt{T-1}\lambda_{n,T} \to 0$, *as* $n, T \to \infty$, *then*

*(i)* $\lim_{n,T\to\infty} P(\tilde{S}_1 = S_1, \ldots, \tilde{S}_p = S_p) = 1$ *and*

*(ii)* $\lim_{n,T\to\infty} P(\tilde{\tau}_{1,1} = \tau_{1,1}, \ldots, \tilde{\tau}_{S_P,P} = \tau_{S_P,P} | \tilde{S}_1 = S_1, \ldots, \tilde{S}_p = S_p) = 1.$

The crucial element for consistently estimating $\tau_{1,1}, \ldots, \tau_{S_P,P}$ is, hence, using a threshold that converges to zero but at a rate slower than $\sqrt{\log(T-1)/(n(T-1))}$.

### 3.4.3 Post-SAW Estimation

For known $\tau_{1,p}, \ldots, \tau_{S_p,p}$, we can rewrite Model (3.22) as

$$\Delta\dot{Y}_{it} = \sum_{p=1}^{P} \sum_{j=1}^{S_p+1} \Delta\dot{X}_{it,p}^{(\tau_{j,p})} \beta_{\tau_{j,p}} + \Delta\dot{e}_{it}, \tag{3.39}$$

where

$$\Delta\dot{X}_{(it,p)}^{(\tau_{j,p})} = \Delta\dot{X}_{it,p}\mathbf{I}\big(\tau_{j-1,p} < t \leq \tau_{j,p}\big),$$

with $\tau_{0,p} = 1$ and $\tau_{S_p+1,p} = T$, for $p \in \{1, \ldots, P\}$. The dot operator transforms the variables as follows: $\dot{u}_{it} = u_{it} - \frac{1}{n}\sum_{i=1}^{n} u_{it}$.

Depending on the set of the jump locations $\tau := \{\tau_{j,p} | j = 1, \ldots, S_p + 1,\ p = 1, \ldots, P\}$, the vector presentation of Model (3.39) can be rewritten as

$$\Delta\dot{Y}_{it} = \Delta\dot{X}_{it,(\tau)}^{'}\beta_{(\tau)} + \Delta\dot{e}_{it}, \tag{3.40}$$

where $\beta_{(\tau)} = (\beta_{\tau_{1,1}}, \ldots, \beta_{\tau_{S_1+1,1}}, \ldots, \beta_{\tau_{1,P}} \ldots, \beta_{\tau_{S_P+1,P}})'$ and $\Delta\dot{X}_{it,(\tau)} = (\Delta\dot{X}_{it,1}^{(\tau_{1,1})}, \ldots, \Delta\dot{X}_{it,1}^{(\tau_{S_1+1,1})}, \ldots, \Delta\dot{X}_{it,P}^{(\tau_{1,P})} \ldots, \Delta\dot{X}_{it,P}^{\tau_{(S_P+1,P)}})'$ .

Let $Z_{it,p}$ denote the instrument chosen for $\Delta\dot{X}_{it,p}$ and $Z_{it,(\tau)} = (Z_{it,1}^{(\tau_{1,1})}, \ldots, Z_{it,1}^{(\tau_{S_1+1,1})}, \ldots, Z_{it,P}^{(\tau_{1,P})}, \ldots, Z_{it,P}^{\tau_{(S_P+1,P)}})'$, with $Z_{it,p}^{(\tau_{j,p})} = Z_{it,p}\mathbf{I}\big(\tau_{j-1,p} < t \leq \tau_{j,p}\big)$. The conventional IV estimator of $\beta_{(\tau)}$ is

$$\hat{\beta}_{(\tau)} = \big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}_{it,(\tau)}^{'}\big)^{-1}\big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{Y}_{it}\big). \tag{3.41}$$

Conditional on $\tilde{S}_1 = S_1, \ldots, \tilde{S}_P = S_P$, we can replace the set of the true jump locations $\tau$ in (3.41) with the detected jump locations $\tilde{\tau} := \{\tilde{\tau}_{j,p} | j \in \{1, \ldots, S_p+1\},\ p \in \{1, \ldots, P\}\}$, to obtain the post-SAW estimator:

$$\hat{\beta}_{(\tilde{\tau})} = \big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}\Delta\dot{X}_{it,(\tilde{\tau})}^{'}\big)^{-1}\big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}\Delta\dot{Y}_{it}\big). \tag{3.42}$$

From (3.26) and (3.42), we can see that the number of parameters to be estimated after detecting the jump locations is much smaller than the number of parameters required to estimate the slope parameters in the SAW regression ($\sum_{p=1}^{P}(S_p+1) < T(\underline{P}-1)$). It is evident that such a gain in terms of regression dimension improves the quality of the estimator.

**Assumption E - Central Limits:** Let $\mathcal{T}_{(\tau)}$ be a $(S+P \times S+P)$ diagonal matrix with the diagonal elements $T_{1,1},\ldots,T_{S_P+1,P}$, where $T_{j,p} = \tau_{j,p} - \tau_{j-1,p} + 1$.

$(i):$ $(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}'_{it,(\tau)} \xrightarrow{p} Q^{\circ}_{(\tau)}$ where $Q^{\circ}_{(\tau)}$ is a full rank finite matrix.

$(ii):$ $(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{j=1}^{n}\sum_{s=2}^{T} Z_{it,(\tau)}Z'_{js,(\tau)}\sigma_{ij,ts} \xrightarrow{p} V^{\circ}_{(\tau)}$, where $V^{\circ}_{(\tau)}$ is a full rank finite matrix.

$(iii):$ $\big(n\mathcal{T}_{(\tau)}\big)^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{e}_{it} \xrightarrow{d} N(0, V^{\circ}_{(\tau)})$.

Assumption E presents standard assumptions that are commonly used in the literature on instrumental variables.

**Theorem 3.8.** *Suppose Assumptions A-E hold. Then conditional on* $\tilde{S}_1 = S_1,\ldots,\tilde{S}_p = S_p$*, we have*

$$\sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\big(\hat{\beta}_{(\tilde{\tau})} - \beta_{(\tau)}\big) \xrightarrow{d} N(0, \Sigma_{(\tau)}),$$

*where* $\Sigma_{(\tau)} = (Q^{\circ}_{(\tau)})^{-1}\big(V^{\circ}_{(\tau)}\big)(Q^{\circ}_{(\tau)})^{-1}$.

If $T \to \infty$ and all $T_{j,p}$ diverge proportionally to $T$, then $\hat{\beta}_{\tau_j,p}$ achieves the usual $\sqrt{nT}$-convergence rate. Based on the asymptotic distribution of $\hat{\beta}_{(\tilde{\tau})}$, we can construct a Chow-type test to examine the statistical significance of the detected jumps and/or a Hotelling test to examine whether a model with constant parameters is more appropriate for the data than a model with jumping parameters.

Because $\Sigma_{(\tau)}$ is unknown, consistent estimators of $Q^{\circ}_{(\tau)}$ and $V^{\circ}_{(\tau)}$ are required to perform inferences. A natural estimator of $Q^{\circ}_{(\tau)}$ is

$$\hat{Q}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}\Delta X'_{it,(\tilde{\tau})}$$

and a consistent estimator of $\Sigma_{(\tau)}$ can be obtained by

$$\hat{\Sigma}_{(\tilde{\tau}),j} = \hat{Q}^{-1}_{(\tilde{\tau})}\hat{V}^{(c)}_{(\tilde{\tau})}\hat{Q}^{-1}_{(\tilde{\tau})},$$

where $\hat{V}^{(c)}_{(\tilde{\tau})}$ a consistent estimator of $V^{\circ}_{(\tilde{\tau})}$ that can be constructed depending on the structure of $\Delta\dot{e}_{it}$. For brevity, we distinguish only four cases:

1. The case of homoscedasticity without the presence of auto- and cross-section correlations:
$$\hat{V}^{(1)}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z'_{it,(\tilde{\tau})} \hat{\sigma}^2,$$
where $\hat{\sigma}^2 = \frac{1}{n(T-1)} \sum_{i=1}^{n}\sum_{t=2}^{T} \Delta\hat{\dot{e}}^2_{it}$, with $\Delta\hat{\dot{e}}_{it} = \Delta\dot{Y}_{it} - \Delta\dot{X}'_{it,(\tilde{\tau})}\hat{\beta}_{(\tilde{\tau})}$.

2. The case of cross-section heteroskedasticity without auto- and cross-section correlations:
$$\hat{V}^{(2)}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z'_{it,(\tilde{\tau})} \hat{\sigma}^2_i,$$
where $\hat{\sigma}^2_i = \frac{1}{T-1} \sum_{t=2}^{T} \Delta\hat{\dot{e}}^2_{it}$.

3. The case of time heteroskedasticity without auto- and cross-section correlations:
$$\hat{V}^{(3)}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z'_{it,(\tilde{\tau})} \hat{\sigma}^2_t,$$
where $\hat{\sigma}^2_t = \frac{1}{n} \sum_{i=1}^{n} \Delta\hat{\dot{e}}^2_{it}$.

4. The case of cross-section and time heteroskedasticity without auto- and cross-section correlations:
$$\hat{V}^{(4)}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z'_{it,(\tilde{\tau})} \Delta\hat{\dot{e}}^2_{it}.$$

**Proposition 3.9.** *Under Assumptions A-E, we have, as* $n, T \to \infty$, $\hat{\Sigma}^{(c)}_{(\tilde{\tau})} = \Sigma_{(\tau)} + o_p(1)$, *for* $c = 1, 2, 3,$ *and* 4.

*Remark* 3.10. If the errors (at the difference level) are autocorrelated, $V^{(c)}_{(\tilde{\tau})}$ can be estimated by applying the standard heteroskedasticity and autocorrelation (HAC) robust limiting covariance estimator to the sequence $\{Z_{it,(\tilde{\tau})}\Delta\hat{\dot{e}}_{it}\}_{i\in\mathbb{N}^*, t\in\mathbb{N}^*\setminus\{1\}}$; see, e.g., Newey and West (1987). In the presence of additional cross-section correlations, one can use the partial sample method together with the Newey-West procedure as proposed by Bai (2009). A formal proof of consistency remains, in this case, to be explored.

## 3.5 SAW with Unobserved Multifactor Effects

If the endogeneity arises from a dynamic model such that one variable on the right hand side is the lag of the explained variable $Y_{it}$, one can follow the existing literature on dynamic panel models and choose one of the commonly used instruments such as $Y_{it-2}$, $Y_{it-3}$, and/or $Y_{it-2} - Y_{it-3}$ ; see, e.g., Anderson and Hsiao (1981), Arellano and Bond (1991), and Kiviet (1995).

In this section, we discuss two possible model extensions: the case in which endogeneity arises from an omitted factor structure; and the case in which endogeneity is due to the presence of simultaneous equations.

### Presence of Multifactor Errors

There is a growing literature on large panel models that allows for the presence of unobserved time-varying individual effects having an approximate factor structure such that

$$e_{it} = \Lambda_i^{'} F_t + \epsilon_{it},$$

where $\Lambda_i$ is a $(d \times 1)$ vector of individual scores (or loadings) $\Lambda_{i1}, \ldots, \Lambda_{id}$ and $F_t$ a $(d \times 1)$ vector of $d$ common factors $F_{1t}, \ldots, F_{dt}$. Note that this extension provides a generalization of panel data models with additive effects and can be very useful in many application areas, especially when the unobserved individual effects are non-static over time; see, e.g., Pesaran (2006), Bai (2009), Ahn et al. (2013), Kneip et al. (2012), and Bada and Kneip (2014).

Leaving the factor structure in the error term and estimating the remaining parameters without considering explicitly the presence of a potential correlation between the observed regressors $X_{1,it}, \ldots, X_{P,it}$ and the unobserved effects $\Lambda_i$ and $F_t$ may lead to an endogeneity problem caused by these omitted model components. The problem with the presence of the factor structure in the error term is that such a structure can not be eliminated by differencing the observed variables or using a simple within-transformation. Owing to the potential correlation between the observable regressors $X_{1,it}, \ldots, X_{P,it}$ and the unobservable heterogeneity effects, we fairly allow for the data generating process of $X_{p,it}$ to have the following form:

$$X_{p,it} = \vartheta_{p,i}^{'} F_t + \Lambda_i^{'} G_{p,t} + a_p \Lambda_i^{'} F_t + \mu_{p,it}, \tag{3.43}$$

where $\vartheta_{p,i}$ is a $(d \times 1)$ vector of unknown individual scores, $G_{p,t}$ is a $(d \times 1)$ vector of unobservable common factors, $a_p$ is a $p$-specific univariate coefficient, and $\mu_{it}$ is an individual specific term that is uncorrelated with $\epsilon_{it}$, $\Lambda_i$, $\vartheta_i$, $F_t$ and $G_t$.

Rearranging (3.43), we can rewrite $X_{p,it}$ as

$$X_{p,it} = \vartheta_{p,i}^{*'} G_{p,t}^{*} + \mu_{p,it}, \tag{3.44}$$

where

$$\vartheta_{p,i}^{*'} = H(a_p \Lambda_i^{'} + \vartheta_{p,i}^{'}, \Lambda_i^{'}), \tag{3.45}$$

and

$$G^*_{p,t} = H^{-1}(F^{'}_t, G^{'}_{p,t})^{'}, \tag{3.46}$$

for some $(2d \times 2d)$ full rank matrix $H$. The role of $H$ is only to ensure orthonormality and identify uniquely (up to a sign change) the elements of the factor structure so that $\sum_{t=1}^{T} G'^{*}_{p,t} G^*_{p,t}/T$ is the identity matrix and $\sum_{i=1}^{n} \vartheta^{*'}_{p,i} \vartheta^*_{p,i}/n$ is a diagonal matrix with ordered diagonal elements.

We can see from (3.43) that a perfect candidate for instrumenting $X_{p,it}$ is $\mu_{p,it}$. Since $\mu_{p,it}$ is unobserved, a feasible instrument can be obtained by

$$Z_{p,it} = X_{p,it} - \hat{\vartheta}^{*'}_{p,i} \hat{G}^*_{p,t}, \tag{3.47}$$

where $\hat{G}^{*'}_{p,t}$ is the $t$-th row element of the $(2d \times 1)$ matrix containing the eigenvectors corresponding to the ordered eigenvalues of the covariance matrix of $X_{p,it}$ and $\hat{\vartheta}^{*'}_{p,i}$ is the projection of $\hat{G}^{*'}_{p,t}$ on $X_{p,it}$. If $d$ is unknown, one can estimate the dimension of $\vartheta^{*'}_{p,i} G^*_{p,t}$ by using an appropriate panel information criterion; see, e.g., Bai and Ng (2002) and Onatski (2010). A crucial assumption about the form of dependency in $\mu_{p,it}$ is that, for all $T$ and $n$, and every $i \leq n$ and $t \leq T$,

1. $\sum_{s=1}^{T} |E(\mu_{p,it}\mu_{p,is})| \leq M$ and
2. $\sum_{k=1}^{n} |E(\mu_{p,it}\mu_{p,kt})| \leq M$.

Bai (2003) proves the consistency of the principal component estimator when additionally $\frac{1}{T}\sum_{t=1}^{T} G^{*'}_{p,t} G^*_{p,t} \xrightarrow{p} \Sigma_{G^*_p}$ for some $(2d \times 2d)$ positive definite matrix $\Sigma_{G^*_p}$, $||\vartheta^*_{p,i}|| \leq M$ for all $i$ and $p$, and $||\frac{1}{n}\sum_{i=1}^{n} \vartheta^{*'}_{p,i}\vartheta^*_{p,i} - \Sigma_{\vartheta^*_p}|| \longrightarrow 0$, as $n \to \infty$ for some $(2d \times 2d)$ positive definite matrix $\Sigma_{\vartheta^*_p}$.

By instrumenting $X_{p,it}$ with $Z_{p,it}$ in (3.47), we can estimate consistently the jumping slope parameters as before. A formal proof remains, of course, to be explored.

### Two-Step SAW for Jumping Reverse Causality

Besides the issues of omitted variables and dynamic dependent variables, another important source of endogeneity is the phenomenon of reverse causality. This occurs when the data is generated by a system of simultaneous equations so that changes in the right-hand side of the model cause changes in the left-hand side variable and changes in the left-hand side variable cause simultaneous changes in the right-hand side variable. A famous example of simultaneous equation models is the partial equilibrium model of supply and demand in a market with perfect competition.

Consider the following two-way simultaneous equation model:

$$Y_{it} = \mu + \sum_{p=1}^{P} X_{p,it}\beta_{t,p} + \alpha_i + \theta_t + e_{it}, \tag{3.48}$$

and

$$X_{q,it} = b_t Y_{it} + \sum_{p\in\{1,\ldots,P\}\setminus\{q\}} X_{p,it} d_{t,p} + v + u_i + \vartheta_t + \nu_{it}, \tag{3.49}$$

for some a $q \in \{1,\ldots,P\}$, where $b_t \neq 1/\beta_{t,q}$, and the parameters $v, u_i$, and $\vartheta_t$ are unknown parameters.

Neglecting the structural form of $X_{q,it}$ in Equation (3.49) and estimating the regression function (3.48) without instrumenting this variable results in an inconsistent estimation since $X_{q,it}$ and $e_{it}$ are correlated (due to the presence of $Y_{it}$ in Equation (3.49)). A natural way to overcome this type of endogeneity problem is to use the fitted variable obtained from Equation (3.49) as an instrument after replacing $Y_{it}$ with its expression in (3.48). In the literature on IV methods, this procedure is termed two-stage least squares method. But note that our method involves an additional complication related to the time-changing character of $\beta_{t,q}$ and the presence of the unobservable heterogeneity effects. Indeed, inserting (3.48) in (3.49) and rearranging it leads to a panel model with time-varying unobservable individual effects:

$$X_{q,it} = \sum_{p\in\{1,\ldots,P\}\setminus\{q\}} X_{p,it} d^*_{t,p} + \vartheta^*_{1t} + u_i\vartheta^*_{2t} + \alpha_i\vartheta^*_{3t} + \varepsilon_{it}, \tag{3.50}$$

where

$$\begin{aligned}
d^*_{t,p} &= b_t\beta_{t,p} + d_{t,p},\\
\vartheta^*_{1t} &= \tfrac{b_t\mu + b_t\theta_t + \vartheta_t + v}{1 - b_t\beta_{t,p}},\\
\vartheta^*_{2t} &= \tfrac{1}{1 - b_t\beta_{t,p}},\\
\vartheta^*_{3t} &= \tfrac{b_t\mu + b_t\theta_t + \vartheta_t + v}{1 - b_t\beta_{t,p}}, \text{ and}\\
\varepsilon_{it} &= b_t e_{it} + \varepsilon_{it}.
\end{aligned}$$

Note that the regression model in (3.50) can be considered a spacial case of the model with multifactor errors discussed above. A potential instrument for $X_{q,it}$ in (3.48) is

$$Z_{q,it} = \sum_{p\in\{1,\ldots,P\}\setminus\{q\}} X_{p,it}\hat{d}^*_{t,p} + \hat{\vartheta}_i'\hat{G}_t, \tag{3.51}$$

where $\hat{d}^*_{t,p}$ and $\hat{\vartheta}_i'\hat{G}_t$ are the estimators of $b_t$ and $\vartheta_i' G_t = \vartheta^*_{1t} + u_i\vartheta^*_{2t} + \alpha_i\vartheta^*_{3t}$, respectively, which can be obtained from (3.50) by using the instruments proposed above to control for the omitted factor structure $\vartheta^*_{1t} + u_i\vartheta^*_{2t} + \alpha_i\vartheta^*_{3t}$.

## 3.6 Monte Carlo Simulations

In this section, we examine, through Monte Carlo simulations, the finite sample performance of our method. Our data generating-processes are based on the following panel data model:

$$Y_{it} = X_{it}\beta_t + \alpha_i + \sqrt{\theta_{it}}e_{it} \text{ for } i \in \{1, \dots, n\} \text{ and } t \in \{1, \dots T\},$$

where

$$\beta_t = \begin{cases} \beta_{\tau_1} & \text{for} \quad t \in \{1, \dots, \tau_1\}, \\ \vdots & \\ \beta_{\tau_S+1} & \text{for} \quad t \in \{\tau_S + 1, \dots, T\}, \end{cases} \tag{3.52}$$

with

$$\beta_{\tau_j} = \frac{2}{3} \cdot (-1)^j \text{ and } \tau_j = \left\lfloor \frac{j}{S+1}(T-1) \right\rfloor, \text{ for } j = 1, \dots, S+1.$$

We examine the situations where the number of jumps is $S = 0, 1, 2, 3$. In the no-jump case ($S = 0$), we compare the performance of our method with the performance of the classical Least Squares Dummy Variable Method (LSDV), the Generalized Least Squares Method for random effect models (GLS), the Iterated Least Squares Method (ILS) of Bai (2009), and the semi-parametric method (KSS) of Kneip et al. (2012). Our thresholding parameter is calculated with $\kappa = 1 - \log(\log(nT))/\log(nT)$. To see how the properties of the estimators vary with $n$ and $T$, we consider 12 different combinations with the sizes $n = 30, 60, 120, 300$ and $T = 2^{L-1} + 1$, for $L = 6, 7, 8$, i.e., $T = 33, 65, 129$. We consider the cases of dyadic (e.g., when $S = 1$ and $\tau_1 = (T-1)/2$) and non-dyadic jump locations (when $S = 2, 3$) as well as models with exogenous and endogenous regressors. In total, our experiments are based on the results of seven different DGP-configurations:

**DGP1** (*exogeneity, and i.i.d. errors*): the dependent variable $X_{it}$ is uncorrelated with $e_{it}$ and generated by

$$X_{it} = 0.5\alpha_i + \xi_{it}, \tag{3.53}$$

with $\xi_{it}, \alpha_i, e_{it} \sim N(0, 1)$ and $\theta_{it} = 1$ for all $i$ and $t$.

**DGP2** (*exogeneity, and cross-section heteroskedasticity*): the DGP of the exogenous regressor $X_{it}$ is of form (3.53); cross-sectionally heteroskedastic errors such that $e_{it} \sim N(0, 1)$ with $\theta_{it} = \theta_i^* \sim U(1, 4)$ for all $t$

**DGP3** (*exogeneity, and heteroskedasticity in time and cross-section dimension*): the DGP of the exogenous regressor $X_{it}$ is of form (3.53); heteroskedastic errors in time and cross-section dimension such that $e_{it} \sim N(0,1)$ and $\theta_{it} \sim U(1,4)$.

**DGP4** (*exogeneity, and serial correlation with cross-section heteroskedasticity*): the DGP of the exogenous regressor $X_{it}$ is of form (3.53); homoscedasticity and autocorrelation in the errors such that

$$e_{it} = \rho_i e_{it-1} + \zeta_{it}, \tag{3.54}$$

with $\rho_i \sim U(0,.5)$, $\zeta_{it} \sim N(0,.5)$, and $\theta_{it} = 1$ for all $i$ and $t$.

**DGP5** (*endogeneity due to a hidden factor structure*): $X_{it}$ and $e_{it}$ are correlated through the presence of a hidden factor structure:

$$\begin{aligned} e_{it} &= \lambda_i f_t + \epsilon_{it} \text{ and} \\ X_{it} &= 0.3\alpha_i + 0.3\nu_t + 0.3\lambda_i f_t + \mu_{it}, \end{aligned} \tag{3.55}$$

with $\lambda_i, f_t \sim N(0,.5)$, $\theta_{it} = 1$ for all $i$ and $t$, and $\alpha_i \sim N(0,1)$.

**DGP6** (*endogeneity due to a hidden approximate factor structure*): $X_{it}$ and $e_{it}$ are correlated as in DGP5, but

$$\begin{aligned} \epsilon_{it} &= \rho_{e,i}\epsilon_{i,t-1} + \zeta_{e,it}, \\ \mu_{it} &= \rho_{\mu,i}\mu_{i,t-1} + \zeta_{\mu,it}, \end{aligned} \tag{3.56}$$

with $\zeta_{e,it}, \zeta_{\mu,it} \sim N(0,.5)$, $\rho_{e,i}, \rho_{\mu,i} \sim U(0,.5)$, $\theta_{it} = 1$ for all $i$ and $t$, and $\alpha_i \sim N(0,1)$.

**DGP7** (*no-jumps, endogeneity, and hidden approximate factor structure*): the slope parameter does not suffer from structural breaks so that $\beta_t = 2$ for all $t$; the regressor and the error are correlated through the presence of an approximate factor structure as in DGP6.

Tables 3.1 -3.4 report the estimation results obtained by averaging the results of 1000 replications. The third, sixth, and ninth columns in Tables 3.1-3.3 report the averages of the estimated number of jumps $\tilde{S}$ detected by (3.37) for $S = 1, 2$, and 3, respectively. The MISE of our estimator is calculated by $\frac{1}{1000}\sum_{r=1}^{1000}\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{\beta}_t^r - \beta_t)^2\right)$, where $\hat{\beta}_t^r$ is the pointwise post-SAW estimate of $\beta_t$ obtained in replication $r$. The fourth, seventh, and tenth columns in Tables 3.1 -3.3 give, on average, the values of a criterion (hereafter called MDCJ) that describes the mean distance between the true jump locations and

the closest post-SAW detected jumps. The MDCJ criterion is calculated as follows:

$$\text{MDCJ} = \frac{1}{S}\sum_{j=1}^{S} \min_{l\in\{1,\dots,\tilde{S}\}} |\tau_j - \tilde{\tau}_l|.$$

We use the `R`-package `phtt` to calculate LSDV, ILS, and KSS and `plm` to calculatd GLS . The corresponding MSEs of LSDV, GLS, ILS, and KSS are obtained by $\frac{1}{1000}\sum_{r=1}^{1000}\big(\hat{\beta}^r_{(M)} - \beta\big)^2$, where $\hat{\beta}^r_{(M)}$ is the estimate of $\beta = \beta_1 = \cdots = \beta_T$ obtained in replication $r$ by using method $M$ = LSDV, ILS, and KSS. The results are reported in Table 3.4.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | DGP1 | | | | | |
| Nbr. of jumps $S$: | | | **1** | | | **2** | | | **3** | |
| $n$ | $T$ | $\tilde{S}$ | MDCJ | MISE | $\tilde{S}$ | MDCJ | MISE | $\tilde{S}$ | MDCJ | MISE |
| 30 | 33 | 1.0 | 0.000 | 0.002 | 2.0 | 0.000 | 0.005 | 3.0 | 0.000 | 0.008 |
| 60 | 33 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.004 | 3.0 | 0.000 | 0.008 |
| 120 | 33 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.004 | 3.0 | 0.000 | 0.007 |
| 300 | 33 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.003 | 3.0 | 0.000 | 0.007 |
| 30 | 65 | 1.0 | 0.000 | 0.001 | 2.1 | 0.000 | 0.001 | 3.1 | 0.000 | 0.002 |
| 60 | 65 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 120 | 65 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 300 | 65 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 30 | 129 | 1.0 | 0.000 | 0.000 | 2.1 | 0.000 | 0.000 | 3.1 | 0.000 | 0.002 |
| 60 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |
| 120 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.000 |
| 300 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.000 |
| | | | | | DGP2 | | | | | |
| 30 | 33 | 0.9 | 3.100 | 0.118 | 1.5 | 2.731 | 0.181 | 2.2 | 2.349 | 0.193 |
| 60 | 33 | 1.0 | 0.000 | 0.003 | 2.0 | 0.111 | 0.011 | 3.0 | 0.053 | 0.013 |
| 120 | 33 | 1.0 | 0.000 | 0.002 | 2.0 | 0.000 | 0.005 | 3.0 | 0.000 | 0.008 |
| 300 | 33 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.003 | 3.0 | 0.000 | 0.008 |
| 30 | 65 | 0.8 | 9.200 | 0.173 | 1.5 | 5.470 | 0.191 | 2.4 | 4.160 | 0.180 |
| 60 | 65 | 1.0 | 0.000 | 0.001 | 1.8 | 0.665 | 0.021 | 2.9 | 0.531 | 0.030 |
| 120 | 65 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 300 | 65 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 30 | 129 | 0.9 | 13.81 | 0.124 | 1.4 | 16.40 | 0.261 | 2.0 | 12.31 | 0.231 |
| 60 | 129 | 1.0 | 2.519 | 0.021 | 2.0 | 0.859 | 0.017 | 2.9 | 0.851 | 0.022 |
| 120 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |
| 300 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |

TABLE 3.1: Simulation results of the Monte Carlo experiments for DGP1-DGP2. The entries are the averages of 1000 replications.

In our examined data configurations, the MISE of the post-SAW estimator and the average of the estimated number of jumps behave properly as both $n$ and $T$ get large as well as when $T$ is fixed and only $n$ gets large. The method performs perfectly in the benchmark case where idiosyncratic errors are independent and identically distributed even when $n$ and $T$ are relatively small (e.g., the combinations where $n = 30$ and/or $T = 33$ in the first part of Table 3.1). In most of the examined cases, where heteroskedasticity in the cross-section and time dimension and/or week serial correlations exist, the method behaves very well, in particular when $n$ is large (see results of DGP3-DGP4 in Tables 3.1 and 3.2). The quality of the estimator seems to be independent of the number and the location of the jumps (i.e., dyadic, for $S = 1$, and non-dyadic for $S = 2, 3$). Not surprisingly, the jump selection method performs poorly when $n$ is fixed and only $T$ is large. In such a case, the threshold under-estimates the true number of jumps and the MDCJ increases with $T$. This effect vanishes properly as $n$ gets large.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | DGP3 | | | | | |
| Nbr. of jumps $S$: | | | **1** | | | **2** | | | **3** | |
| $n$ | $T$ | $\tilde{S}$ | MDCJ | MISE | $\tilde{S}$ | MDCJ | MISE | $\tilde{S}$ | MDCJ | MISE |
| 30 | 33 | 0.9 | 3.000 | 0.117 | 1.5 | 2.730 | 0.180 | 2.3 | 2.347 | 0.190 |
| 60 | 33 | 1.0 | 0.000 | 0.003 | 2.0 | 0.110 | 0.014 | 3.0 | 0.053 | 0.016 |
| 120 | 33 | 1.0 | 0.000 | 0.002 | 2.0 | 0.000 | 0.005 | 3.0 | 0.000 | 0.008 |
| 300 | 33 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.003 | 3.0 | 0.000 | 0.008 |
| 30 | 65 | 0.7 | 9.300 | 0.170 | 1.5 | 5.470 | 0.191 | 2.4 | 4.160 | 0.181 |
| 60 | 65 | 1.0 | 0.000 | 0.001 | 1.9 | 0.660 | 0.025 | 2.9 | 0.533 | 0.031 |
| 120 | 65 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 300 | 65 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 30 | 129 | 0.9 | 13.80 | 0.124 | 1.3 | 16.41 | 0.260 | 2.0 | 12.37 | 0.235 |
| 60 | 129 | 1.0 | 2.520 | 0.023 | 2.0 | 0.860 | 0.016 | 2.9 | 0.853 | 0.023 |
| 120 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |
| 300 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |
| | | | | | DGP4 | | | | | |
| 30 | 33 | 0.2 | 12.60 | 0.477 | 0.5 | 8.250 | 0.431 | 0.6 | 8.673 | 0.499 |
| 60 | 33 | 0.4 | 9.000 | 0.340 | 1.0 | 5.280 | 0.309 | 1.4 | 5.280 | 0.360 |
| 120 | 33 | 0.9 | 1.500 | 0.059 | 1.8 | 1.210 | 0.087 | 2.6 | 1.013 | 0.116 |
| 300 | 33 | 1.0 | 0.000 | 0.002 | 2.0 | 0.000 | 0.004 | 3.0 | 0.000 | 0.009 |
| 30 | 65 | 0.3 | 23.22 | 0.426 | 0.4 | 17.49 | 0.453 | 0.5 | 17.72 | 0.488 |
| 60 | 65 | 0.6 | 13.02 | 0.238 | 0.9 | 11.49 | 0.318 | 1.2 | 11.62 | 0.403 |
| 120 | 65 | 0.9 | 4.340 | 0.080 | 1.8 | 2.190 | 0.073 | 2.6 | 2.560 | 0.110 |
| 300 | 65 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.003 |
| 30 | 129 | 0.1 | 55.44 | 0.496 | 0.3 | 37.41 | 0.472 | 0.4 | 36.05 | 0.516 |
| 60 | 129 | 0.5 | 34.02 | 0.305 | 0.8 | 26.66 | 0.377 | 1.0 | 26.88 | 0.427 |
| 120 | 129 | 0.8 | 10.08 | 0.091 | 1.7 | 6.880 | 0.116 | 2.5 | 6.187 | 0.146 |
| 300 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.001 |

TABLE 3.2: Simulation results of the Monte Carlo experiments for DGP3-DGP4. The entries are the averages of 1000 replications.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | DGP5 | | | | | |
| Nbr. of jumps $S$: | | | **1** | | | **2** | | | **3** | |
| $n$ | $T$ | $\tilde{S}$ | MDCJ | MISE | $\tilde{S}$ | MDCJ | MISE | $\tilde{S}$ | MDCJ | MISE |
| 30 | 33 | 0.8 | 4.100 | 0.187 | 1.3 | 3.230 | 0.250 | 2.1 | 3.367 | 0.291 |
| 60 | 33 | 1.0 | 0.000 | 0.004 | 2.0 | 0.119 | 0.017 | 3.0 | 0.058 | 0.020 |
| 120 | 33 | 1.0 | 0.000 | 0.003 | 2.0 | 0.000 | 0.007 | 3.0 | 0.000 | 0.010 |
| 300 | 33 | 1.0 | 0.000 | 0.002 | 2.0 | 0.000 | 0.002 | 3.0 | 0.000 | 0.009 |
| 30 | 65 | 0.7 | 9.700 | 0.210 | 1.4 | 5.976 | 0.210 | 2.4 | 4.860 | 0.211 |
| 60 | 65 | 1.0 | 0.000 | 0.002 | 1.9 | 0.690 | 0.031 | 2.9 | 0.539 | 0.038 |
| 120 | 65 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 300 | 65 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.002 |
| 30 | 129 | 0.8 | 19.80 | 0.224 | 1.2 | 21.41 | 0.361 | 2.0 | 19.37 | 0.315 |
| 60 | 129 | 1.0 | 2.611 | 0.053 | 2.0 | 0.952 | 0.017 | 2.8 | 1.153 | 0.033 |
| 120 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |
| 300 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |
| | | | | | DGP6 | | | | | |
| 30 | 33 | 0.2 | 11.66 | 0.448 | 0.5 | 8.250 | 0.427 | 0.6 | 8.320 | 0.498 |
| 60 | 33 | 0.5 | 7.200 | 0.272 | 1.0 | 5.390 | 0.306 | 1.5 | 4.907 | 0.370 |
| 120 | 33 | 0.9 | 0.900 | 0.036 | 1.8 | 0.990 | 0.074 | 2.6 | 1.013 | 0.114 |
| 300 | 33 | 1.0 | 0.000 | 0.002 | 2.0 | 0.000 | 0.005 | 3.0 | 0.000 | 0.009 |
| 30 | 65 | 0.3 | 22.94 | 0.419 | 0.3 | 18.77 | 0.484 | 0.6 | 16.93 | 0.480 |
| 60 | 65 | 0.5 | 14.88 | 0.272 | 1.1 | 10.44 | 0.304 | 1.1 | 13.76 | 0.401 |
| 120 | 65 | 0.8 | 5.580 | 0.102 | 1.8 | 1.980 | 0.070 | 2.6 | 2.667 | 0.116 |
| 300 | 65 | 1.0 | 0.000 | 0.001 | 2.0 | 0.000 | 0.001 | 3.0 | 0.000 | 0.003 |
| 30 | 129 | 0.1 | 54.74 | 0.493 | 0.1 | 40.53 | 0.487 | 0.3 | 38.61 | 0.523 |
| 60 | 129 | 0.3 | 42.84 | 0.383 | 0.7 | 27.52 | 0.372 | 1.0 | 26.45 | 0.432 |
| 120 | 129 | 0.8 | 11.34 | 0.102 | 1.8 | 5.160 | 0.091 | 2.4 | 7.680 | 0.157 |
| 300 | 129 | 1.0 | 0.000 | 0.000 | 2.0 | 0.000 | 0.000 | 3.0 | 0.000 | 0.001 |

TABLE 3.3: Simulation results of the Monte Carlo experiments for DGP5-DGP6. The entries are the averages of 1000 replications.

Table 3.3 reports the results of our experiments when the regressors are affected by an omitted factor structure in the error term. The proposed two-step SAW procedure seems to perform very well even when heteroskedasticity in the cross-section and time dimension and/or week serial correlations are present.

The goal of examining DGP7 is to test whether SAW is also able to detect the no-jump case. The answer that can be deciphered from Table 3.4 is: Yes. Our method is slightly inferior in terms of MSE than ILS but better than LSDV, GLS, and KSS. Because LSDV and GLS neglect the presence of the factor structure in the model and KSS is only appropriate for factors that possess smooth patterns over time, the MSEs of these three estimators are affected by a small bias that seems to persist even when $n$ and $T$ get large.

| DGP7 ($S = 0$) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Method | | post-SAW | | LSDV | GLS | ILS | KSS |
| $n$ | $T$ | $\widetilde{S}$ | MISE | MSE | MSE | MSE | MSE |
| 30 | 33 | 0.0000 | 0.0062 | 0.0105 | 0.0148 | 0.0007 | 0.0104 |
| 60 | 33 | 0.0000 | 0.0041 | 0.0105 | 0.0142 | 0.0006 | 0.0101 |
| 120 | 33 | 0.0000 | 0.0012 | 0.0090 | 0.0125 | 0.0002 | 0.0085 |
| 300 | 33 | 0.0000 | 0.0004 | 0.0093 | 0.0128 | 0.0001 | 0.0090 |
| 30 | 65 | 0.0000 | 0.0039 | 0.0105 | 0.0135 | 0.0004 | 0.0099 |
| 60 | 65 | 0.0000 | 0.0015 | 0.0102 | 0.0130 | 0.0002 | 0.0101 |
| 120 | 65 | 0.0000 | 0.0005 | 0.0099 | 0.0126 | 0.0001 | 0.0098 |
| 300 | 65 | 0.0000 | 0.0001 | 0.0108 | 0.0137 | 0.0000 | 0.0108 |
| 30 | 129 | 0.0000 | 0.0015 | 0.0101 | 0.0129 | 0.0002 | 0.0102 |
| 60 | 129 | 0.0000 | 0.0008 | 0.0103 | 0.0127 | 0.0001 | 0.0104 |
| 120 | 129 | 0.0000 | 0.0002 | 0.0101 | 0.0125 | 0.0000 | 0.0101 |
| 300 | 129 | 0.0000 | 0.0000 | 0.0090 | 0.0112 | 0.0000 | 0.0090 |

TABLE 3.4: Simulation results of the Monte Carlo experiments for DGP7. The entries are the averages of 1000 replications.

The Monte Carlo experiments show that, in many configurations of the data, our method performs very well even when the idiosyncratic errors are weakly affected by serial-autocorrelation and/or heteroskedasticity and this independently of the number and locations of the jumps.

## 3.7 Application: Algorithmic Trading and Market Quality

An issue of increasing debate, both academically and politically, is the impact of algorithmic trading (AT) on standard measures of market quality such as liquidity and volatility. Proponents, including many of the exchanges themselves, argue that AT provides added liquidity to markets and is beneficial to investors. Opponents instead caution that algorithmic trading increases an investor's perception that an algorithmic

trading partner possesses an informational advantage. Furthermore, incidents such as the ”flash crash”, although circumstantial in nature, do nothing to alleviate these fears.

Recent work examining the effects of AT on market quality have generally found its presence to be beneficial in the sense that standard measures of liquidity such as bid-ask spreads and price-impact are reduced as a consequence of the increase in AT. For example Hendershott et al. (2011) find that, with the exception of the smallest quintile of NYSE stocks, AT almost universally reduces quoted and effective spreads in the remaining quintiles. Hasbrouck and Saar (2013) find similarly compelling evidence using a measure of AT constructed from order level data. A drawback of both approaches and more specifically of the standard panel regression approach is that estimates of the marginal effect of increasing are necessarily averaged over all possible states of the market. This is problematic from an asset pricing perspective.

Of particular importance to the concept of liquidity is the timing of its provision. The merits of added liquidity during stable market periods at the expense of its draw back during periods of higher uncertainty are ambiguous without a valid welfare analysis and can potentially leave investors worse off. The issue of timing is particularly important for empirical work examining the effects of AT on market quality. Samples are often constrained in size due to limitations on the availability of data and computational concerns. As noted by Hendershott et al. (2011), it may be because samples often used do not cover large enough periods of market turbulence to detect possible negative effects. Additionally, standard subsample analysis requires the econometrician to diagnose market conditions as well their start and end dates, in effect imposing their own prior beliefs on the factors that might cause variation in the marginal effects. Because of this we propose the use of our estimator to automatically detect jumps in slope parameters. Indeed, our methodology alleviates concerns about ad-hoc subsample selection. Furthermore, we believe analysis of periods where the effects vary may provide valuable insight for future studies (both theoretical and empirical) and policy recommendations regarding the regulation of trading in financial markets.

### 3.7.1 Liquidity and Asset Pricing

Before discussing the effects of liquidity on asset pricing, we first examine conventional tests that assume constant parameters. In this simple example, we regress a measure of market quality on an AT proxy for an individual stock using the following model:

$$MQ_t = \alpha + AT_t\overline{\beta} + e_t, \tag{3.57}$$

where the time index $t \in \{1, \ldots, T\}$. If the slope parameter is time varying then $\overline{\beta}$ in (3.57) presents only the time average of the true parameter, say $\beta_t$. In this case the conventional estimator of $\overline{\beta}$ is consistent only under the assumption $\sum_{t=1}^{T} AT_t^2(\beta_t - \overline{\beta})/T \xrightarrow{p} 0$, as $T$ get large. Even when such a requirement is satisfied, the *average* effect is not the correct measure to consider when the question is whether AT is beneficial to market quality.

A general result in asset pricing that is a consequence of no arbitrage is that there exists a strictly positive stochastic discount factor (SDF) such that,

$$1 = E_t(M_{t+1}R_{t+1}),$$

where $M_{t+1}$ is the SDF and $R_{t+1}$ is the return on a security. This expression can be expanded and rewritten as,

$$E_t(R_{t+1}) = \frac{1}{E_t(M_{t+1})} - \frac{1}{E_t(M_{t+1})} cov_t(M_{t+1}, R_{t+1}).$$

Expected security returns (i.e., its risk premium) are a function of covariance with the SDF. While the form of the SDF depends on the asset pricing model one is considering, it can in general be thought of as the ratio of marginal value of wealth between time $t+1$ and $t$. Therefore, holding the expectation of $M_{t+1}$ constant, if a security pays off more in states where the marginal value of wealth is relatively low and less in states where the marginal value of wealth is relatively high ($cov_t(M_{t+1}, R_{t+1}) < 0$) then that security earns a premium for this undesirable property. In light of the above analysis, if a security's returns contain a stochastic liquidity component then its covariance with the SDF can have a substantial impact on expected returns.

The model of Acharya and Pedersen (2005) is particularly relevant as it exemplifies the many avenues through which time varying liquidity can affect expected returns. Using an overlapping generations model they decompose conditional security returns into five components: one related to the expected level of illiquidity and four others related to covariance terms between market return, market illiquidity, security returns and security illiquidity. They show that portfolio returns are increasing in the covariance between portfolio illiquidity and market illiquidity and decreasing in the covariance between security illiquidity and the market return. A consequence of this is that if AT intensifies these dynamics in the liquidity for a particular security then the effect will be to increase the risk premium associated with that security. Increased risk premiums represent higher costs of capital for firms, thus increased AT can potentially decrease firm investment (relative to a market with no AT) through its effects on liquidity dynamics.

### 3.7.2 Data

Our sample consists of a balanced panel of stocks whose primary exchange is the New York Stock Exchange (NYSE) and covers the calendar period $2003 - 2008$. The choice of this sample period reflects our desire to include both relatively stable and turbulent market regimes. We are limited in our choice of sample periods by the fact that *AT* is a recent phenomenon and that our estimation procedure requires a balanced panel. In this six-year period we consider results in a total of 378 firms.

To build measures of market quality, we use the NYSE Trade and Quotation Database (TAQ) provided by Wharton Research Data Services (WRDS) to collect intra-day data on securities. We aggregate intra-day up to a monthly level to construct our sample, which consists of 65 months for each firm. We merge the TAQ data with information on price and shares outstanding from the Center for Research in Security Prices (CRSP). A discussion of our algorithmic trading proxy and measures of market quality follows.

#### 3.7.2.1 The Algorithmic Trading Proxy

Our AT proxy is motivated by Hendershott et al. (2011) and Boehmer et al. (2012), who note that AT is generally associated with an increase in order activity at smaller dollar volumes. Thus the proxy we consider is the negative of dollar volume (in hundreds of dollars, $\text{Vol}_{it}$) over time period $t$ divided by total order activity over time period $t$. We define order activity as the sum of trades ($\text{Tr}_{it}$) and updates to the best prevailing bid and offer ($q_{it}$) on the securities' primary exchange:

$$\text{AT}_{it} = -\frac{\text{Vol}_{it}}{\text{Tr}_{it} + q_{it}}.$$

An increase in $\text{AT}_{it}$ represents a decrease in the average volume per instance of order activity and represents an increase in the AT in the particular security. For example, an increase of 1 unit of $\text{AT}_{it}$ represents a decrease of \$100 of trading volume associated with each instance of order activity (trade or quote update).

Our proxy, like that in Boehmer et al. (2012), differs from the proxy in Hendershott et al. (2011) since they have access to the full order book of market makers whereas we only have access to the trades and the best prevailing bid and offers of market makers through TAQ. We appeal to the same argument as Boehmer et al. (2012) in that many AT strategies are generally executed at the best bid and offer rather than behind it. Therefore, we feel our proxy is in general representative of the full order book.

#### 3.7.2.2 Market Quality Measures

We consider several common measures of market quality to assess the impact of AT on markets for individual securities.

**Proportional Quoted Spread**

The proportional quoted spread ($\text{PQS}_{it}$) measures the quoted cost as a percentage of price (Bid-Offer midpoint) of executing a trade in security $i$ and is defined as,

$$\text{PQS}_{it} = 100 \left( \frac{\text{Ofr}_{it} - \text{Bid}_{it}}{0.5(\text{Ofr}_{it} + \text{Bid}_{it})} \right).$$

We multiply by 100 in order to place this metric in terms of percentage points. We aggregate this metric to a monthly quality by computing a share volume-weighted average over the course of each month. An increase in $\text{PQS}_{it}$ represents a decrease in the amount of liquidity in the market for security $i$ due to increased execution costs.

**Proportional Effective Spread**

The proportional effective spread ($\text{PES}_{it}$) is quite similar to ($\text{PQS}_{it}$) but accommodates potentially hidden liquidity or stale quotes by evaluating the actual execution costs of a trade. It is defined as,

$$\text{PES}_{it} = 100 \left( \frac{|P_{it} - M_{it}|}{M_{it}} \right),$$

where $P_{it}$ is the price paid for security $i$ at time $t$ and $M_{it}$ is the midpoint of the prevailing bid and ask quotes for security $i$ at time $t$. Thus, $\text{PES}_{it}$ is the actual execution cost associated with every trade. We again aggregate this measure up to a monthly quantity in the same way as we do for quoted spreads. Like $\text{PQS}_{it}$, $\text{PES}_{it}$ is also in terms of percentage points. An increase in $\text{PES}_{it}$ represents a decrease in the amount of liquidity in the market for security $i$ due to increased execution costs.

**Measures of Volatility**

We also consider two different measures of price volatility in security $i$ over time period $t$. The first is the daily high-low price range given by,

$$\text{H-L}_{it} = 100 \left( \frac{\max_{\tau \in t}(P_{it}) - \min_{\tau \in t}(P_{it})}{P_{it}} \right),$$

which represents the extreme price disparity over the course of a trading day. We also consider the realized variance of returns over each day computed using log returns over 5-minute intervals:

$$RV_{it} = 100 \left( \sum_{\tau \in t} r_{i\tau}^2 \right).$$

Realized variance is a nonparametric estimator of the integrated variance over the course of a trading day (see, for example, Andersen et al. (2003)). We aggregate both measures up to a monthly level by averaging over the entire month. We additionally multiply both variables by 100, thus $H - L_{it}$ is the price range as a percentage of the daily closing price and $RV_{it}$ is an estimate of the integrated variance of log returns in percentages. Both measures represent a measure of the price dispersion over the course of the trading month.

**Additional Control Variables**

While we attempt to determine the effect our AT proxy has on measures of market quality we include in all our regressions a vector of control variables to isolate the effects of AT independent of the state of the market. We lag the control variables by one month so they represent the state of the market at the beginning of the trading month in question. The control variables are: (1) Share Turnover ($\mathrm{ST}_{it}$), which is the number of shares traded over the course of a day in a particular stock relative to the total amount of shares outstanding; (2) Inverse price, which represents transaction costs due to the fact that the minimum tick size is 1 cent; (3) Log of market value of equity to accommodate effects associated with micro-cap securities; (4) Daily price range to accommodate any effects from large price swings in the previous month.

To avoid adding lagged dependent variables in the model, for regressions where the daily price range is the dependent variable we replace it in the vector of controls with the previous month's realized variance.

We additionally include security and time period fixed effects to proxy for any time period or security related effects not captured by our included variables. We consider panel regressions of the form

$$\mathrm{MQ}_{it} = \alpha_i + \theta_t + \mathrm{AT}_{it}\beta_t + W_{it}^{'}\vartheta + e_{it}, \tag{3.58}$$

where $\mathrm{MQ}_{it}$ is the market quality measure under consideration, $\alpha_i$ and $\theta_t$ are the security and time period fixed effects, $W$ is a vector of the lagged control variables listed above, and $e_{it}$ is the innovation to $\mathrm{MQ}_{it}$, which we assume to be dependent of the fixed effects and the control variables. The time subscript on the parameter beta allows for a possibly

time varying effect of AT on market quality. Thus we are able to test the null hypothesis of a constant effect versus the alternative of a time jumping effect. We are also able to measure the magnitude and direction of any possible change.

**Potential Endogeneity Issue**

Absent a theoretical model of AT, an issue on which the literature is still somewhat agnostic, it is uncertain whether AT strategies attempt to time shocks to market quality. This creates the potential problem of endogeneity with our AT proxy. That is, when estimating the regression equation (3.58), $E(\mathrm{AT}_{it}e_{it}) \neq 0$.

To overcome this potential issue we use the approach of Hasbrouck and Saar (2013) (albeit with different variables) and choose as an instrument the average value of algorithmic trading over all other firms not in the same industry as firm $i$. To this end, we define industry groups using 4-digit SIC codes and define these new variables $\mathrm{AT}_{-IND,it}$. The use of this IV assumes that there is some commonality in the level of AT across all stocks that is sufficient to pick up some exogenous variation. It further rules out trading strategies by AT across firms in different industry groups. Lacking complete knowledge of the algorithms used by AT firms we view this assumption to be reasonable.

To estimate the model we use a two-stage approach and first fit the regression model,

$$\mathrm{AT}_{it} = a_i + g_t + b\mathrm{AT}_{-IND,it} + dW_{it} + \epsilon_{it} \tag{3.59}$$

to obtain an instrument, $Z_{it}$, for $\mathrm{AT}_{it}$ given by the fitted values from (3.59), i.e., $Z_{it} := \hat{\mathrm{AT}}_{it} = \hat{a}_i + \hat{g}_t + \hat{b}\mathrm{AT}_{-IND,it} + \hat{d}W_{it}$, where $\hat{a}_i$, $\hat{g}_t$, $\hat{b}$, and $\hat{d}$ are the conventional estimates of $a_i$, $g_t$, $b$, and $d$.

We then carry out the second stage regression using equation (3.58) as described in Section 3.3. For comparison purposes, we additionally apply the conventional panel data model assuming constant slope parameter, i.e., $\beta_1 = \beta_2 = \cdots = \beta_T$.

### 3.7.3 Results

Table 3.5 presents the results from a baseline model that assumes the slope parameters are constant over time.[1] These results are largely consistent with previous studies that find a positive (in terms of welfare) *average* relationship between AT and measures of market quality over the time period considered. The coefficient estimates on the

[1]For the purpose of readability we divide the AT variable by 100 to reduce trailing zeros after the decimal.

AT proxy are negative and significant for all four measures of market quality that we consider. That is, increases in AT generally reduce both of the spread measures and both of the variance measures we consider.

| Dependent | Variable | $\hat{\text{AT}}_{it}$ | $\ln(\text{ME})_{it-1}$ | $\text{T/O}_{it-1}$ | $1/\text{P}_{it-1}$ | $\text{H-L}_{it-1}$ | $\text{RV}_{it-1}$ |
|---|---|---|---|---|---|---|---|
| $\text{PQS}_{it}$ | Coef. | -0.013 | -0.003 | 0.027 | 0.619 | 0.002 | |
| | $t$-value | -3.61 | -2.65 | 0.73 | 15.38 | 5.67 | |
| $\text{PES}_{it}$ | Coef. | -0.006 | -0.001 | -0.077 | 0.517 | 0.004 | |
| | $t$-value | -3.62 | -1.83 | -3.42 | 18.51 | 16.96 | |
| $\text{RV}_{it}$ | Coef. | -0.691 | 0.046 | -1.575 | 7.25 | 0.415 | |
| | $t$-value | -12.73 | 2.39 | -2.45 | 10.19 | 44.46 | |
| $\text{H-L}_{it}$ | Coef. | -1.404 | -0.038 | -6.15 | 7.88 | | 1.151 |
| | $t$-value | -11.6 | -1.04 | -4.71 | 6.04 | | 35.78 |
| | | | | | | N=378 | T=71 |

TABLE 3.5: Instrumental variable panel data model with constant parameters
This table shows the results of the 2SLS panel regression of our measures of market quality on our AT proxy. The dependent variables are proportional quoted spread, proportional effective spread, daily high-low price range and daily realized variance. In addition to AT, additional regressors included as control variables are the previous month's log of market Cap (ln(ME)), share turnover (T/O), inverse price (1/P) and high-low price range (H-L). When the dependent variable is the current month's high-low price range, last month's value of realized variance (RV) is used instead to avoid a dynamic panel model. Standard errors are corrected for heteroskedasticity.

To gauge the size of this effect we note that the within-standard deviation of our AT proxy, after being scaled by 100, is 0.18. Combining this with the coefficient estimates from Table 3.5 implies that a one standard deviation increase in AT results in quoted spreads (effective spreads) being lowered by approximately 0.002% (0.001%). On an absolute level these effects are small. For example, given a hypothetical stock with an initial price of \$100, a one standard deviation increase in AT would reduce the quoted spread by less than a penny.[2] These results differ from those in Hendershott et al. (2011). We attribute this to a combination of the differences in our AT proxies as well as our inclusion of a more recent sample period. One possible explanation is that the initial increase in AT during its inception has been far larger in terms of effects than subsequent increases. For the variance measures, a one standard deviation increase in our AT proxy results in a decrease in the proportional daily high-low spread of approximately 0.25% and a decrease in realized variance associated with percentage log returns of approximately 0.12 (or equivalently a reduction in realized daily volatility of approximately 0.35%).

From a welfare perspective the magnitude of the effect is important. As mentioned above and further investigated below, if AT amplifies variation in liquidity this is likely

[2]It should be noted that this is technically impossible.

to demand a premium from investors and increase the cost of a capital for firms using markets in which AT is present. Because of this, any benefits in terms of increased liquidity on average, needs to be evaluated against the costs associated with increased variation.

Tables 3.6 through 3.9 present the results when we allow the parameter to jump discretely over time. The coefficient estimates in Tables 3.6 through 3.9 represent the size of the estimated jump in the coefficient and a test of its significance, as outlined the previous section. Figures 4 through 7 plot both the estimated SAW coefficients and the results from period by period cross-sectional regressions. The effect of AT on our measures of market quality is stable prior to the 2007-2008 period. Of course, the 2007-2008 period covers the financial crisis, a time during which liquidity in many markets tightened substantially. During the financial crisis period we find significant evidence of both positive and negative jumps in the coefficient on AT.

For the two spread measures we find evidence of two large positive jumps in the coefficients in April and September/October of 2008 and other smaller jumps around those two time periods. A positive jump in the coefficient represents a reduction in the benefit of AT on spreads and potentially a reversal in its effects on spreads. Such is the case for the two large positive jumps mentioned above. We find that during these two months increases in AT lead to an increase in spreads and thus transacting in the securities with high AT is, other things being equal, costlier than in low AT securities. April and September/October of 2008 represent two particularly volatile periods for equity markets (and markets in general) in the US. In April markets were still rebounding from the bailout of Bear Stearns and its eventual sale to JP Morgan. This all occurred during a period when the exposure of many banks to US housing markets through various structured financial products was beginning to be understood by investors. Similarly, the failure of Lehman Brothers in September was another event that rattled financial markets. The results for our variance measures are similar, as we also find evidence of both positive and negative jumps during the 2007-2008 period. Of note is that for realized variance we find the jumps to be, in general, beneficial for investors. That is, we find that increases in AT cause a larger reduction in realized variance. Some caution should be taken with respect to the interpretation of these results due to the fact that variance is generally found to be strongly autocorrelated. Although we attempt to control for this using the lagged value of the high-low price range, it is possible the use of this variable is not sufficient.

A potential explanation for the variation in the marginal effect of AT is the presence of increased uncertainty. From both a valuation and a regulatory/policy perspective, the periods following large, unpredictable shocks to asset markets can be associated

with heightened uncertainty among investors. If investors fear that algorithmic traders possess an informational advantage then it would be precisely during these periods when an increase in AT would cause investors to be most at risk. Although a model of the dynamic effects of AT and uncertainty is beyond the scope of this chapter, the above results clearly point to a time varying relationship between the effects of AT on various measures of market quality.

| | Coef. | $Z$-value on the difference | $p$-value | |
|---|---|---|---|---|
| from 2003-09-01 to 2008-02-01 | 6.49e-05 | - | - | - |
| from 2008-03-01 to 2008-03-01 | 6.51e-04 | 1.650 | 0.0998 | . |
| from 2008-04-01 to 2008-04-01 | 4.13e-03 | 6.420 | 1.36e-10 | *** |
| from 2008-05-01 to 2008-08-01 | 7.66e-04 | -7.420 | 1.16e-13 | *** |
| from 2008-09-01 to 2008-10-01 | 1.03e-03 | 0.932 | 0.3510 | |
| from 2008-11-01 to 2008-12-01 | -1.46e-04 | -4.620 | 3.83e-06 | *** |

TABLE 3.6: Post-wavelet estimates for the proportional quoted spread.
This table presents the Post-SAW estimates for the parameters and the results of tests for jump significance for the coefficient of AT when the dependent variable is $PQS$. The column labeled estimate is the Post-SAW estimate for the parameter and the Z statistic represents a test of the significance of the change from the previous time period (set equal to 0 for the first period). All tests are asymptotic. *** denotes significance at the 0.1% level, ** denotes significance at the 1% level, * denotes significance at the 5% level and . denotes significance at the 10% level.

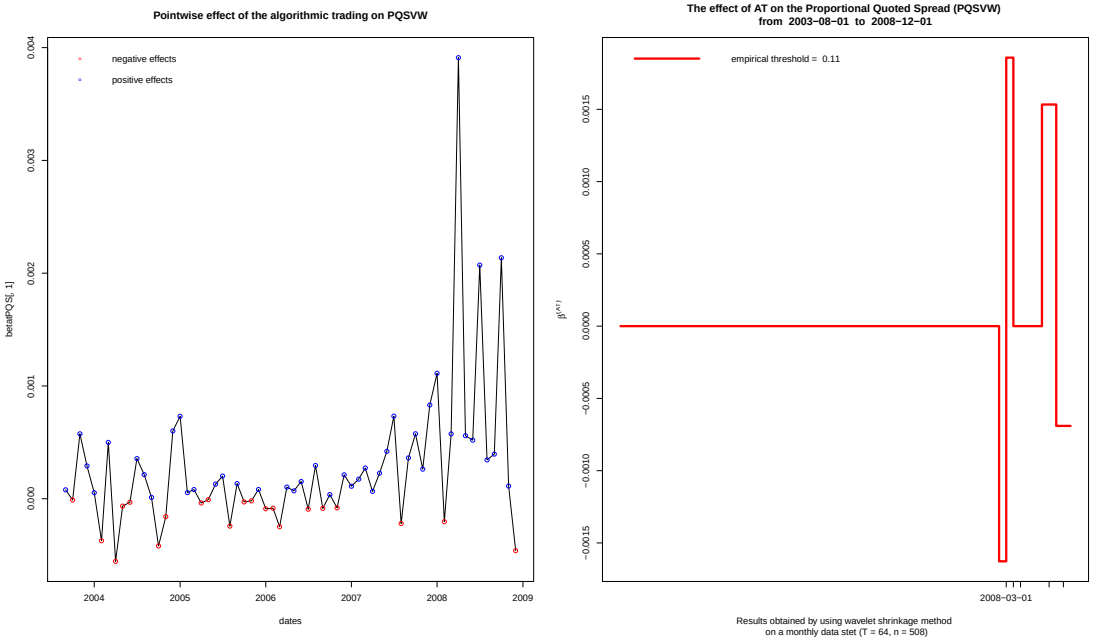


FIGURE 3.3: Time varying effect of algorithmic trading on the proportional quoted spread.

| | Coef. | $Z$-value on the difference | $p$-value | |
|---|---|---|---|---|
| from 2003-09-01 to 2007-08-01 | 9.06e-06 | - | - | |
| from 2007-09-01 to 2007-12-01 | 6.73e-04 | 3.750 | 0.000176 | *** |
| from 2008-01-01 to 2008-02-01 | 1.35e-04 | -2.000 | 0.045800 | * |
| from 2008-03-01 to 2008-03-01 | 5.31e-04 | 1.160 | 0.248000 | |
| from 2008-04-01 to 2008-04-01 | 4.19e-03 | 10.700 | $<$ 2.2e-16 | *** |
| from 2008-05-01 to 2008-08-01 | 4.15e-04 | -15.600 | $<$ 2.2e-16 | *** |
| from 2008-09-01 to 2008-09-01 | -1.74e-03 | -7.540 | 4.77e-14 | *** |
| from 2008-10-01 to 2008-10-01 | 1.79e-03 | 11.700 | $<$ 2.2e-16 | *** |
| from 2008-11-01 to 2008-12-01 | 3.55e-06 | -9.610 | $<$ 2.2e-16 | *** |

TABLE 3.7: Post-wavelet estimates for the proportional effective spread.
This table presents the Post-SAW estimates for the parameters and the results of tests for jump significance for the coefficient of AT when the dependent variable is $PES$. The column labeled estimate is the Post-SAW estimate for the parameter and the Z statistic represents a test of the significance of the change from the previous time period (set equal to 0 for the first period). All tests are asymptotic. *** denotes significance at the 0.1% level, ** denotes significance at the 1% level, * denotes significance at the 5% level and . denotes significance at the 10% level.

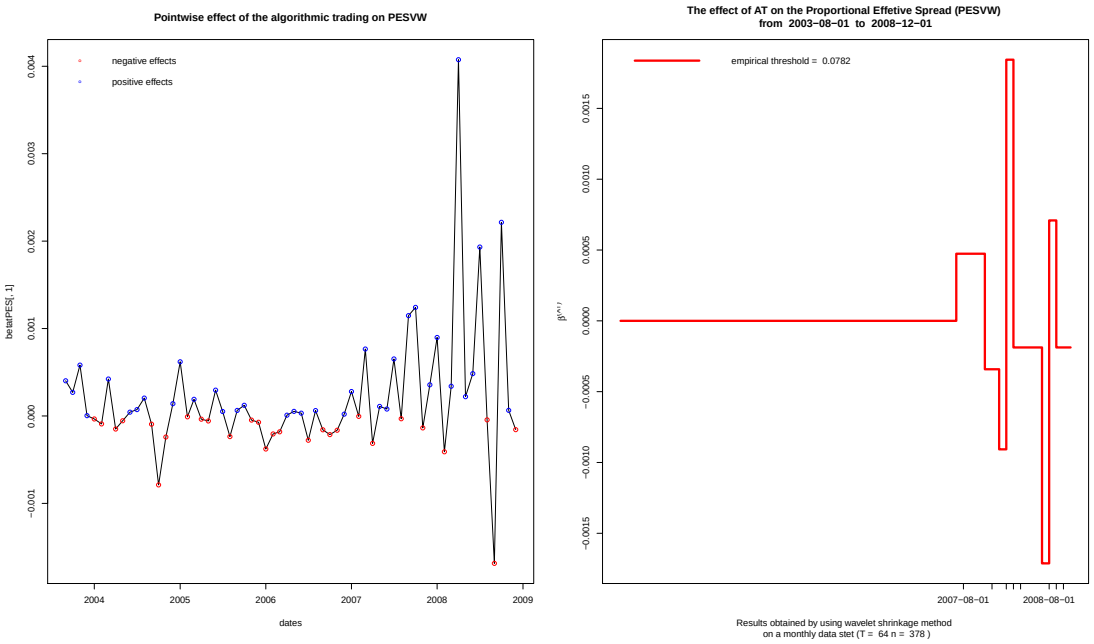


FIGURE 3.4: Time varying effect of algorithmic trading on the proportional effective spread.

| | Coef. | $Z$-value on the difference | $p$-value | |
|---|---|---|---|---|
| from 2003-09-01 to 2007-06-01 | -0.017100 | - | - | - |
| from 2007-07-01 to 2007-07-01 | -0.048500 | -1.08 | 0.28200 | |
| from 2007-08-01 to 2007-08-01 | -0.154000 | -2.68 | 0.00745 | ** |
| from 2007-09-01 to 2008-08-01 | -0.012400 | 5.23 | 1.65e-07 | *** |
| from 2008-09-01 to 2008-09-01 | -0.107000 | -4.40 | 1.07e-05 | *** |
| from 2008-10-01 to 2008-10-01 | -0.000913 | 4.59 | 4.33e-06 | *** |
| from 2008-11-01 to 2008-12-01 | -0.021400 | -1.60 | 0.11000 | |

TABLE 3.8: Post-wavelet estimates for the daily high-low price range.
This table presents the Post-SAW estimates for the parameters and the results of tests for jump significance for the coefficient of AT when the dependent variable is $H - L$. The column labeled estimate is the Post-SAW estimate for the parameter and the Z statistic represents a test of the significance of the change from the previous time period (set equal to 0 for the first period). All tests are asymptotic. *** denotes significance at the 0.1% level, ** denotes significance at the 1% level, * denotes significance at the 5% level and . denotes significance at the 10% level.

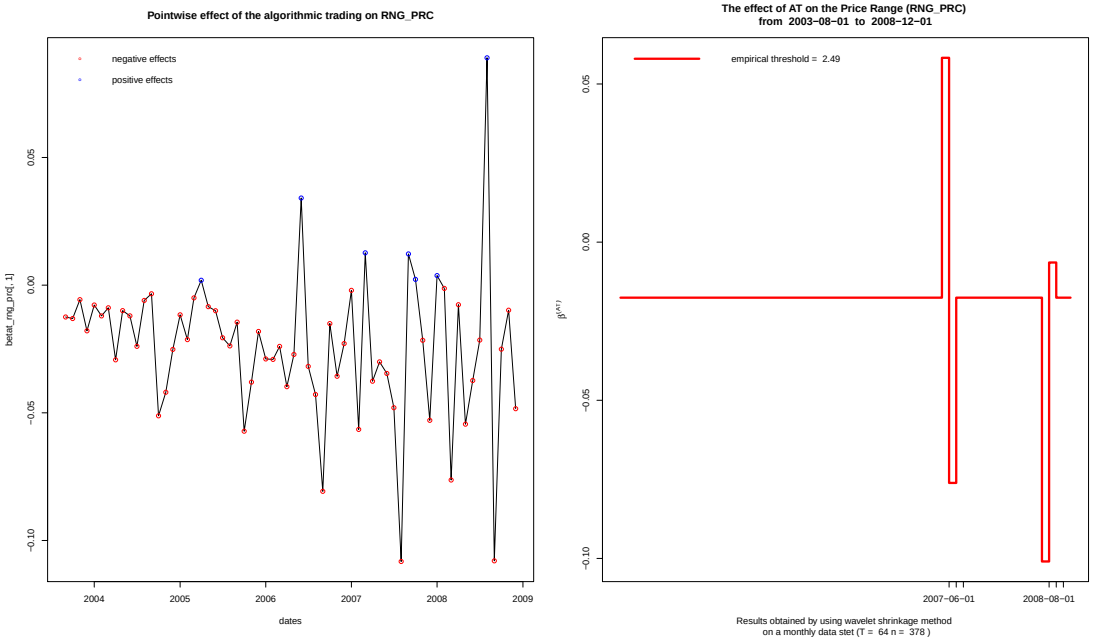


FIGURE 3.5: Time varying effect of algorithmic trading on the daily high-low price range.

| | Coef. | $Z$-value on the difference | $p$-value | |
|---|---|---|---|---|
| from 2003-09-01 to 2008-08-01 | -0.008080 | -14.20 | $<$ 2.2e-16 | *** |
| from 2008-09-01 to 2008-09-01 | -0.063100 | -5.09 | 3.57e-07 | *** |
| from 2008-10-01 to 2008-10-01 | 0.000888 | 5.29 | 1.21e-07 | *** |
| from 2008-11-01 to 2008-12-01 | -0.007830 | -1.26 | 0.208 | |

TABLE 3.9: Post-wavelet estimates for the realized variance.
This table presents the Post-SAW estimates for the parameters and the results of tests for jump significance for the coefficient of AT when the dependent variable is $RV$. The column labeled estimate is the Post-SAW estimate for the parameter and the Z statistic represents a test of the significance of the change from the previous time period (set equal to 0 for the first period). All tests are asymptotic. *** denotes significance at the 0.1% level, ** denotes significance at the 1% level, * denotes significance at the 5% level and . denotes significance at the 10% level.

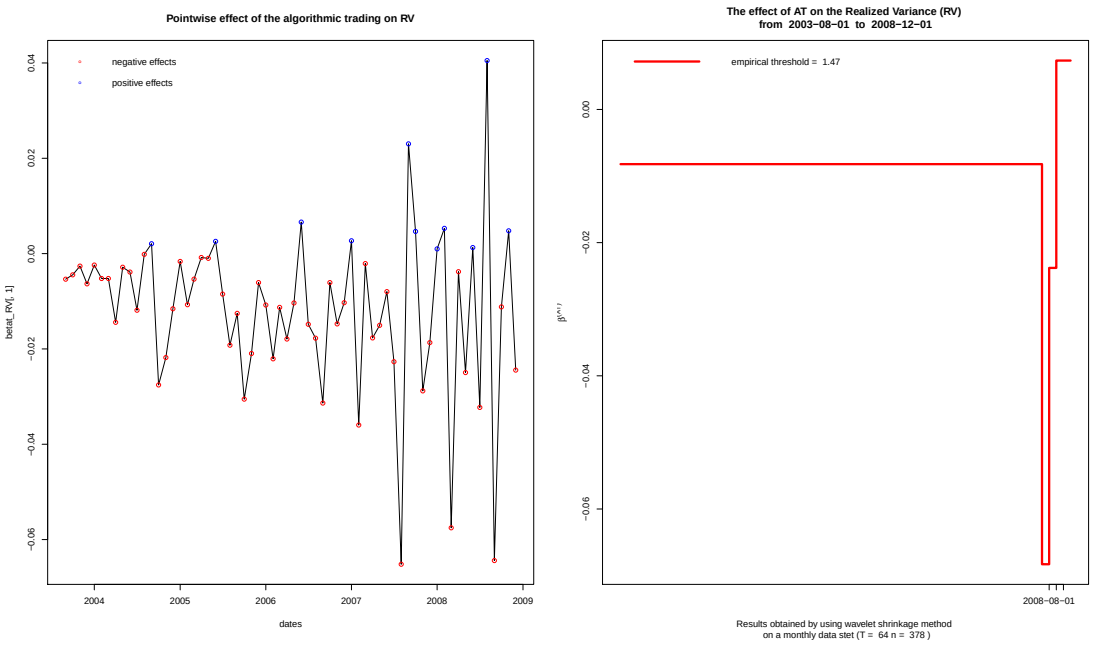


FIGURE 3.6: Time varying effect of algorithmic trading on the realized variance.

## 3.8 Conclusion

This chapter generalizes the special panel model specifications in which the slope parameters are either constant over time or extremely time heterogeneous to allow for panel models with multiple structural changes that occur at unknown date points and may affect each slope parameter individually. Consistency under weak forms of dependency and heteroscedasticity in the idiosyncratic errors is established and convergence rates are derived. Our empirical vehicle for highlighting this new methodology addresses the stability of the relationship between Algorithmic Trading (AT) and Market Quality (MQ). We find evidence that the relationship between AT and MQ was disrupted during the

time between 2007 and 2008. This period coincides with the beginning of the subprime crisis in the US market and the bankruptcy of the big financial services firm Lehman Brothers.

# Appendix A

# Appendix of Chapter 1

## A.1 Theoretical Results and Proofs

Before beginning with proofs, we now define these quantities and show how to construct consistent estimate of the bias term. Following again Bai et al. (2009) we first use kernel estimators to approximate the long-run covariance matrices of $w_{it}$. Estimates of $\epsilon_{it}$, $\zeta_{it}$ and $\eta_t$ are given by the regression residuals $\widehat{\epsilon}_{it}$, $\hat{\zeta}_{it} = \Delta X_{it}$ and $\widehat{\eta}_t = \Delta \widehat{F}_t$. For all $h = -T+1, \ldots, T-1$ and all $i = 1, \ldots, n$ let $\widehat{\Gamma}_{\epsilon,i}(h)$, $\widehat{\Gamma}_{\epsilon,\zeta,i}(h)$, $\widehat{\Gamma}_{\epsilon,\eta,i}(h)$, $\widehat{\Gamma}_{\epsilon,b,i}(h)$ and $\widehat{\Gamma}_{b,i}(h)$ denote the $(1\times 1(, (1\times p), (1\times d)$ , $(1\times(p+d))$ and $((p+d)\times(p+d))$ empirical lag $h$ autocovariance matrices of $(\widehat{\epsilon}_{it}, \widehat{\epsilon}_{i,t+h})$, $(\widehat{\epsilon}_{it}, \hat{\zeta}_{i,t+h})$, $(\widehat{\epsilon}_{it}, \widehat{\eta}_{t+h})$, $(\widehat{\epsilon}_{it}, (\hat{\zeta}'_{i,t+h}, \eta'_{t+h})')$ as well as $((\hat{\zeta}'_{i,t}, \widehat{\eta}'_t)', (\hat{\zeta}'_{i,t+h}, \widehat{\eta}'_{t+h})')$, $t = 1, \ldots, T$. Then define

$$\widehat{\Omega}_{\epsilon,i} = \sum_{j=-T+1}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,i}(j), \quad \widehat{\Omega}_{\epsilon,b,i} = \sum_{j=-T+1}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,b,i}(j)$$

$$\widehat{\Omega}_{b,i} = \sum_{j=-T+1}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{b,i}(j), \quad \widehat{\Omega}_{\epsilon|b,i} = \widehat{\Omega}_{\epsilon,i} - \widehat{\Omega}'_{\epsilon,b,i}\widehat{\Omega}^{-1}_{b,i}\widehat{\Omega}_{\epsilon,b,i},$$

$$\begin{pmatrix} \widehat{\Delta}^{+}_{\zeta,\epsilon,i} \\ \widehat{\Delta}^{+}_{\eta,\zeta,i} \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,\zeta,i}(h) \\ \sum_{j=0}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,\eta,i}(h) \end{pmatrix} - \sum_{j=0}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{b,i}(h)\widehat{\Omega}^{-1}_{b,i}\widehat{\Omega}_{\epsilon,b,i}.$$

Here, the kernel function $\omega(.)$ satisfies the following assumption:

**Assumption 4. The kernel function $\omega(.)$:**

> **The kernel function** $\omega(.) : R \to [-1,1]$ satisfies (i) $\omega(0) = 1$, $\omega(x) = \omega(-x)$, (ii) $\int_{-1}^{1} \omega(x)^2 dx < \infty$ and with Parzen'exponent $q \in (0,\infty)$ such that $\lim \frac{1-\omega(x)}{|x|^q} < \infty$

and $\liminf_{(N,T)\leftrightarrow\infty}\left(\frac{\log(T)}{\log(N)}\right) > 1$; the bandwidth parameter $\kappa \sim N^b$ where $b \in (\frac{1}{2q}, \liminf\left(\frac{\log(T)}{\log(N)}\right) - 1)$.

$\widehat{\Omega}_{\epsilon,i}$, $\widehat{\Omega}_{\epsilon,b,i}$, $\widehat{\Omega}_{b,i}$, $\widehat{\Omega}_{\epsilon|b,i}$, $\widehat{\Delta}^{+}_{\zeta,\epsilon,i}$ and $\widehat{\Delta}^{+}_{\eta,\zeta,i}$ estimate their theoretical analogues $\Omega_{\epsilon,i}$, $\Omega_{\epsilon,b,i}$, $\Omega_{b,i}$, $\Omega_{\epsilon|b,i}$, $\Delta^{+}_{\zeta,\epsilon,i}$ and $\Delta^{+}_{\eta,\zeta,i}$ which are defined by replacing in the above equation the terms $\omega(\frac{\cdot}{\kappa})\widehat{\Gamma}.(\cdot)$ by the corresponding true autocovariance matrices of $w_{it} = (\varepsilon_{it}, \zeta^{*}_{it}, \eta_t)$. In addition, summation then ranges from $-\infty$ to $\infty$ (instead of $-T+1$ to $T-1$) and 0 to $\infty$ (instead of 0 to $T-1$).

Now define the projection matrices of $F^{\circ}$ as $M_F = I_T - F^{\circ}(F^{\circ\prime}F^{\circ})^{-1}F^{\circ\prime}$ and the scalar $a_{ik}$ as the element $i,k$ of the projection matrix $A_\Lambda = \Lambda^{\circ}(\Lambda^{\circ\prime}\Lambda^{\circ})^{-1}\Lambda^{\circ\prime}$. Corresponding estimates $\widehat{M_F}$ and $\hat{a}_{ik}$ are obtained by replacing $F^{\circ}$ and $\Lambda^{\circ}$ by $\widehat{F}$ and $\widehat{\Lambda}$. Then let

$$Z_i = M_F X^{*}_i - \frac{1}{N}\sum_{j=1}^{N} M_F X^{*}_j a_{ij}, \quad \widehat{Z}_i = \widehat{M_F} X_i - \frac{1}{N}\sum_{j=1}^{N} \widehat{M_F} X_j \hat{a}_{ij}$$

Conditional on $F^{\circ}$, the bias term $\phi$ is then given by

$$\begin{aligned}
\phi &= \left(\frac{1}{NT^2}\sum_{i=1}^{N} Z_i' Z_i\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}\theta_i, \\
\theta_i &= Z_i'(\Delta b_i)\Omega^{-1}_{b,i}\Omega_{\epsilon,b,i} + \Delta^{+}_{\zeta,\epsilon,i} - \delta_i'\Delta^{+}_{\eta,\zeta,i}, \\
\Delta\hat{b}_i &= (\Delta X^{*}_i - \frac{1}{N}\sum_{j=1}^{N}\Delta X^{*}_j a_{ij}, \Delta F^{\circ}), \quad \delta_i = (F^{\circ\prime}F^{\circ})^{-1}F^{\circ\prime}X_i,
\end{aligned}$$

and a consistent estimator can be determined by

$$\begin{aligned}
\widehat{\phi}_{NT} &= \left(\frac{1}{NT^2}\sum_{i=1}^{N} \hat{Z}_i' \hat{Z}_i\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_i, \\
\widehat{\theta}_i &= \hat{Z}_i'(\Delta\hat{b}_i)\widehat{\Omega}^{-1}_{b,i}\widehat{\Omega}_{\epsilon,b,i} + \widehat{\Delta}^{+}_{\zeta,\epsilon,i} - \hat{\delta}_i'\widehat{\Delta}^{+}_{\eta,\zeta,i}, \\
\Delta\hat{b}_i &= (\Delta X_i - \frac{1}{N}\sum_{j=1}^{N}\Delta X_j \hat{a}_{ij}, \Delta\widehat{F}), \quad \hat{\delta}_i = (\widehat{F}'\widehat{F})^{-1}\widehat{F}' X_i.
\end{aligned}$$

Conditional on $F^{\circ}$, Bai et al. (2009) show that there exists random matrices $R_{Ci}$, defined as conditional expectations of integrated Brownian motions with individually different covariance structure, such that as $(N,T) \to \infty$ we have $\frac{1}{NT^2}\sum_{i=1}^{N} Z_i Z_i' \to_d \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} R_{Ci}$. The covariance matrix $\Sigma_c$ is then defined by

$$\Sigma_c = \left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} R_{Ci}\right)^{-1}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Omega_{\epsilon|b,i}R_{Ci}\right)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} R_{Ci}\right)^{-1}, \quad \text{(A.1)}$$

Bai et al. (2009) do not propose an estimator of $\Sigma_c$. However, following their arguments it is straightforward to show that a consistent estimate of the covariance matrix $\Sigma_c$ is given by

$$\widehat{\Sigma}_c = \left(\frac{1}{NT^2}\sum_{i=1}^{N}\hat{Z}_i'\hat{Z}_i\right)^{-1}\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\Omega}_{\epsilon|b,i}\hat{Z}_i'\hat{Z}_i\left(\frac{1}{NT^2}\sum_{i=1}^{N}\hat{Z}_i'\hat{Z}_i\right)^{-1},$$

We then obtain the following proposition:

**Proposition A.1.** *Under Assumptions 1-3, we have, as* $(N,T)\to\infty$*,*

$$\Sigma_c^{1/2}\left(\sqrt{N}T(\hat{\beta}(d)-\beta^\circ)-\sqrt{N}\phi\right)\xrightarrow{d}N(0,I_p),$$

*where* $\hat{\beta}(d)$ *is the iterative least squares estimator obtained after transforming Model* (1.17) *by using* $\mathcal{T}(.)$*.*

**Proof of A.1:** Let $\tilde{Z}_i=\widehat{M}_F X_i-\frac{1}{N}\sum_{j=1}^{N}\widehat{M}_F X_j a_{ij}$. In view of Proposition 4 and Lemma A.2 of Bai et al. (2009) we only have to show that $\frac{1}{NT^2}\sum_{i=1}^{N}\tilde{Z}_i'\tilde{Z}_i$ and $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{Z}_i'\epsilon_i$ have the same limit distributions as $\frac{1}{T^2}Z_i'Z_i$ and $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}Z_i'\varepsilon_i$. Note that $X_{it}-X_{it}^*=X_{i,t-1}-X_{i,t-1}^*+(\sum_{k=1}^{K}\delta_{ik}(\bar{\zeta}_{kt}-\zeta_{kt}^0))$ is also an $I(1)$-process. But the innovations $(\sum_{k=1}^{K}\delta_{ik}(\bar{\zeta}_{kt}-\zeta_{kt}^0))$ are averages over $m_k$ individuals and hence $var_C((\sum_{k=1}^{K}\delta_{ik}(\bar{\zeta}_{kt}-\zeta_{kt}^0))\leq M_1/N$, where $M_1<\infty$ is some constant independent of $i,t$. Therefore, as $(N,T)\to\infty$ $\|\frac{1}{NT^2}\sum_{i=1}^{N}\tilde{Z}_i\tilde{Z}_i'-\frac{1}{NT^2}\sum_{i=1}^{N}Z_iZ_i'\|=O_P(N^{-1/2})$. Our assumptions imply that conditional on $\{\eta_t\}$ the random variables $\zeta_{it}^*$ and $\varepsilon_{it}$ are independent. Consequently, as $(N,T)\to\infty$, we have $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{Z}_i'(\sum_{k=1}^{K}\delta_{ik}\bar{\varepsilon}_k)=\sum_{k=1}^{K}(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\delta_{ik}\tilde{Z}_i')\bar{\varepsilon}_k=o_P(1)$ as well as $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(Z_i'-\tilde{Z}_i')\varepsilon_i=o_P(1)$. One can conclude that $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{Z}_i'\epsilon_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}Z_i'\varepsilon_i+o_p(1)$. The proposition is an immediate consequence. □

**Corollary A.2.** *Under Assumptions 1-4, we have, as* $(N,T)\to\infty$*,*

$$\Sigma_c^{1/2}\left(\sqrt{N}T(\left(\hat{\beta}(d)-\frac{1}{T}\widehat{\phi}_{NT}\right)-\beta^\circ)\right)\xrightarrow{d}N(0,I_p),$$

*where* $\hat{\beta}(d)$ *is the iterative least squares estimator obtained after using the transformation operator* $\mathcal{T}(.)$ *and* $\widehat{\phi}_{NT}$ *is estimated as above.*

**Proof of Corollary A.2:** Corollary 1 follows from a straightforward generalization of the arguments used in the proof of Theorem 2 in Bai et al. (2009). Note that, all additional terms induced by the differences $X_{it}^*-X_{it}$ and $\varepsilon_{it}-\epsilon_{it}$ are asymptotically negligible. □

The main difference between our approach and the methodology of Bai et al. (2009) consists in the fact that our estimation procedure directly incorporates a dimension estimate. Our final estimator $\hat{\beta}_{EupBC} = \hat{\beta}(\hat{d}_{Eup}) - \frac{1}{T}\widehat{\phi}_{NT}$, thus, relies on the estimated dimension $\hat{d}_{Eup}$. The following theorem shows that with high probability $\hat{d}_{Eup}$ will asymptotically coincide with the true dimension $d$. The asymptotic distributions derived in Corollary 1 thus remain valid when replacing $d$ by $\hat{d}_{Eup}$. Furthermore, the final estimator $\widehat{F}_{EupBC}$ yields a consistent estimator of the true factor structure (up to rotations).

**Theorem A.3.** *Under assumptions 1-3, we have, as* $(N,T) \to \infty$*,*

*a)* $P(\hat{d}_{Eup} = d) \to 1$*, if the starting estimate* $d_{max} \geq d$ *and* $g$ *is of the form* $g = cp_{NT}$ *such that* $(i)$ $c = O_P(1)$ *and strict positive,* $(ii)$ $p_{NT} \to \infty$*, and* $(iii)$ $\frac{\log\log(T)}{T} p_{NT} \to 0$*,*

*b) with the additional Assumption 4,*

$$\Sigma_c^{1/2}\sqrt{NT}(\hat{\beta}_{EupBC} - \beta^\circ) \xrightarrow{d} N(0, I_p),$$

*c) and for some* $(d \times d)$ *invertible matrix* $H$*,*

$$\frac{1}{T}\sum_{t=1}^{T} \|\widehat{F}_{EupBC,t} - F_t^\circ H\|^2 = O_P(\frac{1}{N}) + O_P(\frac{1}{T}).$$

**Proof of Theorem A.3:** Assertion a). We can infer from the theoretical results of Bai et al. (2009) that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d))^2 = O_P(1), \tag{A.2}$$

where $\hat{Y}_{it}(d)$ denotes the fitted variable for the true dimension $d$.

Since our estimates are obtained by minimizing least squares objective function, we have, $\forall l \in \{d+1, d+2, \ldots, d_{max}\}$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(l))^2 \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d))^2 = O_P(1). \tag{A.3}$$

On the other hand, straightforward generalization of the arguments of Bai (2004) shows that, $\forall l \in \{0, \ldots, d-1\}$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(l))^2 - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d))^2 = O_p(T/\log\log(T)). \tag{A.4}$$

Note that (A.4) is strict positive.

By construction of the estimation algorithm, we know that the number of outer iterations $r_m$ is maximal $d_{max}+1$, where $d_{max}$ is larger than $d$. To prove that $\lim_{(N,T)\to\infty} P(\hat{d}_{Eup} = d) = 1$, we use inductive reasoning and show that, for all outer iterations $r \in \{1, \ldots, r_m | \hat{d}^{(r_m)} = \hat{d}^{(r_m-1)}\}$,

$$\begin{aligned} \lim_{(N,T)\to\infty} P(\hat{d}^{(1)} = d | d_{max} \geq d) &= 1, \text{ and} \\ \lim_{(N,T)\to\infty} P(\hat{d}^{(r)} = d) &= 1, \; r \in \{2, \ldots, r_m | \hat{d}^{(r_m)} = \hat{d}^{(r_m-1)}\}, \end{aligned}$$

as long as $g$ satisfies Conditions $(i) - (iii)$.

First of all, we show that the optimal dimension $\hat{d}^{(1)}$ obtained by using the stating scaling coefficient $\hat{\sigma}^2(d_{max}) = \frac{1}{NT}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d_{max}))^2$ in $g$ is consistent.

According to (A.3), $\hat{\sigma}^2(d_{max}) = O_p(1)$.

Since

$$S_4^{(1)}(l) - S_4^{(1)}(d) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(l))^2 - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d))^2 + (l-d)g^{(1)},$$

where $g^{(1)} = \hat{\sigma}^2(d_{max})p_{NT}$ such that $p_{NT} \to \infty$ and $\frac{\log\log(T)}{T}p_{NT} \to 0$, as $(N,T) \to \infty$, we can conclude from (A.3) and (A.4) that

$$\lim_{(N,T)\to\infty} P(S_4^{(1)}(l) - S_4^{(1)}(d) > 0) = 0 \; \forall l \in \{0, \ldots, d_{max} | l \neq d\}.$$

Because $\hat{d}^{(1)}$ is the minimizer of $S_4^{(1)}(l)$, it follows that

$$\lim_{(N,T)\to\infty} P(\hat{d}^{(1)} = d) = 1. \tag{A.5}$$

At the outer iteration stage $r+1$, the law of total probability implies

$$\begin{aligned} \lim_{(N,T)\to\infty} P(\hat{d}^{(r+1)} = d) &= \lim_{(N,T)\to\infty} \Big( P(\hat{d}^{(r+1)} = d | \hat{d}^{(r)} = d) P(\hat{d}^{(r)} = d) \\ &\quad + P(\hat{d}^{(r+1)} = d | \hat{d}^{(r)} \neq d) P(\hat{d}^{(r)} \neq d) \Big). \end{aligned}$$

To complete the proof, we have, therefore, to verify that

$$\lim_{(N,T)\to\infty} P(\hat{d}^{(r+1)} = d | \hat{d}^{(r)} = d) = 1, \quad \text{for} \quad r = 2, \ldots, r_m. \tag{A.6}$$

According to (A.3), we have, for $\hat{d}^{(r)} = d$, $\hat{\sigma}^2(\hat{d}^{(r)}) = O_p(1)$. From (A.3) and (A.4), we can conclude that

$$\lim_{(N,T)\to\infty} P(S_4^{(r+1)}(l) - S_4^{(r+1)}(d) > 0) = 0 \; \forall l \in \{0, \ldots, d_{max} | l \neq d\},$$

with $g^{(r)} = \hat{\sigma}^2(\hat{d}^{(r)})p_{NT}$, and hence

$$\lim_{(N,T)\to\infty} P(\hat{d}^{(r+1)} = d | \hat{d}^{(r)} = d) = 1, \quad \text{for} \quad r = 2, \ldots, r_m.$$

Induction implies

$$\lim_{(N,T)\to\infty} P(\hat{d}^{(r+1)} = d | \hat{d}^{(r)} = d) = \lim_{(N,T)\to\infty} P(\hat{d}^{(r+1)} = d) = 1, \quad \text{for} \quad r = 2, \ldots, r_m.$$

Assertion a) is an immediate consequence.

Now, we have

$$\begin{aligned}
\Sigma_c^{-1/2} \left( \sqrt{N}T(\hat{\beta}_{CupBC}(\hat{d}_{Eup}) - \beta) \right) &= \Sigma_c^{-1/2} \left( \sqrt{N}T(\hat{\beta}_{CupBC}(d) - \beta) \right) \\
&\quad + \Sigma_c^{-1/2} \left( \sqrt{N}T(\hat{\beta}_{CupBC}(\hat{d}_{Eup}) - \hat{\beta}_{CupBC}(d) \right) \\
&= \Sigma_c^{-1/2} \left( \sqrt{N}T(\hat{\beta}_{CupBC}(d) - \beta) \right) + o_P(1).
\end{aligned}$$

By Corollary 1, Assertion b) is, hence, an immediate consequence. Assertion c) follows from Assertion a) and Proposition 5 of Bai et al. (2009). □

# Appendix B

# Appendix of Chapter 3

## B.1 Proofs of Section 3.2

**Lemma B.1.** *Let $T = 2^{L-1}$ for some integer $L \geq 2$ and $\beta = (\beta_1, \ldots, \beta_T)' \in \mathbb{R}^T$ a vector that possesses exactly one jump at $\tau \in \{1, \ldots, T\}$ such that*

$$\beta_t = \begin{cases} \beta_\tau & \text{for} \quad t \in \{1, \ldots, \tau\} \\ \beta_{\tau+1} \neq \beta_\tau & \text{for} \quad t \in \{\tau+1, \ldots, T\}. \end{cases}$$

*Let $w_{lk}(t)$ be defined as (3.8) and $h_{lk}(t)$ as (3.9), where $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ are positive real values for all $l \in \{1, \ldots, L\}$, and $k \in \{1, \ldots, K_l\}$. There then exists unique $l_\tau$ non-zero coefficients $\{b_{lk_l} | l_\tau \leq L\}$, where $k_l \in \{1, \ldots, K_l\}$, such that*

$$\beta_t = \sum_{l=1}^{l_\tau} w_{lk_l}(t) b_{lk_l}.$$

**Proof of Lemma B.1:** To prove the proposition, we show that $\beta_t$ can be reconstructed by using at most $L$ wavelet basis if it processes exactly one jump, say at $\tau \in \{1, \ldots, T\}$. To simplify the exposition, we re-define the wavelet basis $w_{l,k}(t)$, for $l > 1$ as follows:

$$w_{l,k}(t) = a^*_{l,2k-1} h^*_{l,2k-1}(t) - a^*_{l,2k} h^*_{l,2k}(t),$$

where

$$h^*_{l,k}(t) = \begin{cases} 1 & \text{for} \quad t \in \left\{ \left(2^{L-l-1}(k-1)+1\right), \ldots, \left(2^{L-l-1}k\right) \right\} \\ 0 & \text{else.} \end{cases}$$

This is equivalent to (3.8). The unique difference is that the coefficients $a^*_{l,2k-1}$ and $a^*_{l,2k}$ are scaled by $\sqrt{2^l}$ in order to simplify the construction of $h^*_{l,k}(t)$ and let it be either 1 or 0.

Note that by construction, there exists a unique $l_\tau \in \{2, \ldots, L\}$ and a unique $k_{l_\tau} \in \{1, \ldots, 2^{l_\tau - 2}\}$ such that

$$w_{l_\tau k_{l_\tau}}(\tau) = a^*_{l_\tau, 2k_{l_\tau}-1} \quad \text{and} \quad w_{l_\tau k_{l_\tau}}(\tau+1) = -a^*_{l_\tau, 2k_{l_\tau}}.$$

Moreover, there exists in each level $l \in \{1, \ldots, L | l < l_\tau\}$ at most one basis $w_{lk_l}(t)$ that satisfies the following condition:

$$w_{lk_l}(\tau) = w_{l,k_l}(\tau+1) \neq 0.$$

Define the time interval $\mathcal{I}_l$, for each $l = 1, \ldots, l_\tau$, as follows:

$$\mathcal{I}_l = \{t \in \{1, \ldots, T\} | w_{l,k_l}(t) \neq 0\}.$$

such that

$$\bigcup_{l=1}^{l_\tau} \mathcal{I}_l = \{1, \ldots, T\}$$

and

$$\mathcal{I}_{l_\tau} \subset \mathcal{I}_{l_\tau - 1} \subset \cdots \subset \mathcal{I}_2 \subseteq \mathcal{I}_1 = \{1, \ldots, T\}.$$

We now begin with the thinnest interval $\mathcal{I}_{l\tau}$ that contains the jump. Define

$$\beta_t^{(l_\tau)} = \begin{cases} \beta_t = \beta_\tau & \text{if} \quad t \leq \tau \text{ and } t \in \mathcal{I}_{l\tau} \cap \{t | t \leq \tau\} \\ \beta_t = \beta_{\tau+1} & \text{if} \quad t > \tau \text{ and } t \in \mathcal{I}_{l\tau} \cap \{t | t > \tau\} \\ 0 & \text{else.} \end{cases}$$

Because $\beta_\tau \neq \beta_{\tau+1}$ and $a^*_{l_\tau, 2k_{l_\tau}-1}, a^*_{l_\tau, 2k_{l_\tau}} > 0$, there exists a non-zero coefficient $b_{l_\tau, k_{l_\tau}} = \frac{\beta_\tau - \beta_{\tau+1}}{a^*_{l_\tau, 2k_{l_\tau}-1} + a^*_{l_\tau, 2k_{l_\tau}}}$ and a constant $\beta^{(l_\tau)} \neq \{\beta_\tau, \beta_{\tau+1}\}$ such that

$$\beta_t^{(l_\tau)} = \begin{cases} \beta_\tau = \beta^{(l_\tau)} + a^*_{l_\tau, 2k_{l_\tau}-1} b_{l_\tau, k_{l_\tau}} & \text{if} \quad t \leq \tau \text{ and } t \in \mathcal{I}_{l\tau} \\ \beta_{\tau+1} = \beta^{(l_\tau)} - a^*_{l_\tau, 2k_{l_\tau}} b_{l_\tau, k_{l_\tau}} & \text{if} \quad t > \text{ and } t \in \mathcal{I}_{l\tau} \\ 0 & \text{else.} \end{cases} \tag{B.1}$$

Using the definition of $w_{lk}(t)$, we can rewrite (B.1) as

$$\beta_t^{(l_\tau)} = \begin{cases} \beta_t = \beta^{(l_\tau)} + w_{l_\tau, k_{l_\tau}}(t) b_{l_\tau, k_{l_\tau}} & \text{if} \quad t \in \mathcal{I}_{l\tau} \\ 0 & \text{else.} \end{cases} \tag{B.2}$$

Consider the second thinnest interval $\mathcal{I}_{l_{\tau-1}}$. Let

$$\beta_t^{(l_\tau-1)} = \begin{cases} \beta_t & \text{if} \quad t \in \mathcal{I}_{l_{\tau-1}} \setminus \mathcal{I}_{l_\tau} \\ \beta^{(l_\tau)} & \text{if} \quad t \in \mathcal{I}_{l_\tau} \\ 0 & \text{else.} \end{cases}$$

Note that $\beta_t$ is constant over $\mathcal{I}_{l_{\tau-1}} \setminus \mathcal{I}_{l_\tau}$; it can be either $\beta_\tau$ or $\beta_{\tau+1}$. Now, because $\beta^{(l_\tau)} \neq \{\beta_\tau, \beta_{\tau+1}\}$, we can determine a second unique non-zero coefficient $b_{l_\tau-1,k_{l_\tau-1}}$ and a second unique constant $\beta^{(l_{\tau-1})} \neq \{\beta_\tau, \beta_{\tau+1}\}$ such that

$$\beta_t^{(l_\tau-1)} = \begin{cases} \beta^{(l_{\tau-1})} + w_{l_\tau-1,k_{l_\tau-1}}(t) b_{l_\tau-1,k_{l_\tau-1}} = \beta_t & \text{if} \quad t \in \mathcal{I}_{l_{\tau-1}} \setminus \mathcal{I}_{l_\tau} \\ \beta^{(l_{\tau-1})} + w_{l_\tau-1,k_{l_\tau-1}}(t) b_{l_\tau-1,k_{l_\tau-1}} = \beta^{(l_\tau)} & \text{if} \quad t \in \mathcal{I}_{l_\tau} \\ 0 & \text{else.} \end{cases}$$

Because $w_{l_\tau,k_{l_\tau}}(t) = 0$ for all $t \notin \mathcal{I}_{l_{\tau-1}}$ and all $t \in \mathcal{I}_{l_{\tau-1}} \setminus \mathcal{I}_{l_\tau}$, adding $w_{l_\tau,k_{l_\tau}}(t) b_{l,k_l}$ on both sides, gives

$$\beta_t^{(l_\tau-1)} + w_{l_\tau,k_{l_\tau}}(t) b_{l,k_l} = \begin{cases} \beta_t + w_{l_\tau,k_{l_\tau}}(t) b_{l,k_l} & \text{if} \quad t \in \mathcal{I}_{l_{\tau-1}} \setminus \mathcal{I}_{l_\tau} \\ \beta^{(l_\tau)} + w_{l_\tau,k_{l_\tau}}(t) b_{l,k_l} & \text{if} \quad t \in \mathcal{I}_{l_\tau} \\ 0 & \text{else.} \end{cases}$$

Moreover, because $\beta^{(l_\tau)} + w_{l_\tau,k_{l_\tau}}(t) b_{l,k_l} = \beta_t$ for all $t \in \mathcal{I}_{l_\tau}$, we can write

$$\beta_t^{(l_\tau-1)} + w_{l_\tau,k_{l_\tau}}(t) b_{l,k_l} = \begin{cases} \beta^{(l_{\tau-1})} + \sum_{l=l_\tau-1}^{l_\tau} w_{l,k_l}(t) b_{l,k_l} = \beta_t & \text{if} \quad t \in \mathcal{I}_{l_{\tau-1}} \\ 0 & \text{else.} \end{cases}$$

Replacing $\beta_t^{(t_\tau-1)}$ by $\beta_t^{(t_\tau-2)}$ and proceeding with the recursion until $\beta_t^{(l_\tau-l)}$, for $l \in \{2, \ldots, l_\tau\}$, we end up with

$$\beta_t^{(l_\tau-l)} + w_{l_\tau-l+1,k_{l_\tau-l+1}}(t) b_{l_\tau-l+1,k_{l_\tau-l+1}} = \begin{cases} \beta^{(l_\tau-l)} + \sum_{s=l_\tau-l}^{l_\tau} w_{s,k_s}(t) b_{s,k_s} = \beta_t & \text{if} \quad t \in \mathcal{I}_{l_\tau-l} \\ 0 & \text{else.} \end{cases} \tag{B.3}$$

where $\beta^{(l_\tau-l)}$ is constant over $\mathcal{I}_{l_\tau-l}$. Finally, from (B.3), we can infer that, for all $t \in \mathcal{I}_1 = \{1, \ldots, T\}$,

$$\beta_t = \beta^{(1)} + \sum_{l=2}^{l_\tau} w_{l,k_l}(t) b_{l,k_l} \qquad \forall t \in \{1, \ldots, T\}.$$

Because $\beta^{(1)}$ is a constant and $w_{11}(t) = a_{11} \neq 0,\ \forall t \in \{1, \ldots, T\}$, we can express $\beta_t$ in terms of $l_\tau \leq L$ basis such that

$$\beta_t = \sum_{l=1}^{l_\tau} w_{l,k_l}(t) b_{l,k_l} \qquad \forall t \in \{1, \ldots, T\}.$$

This completes the proof. □

**Proof of Proposition** 3.1: To prove the assertion, we expand the original vector in a series of $S$ vectors so that each new vector contains only one jump, and make use of Proposition B.1. Let $\beta$ be a $T \times 1$ vector such that

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{\tau_1} \\ \beta_{\tau_1+1} \\ \vdots \\ \beta_{\tau_2} \\ \beta_{\tau_2+1} \\ \vdots \\ \beta_{\tau_S+1} \\ \vdots \\ \beta_T \end{pmatrix} = \begin{pmatrix} \beta_{\tau_1} \\ \beta_{\tau_1} \\ \vdots \\ \beta_{\tau_1} \\ \beta_{\tau_2} \\ \vdots \\ \beta_{\tau_2} \\ \beta_{\tau_3} \\ \vdots \\ \beta_{\tau_{S+1}} \\ \vdots \\ \beta_{\tau_{S+1}} \end{pmatrix}$$

where $\{\tau_s \in \{1,\dots,T\} | \tau_1 < \dots < \tau_S\}$. We can transform $\beta$ in a series of $S+1$ Vectors, $\overline{\beta}_{\tau_1},\dots,\overline{\beta}_{\tau_S}$ as follows:

$$\underbrace{\begin{pmatrix} \beta_{\tau_1} \\ \vdots \\ \beta_{\tau_1} \\ \beta_{\tau_2} \\ \vdots \\ \beta_{\tau_2} \\ \beta_{\tau_3} \\ \vdots \\ \beta_{\tau_S} \\ \beta_{\tau_{S+1}} \\ \vdots \\ \beta_{\tau_{S+1}} \end{pmatrix}}_{\beta} = \underbrace{\begin{pmatrix} \beta_{\tau_1}-\beta_{\tau_2} \\ \vdots \\ \beta_{\tau_1}-\beta_{\tau_2} \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\overline{\beta}_{\tau_1}} + \dots + \underbrace{\begin{pmatrix} \beta_{\tau_{S-1}}-\beta_{\tau_S} \\ \vdots \\ \beta_{\tau_{S-1}}-\beta_{\tau_S} \\ \beta_{\tau_{S-1}}-\beta_{\tau_S} \\ \vdots \\ \beta_{\tau_{S-1}}-\beta_{\tau_S} \\ \beta_{\tau_{S-1}}-\beta_{\tau_S} \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\overline{\beta}_{\tau_S}} + \underbrace{\begin{pmatrix} \beta_{\tau_S} \\ \vdots \\ \beta_{\tau_S} \\ \beta_{\tau_S} \\ \vdots \\ \beta_{\tau_S} \\ \beta_{\tau_S} \\ \vdots \\ \beta_{\tau_S} \\ \beta_{\tau_{S+1}} \\ \vdots \\ \beta_{\tau_{S+1}} \end{pmatrix}}_{\overline{\beta}_{\tau_S+1}},$$

so that each new vector processes exactly one jump (except $\overline{\beta}_{\tau_S+1}$, which is constant over all). From Proposition B.1, we know that each vector $\overline{\beta}_{\tau_s}$, $s = 1,\dots,S$, processes a unique expansion of the form

$$\overline{\beta}_{\tau_s} = \sum_{l=1}^{L} \sum_{k=1}^{K_l} w_{lk} b_{lk}^{(s)}$$

with at most $L$ non-zero coefficients in $\{b_{lk}^{(s)}\}_{l=1,\dots,L;k=1,\dots,K_l}$, where

$$K_l = \begin{cases} 1 & \text{if} \quad l = 1 \\ 2^{l-2} & \text{if} \quad l = 2,\dots,L. \end{cases}$$

The fact that $\beta = \sum_{s=1}^{S+1} \overline{\beta}_{\tau_s}$ completes the proof. □

**Proposition B.2.** *If $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ are chosen for each $l \in \{1,\dots,L\}$ and $k \in \{1,\dots,K_l\}$ such that*

*(i)* $a_{l,2k-1}^2 \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Z_{it} h_{l,2k-1}^2(t) + a_{l,2k}^2 \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Z_{it} h_{l,2k}^2(t) = 1,$

*(ii)* $a_{l,2k-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Z_{it} h_{l,2k-1}^2(t) - a_{l,2k} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Z_{it} h_{l,2k}^2(t) = 0$

*(iii)* $a_{1,1}^2 \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Z_{it} = 1$

*then* $(a)$ *and* $(b)$ *are satisfied for all* $l, l^{'} \in \{1, \ldots, L\}$, $k, \in \{1, \ldots, K_l\}$, *and* $k^{'}, \in \{1, \ldots, K_{l'} | l \neq l^{'}\}$, *or* $k, k^{'}, \in \{1, \ldots, K_l | k \neq k^{'}; l = l^{'}\}$.

**Proof of Proposition B.2:** To prove that $(i) - (iii)$ imply the orthonormality Conditions $(a)$ and $(b)$, for all $l, l^{'} \in \{1, \ldots, L\}$, $k \in \{1, \ldots, K_l\}$, and $k^{'} \in \{1, \ldots, K_{l'}\}$, it is sufficient to verify the following three statements:

**(S.1):** condition $(b)$ holds if $l = l^{'}$ and $k^{'} \neq k$.

**(S.2):** condition $(b)$ holds if $(ii)$ is satisfied for all $l^{'} < l$, and

**(S.3):** condition $(a)$ holds if $(i)$ and $(iii)$ are satisfied for all $(l, k) = (l^{'}, k^{'})$.

Before checking S.1-S.3, we begin with examining the product $\mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it}$. If $(l, k) \neq (l^{'}, k^{'})$,

$$\begin{aligned}
\mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it} &= Z_{it,lk}Z_{it,l^{'}k^{'}} \\
&= X_{it}Z_{it}\left(w_{lk}(t)w_{l^{'}k^{'}}(t)\right) \\
&= X_{it}Z_{it}\left(a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t)\right)\left(a_{l^{'},2k^{'}-1}h_{l,2k^{'}-1}(t) - a_{l^{'},2k^{'}}h_{l^{'},2k^{'}}(t)\right) \\
&= X_{it}Z_{it}\Big(a_{l,2k-1}a_{l^{'},2k^{'}-1}h_{l,2k-1}(t)h_{l^{'},2k^{'}-1}(t) - a_{l,2k-1}a_{l^{'},2k^{'}}h_{l,2k-1}(t)h_{l^{'},2k^{'}}(t) \\
&\quad -a_{l,2k}a_{l^{'},2k^{'}-1}h_{l,2k}(t)h_{l^{'},2k^{'}-1}(t) + a_{l,2k}a_{l^{'},2k^{'}}h_{l,2k}(t)h_{l^{'},2k^{'}}(t)\Big)
\end{aligned}$$

If $(l, k) = (l^{'}, k^{'})$,

$$\begin{aligned}
\mathcal{Z}_{l,k,it}\mathcal{X}_{l,k,it} &= X_{it}Z_{it}\left(w_{lk}(t)\right)^2 \\
&= X_{it}Z_{it}\left(a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t)\right)^2 \\
&= X_{it}Z_{it}\left(a_{l,2k-1}^2h_{l,2k-1}^2(t) + a_{l,2k}^2h_{l,2k}^2(t) - 2a_{l,2k-1}a_{l,2k}\underbrace{h_{l,2k-1}(t)h_{l,2k}(t)}_{0}\right) \\
&= X_{it}Z_{it}\left(a_{l,2k-1}^2h_{l,2k-1}^2(t) + a_{l,2k}^2h_{l,2k}^2(t)\right),
\end{aligned} \tag{B.4}$$

The product $h_{l,2k-1}(t)h_{l,2k}(t)$ (in the third line) is zero because $h_{l,2k}(t) = 0$, for all $t \in \left\{\left((2k-2)2^{L-l}+1\right), \ldots, \left((2k-1)2^{L-l}\right)\right\}$, $h_{l,2k-1}(t) = 0$, for all $t \in \left\{\left((2k-1)2^{L-l}+1\right), \ldots, \left(2k2^{L-l}\right)\right\}$ and both $h_{l,2k}(t) = h_{l,2k-1}(t) = 0$ else.

Consider **(S.1)**. If $l = l^{'}$, and $k^{'} \neq k$, we have, for all $t \in \{1, \ldots, T\}$,

$$\begin{aligned}
\mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it} &= Z_{it}X_{it}\Bigg(a_{l,2k-1}a_{l,2k^{'}-1}\underbrace{h_{l,2k-1}(t)h_{l,2k^{'}-1}(t)}_{=0} - a_{l,2k-1}a_{l,2k^{'}}\underbrace{h_{l,2k-1}(t)h_{l,2k^{'}}(t)}_{=0} \\
&\quad -a_{l,2k}a_{l,2k^{'}-1}\underbrace{h_{l,2k}(t)h_{l,2k^{'}-1}(t)}_{=0} + a_{l,2k}a_{l,2k^{'}}\underbrace{h_{l,2k}(t)h_{l,2k^{'}}(t)}_{=0}\Bigg) \\
&= 0
\end{aligned}$$

This implies $(b)$, for all $l, l^{'} \in \{2, \ldots, L | l = l^{'}\}$ and $k, k^{'} \in \{1, \ldots, 2^{l-2} | k^{'} \neq k\}$.
Consider **(S.2)**. If $l^{'} < l$, we have by construction either

$$\begin{aligned} \mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it} &= Z_{it}X_{it}a_{l^{'},2k^{'}}h_{l^{'},2k^{'}}(t)\left(a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t)\right) \\ &= a_{l^{'},2k^{'}}\Big(Z_{it}X_{it}a_{l,2k-1}h_{l,2k-1}(t)h_{l^{'},2k^{'}}(t) - Z_{it}X_{it}a_{l,2k}h_{l,2k}(t)h_{l^{'},2k^{'}}(t)\Big) \end{aligned}$$

or

$$\begin{aligned} \mathcal{Z}_{l,k,it}\mathcal{X}_{l^{'},k^{'},it} &= Z_{it}X_{it}a_{l^{'},2k^{'}-1}h_{l^{'},2k^{'}-1}(t)\left(a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t)\right) \\ &= a_{l^{'},2k^{'}-1}\Big(Z_{it}X_{it}a_{l,2k-1}h_{l,2k-1}(t)h_{l^{'},2k^{'}-1}(t) - Z_{it}X_{it}a_{l,2k}h_{l,2k}(t)h_{l^{'},2k^{'}-1}(t),\Big) \end{aligned}$$

If $h_{l^{'},2k^{'}}(t) = \sqrt{2^l}$, then $h_{l^{'},2k^{'}-1}(t) = 0$ and if $h_{l^{'},2k^{'}-1}(t) = \sqrt{2^l}$, then $h_{l^{'},2k^{'}}(t) = 0$, otherwise both $h_{l^{'},2k^{'}}(t)$ and $h_{l^{'},2k^{'}-1}(t)$ are zeros. Thus condition $(ii)$ ensures $(b)$.

Consider **(S.3)**. From (B.4), we can easily verify that $(a)$ is a direct result of $(i)$ for all $l \in \{2, \ldots, L\}$ and $k \in \{1, \ldots, K_l\}$.□

## B.2 Proofs of Section 3.3

**Proof of Lemma 3.4:** The IV estimator of our (modified) wavelets coefficients is given by

$$\begin{aligned} \tilde{b}_{l,k,p} &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\mathcal{Z}_{lk,it,p}\Delta y_{it}, \\ &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\mathcal{Z}_{lk,it,p}(\sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}}\mathcal{Z}_{lk,it,q}b_{l,k,q} + \Delta e_{it}), \\ &= b_{l,k,p} + \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{q=1}^{\underline{P}}\mathcal{Z}_{lk,it,p}\Delta e_{it}. \end{aligned}$$

The last equality is due to the orthonormality conditions $(A)$ and $(B)$. Subtracting $b_{l,k,p}$ from both sides and multiplying by $\sqrt{n(T-1)}$, we get, for $l > 1$,

$$\begin{aligned}
\sqrt{n(T-1)}(\tilde{b}_{l,k,p} - b_{l,k,p}) &= \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{\underline{P}} \mathcal{Z}_{lk,it,q} \Delta e_{it}, \\
&= \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{\underline{P}} W_{lk,pq}(t) Z_{it,q} \Delta e_{it}, \\
&= \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{\underline{P}} A_{l,2k,pq} h_{l,2k}(t) Z_{it,q} \Delta e_{it} \\
&\quad - \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{\underline{P}} A_{l,2k-1,pq} h_{l,2k-1}(t) Z_{it,q} \Delta e_{it}, \\
&= \frac{1}{\sqrt{n(2^{L-l-1}-1)}} \sum_{q=1}^{\underline{P}} A_{l,2k,pq} \sum_{i=1}^{n} \sum_{t \in \{h_{l,2k}(t) \neq 0\}} Z_{it,q} \Delta e_{it} \\
&\quad - \frac{1}{\sqrt{n(2^{L-l-1}-1)}} \sum_{q=1}^{\underline{P}} A_{l,2k-1,pq} \sum_{i=1}^{n} \sum_{t \in \{h_{l,2k-1}(t) \neq 0\}} Z_{it,q} \Delta e_{it},
\end{aligned}$$

where $W_{lk,pq}(t)$ and $A_{l,m,pq}$ are the $(p,q)$- elements of the matrices $W_{l,k}(t)$ and $A_{l,m}$, respectively. and, for $l = 1$,

$$\sqrt{n(T-1)}(\tilde{b}_{1,1,p} - b_{1,1,p}) = \frac{1}{\sqrt{n(2^{L}-1)}} \sum_{q=1}^{\underline{P}} A_{1,1,pq} \sum_{i=1}^{n} \sum_{t=2}^{T} Z_{it,q} \Delta e_{it}.$$

By Assumption B.(i), we know that $E_c(Z_{it}\Delta e_{it}) = 0$, for all $i$ and $t$. The law of total expectation implies

$$E\big(\sqrt{n(T-1)}(\tilde{b}_{l,k,p} - b_{l,k,p})\big) = 0,$$

for all $l$ and $k$. The total variance, for $l > 1$, can be written as

$$\begin{aligned}
\Sigma_{l,k,p} &= E\big((\sqrt{n(T-1)}(\tilde{b}_{l,k,p} - b_{l,k,p}))^2\big), \\
&= E\left( \frac{1}{n(2^{L-l-1}-1)} \sum_{q,r=1}^{\underline{P}} A_{l,2k,pq} A_{l,2k,pr} \sum_{i,j=1}^{n} \sum_{t,s \in H} Z_{it,q} Z_{js,r} E_c\big(\Delta e_{it} \Delta e_{js}\big) \right) \\
&\quad + E\left( \frac{1}{n(2^{L-l-1}-1)} \sum_{q,r=1}^{\underline{P}} A_{l,2k-1,pq} A_{l,2k-1,pr} \sum_{i,j=1}^{n} \sum_{t,s \in H} Z_{it,q} Z_{js,r} E_c\big(\Delta e_{it} \Delta e_{js}\big) \right), \\
&= \Pi_{l,k,1} + \Pi_{l,k,2},
\end{aligned}$$

where $\sum_{q,r=1}^{\underline{P}}$, $\sum_{i,j=1}^{n}$ and $\sum_{t,s\in H}$ denote the double summations $\sum_{q=1}^{\underline{P}}\sum_{q=1}^{\underline{P}}\sum_{i=1}^{n}\sum_{j=1}^{n}$ and $\sum_{t\in\{H_{l,2k}(t)\neq 0\}}\sum_{s\in\{H_{l,2k}(s)\neq 0\}}$, respectively.

For $l=1$,

$$\begin{aligned}\Sigma_{1,1,p} &:= E\big((\sqrt{n(T-1)}(\tilde{b}_{1,1,p}-b_{1,1,p}))^2\big)\\ &= E\left(\sum_{q,r=1}^{\underline{P}}\frac{1}{n(2^L-1)}A_{1,1,pq}A_{1,1,pr}\sum_{i,j=1}^{n}\sum_{t,s=2}^{T}Z_{it,q}Z_{js,r}E_c\big(\Delta e_{it}\Delta e_{js}\big)\right).\end{aligned}$$

By using Assumption C, we can infer

$$\begin{aligned}\Pi_{l,k,1} &= E\left(\frac{1}{n(2^{L-l-1}-1)}\sum_{q,r=1}^{\underline{P}}A_{l,2k,pq}A_{l,2k,pr}\sum_{i,j=1}^{n}\sum_{t,s\in H}Z_{it,q}Z_{js,r}\sigma_{ij,ts}\right),\\ &\leq E\left(\frac{1}{n(2^{L-l-1}-1)}\sum_{q,r=1}^{\underline{P}}A_{l,2k,pq}A_{l,2k,pr}\sum_{i,j=1}^{n}\sum_{t,s\in H}Z_{it,q}Z_{js,r}|\sigma_{ij,ts}|\right),\\ \Pi_{l,k,2} &\leq E\left(\frac{1}{n(2^{L-l-1}-1)}\sum_{q,r=1}^{\underline{P}}A_{l,2k-1,pq}A_{l,2k-1,pr}\sum_{i,j=1}^{n}\sum_{t,s\in H}Z_{it,q}Z_{js,r}|\sigma_{ij,ts}|\right), \text{ and}\\ \Sigma_{1,1,p} &\leq E\left(\frac{1}{n(2^{L}-1)}\sum_{q,r=1}^{\underline{P}}A_{1,1,pq}A_{1,1,pr}\sum_{i,j=1}^{n}\sum_{t,s=2}^{T}Z_{it,q}Z_{js,r}|\sigma_{ij,ts}|\right).\end{aligned}$$

Because $E(||A_{l,2k}||^4)$ and $E(||A_{l,2k-1}||^4)$ are bounded uniformly in $l,k$, and $E(||Z_{it}||^4)$, and $|\sigma_{ij,ts}|$ is bounded uniformly in $i,j,t,s$ (see Assumptions B and C), we can easily show (by Cauchy-Schwarz inequality) that $\Sigma_{l,k,p}\leq M$ is bounded uniformly in $l,k,p$. Using Assumption B(iii), we can write

$$\begin{aligned}P\left(\left|\tilde{b}_{l,k,p}-b_{l,k,p}\right|>M^{\frac{1}{2}}\frac{c}{\sqrt{n(T-1)}}\right) &\leq P\left(\Sigma_{l,k,p}^{-\frac{1}{2}}\sqrt{n(T-1)}\left|\tilde{b}_{l,k,p}-b_{l,k,p}\right|>c\right),\\ &\leq \frac{1}{c}\exp(-\frac{c^2}{2}).\end{aligned}\tag{B.5}$$

Using Boole's inequality and (B.5), we get

$$\begin{aligned}P\left(\sup_{l,k,p}\left|\tilde{b}_{l,k,p}-b_{l,k,p}\right|>M^{\frac{1}{2}}\frac{c}{\sqrt{n(T-1)}}\right) &\leq \sum_{l,k,p}P\left(\left|\tilde{b}_{l,k,p}-b_{l,k,p}\right|>M^{\frac{1}{2}}\frac{c}{\sqrt{n(T-1)}}\right),\\ &\leq (2^{L-1}\underline{P})\frac{1}{c}\exp(-\frac{c^2}{2}),\\ &= (T-1)\underline{P}\frac{1}{c}\exp(-\frac{c^2}{2}),\end{aligned}$$

where $\sum_{l,k,p}$ denotes the triple summation $\sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{p=1}^{\underline{P}}$. The assertion of the theorem follows by replacing $c$ with $\sqrt{2\log((T-1)\underline{P})}c^*$ for any $c^*>0$. □

**Proof of Theorem 3.5:** We have first to prove that $(i)$ : $\sup_t |\tilde{\gamma}_{t,p} - \gamma_{t,p}| = o_p(1)$ for all $p \in \{1, \ldots, \underline{P}\}$ if $\sqrt{T-1}\lambda_{n,T} \to 0$, as $n, T \to \infty$ or $n \to \infty$ and $T$ is fixed, and then conclude that $(ii)$ : $\frac{1}{T-1}\sum_{t=2}^{T} ||\tilde{\gamma}_t - \gamma_t||^2 = O_p((\log(T-1)/n)^{\kappa})$, if $\sqrt{T-1}\lambda_{n,T} \sim (\log(T-1)/n)^{\kappa/2}$, for $\kappa \in ]0, 1]$.

By construction,

$$\tilde{\gamma}_{t,p} - \gamma_{t,p} = \sum_{q=1}^{\underline{P}} \sum_{l=1}^{L} \sum_{k=1}^{K_l} W_{lk,pq}(t)\hat{b}_{l,k,q} - \sum_{q=1}^{\underline{P}} \sum_{l=1}^{L} \sum_{k=1}^{K_l} W_{lk,pq}(t) b_{l,k,q}, \tag{B.6}$$

where

$$\hat{b}_{l,k,q} = \tilde{b}_{l,k,q} - \tilde{b}_{l,k,q}(|\tilde{b}_{l,k,q}| < \lambda_{n,T}). \tag{B.7}$$

and

$$\begin{aligned} W_{lk,pq}(t) &= A_{l,2k,pq}(t)H_{l,2k}(t) - A_{l,2k-1,pq}(t)H_{l,2k-1}(t), \qquad (B.8) \\ &= \sqrt{2^{l-2}}A_{l,2k,pq}\mathbf{I}(H_{l,2k}(t) \neq 0) - \sqrt{2^{l-2}}A_{l,2k-1,pq}\mathbf{I}(H_{l,2k-1}(t) \neq 0). \end{aligned}$$

Plugging (B.7) and (B.8) in (B.6) and using the absolute value inequality, we get

$$\begin{aligned} |\tilde{\gamma}_{t,p} - \gamma_{t,p}| \leq & \sum_{q=1}^{\underline{P}} \sum_{l=1}^{L} \sum_{k=1}^{K_l} \sqrt{2^{l-2}} |A_{l,2k,pq}\mathbf{I}(H_{l,2k}(t) \neq 0)(\tilde{b}_{l,k,q} - b_{l,k,q})| \\ & + \sum_{q=1}^{\underline{P}} \sum_{l=1}^{L} \sum_{k=1}^{K_l} \sqrt{2^{l-2}} |A_{l,2k,pq}\mathbf{I}(H_{l,2k}(t) \neq 0)\tilde{b}_{l,k,q}\mathbf{I}(|\tilde{b}_{l,k,q}| < \lambda_{n,T})| \\ & + \sum_{q=1}^{\underline{P}} \sum_{l=1}^{L} \sum_{k=1}^{K_l} \sqrt{2^{l-2}} |A_{l,2k-1,pq}\mathbf{I}(H_{l,2k-1}(t) \neq 0)(\tilde{\underline{b}}_{l,k,q} - \underline{b}_{l,k,q})| \\ & + \sum_{q=1}^{\underline{P}} \sum_{l=1}^{L} \sum_{k=1}^{K_l} \sqrt{2^{l-2}} |A_{l,2k,pq}\mathbf{I}(H_{l,2k-1}(t) \neq 0)\tilde{b}_{l,k,q}\mathbf{I}(|\tilde{b}_{l,k,q}| < \lambda_{n,T})|, \\ = & \; a + b + c + d. \end{aligned}$$

Because $\tilde{b}_{l,k,p}\mathbf{I}(|\tilde{b}_{l,k,p}| < \lambda_{n,T}) < \lambda_{n,T}$ and $|\tilde{b}_{l,k,p} - b_{l,k,p}| \leq \sup_{l,k,p}|\tilde{b}_{l,k,p} - b_{l,k,p}|$ for all $p \in \{1, \ldots, \underline{P}\}$, we can write

$$\begin{aligned}
a &\leq \sup_{l,k,p}|\tilde{b}_{l,k,p} - b_{l,k,p}| \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l}|A_{l,2k,pq}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k}(t) \neq 0)|, \\
b &\leq \lambda_{n,T}\sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l}|A_{l,2k,pq}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k}(t) \neq 0)|, \\
c &\leq \sup_{l,k,p}|\tilde{b}_{l,k,p} - b_{l,k,p}| \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l}|A_{l,2k-1,pq}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k}(t) \neq 0)|, \text{ and} \\
d &\leq \lambda_{n,T}\sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l}|A_{l,2k-1,pq}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k-1}(t) \neq 0)|.
\end{aligned}$$

By Assumption B, $E(||A_{l,2k}||^4)$ and $E(||A_{l,2k-1}||^4)$ are bounded uniformly in $l$ and $k$. We can deduce that

$$\begin{aligned}
\sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l}|A_{l,2k,pq}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k}(t) \neq 0)| &= O_p(1)\sum_{l=1}^{L}\sum_{k=1}^{K_l}|\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k}(t) \neq 0)| \text{ and} \\
\sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l}|A_{l,2k-1,pq}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k-1}(t) \neq 0)| &= O_p(1)\sum_{l=1}^{L}\sum_{k=1}^{K_l}|\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k-1}(t) \neq 0)|.
\end{aligned}$$

Moreover, from the construction of $H_{l,2k}(t)$ and $H_{l,2k-1}(t)$, we can easily verify that

$$\sup_t \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sqrt{2^{l-2}}\mathbf{I}(H_{l,2k-1}(t) \neq 0) = \sum_{l=1}^{L}\sqrt{2^{l-2}} = O(\sqrt{2^{L-1}}) = O(\sqrt{T-1})$$

By Lemma 3.4, we can infer that

$$\begin{aligned}
\sup_{t,p}|\tilde{\gamma}_{t,p} - \gamma_{t,p}| &= \sup_{l,k,p}|\tilde{b}_{l,k,p} - b_{l,k,p}| \times O_p(\sqrt{T-1}) + \lambda_{n,T} \times O_p(\sqrt{T-1}), \\
&= O_p\Big(\sqrt{\frac{\log(T-1)}{n}} + \sqrt{T-1}\lambda_{n,T}\Big). \qquad \text{(B.9)}
\end{aligned}$$

Assertion $(i)$ follows immediately if $\sqrt{T-1}\lambda_{n,T} \to 0$ with $\log(T-1)/n \to 0$, as $n, T \to \infty$.

Consider Assertion $(ii)$. Let $\mathcal{L}_p := \{(l,k)|b_{l,k,p} = 0\}$ denote the set of double indexes corresponding to the non-zero true wavelet coefficients so that $\gamma_{t,p} = \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{l,k,pq}(t)b_{l,k,q}$ can be written as

$$\gamma_{t,p} = \sum_{q=1}^{\underline{P}}\sum_{(l,k)\in\mathcal{L}_p} W_{l,k,pq}(t)b_{l,k,q},$$

and $\tilde{\gamma}_{t,p} = \sum_{q=1}^{P}\sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{lk,pq}(t)\hat{b}_{l,k,q}$ as

$$\tilde{\gamma}_{t,p} = \sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p} W_{lk,pq}(t)\hat{b}_{l,k,q} + \sum_{q=1}^{P}\sum_{(l,k)\notin\mathcal{L}_p} W_{lk,pq}(t)\hat{b}_{l,k,q}.$$

The difference, can be written as

$$\tilde{\gamma}_{t,p} - \gamma_{t,p} = \sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p} W_{lk,pq}(t)(\hat{b}_{l,k,q} - b_{l,k,q}) + \sum_{q=1}^{P}\sum_{(l,k)\notin\mathcal{L}_p} W_{lk,pq}(t)\hat{b}_{l,k,q}.$$

Averaging the square, we get

$$\begin{aligned}
\frac{1}{T-1}\sum_{t=2}^{T-1}(\tilde{\gamma}_{t,p} - \gamma_{t,p})^2 &= \frac{1}{T-1}\sum_{t=2}^{T-1}\left(\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p} W_{lk,pq}(t)(\hat{b}_{l,k,q} - b_{l,k,q})\right)^2 \\
&\quad + \frac{1}{T-1}\sum_{t=2}^{T-1}\left(\sum_{q=1}^{P}\sum_{(l,k)\notin\mathcal{L}_p} W_{lk,pq}(t)\hat{b}_{l,k,q}\right)^2 \\
&\quad - \frac{1}{T-1}\sum_{t=2}^{T-1}\left(\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p} W_{lk,pq}(t)(\hat{b}_{l,k,q} - b_{l,k,q})\right) \times \\
&\quad \left(\sum_{q=1}^{P}\sum_{(l,k)\notin\mathcal{L}_p} W_{lk,pq}(t)\hat{b}_{l,k,q}\right), \\
&= \frac{1}{T-1}\sum_{t=2}^{T-1} e_t^2 + \frac{1}{T-1}\sum_{t=2}^{T-1} f_t^2 - \frac{1}{T-1}\sum_{t=2}^{T-1} e_t f_t.
\end{aligned}$$

From the analysis of assertion $(i)$, we can see that

$$\begin{aligned}
e_t &= \sup_{l,k,p}|\hat{b}_{l,k,p} - b_{l,k,p}|O_p(1)\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p}\sqrt{2^{l-1}}\mathbf{I}(H_{l,2k-1}(t)\neq 1; H_{l,2k}(t)\neq 1) \\
&= O_p\Big(\sqrt{\frac{\log(T-1)}{n(T-1)}} + \lambda_{n,T}\Big)\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p}\sqrt{2^{l-1}}\mathbf{I}(H_{l,2k-1}(t)\neq 1; H_{l,2k}(t)\neq 1),
\end{aligned}$$

and

$$f_t = \sup_{(l,k)\in\mathcal{L}_p,p}|\hat{b}_{l,k,p}|O_p(1)\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p}\sqrt{2^{l-1}}\mathbf{I}(H_{l,2k-1}(t)\neq 1; H_{l,2k}(t)\neq 1).$$

Using Cauchy-Schwarz inequality to $(\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p}\sqrt{2^{l-1}}\mathbf{I}(H_{l,2k-1}(t)\neq 1; H_{l,2k}(t)\neq 1))^2$ over $(l,k)$, we can infer that

$$e_t^2 \;\leq\; O_p\big(\frac{\log(T-1)}{n(T-1)}+\lambda_{n,T}^2\big)\sum_{q=1}^{P}\sum_{(l,k)\in\mathcal{L}_p}2^{l-1}\mathbf{I}(H_{l,2k-1}(t)\neq 1; H_{l,2k}(t)\neq 1),$$

and

$$\frac{1}{T-1}\sum_{t=2}^{T-2}f_t^2 \;\leq\; (\sup_{(l,k)\in\mathcal{L}_p,p}|\hat{b}_{l,k,p}|)^2 O_p(T-1).$$

If $\sqrt{T-1}\lambda_{n,T}\sim(\log(T-1)/n)^{\kappa/2}$, then $\text{plim}(\frac{1}{T-1}\sum_{t=2}^{T-2}f_t^2)=0$ as $T$ and\or $n$ pass to infinity, for any $\kappa\in]0,1[$.

Let us now examine the average of $e_t^2$ over $t$. If, in total, the maximal number of jumps is $S^*=\sum_p^P S_p$, then by Proposition 3.1 the number of non-zero coefficients is at most $(S^*+1)L$. By taking the average of $e_t^2$ over $t$, we can hence infer that

$$\frac{1}{T-1}\sum_{t=2}^{T-1}e_t^2 \;\leq\; O_p\big(\frac{\log(T-1)}{n(T-1)}+\lambda_{n,T}^2\big)\left(\min\{(S^*+1)\log(T-1),(T-1)\}\right).$$

Finally, because $\text{plim}(\frac{1}{T-1}\sum_{t=2}^{T-2}f_t^2)=0$, by Cauchy-Schwarz inequality, we can infer that $\frac{1}{T-1}\sum_{t=2}^{T-1}e_t f_t$ also can be neglected. Thus

$$\frac{1}{T-1}\sum_{t=2}^{T-1}(\tilde{\gamma}_{t,p}-\gamma_{t,p})^2 = O_p\big(\frac{J^*(\log(T-1)/n)^\kappa}{(T-1)}\big),$$

where $J^*=\min\{(S^*+1)\log(T-1),(T-1)\}$. This completes the proof. □

## B.3 Proofs of Section 3.4

**Proof of Lemma 3.6:** We have to show that

$$\sup_{k,p\in\{1,\ldots,P\}}\left|\tilde{c}_{L,k,p}^{(m)}-c_{L,k,p}^{(m)}\right| = O_p\big(\sqrt{\log(T-1)/(n(T-1))}\big),$$

for $m=s,u$.

For $p \in \{1, \ldots, P\}$ and $m = s$, we have by construction

$$\begin{aligned}
\tilde{c}^{(s)}_{L,k,p} - c^{(s)}_{L,k,p} &= \frac{1}{T-1} \sum_{t=2}^{T} \psi_{L,k}(t-1)(\tilde{\gamma}_{t,p} - \gamma_{t,p}), \\
&= \frac{1}{T-1} \sum_{t=2}^{T} \psi_{L,k}(t-1) \sum_{l,m,q} W_{l,m,p,q}(t)(\tilde{b}_{l,m,q} - b_{l,m,q}), \\
&= \frac{1}{T-1} \sum_{t \in \{\psi_{L,k}(t-1) \neq 0\}} \psi_{L,k}(t-1) \sum_{l,m,q} W_{l,m,p,q}(t)(\tilde{b}_{l,m,q} - b_{l,m,q}),
\end{aligned}$$

where $\sum_{l,m,q}$ denotes the triple summation $\sum_{l=1}^{L} \sum_{k=1}^{K_l} \sum_{q=1}^{P}$.

Taking the absolute value, we obtain

$$|\tilde{c}^{(s)}_{L,k,p} - c^{(s)}_{L,k,p}| \leq \sup_{l,k,p} |\tilde{b}_{l,k,p} - b_{l,k,p}| \frac{1}{T-1} \sum_{t \in \{\psi_{L,k}(t-1) \neq 0\}} \left| \psi_{L,k}(t-1) \sum_{l,m,q} W_{l,m,p,q}(t) \right|.$$

Recall that $\frac{1}{T-1} \sum_{t \in \{\psi_{L,k}(t-1) \neq 0\}} \psi_{L,k}(t-1)^2 = 1$. By using Cauchy-Schwarz inequality, we can easily verify that

$$\frac{1}{T-1} \sum_{t \in \{\psi_{L,k}(t-1) \neq 0\}} \left| \psi_{L,k}(t-1) \sum_{l,m,q} W_{l,m,p,q}(t) \right| \leq \left( \frac{1}{T-1} \sum_{t \in \{\psi_{L,k}(t-1) \neq 0\}} (\sum_{l,m,q} W_{l,m,p,q}(t))^2 \right)^{1/2}.$$

Because the support of $\psi_{L,k}(t-1)$ is of length 2 ($\sum_t \mathbf{I}(t \in \{\psi_{L,k}(t-1) \neq 0\}) = 2$ ), by using a similar analysis to that used in the proof of Theorem 3.5, we can easily verify that the term in the last inequality is $O_p(1)$. By Lemma 3.4, we can hence infer that

$$|\tilde{c}^{(s)}_{L,k,p} - c^{(s)}_{L,k,p}| \leq \sup_{l,k,p} |\tilde{b}_{l,k,p} - b_{l,k,p}| O_p(1) = O_p(\sqrt{\log(T-1)/n(T-1)}).$$

The proof of $\sup_{L,k,p} |\tilde{c}^{(u)}_{L,k,p} - c^{(u)}_{L,k,p}|$ being $O_p(\sqrt{\log(T-1)/n(T-1)})$ is similar and thus omitted. □

**Proof of Theorem 3.7:** To prove the assertion, we show, in a first part, that asymptotically no jump can be detected in the stability intervals if $\lambda_{n,T}$ satisfies Condition *c*.1. In a second part, we show that all existing jumps must be asymptotically identified if $\lambda_{n,T}$ satisfies Condition *c*.2.

We begin with defining the following sets for each $p \in \{1, \ldots, \underline{P}\}$:

$$\begin{aligned}
\mathcal{J}_p &:= \{\tau_{1,p}, \ldots, \tau_{S_p,p}\}, \\
\mathcal{J}_p^c &:= \{1, \ldots, T\} \setminus \mathcal{J}_p, \\
\overline{\mathcal{J}}_p &:= \{2, 4, \ldots, T-1\} \cap \mathcal{J}_p, \\
\underline{\mathcal{J}}_p &:= \{3, 5, \ldots, T\} \cap \mathcal{J}_p, \\
\overline{\mathcal{J}}_p^c &:= \{2, 4, \ldots, T-1\} \setminus \overline{\mathcal{J}}_p, \text{ and} \\
\underline{\mathcal{J}}_p^c &:= \{3, 5, \ldots, T\} \setminus \underline{\mathcal{J}}_p.
\end{aligned}$$

Here, $\mathcal{J}_p$ is the set of all jump locations for parameter $\beta_{t,p}$, $\mathcal{J}_p^c$ is its complement, which contains only the stability intervals, $\overline{\mathcal{J}}_p$ is the set of all even jump locations and $\underline{\mathcal{J}}_p$ is the set of all odd jump locations so that $\overline{\mathcal{J}}_p \cap \underline{\mathcal{J}}_p = \emptyset$ and $\overline{\mathcal{J}}_p \cup \underline{\mathcal{J}}_p = \mathcal{J}_p$. Finally, the sets $\overline{\mathcal{J}}_p^c$ and $\underline{\mathcal{J}}_p^c$ define the complements of $\overline{\mathcal{J}}_p$ and $\underline{\mathcal{J}}_p$, respectively.

Define the event

$$\omega_{n,T} := \{ \sup_{t \in \mathcal{J}_p^c,\, p \in \{1,\ldots,P\}} \{|\Delta\tilde{\beta}_{t,p}^{(u)}| \mathbf{I}_{\overline{\mathcal{J}}_p^c} + |\Delta\tilde{\beta}_{t,p}^{(s)}| \mathbf{I}_{\underline{\mathcal{J}}_p^c}\} = 0\},$$

where $\mathbf{I}_{\overline{\mathcal{J}}_p^c} = \mathbf{I}(t \in \overline{\mathcal{J}}_p^c)$, $\mathbf{I}_{\underline{\mathcal{J}}_p^c} = \mathbf{I}(t \in \underline{\mathcal{J}}_p^c)$ and $\mathbf{I}(.)$ is the indicator function.

To prove that no jump can be identified in the stability intervals, we have to show, that $P(\omega_{n,T}) \to 1$, if $\sqrt{\frac{n(T-1)}{\log(T-1)}}\lambda_{n,T} \to \infty$, as $n, T \to \infty$ or as $n \to \infty$ and $T$ is fixed. Note that $\overline{\mathcal{J}}_p^c$ and $\underline{\mathcal{J}}_p^c$ are adjacent.

Let's now start with the no-jump case in $\overline{\mathcal{J}}_p^c$. By construction, we have, for all $t \in \{2, 4, \ldots, T-1\}$,

$$\Delta\tilde{\beta}_{t,p}^{(u)} = \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t)\hat{c}_{L,k,p}^{(u)}$$

Recall that at $l = L$, the construction of the wavelets basis implies that at each $t \in \{2, 4, \ldots, T-1\}$ there is only one differenced basis $\Delta\psi_{L,k}(t)$ that is not zero. Let $\mathcal{K}_p^c = \{k | \Delta\psi_{L,k}(t) \neq 0,\, t \in \overline{\mathcal{J}}_p^c\} = \{k | \Delta\psi_{L,k}(t-1) \neq 0,\, t \in \underline{\mathcal{J}}_p^c\}$. We can infer that $\{\sup_{t \in \overline{\mathcal{J}}_p^c} | \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t)\hat{c}_{L,k,p}^{(u)}| = 0\}$ occurs only if $\{\sup_{k \in \mathcal{K}_p^c} |c_{L,k,p}^{(u)}| = 0\}$ occurs.

By analogy, we can show the same assertion for the complement set $\underline{\mathcal{J}}_p^c$, i.e., $\{\sup_{t \in \underline{\mathcal{J}}_p^c} |\Delta\tilde{\beta}_{t,p}^{(s)}| = 0\}$ occurs only if $\{\sup_{k \in \mathcal{K}_p^c} |\hat{c}_{L,k,p}^{(s)}| = 0\}$ occurs.

To study $P(\omega_{n,T})$, it is hence sufficient to study

$$P(\sup_{k \in \mathcal{K}_p^c, m, p \in \{1,\ldots,P\}} |\hat{c}_{L,k,p}^{(m)}| = 0) \quad = \quad P(\sup_{k \in \mathcal{K}_p^c, m, p \in \{1,\ldots,P\}} |\tilde{c}_{L,k,p}^{(m)}| < \lambda_{n,T}).$$

By Lemma 3.6, $\sup_{k\in\mathcal{K}_p^c,m,p\in\{1,\ldots,P\}} |\tilde{c}_{L,k,p}^{(m)}| = O_p\big(\sqrt{\log(T-1)/n(T-1)}\big)$, since $c_{L,k,p}^{(m)} = 0$, for all $k \in \mathcal{K}_p^c$, and $p \in \{1,\ldots,P\}$. Thus, if $\sqrt{\frac{n(T-1)}{\log(T-1)}}\lambda_{n,T} \to \infty$, as $n, T \to \infty$ or $n \to \infty$ and $T$ is fixed, then $P(\omega_{n,T}) \to 1$.

To complete the proof and demonstrate that all true jumps will be asymptotically identified, we suppose that there exists a jump location $\tau_{j,p} \in \overline{\mathcal{J}}_p \cup \underline{\mathcal{J}}_p$ for at least one $p \in \{1,\ldots,P\}$ that is not detected and show the contradiction. If $\tau_{j,p} \in \overline{\mathcal{J}}_p$, then

$$|\Delta\tilde{\beta}_{\tau_{j,p},p}^{(u)}|\mathbf{I}_{\overline{\mathcal{J}}_p} + |\Delta\tilde{\beta}_{\tau_{j,p},p}^{(s)}|\mathbf{I}_{\underline{\mathcal{J}}_p} = |\Delta\tilde{\beta}_{\tau_{j,p},p}^{(u)}|.$$

Adding and subtracting $\Delta\beta_{\tau_{j,p},p}^{(u)}$, we get

$$\begin{aligned}\Delta\tilde{\beta}_{\tau_{j,p},p}^{(u)} &= \sum_{k=1}^{K_L}\Delta\psi_{L,k}(\tau_{j,p})(\tilde{c}_{L,k,p}^{(u)} - c_{L,k,p}^{(u)}) - \sum_{k=1}^{K_L}\Delta\psi_{L,k}(\tau_{j,p})\tilde{c}_{L,k,p}^{(u)}\mathbf{I}(\tilde{c}_{L,k,p}^{(u)} < \lambda_{n,T}) \\ &\quad + \sum_{k=1}^{K_L}\Delta\psi_{L,k}(\tau_{j,p})c_{L,k,p}^{(u)}, \\ &= I + II + III.\end{aligned}$$

By Lemma 3.6, $I = o_p(1)$, $II = o_p(1)$ as long as $\sqrt{T-1}\lambda_{n,T} \to 0$, and $III \neq 0$ because $\sum_{k=1}^{K_L}\Delta\psi_{L,k}(t)c_{L,k,p}^{(u)} = \Delta\beta_{\tau_{j,p},p}^{(u)} \neq 0$. The probability of getting $\Delta\tilde{\beta}_{\tau_{j,p},p}^{(u)} = 0$ converges hence to zero.

If $\tau_{j,p} \in \underline{\mathcal{J}}_p$, then

$$|\Delta\tilde{\beta}_{\tau_{j,p},p}^{(u)}|\mathbf{I}_{\overline{\mathcal{J}}_p} + |\Delta\tilde{\beta}_{\tau_{j,p},p}^{(s)}|\mathbf{I}_{\underline{\mathcal{J}}_p} = |\Delta\tilde{\beta}_{\tau_{j,p},p}^{(s)}|.$$

The prove is similar to the case of $\tau_{j,p} \in \overline{\mathcal{J}}_p$ and thus omitted. This completes the proof. □

**Proof of Theorem 3.8:** Recall that the post-Wavelet estimator is obtained by replacing the set of the true jump locations $\tau_{1,1},\ldots,\tau_{S_1+1,1},\ldots,\tau_{1,P},\ldots,\tau_{S_P+1,P}$ in $\hat{\beta}_{(\tau)} = (\hat{\beta}_{\tau_{1,1}},\ldots,\hat{\beta}_{\tau_{S_1+1,1}},\ldots,\hat{\beta}_{\tau_{1,P}},\ldots,\hat{\beta}_{\tau_{S_P+1,P}})'$ by the estimated jump locations $\tilde{\tau} := \{\tilde{\tau}_{j,p}|j \in \{1,\ldots,S_p+1\},\ p \in \{1,\ldots,P\}\}$, given $\tilde{S}_1 = S_1,\ldots,\tilde{S}_p = S_p$. By using Theorem 3.7, we can infer that, conditional on $\tilde{S}_1 = S_1,\ldots,\tilde{S}_p = S_p$,

$$\sqrt{n}\mathcal{T}_{(\tilde{\tau})}^{\frac{1}{2}}\hat{\beta}_{(\tilde{\tau})} = \sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\hat{\beta}_{(\tau)} + o_p(1).$$

To study the asymptotic distribution of $\sqrt{n}\mathcal{T}_{(\tilde{\tau})}^{\frac{1}{2}}\hat{\beta}_{(\tilde{\tau})}$ it is hence sufficient to study $\sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\hat{\beta}_{(\tau)}$.

$$\begin{aligned}\hat{\beta}_{(\tau)} &= \big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}'_{it,(\tau)}\big)^{-1}\big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{Y}_{it}\big)\\ &= \beta_{(\tau)} + \big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}'_{it,(\tau)}\big)^{-1}\big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{e}_{it}\big).\end{aligned}$$

Scaling by $\sqrt{n}\mathcal{T}_{(\tilde{\tau})}^{\frac{1}{2}}$ and rearranging, we get

$$\sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\big(\hat{\beta}_{(\tilde{\tau})} - \beta_{(\tau)}\big) = \left((n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}'_{it,(\tau)}\right)^{-1}\left(\big(n\mathcal{T}_{(\tau)}\big)^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{e}_{it}\right).$$

By Assumption E, the first term on the right hand side converges in probability to $Q^{\circ}_{(\tau)}$ and the second term converges in distribution to $N(0, V^{\circ}_{(\tau)})$. Slutsky's rule implies

$$\sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\big(\hat{\beta}_{(\tau)} - \beta_{(\tau)}\big) \xrightarrow{d} N(0, (Q^{\circ}_{(\tau)})^{-1}\big(V^{\circ}_{(\tau)}\big)(Q^{\circ}_{(\tau)})^{-1}).$$

It follows

$$\sqrt{n}\mathcal{T}_{(\tilde{\tau})}^{\frac{1}{2}}\big(\hat{\beta}_{(\tilde{\tau})} - \beta_{(\tau)}\big) = \sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\big(\hat{\beta}_{(\tau)} - \beta_{(\tau)}\big) + o_p(1) \xrightarrow{d} N(0, (Q^{\circ}_{(\tau)})^{-1}\big(V^{\circ}_{(\tau)}\big)(Q^{\circ}_{(\tau)})^{-1}).$$

This completes the Proof. □

**Proof of Proposition 3.9** Consider $c = 1$ (the case of homoscedasticity without presence of auto- and cross-section correlation). Because by Assumption E, we know that

$$(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}'_{it,(\tau)} \xrightarrow{p} Q^{\circ}_{(\tau)} \text{ and}$$

$$(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{j=1}^{n}\sum_{s=2}^{T} Z_{it,(\tau)}Z'_{js,(\tau)}\sigma_{ij,ts} \xrightarrow{p} V^{\circ}_{(\tau)},$$

it suffices to prove that

$$\hat{V}^{(1)}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})}\hat{\sigma}^2 \xrightarrow{p} V^{(1)}_{(\tau)},$$

where $V_{(\tau)}^{(1)} = (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}Z_{it,(\tau)}'\sigma^2$, with $\sigma^2 = E_c(\Delta\dot{e}_{it})$.

$$\begin{aligned}
\hat{V}_{(\tilde{\tau})}^{(1)} - V_{(\tau)}^{(1)} &= (\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\hat{e}_{it}^2 - \sigma^2)(n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})}', \\
&= +\sigma^2\left((n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})}' - (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}Z_{it,(\tau)}'\right), \\
&= a + b.
\end{aligned}$$

From Assumption B(ii), we can infer

$$\begin{aligned}
||a|| &\leq (\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\hat{e}_{it}^2 - \sigma^2)\frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}||(\mathcal{T}_{(\tilde{\tau})})^{-1/2}Z_{it,(\tilde{\tau})}||^2, \\
&= \left(\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}((\Delta\hat{e}_{it}^2 - \Delta e_{it}^2) + (\Delta e_{it}^2 - \sigma^2))\right)\frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}||(\mathcal{T}_{(\tilde{\tau})})^{-1/2}Z_{it,(\tilde{\tau})}||^2,
\end{aligned}$$

From

$$\begin{aligned}
\Delta\hat{\dot{e}}_{it} &= \Delta\dot{Y}_{it} - \Delta\dot{X}_{it,(\tilde{\tau})}'\hat{\beta}_{(\tilde{\tau})}, \\
&= \Delta\dot{e}_{it} + \Delta\dot{X}_{it,(\tilde{\tau})}'(\beta_{(\tilde{\tau})} - \hat{\beta}_{(\tilde{\tau})}),
\end{aligned} \tag{B.10}$$

and by using Theorem 3.8 together with Assumption B(ii), we can show that

$$\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\hat{\dot{e}}_{it} - \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\dot{e}_{it} = o_p(1). \tag{B.11}$$

By the law of large numbers,

$$\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\dot{e}_{it} - \sigma^2 = o_p(1).$$

Thus, $||a|| = (o_p(1) + o_p(1))O_p(1) = o_p(1)$. Moreover, from Theorem 3.7, we can infer that, given $\tilde{S}_1 = S_1, \ldots, \tilde{S}_P = S_P$,

$$(n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})} = (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}Z_{it,(\tau)} + o_p(1).$$

Thus,

$$\hat{V}_{(\tilde{\tau})}^{(1)} - V_{(\tau)}^{(1)} = o_p(1).$$

Consider $c = 2$ (the case of cross-section heteroskedasticity without auto- and cross-section correlations). Because of Assumption E, it suffices to prove that

$$\hat{V}^{(2)}_{(\tilde{\tau})} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z_{it,(\tilde{\tau})} \hat{\sigma}_i^2 \xrightarrow{p} V^{(2)}_{(\tau)},$$

where $V^{(2)}_{(\tau)} = (n\mathcal{T}_{(\tau)})^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} Z_{it,(\tau)} Z'_{it,(\tau)} \sigma_i^2$, with $\sigma_i^2 = E_c(\Delta \dot{e}_{it})$.

$$\begin{aligned}
\hat{V}^{(2)}_{(\tilde{\tau})} - V^{(2)}_{(\tau)} &= \frac{1}{n} \sum_{i=1}^{n} (\hat{\sigma}_i^2 - \sigma_i^2) (\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z'_{it,(\tilde{\tau})}, \\
&= +\frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \left( (\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z'_{it,(\tilde{\tau})} - (\mathcal{T}_{(\tau)})^{-1} \sum_{t=2}^{T} Z_{it,(\tau)} Z'_{it,(\tau)} \right), \\
&= d + e.
\end{aligned}$$

$$\begin{aligned}
||d|| &\leq \frac{1}{n} \sum_{i=1}^{n} (\hat{\sigma}_i^2 - \sigma_i^2) \sum_{t=2}^{T} ||(\mathcal{T}_{(\tilde{\tau})})^{-1/2} Z_{it,(\tilde{\tau})}||^2, \\
&= \frac{1}{n} \sum_{i=1}^{n} ((\hat{\sigma}_i^2 - \frac{1}{(T-1)} \sum_{t=2}^{T} \Delta \dot{e}_{it}) - (\sigma_i^2 - \frac{1}{(T-1)} \sum_{t=2}^{T} \Delta \dot{e}_{it}) \frac{1}{n} \sum_{t=2}^{T} ||(\mathcal{T}_{(\tilde{\tau})})^{-1/2} Z_{it,(\tilde{\tau})}||^2.
\end{aligned}$$

From Equation (B.10), and Theorem 3.8, we can infer

$$\frac{1}{(T-1)} \sum_{t=2}^{T} \Delta \hat{\dot{e}}_{it} - \frac{1}{(T-1)} \sum_{t=2}^{T} \Delta \dot{e}_{it} = o_p(1) \nu_i, \tag{B.12}$$

where $\frac{1}{n} \sum_{i=1}^{n} |\nu_i| = O_p(1)$. Moreover,

$$\sigma_i^2 - \frac{1}{(T-1)} \sum_{t=2}^{T} \Delta \dot{e}_{it} = o_p(1) \mu_i, \tag{B.13}$$

where $\frac{1}{n} \sum_{i=1}^{n} |\mu_i| = O_p(1)$. Note that the first terms in (B.12) and (B.13) do not depend on $i$. By using Assumption B(ii), we can infer

$$\begin{aligned}
||d|| &\leq o_p(1) \frac{1}{n} \sum_{i=1}^{n} |\nu_i| \sum_{t=2}^{T} ||(\mathcal{T}_{(\tilde{\tau})})^{-1/2} Z_{it,(\tilde{\tau})}||^2 + o_p(1) \frac{1}{n} \sum_{i=1}^{n} |\mu_i| \sum_{t=2}^{T} ||(\mathcal{T}_{(\tilde{\tau})})^{-1/2} Z_{it,(\tilde{\tau})}||^2, \\
&= o_p(1) O_p(1) + o_p(1) O_p(1).
\end{aligned}$$

The proof of $e$ being $o_p(1)$ is similar to the proof of $b$ in the first part. This is because $\sigma_i^2$ does not affect the analysis.

The proof of $\hat{V}_{\tilde{\tau}}^{(3)}$ being $(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tau)}Z_{it,(\tau)}^{'}\sigma_t^2+o_p(1)$, with $\sigma_t^2=E_c(\Delta\dot{e}_{it})$ is conceptually similar and thus omitted.

Finally, consider $c=4$ (The case of cross-section and time heteroskedasticity without auto- and cross-section correlations). As in the previous cases, all we need is to prove that

$$\hat{V}_{(\tilde{\tau})}^{(4)}=(n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})}^{'}\Delta\hat{\dot{e}}_{it}^2\xrightarrow{p}V_{(\tau)}^{(4)},$$

where

$$V_{(\tau)}^{(4)}=(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tau)}Z_{it,(\tau)}^{'}\sigma_{it}^2,$$

with $\sigma_{it}^2=E_c(\Delta\dot{e}_{it})$.

$$\begin{aligned}\hat{V}_{(\tilde{\tau})}^{(4)}-V_{(\tau)}^{(4)} &= (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})}^{'}(\Delta\hat{\dot{e}}_{it}^2-\Delta\dot{e}_{it}^2)\\ &\quad+(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}(Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})}^{'}-Z_{it,(\tau)}Z_{it,(\tau)}^{'})\Delta\dot{e}_{it}^2\\ &\quad+(n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tau)}Z_{it,(\tau)}^{'}(\Delta\dot{e}_{it}^2-\sigma_{it}^2).\\ &= f+g+h.\end{aligned}$$

Cauchy-Schwarz inequality implies

$$||f||\leq\left((n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}||Z_{it,(\tau)}||^2\right)^{1/2}\left((n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta\hat{\dot{e}}_{it}^2-\Delta\dot{e}_{it}^2)\right)^{1/2}=o_p(1).$$

By using Theorem 3.8, we can also verify that $||g||=o_p(1)$. Finally, Cauchy-Schwarz, Assumption B(ii), the law of large numbers implies that $||h||=o_p(1)$. It follows

$$\hat{V}_{(\tilde{\tau})}^{(4)}\xrightarrow{p}V_{(\tau)}^{(4)}.$$

This completes the proof. □

# Bibliography


Acharya, V. and Pedersen, L. (2005). Asset Pricing with Liquidity Risk. *Journal of Financial Economics*, 77(2):375–410.

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

Ahn, S., Lee, Y. H., and Schmidt, P. (2013). Panel Data Models with Multiple Time-varying Individual Effects. *Journal of Econometrics*, 174(1):1 – 14.

Ahn, S. C. and Horenstein, A. R. (2009). Eigenvalue Ratio Test for the Number of Factors.

Aït-Sahalia, Y. and Jacod, J. (2009). Testing for Jumps in a Discretely Observed Process. *The Annals of Statistics*, 37(1):184–222.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved Penalization for Determining the Number of Factors in Approximate Factor Models. *Statistics and Probability Letters*, 80(23-24):1806–1813.

Amato, J. D. and Remolona, E. M. (2003). The Credit-spread Puzzle. *BIS Quarterly Review*, (December):51–64.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and Forcasting Realized Volatility. *Econometrica*, 71:529–626.

Anderson, T. W. and Hsiao, C. (1981). Estimation of Dynamic Models With Estimation Error Components. *Journal of the American Statistical Association*, 76(375):598–606.

Andrews, D. W. (1993). Tests For Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, 61(4):821–856.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Carlo Application to Data : Evidence and an Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Bada, O. and Kneip, A. (2014). Parameter Cascading for Panel Models with Unknown Number of Heterogeneous Factors : An Application to the Credit Spread Puzzle. *Computational Statistics and Data and Analysis*, 76:95–115.

Bada, O. and Liebl, D. (2014a). phtt: Panel Data Analysis with Heterogeneous Time Trends Version. *R package version 2.07. 98*.

Bada, O. and Liebl, D. (2014b). The R-package phtt: Panel Data Analysis with Hererogeneous Time Trends. *Journal of Statistical Software*, 59(6).

Bai, J. (1997). Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics*, 79:551–563.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1):135–171.

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. (2010). Common Breaks in Means and Variances for Panel Data. *Journal of Econometrics*, 157(1):78–92.

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

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

Bai, J. and Ng, S. (2007). Determining the Number of Primitive Shocks in Factor Models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66(1):47–78.

Bai, J. and Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18(1):1–22.

Baltagi, B. H. and 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*, 68(1):148–155.

Baltagi, B. H. and Li, D. (2006). Prediction in the Panel Data Model with Spatial Correlation: the Case of Liquor. In *Spatial Economic Analysis*, volume 1, pages 175–185.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., and Dai, B. (2014). lme4: Linear Mixed-effects Models Using Eigen and S4 Maintainer.

Bessembinder, H., Maxwell, W., and Venkataraman, K. (2006). Optimal Market Transparency : Evidence from the Initiation of Trade Reporting in Corporate Bonds. *Journal of Financial Economics*, 82:251–288.

Boehmer, E., Fong, K., and Wu, J. (2012). International Evidence on Algorithmic Trading. *Working paper*, 2012.

Brown, R., Durbin, J., and Evans, J. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192.

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

Carr, P. and Wu, L. (2003). What Type of Process Underlies Options ? A Simple Robust Test. *The Journal of Finance*, LVIII(6):2581–2610.

Castagnetti, C. and Rossi, E. (2011). Euro Corporate Bond Risk Factors. *Journal of Applied Econometrics*, 28(3):1099–1255.

Chan, L. K. C., Jegadeesh, N., and Lakonishok, J. (1996). Momentum Strategies. *The Journal of Finance*, 51(5):1681.

Chen, L., Lesmond, D. A., and Wei, J. (2002). Bond Liquidity Estimation and the Liquidity Effect in Yield Spreads. *Working paper.*

Chen, Y.-P., Huang, H.-C., and Tu, I.-P. (2010). A New Approach for Selecting the Number of Factors. *Computational Statistics & Data Analysis*, 54(12):2990–2998.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3).

Coakley, J., Fuertes, A.-m., and Smith, R. (2002). A Principal Components Approach to Cross-Section Dependence in Panels. *Working paper.*

Collin-Dufresne, P., Goldstein, R. S., and Martin, J. S. (2001). The Determinants of Credit Spread Changes. *The Journal of Finance*, 56(6):2177–2207.

Craven, P. and Wahba, G. (1979). Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerische Mathematik*, 403:377–403.

Croissant, Y. and Millo, G. (2008). Panel Data Econometrics in R : The plm Package. *Journal Of Statistical Software*, 27(2).

de Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, revised ed edition.

De Wachter, S. and Tzavalis, E. (2012). Detection of Structural Breaks in Linear Dynamic Panel Data Models. *Computational Statistics & Data Analysis*, 56(11):3020–3034.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, 81(3):425–55.

Dray, S. (2008). On the Number of Principal Components: A Test of Dimensionality Based on Measurements of Similarity between Matrices. *Computational Statistics & Data Analysis*, 52(4):2228–2237.

Elton, E. J., Gruber, M. J., Agrawal, D., and Mann, C. (2001). Explaining the Rate Spread on Corporate Bonds. *The Journal of Finance*, 56(1):247–277.

Gebhardt, W. R., Hvidkjaer, S., and Swaminathan, B. (2005). Stock and Bond Market Interaction: Does Momentum Spill over? *Journal of Financial Economics*, 75:651–690.

Greenaway-mcgrevy, R. (2012). Asymptotic Distribution of Factor Augmented Estimators for Panel Regression. *Journal of Econom*, 169(1):48–53.

Greenaway-McGrevy, R., Han, C., and Sul, D. (2012). Estimating the Number of Common Factors in Serially Dependent Approximate Factor Models. *Economics Letters*, 116(3):531 – 534.

H., B. B. (2005). *Econometric Analysis of Panel Data*. Sons, John Wiley & Sons Ltd The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, third edition.

Hallin, M. and Liška, R. (2007). Determining the Number of Factors in the General Dynamic Factor Model. *Journal of the American Statistical Association*, 102:603–617.

Han, A. K. and Park, D. (1989). Testing for Structural Change in Panel Data: Application to a Study of U.S. Foreign Trade in Manufacturing Goods. *The Review of Economics and Statistics*, 71(1):135–142.

Hasbrouck, J. and Saar, G. (2013). Low-latency Trading. *Journal of Financial Markets*, 16(4):646–679.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does Algorithmic Trading Improve Liquidity ? *Journal of finance*, LXVI(1):1–34.

Houweling, P., Mentink, A., and Vorst, T. (2005). Comparing Possible Proxies of Corporate Bond Liquidity. *Journal of Banking and Finance*, 29(6):1331–1358.

Huang, J.-z. and Huang, M. (2012). How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk ? 1. *Review of Asset Pricing Studies*, 2(2).

Inoue, A. and Rossi, B. (2011). Identifying the Sources of Instabilities in Macroeconomic Fluctuations. *The Review of Economics and Statistics*, 93(4):1186–1204.

Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91.

Josse, J. and Husson, F. (2012). Selecting the Number of Components in Principal Component Analysis Using Cross-Validation Approximations. *Computational Statistics & Data Analysis*, 56(6):1869–1879.

Kagraoka, Y. (2010). A Time-varying Common Risk Factor Affecting Corporate Yield Spreads. *The European Journal of Finance*, 16(6):527–539.

Kapetanios, G. (2010). A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets. *Journal of Business & Economic StatisticsEconomic Statistics*, 28(3).

Karavias, Y. and Tzavalis, E. (2012). Testing for Unit Roots in Short Panels Allowing for a Structural Break. *Computational Statistics & Data Analysis*.

Kim, D. (2014). Common Breaks in Time Trends for Large Panel Data with a Factor Structure. *The Econometrics Journal*.

Kiviet, J. F. (1995). On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics*, 68:53–78.

Kneip, A., Sickles, R. C., and Song, W. (2012). A New Panel Data Treatment for Heterogeneity in Time Trends. *Econometric Theory*, 28:590–628.

Lesmond, D. a. (1999). A New Estimate of Transaction costs. *Review of Financial Studies*, 12(5):1113–1141.

Longstaff, F. A., Mithal, S., and Neis, E. (2005). Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit-Default Swap Market. *The Journal of Finance*, LX(5).

Lucas, R. E. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

Millo, G. and Piras, G. (2012). splm : Spatial Panel Data Models in R. *Journal Of Statistical Software*, 47(1).

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708.

Onatski, A. (2009). Testing Hypotheses About the Number of Factors in Large Factor Models. *Econometrica*, 77(5):1447–1479.

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

Onatski, A. (2012). Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors. *Journal of Econometrics*, 168:244–258.

Pauwels, L. L., Chan, F., and Mancini Griffoli, T. (2012). Testing for Structural Change in Heterogeneous Panels with an Application to the Euro's Trade Effect. *Journal of Time Series Econometrics*, 4(2).

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

Pesaran, M. H., Pick, A., and Pranovich, M. (2011). Optimal Forecasts in the Presence of Structural Breaks. *Working paper*, (327).

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and Team, R. C. (2014). nlme: Linear and Nonlinear Mixed Effects Models.

Qu, Z. and Perron, P. (2007). Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica*, 75(2):459–502.

Stock, J. H. and Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2005). Implications of Dynamic Factor Models for VAR Analysis. *Working paper*.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.