Panel Data Models with Multiple Jump Discontinuities in the Parameters

Bada O.*[1], Gualtieri J.†[2], Kneip A. ‡[1] and Sickles R. C.§[2]

[1]Institute for Financial Economics and Statistics, University of Bonn
[2]Department of Economics, Rice University

This version: March 27, 2015

Abstract

In recent years, many large panel data models have been developed to make greater use of the information available in such data structures. While a vast literature on structural break change point analysis exists for univariate time series, research on large panel data models has not been as extensive. In this paper, 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. Our method is related to the Haar wavelet technique; we adjust it according to the structure of the observed variables in order to detect the change points of the parameters consistently. The asymptotic property of our estimator is established. In our application, we examine the impact of algorithmic trading on standard measures of market quality such as liquidity and volatility over a large time period. We propose to detect jumps in regression slope parameters automatically and to alleviate concerns about ad-hoc subsample selection so as to examine the effect of algorithmic trading on market quality in different market situations.

*bada@uni-bonn.de
†james.n.gualtieri@gmail.com
‡akneip@uni-bonn.de
§rsickles@rice.edu

1 Introduction

Panel datasets with large cross-sectional dimensions and large periods of time observations are becoming increasingly 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 rarely addressed in the existing literature and that is 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 considered by Lucas (1976), who points out the risk of naively predicting 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. Of course, the larger the observation period, the more likely the occurrence of such shocks.

In this paper, 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. Given a dependent variable Y_{it} observed for $i = 1, \dots, n$ individuals at $t = 1, \dots, T$ time points, we consider the model

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

where $\mathbf{I}(.)$ is the indicator function, $X_{it,p}$, $p = 1, \dots, P$, are explanatory variables, α_i is an individual specific effect, θ_t is a common time parameter, and ε_{it} is an unobserved idiosyncratic term that may be correlated with one or more explaining variables. For each variable $X_{it,p}$, $p = 1, \dots, P$, the corresponding slope parameter is piecewise constant with an unknown set of jump points $\{\tau_{0,p}, \tau_{1,p}, \dots, \tau_{S_p+1,p} | \ \tau_{0,p} = 1 < \tau_{1,p} < \dots < \tau_{S_p+1,p} = T\} \subseteq \{1, \dots, T\}$ for some unknown $S_p \geq 1$.

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 improved statistical efficiency, the determination of the change point locations can be of particular importance in many applications. Indeed, estimating the number and locations of the structural breaks alleviates concerns about ad-hoc subsample selection, enables interpretation of historical events that are not explicitly considered in the model, and avoids related issues of statistical under- or over-parametrization.

One of the earliest contributions to the literature on testing for structural breaks in panel data 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 likelihood ratio-type statistics based on Gaussian errors. The method requires, however, that the number of equations 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. However, 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 heteroskedasticity in the idiosyncratic errors is established and the convergence rate of our slope estimator is derived. To consistently detect 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 paper is organized as follows. Section 2 explains the basic idea of our estimation procedure by using a relatively straightforward centered univariate panel model. In Section 3, we consider panel models with unobserved effects and multiple jumping slope parameters, present our model assumptions, and derive the main asymptotic results. Section 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 5, we discuss models with an issue of omitted common factors and endogenous models arising from structural simultaneous equation systems. Section 6 presents the simulation results of our Monte Carlo experiments. Section 7 focuses on the empirical application. The conclusion follows in Section 8. The mathematical proofs are collected in Appendix 9.

2 Basic Concepts

A Simple Panel Model with one Jumping Parameter

To simplify exposition of our basic approach, we begin with a special case of model (1). We consider a centered univariate panel data model of the form

$$Y_{it} = X_{it}\beta_t + e_{it} \quad \text{for} \quad i \in \{1, \dots, n\} \quad \text{and} \quad t \in \{1, \dots, T\}, \tag{2}$$

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

$$\beta_t = \sum_{j=1}^{S+1} \mathbf{I}\big(\tau_{j-1} < t \leq \tau_j\big)\beta_{\tau_j}, \quad t = 1, \dots, T \tag{3}$$

for some $\tau_0 = 1 < \tau_1 < \ldots < \tau_{S+1} = T$, where $S \geq 1$, $\tau_1, \ldots, \tau_S$, as well as the coefficients β_{τ_j}, $j = 1, \ldots, S+1$, are unknown.

Some Fundamental Concepts of Wavelet Transform

The idea behind our approach basically consists of using the Haar wavelet expansion of β_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

$$\psi_{l,k}(t) = a_l^{\psi} I_{l,2k-1}(t) - a_l^{\psi} I_{l,2k}(t) \quad \text{and} \tag{4}$$

$$\varphi_{l_0,k}(t) = a_{l_0}^{\varphi} I_{l+1,2k-1}(t) + a_{l_0}^{\varphi} I_{l+1,2k}(t), \tag{5}$$

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 β_t can be presented as follows:

$$\beta_t = \sum_{k=1}^{K_{l_0}} \varphi_{l_0,k}(t) d_{l_0,k} + \sum_{l=l_0+1}^{L} \sum_{k=1}^{K_l} \psi_{l,k}(t) c_{l,k}, \qquad \text{for } \ t \in \{1, \ldots, T\}, \tag{6}$$

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 β_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 sudden piecewise changes.

Orthonormalization and Estimation

Note that the collection of functions, in (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 β_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:

$$\beta_t = \sum_{l=1}^{L}\sum_{k=1}^{K_l} w_{l,k}(t) b_{l,k} \qquad \text{for } \ t \in \{1, \ldots, T\}, \tag{7}$$

where

$$w_{l,k}(t) = \begin{cases} a_{1,1} = a_{2,1}h_{2,1}(t) + a_{2,2}h_{2,2}(t) & \text{if} \quad l = 1, \text{ and} \\ a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t) & \text{if} \quad l > 1, \end{cases} \tag{8}$$

for some positive standardizing scales $a_{l,2k-1}$ and $a_{l,2k}$ that, unlike the conventional wavelets, depend on both the dilation level l and translation index k. Their exact form will be discussed in detail below. We define the function $h_{l,m}(t)$ as follows:

$$h_{l,m}(t) = \sqrt{2^{l-2}} I_{l,m}(t). \tag{9}$$

The most appealing feature of the expansion (7) (and (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 β_t can be sensed at each dilation level by at most one translation function. Proposition 1 states the existence of (7) for any arbitrary positive real scales $a_{l,2k}$ and $a_{l,2k-1}$.

Proposition 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)*. 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* (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 (7), we can rewrite Model (2) as

$$Y_{it} = \sum_{l=1}^{L}\sum_{k=1}^{K_l} \mathcal{X}_{l,k,it} b_{l,k} + e_{it}, \tag{10}$$

where

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

In vector notation,

$$Y_{it} = \mathcal{X}_{it}' b + e_{it}, \tag{11}$$

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 $\mathcal{Z}_{it} = (\mathcal{Z}_{1,1,it}, \dots, \mathcal{Z}_{L,K_L,it})'$. 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

$$E\big(\mathcal{Z}_{it}(Y_{it} - \mathcal{X}_{it}^{'}b)\big) = 0. \tag{12}$$

The IV estimator of b (hereafter, denoted by $\tilde{b}$) is obtained by solving the empirical counterpart of (12), i.e.,

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\big(\mathcal{Z}_{it}(Y_{it} - \mathcal{X}_{it}^{'}\tilde{b})\big) = 0. \tag{13}$$

Remark 1 *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. Hence, our estimator of b can be seen as the GMM estimator:*

$$\tilde{b} = \arg\min_{b} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \mathcal{X}_{it}^{'}b)\mathcal{Z}_{it}^{'}\mathcal{W}_T\mathcal{Z}_{it}(Y_{it} - \mathcal{X}_{it}^{'}b), \tag{14}$$

where $\mathcal{W}_T$ is an arbitrary symmetric $(T \times T)$ full rank matrix. Since the choice of $\mathcal{W}_T$ in the just-identified case is irrelevant, we can use the identity matrix to solve (14).

Under general assumptions, we can state the consistency of $\tilde{b}$ for any arbitrary collection of wavelet functions. However, the problem with naively using the conventional basis functions is that the identification of the zero- and non-zero coefficients will be ambiguous. Not only will the presence of the error term in (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 solution consists of adjusting the scales $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ in (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 3, in Appendix 10, gives the mathematical conditions for $a_{1,1}, a_{l,2k-1}$ and $a_{l,2k}$ to ensure (a) and (b). The solution is

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

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 (13) (or (14)) with respect to $b_{l,k}$ under Restrictions (a) and (b), we obtain

$$\tilde{b}_{l,k} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \mathcal{Z}_{l,k,it}Y_{it}. \tag{15}$$

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

$$\tilde{\beta}_t = \sum_{l=1}^{L}\sum_{k=1}^{K_l} w_{l,k}(t)\hat{b}_{l,k}, \tag{16}$$

where

$$\hat{b}_{l,k} = \begin{cases} \tilde{b}_{l,k} & \text{if} \quad |\tilde{b}_{l,k}| > \lambda_{n,T} \text{ and} \\ 0 & \text{else,} \end{cases} \tag{17}$$

for some threshold $\lambda_{n,T}$ that depends on n and T. Theorems 1, 2 give the necessary conditions for $\lambda_{n,T}$ to ensure consistency under Assumptions A-C presented in Section 3.

Remark 2 *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,*

$$\hat{b} = \arg\min_{b} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \mathcal{X}_{it}'b)\mathcal{Z}_{it}'\mathcal{W}_T\mathcal{Z}_{it}(Y_{it} - \mathcal{X}_{it}'b) + \lambda_{n,T}|b|, \tag{18}$$

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 (18) *and* (17) *lead to the same result independently of the choice of $\mathcal{W}_T$.*

Now that we have developed procedures for first step SAW estimation for a straightforward centered panel model we turn to generalizations for multivariate models with unobserved heterogeneity effects as well as corresponding post-SAW procedures.

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 rewrite Model (1) as

$$Y_{it} = \mu + X_{it}^{'}\beta_t + \alpha_i + \theta_t + e_{it}, \tag{19}$$

where $X_{it} = (X_{1,it}, \ldots, X_{P,it})^{'}$ is the $(P \times 1)$ vector of regressors, $\beta_t = (\beta_{t,1} \ldots, \beta_{t,P})^{'}$ is a $(P \times 1)$ is the unknown vector of slope parameters, and for each $\beta_{t,p}$, $p \in \{1, \ldots, P\}$, we have

$$\beta_{t,p} = \sum_{j=1}^{S_p+1} \mathbf{I}\big(\tau_{j-1,p} < t \leq \tau_{j,p}\big)\beta_{\tau_j,p}, \quad t = 1, \ldots, T \tag{20}$$

The estimation procedure for this model is conceptually analogous to the univariate method discussed in Section 2. However, besides the need to deal with multivariate wavelets, we have to control for the additional unknown parameters μ, α_i, and θ_t.

From the literature on panel models, we know that uniqueness of μ, α_i, and θ_t requires the following identification conditions:

$$\begin{array}{lrcl} \text{C.1:} & \sum_{i=1}^{n} \alpha_i & = & 0, \text{ and} \\ \text{C.2:} & \sum_{t=1}^{T} \theta_t & = & 0. \end{array} \tag{21}$$

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 (19), we have the expression

$$(Y_{it} - Y_{it-1}) = X_{it}^{'}\beta_t - X_{it-1}^{'}\beta_{t-1} + (\theta_t - \theta_{t-1}) + (e_{it} - e_{it-1}), \tag{22}$$

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

We can eliminate the term $\Delta\theta_t = \theta_t - \theta_{t-1}$ 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 β_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 θ_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 (22) as

$$\begin{aligned} \Delta Y_{it} &= (X_{it}^{'}, -X_{it-1}^{'}, 1)\begin{pmatrix} \beta_t \\ \beta_{t-1} \\ \Delta\theta_t \end{pmatrix} + \Delta e_{it}, \\ &= \underline{X}_{it}^{'}\gamma_t + \Delta e_{it}, \end{aligned} \tag{23}$$

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

By using multivariate structure adapted wavelet functions, we can estimate γ_t using similar methods to those discussed in Section 2.

The multivariate structure adapted wavelet expansion of γ_t can be written as:

$$\gamma_t = \sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{lk}(t)\underline{b}_{l,k} \qquad \text{for } \; t \in \{2, \ldots, T\}, \tag{24}$$

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

$$W_{l,k}(t) = \begin{cases} A_{1,1} = A_{2,1}H_{2,1}(t) + A_{2,2}H_{2,2}(t) & \text{if} \quad l = 1, \;\text{ and} \\ A_{l,2k-1}H_{l,2k-1}(t) - A_{l,2k}H_{l,2k}(t) & \text{if} \quad l > 1, \end{cases} \tag{25}$$

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 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

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

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

$$
\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left(\underline{\mathcal{Z}}_{l,k,it}(\Delta Y_{it} - \underline{\mathcal{X}}_{l,kit}'\tilde{\underline{b}}_{l,k})\right) = 0. \tag{26}
$$

Solving (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 pth 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

$$
\hat{\gamma}_{t,p} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}} W_{lk,p,q}(t)\hat{b}_{l,k,q}, \tag{27}
$$

or

$$
\hat{\gamma}_{t+1,p+P} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{q=1}^{\underline{P}} W_{lk,p+P,q}(t+1)\hat{b}_{l,k,q}, \tag{28}
$$

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

$$
\hat{b}_{l,k,q} = \begin{cases} \tilde{b}_{l,k,q} & \text{if} \quad |\tilde{b}_{l,k,q}| > \lambda_{n,T} \text{ and} \\ 0 & \text{else.} \end{cases} \tag{29}
$$

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. Δ 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 $k\in\{1,\dots,K_l\}$,

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

where $\underline{Q}_{l,k}^{\circ}$ 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 $A_{l,2k} = \underline{Q}_{l,2k}^{-1}(\underline{Q}_{l,2k}^{-1}+\underline{Q}_{l,2k-1}^{-1})^{-1/2}$ and $A_{l,2k-1} = \underline{Q}_{l,2k-1}^{-1}(\underline{Q}_{l,2k}^{-1}+\underline{Q}_{l,2k-1}^{-1})^{-1/2}$, 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.,

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

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 1 *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 1 establishes the uniform and the mean square consistency of $\tilde{\gamma}_{t,p}$.

Theorem 1 *Assume Assumptions A-C, then the following statements hold:*

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

(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 1, can be constructed as follows:

$$\lambda_{nT}=\hat{V}_{nT}^{\frac{1}{2}}\left(\frac{2\log((T-1)\underline{P})}{n(T-1)^{1/\kappa}}\right)^{\kappa/2}, \text{ for some } \kappa\in]0,1[, \tag{30}$$

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 κ 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).

4 Post-SAW Procedures

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.

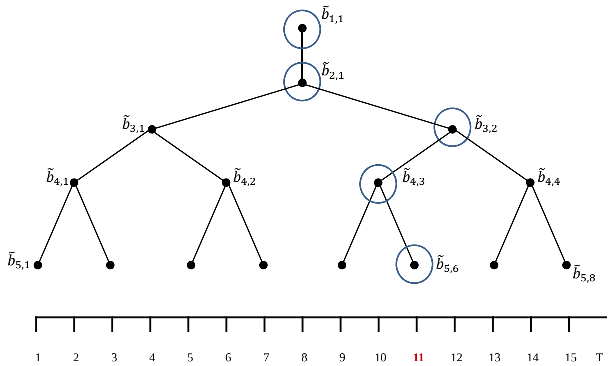

Figure 1: An illustrating example of a tree-structured representation for the wavelet coefficients.

As an illustration, consider the tree-structured representation in Figure 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). Our allowance for a heteroskedastic error process is in part to allow us to circumvent the size and power distortions of the Chow test discussed in Schmidt and Schmidt and Sickles (1977), among others.

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 1, we can see that the tree-structured representations of the shifted and non-shifted coefficients presented in Figure 2 support the hypothesis of only one jump at 11.

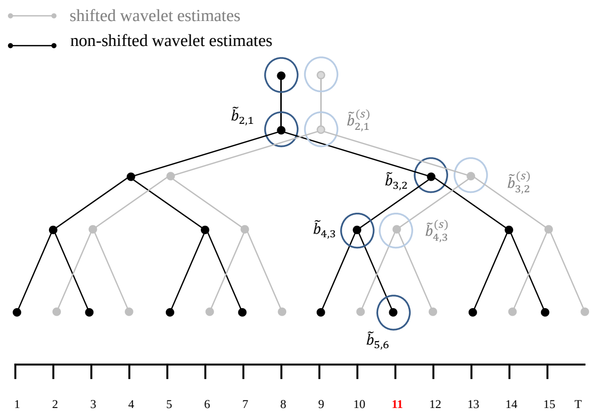

Figure 2: An illustrating example of a tree-structured representation for the shifted and non-shifted coefficients.

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 γ_t with the conventional univariate wavelet basis functions defined in (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, β_p, two sets of univariate wavelet coefficients:

$$c^{(s)}_{l,k,p} \quad = \quad \frac{1}{T-1} \sum_{t=2}^{T} \psi_{l,k}(t-1)\gamma_{t,p}, \tag{31}$$

and

$$c^{(u)}_{l,k,p} = \frac{1}{T-1}\sum_{t=1}^{T-1}\psi_{l,k}(t)\gamma_{t+1,p+P}. \tag{32}$$

We use the superscripts (s) and (u) in (31) and (32) to denote the shifted and non-shifted coefficients, respectively.

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

$$\tilde{c}^{(s)}_{l,k,p} = \frac{1}{T-1}\sum_{t=2}^{T}\psi_{l,k}(t-1)\tilde{\gamma}_{t,p}, \tag{33}$$

and

$$\tilde{c}^{(u)}_{l,k,p} = \frac{1}{T-1}\sum_{t=1}^{T-1}\psi_{l,k}(t)\tilde{\gamma}_{t+1,p+P}. \tag{34}$$

Having an appropriate threshold for $\tilde{c}^{(u)}_{l,k,p}$, 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 $\tilde{c}^{(s)}_{L,k,p}$ and $\tilde{c}^{(u)}_{L,k,p}$ and states their order of magnitude in probability.

Lemma 2 *Suppose Assumptions A-C hold, then, for all* $p \in \{1, \dots, P\}$ *and* $m \in \{m, s\}$

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

From Lemma 2, 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 2 gives the necessary asymptotic conditions to ensure consistency of the jump selection method.

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 (33) and (34) to construct a consistent selection method for detecting the jump locations.

We use (33) and (34) to obtain the following two estimators of $\Delta\beta_t$:

$$\Delta\tilde{\beta}^{(u)}_{t,p} = \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t)\hat{c}^{(u)}_{L,k,p}, \quad \text{for } t \in \mathcal{E}, \tag{35}$$

and

$$\Delta\tilde{\beta}^{(s)}_{t,p} = \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t-1)\hat{c}^{(s)}_{l,k,p}, \quad \text{for } t \in \mathcal{E}^c, \tag{36}$$

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

The number of jumps of each parameter can be estimated by

$$\tilde{S}_p = \sum_{t\in\mathcal{E}} \mathbf{I}(\Delta\tilde{\beta}^{(u)}_{t,p} \neq 0) + \sum_{t\in\mathcal{E}^c} \mathbf{I}(\Delta\tilde{\beta}^{(s)}_{t,p} \neq 0). \tag{37}$$

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

$$\tilde{\tau}_{j,p} = \min\left\{ s \,\middle|\, j = \sum_{t=2}^{s} \mathbf{I}\left(\Delta\tilde{\beta}^{(u)}_{t,p} \neq 0,\, t \in \mathcal{E}\right) + \sum_{t=3}^{s} \mathbf{I}\left(\Delta\tilde{\beta}^{(s)}_{t,p} \neq 0,\, t \in \mathcal{E}^c\right) \right\}, \tag{38}$$

for $j \in \{1, \dots, \tilde{S}_p\}$. 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 2 *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)* $\lim_{n,T\to\infty} P(\tilde{S}_1 = S_1, \dots, \tilde{S}_p = S_p) = 1$ *and*
- *(ii)* $\lim_{n,T\to\infty} P(\tilde{\tau}_{1,1} = \tau_{1,1}, \dots, \tilde{\tau}_{S_P,P} = \tau_{S_P,P} | \tilde{S}_1 = S_1, \dots, \tilde{S}_p = S_p) = 1$.

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

4.3 Post-SAW Estimation

For known $\tau_{1,p}, \ldots, \tau_{S_p,p}$, we can rewrite Model (22) as

$$\Delta\dot{Y}_{it} = \sum_{p=1}^{P}\sum_{j=1}^{S_p+1} \Delta\dot{X}_{it,p}^{(\tau_{j,p})}\beta_{\tau_{j,p}} + \Delta\dot{e}_{it}, \tag{39}$$

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 (39) can be rewritten as

$$\Delta\dot{Y}_{it} = \Delta\dot{X}_{it,(\tau)}^{\prime}\beta_{(\tau)} + \Delta\dot{e}_{it}, \tag{40}$$

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

$$\hat{\beta}_{(\tau)} = \big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{X}_{it,(\tau)}^{\prime}\big)^{-1}\big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{Y}_{it}\big). \tag{41}$$

Conditional on $\tilde{S}_1 = S_1, \ldots, \tilde{S}_P = S_P$, we can replace the set of the true jump locations τ in (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:

$$\hat{\beta}_{(\tilde{\tau})} = \big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}\Delta\dot{X}_{it,(\tilde{\tau})}^{\prime}\big)^{-1}\big(\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}\Delta\dot{Y}_{it}\big). \tag{42}$$

From (26) and (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) : $\left(n\mathcal{T}_{(\tau)}\right)^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}\Delta\dot{e}_{it} \xrightarrow{d} N(0, V^{\circ}_{(\tau)})$.

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

Theorem 3 *Suppose Assumptions A-E hold. Then conditional on* $\tilde{S}_1 = S_1, \ldots, \tilde{S}_p = S_p$*, we have*

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

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:

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

where $\hat{\sigma}_i^2 = \frac{1}{T-1} \sum_{t=2}^{T} \Delta\hat{e}_{it}^2$.

3. The case of time heteroskedasticity without auto- and cross-section correlations:

$$\hat{V}_{(\tilde{\tau})}^{(3)} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z_{it,(\tilde{\tau})}^{'} \hat{\sigma}_t^2,$$

where $\hat{\sigma}_t^2 = \frac{1}{n} \sum_{i=1}^{n} \Delta\hat{e}_{it}^2$.

4. The case of cross-section and time heteroskedasticity without auto- and cross-section correlations:

$$\hat{V}_{(\tilde{\tau})}^{(4)} = (n\mathcal{T}_{(\tilde{\tau})})^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} Z_{it,(\tilde{\tau})} Z_{it,(\tilde{\tau})}^{'} \Delta\hat{e}_{it}^2.$$

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

Remark 3 *If the errors (at the difference level) are autocorrelated,* $V_{(\tilde{\tau})}^{(c)}$ *can be estimated by applying the standard heteroskedasticity and autocorrelation (HAC) robust limiting covariance estimator to the sequence* $\{Z_{it,(\tilde{\tau})}\Delta\hat{e}_{it}\}_{i,t}$ *for* $i \in \mathbb{N}^*$ *and* $t \in \mathbb{N}^* \setminus \{1\}$*; 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.*

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 Λ_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 Λ_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 allow for the data generating process of $X_{p,it}$ to have the following rather general form:

$$X_{p,it} = \vartheta_{p,i}^{'} F_t + \Lambda_i^{'} G_{p,t} + a_p \Lambda_i^{'} F_t + \mu_{p,it}, \tag{43}$$

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 μ_{it} is an individual specific term that is uncorrelated with ϵ_{it}, Λ_i, ϑ_i, F_t and G_t.

Rearranging (43), we can rewrite $X_{p,it}$ as

$$X_{p,it} = \vartheta_{p,i}^{*'} G_{p,t}^{*} + \mu_{p,it}, \tag{44}$$

where

$$\vartheta_{p,i}^{*'} = H(a_p \Lambda_i^{'} + \vartheta_{p,i}^{'}, \Lambda_i^{'}), \tag{45}$$

and

$$G_{p,t}^{*} = H^{-1}(F_t^{'}, G_{p,t}^{'})^{'}, \tag{46}$$

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 (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

$$Z_{p,it} = X_{p,it} - \hat{\vartheta}^{*'}_{p,i}\hat{G}^{*}_{p,t}, \tag{47}$$

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 (47), we can estimate consistently the jumping slope parameters as before. A formal proof remains, of course, to be explored.

Two-Step SAW for Jump 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, e.g., is generated by a system of simultaneous equations.

Consider the following two-equation simultaneous equation system:

$$Y_{it} = \mu + \sum_{p=1}^{P} X_{p,it}\beta_{t,p} + \alpha_i + \theta_t + e_{it}, \tag{48}$$

and

$$X_{q,it} = b_t Y_{it} + \sum_{p \in \{1,\ldots,P\}\setminus\{q\}} X_{p,it}d_{t,p} + v + u_i + \vartheta_t + \nu_{it}, \tag{49}$$

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

Neglecting the structural form of $X_{q,it}$ in Equation (49) and estimating the regression function (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 (49)). A natural way to overcome this type of endogeneity problem is to use the fitted variable obtained from Equation (49) as an instrument after replacing Y_{it} with its expression in (48). However, our model involves an additional complication related to the time-changing character of $\beta_{t,q}$ and the presence of the unobservable heterogeneity effects that render such two-stage least squares estimators problematic. Inserting (48) in (49) and rearranging it leads to a panel model with time-varying unobservable individual effects:

$$X_{q,it} = \sum_{p\in\{1,\dots,P\}\setminus\{q\}} X_{p,it} d^*_{t,p} + \vartheta^*_{1t} + u_i \vartheta^*_{2t} + \alpha_i \vartheta^*_{3t} + \varepsilon_{it}, \tag{50}$$

where

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

Note that the regression model in (50) can be considered a spacial case of the model with multifactor errors discussed above. A potential instrument for $X_{q,it}$ in (48) is then

$$Z_{q,it} = \sum_{p\in\{1,\dots,P\}\setminus\{q\}} X_{p,it} \hat{d}^*_{t,p} + \hat{\vartheta}_i^{'} \hat{G}_t, \tag{51}$$

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, and which can be obtained from (50) by using the instruments proposed above to control for the omitted factor structure $\vartheta^*_{1t}+u_i\vartheta^*_{2t}+\alpha_i\vartheta^*_{3t}$.

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

$$\beta_t = \begin{cases} \beta_{\tau_1} & \text{for} \quad t \in \{1,\dots,\tau_1\}, \\ \vdots & \\ \beta_{\tau_S+1} & \text{for} \quad t \in \{\tau_S+1,\dots,T\}, \end{cases} \tag{52}$$

with

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

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

$$X_{it} = 0.5\alpha_i + \xi_{it}, \tag{53}$$

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 (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 (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 (53); homoscedasticity and autocorrelation in the errors such that

$$e_{it} = \rho_i e_{it-1} + \zeta_{it}, \tag{54}$$

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:

$$\begin{aligned} e_{it} &= \lambda_i f_t + \epsilon_{it} \text{ and} \\ X_{it} &= 0.3\alpha_i + 0.3\nu_t + 0.3\lambda_i f_t + \mu_{it}, \end{aligned} \tag{55}$$

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

$$\begin{array}{rcl} \epsilon_{it} & = & \rho_{e,i}\epsilon_{i,t-1} + \zeta_{e,it}, \\ \mu_{it} & = & \rho_{\mu,i}\mu_{i,t-1} + \zeta_{\mu,it}, \end{array} \tag{56}$$

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 1 -4 report the estimation results obtained by averaging the results of 1000 replications. The third, sixth, and ninth columns in Tables 1-3 report the averages of the estimated number of jumps $\tilde{S}$ detected by (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 β_t obtained in replication r. The fourth, seventh, and tenth columns in Tables 1 -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 $\frac{1}{1000}\sum_{r=1}^{1000}\left(\hat{\beta}_{(M)}^r - \beta\right)^2$, where $\hat{\beta}_{(M)}^r$ 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 4.

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 become large as well as when T is fixed and only n becomes large. The method performs quite well 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 1). In most of the examined cases, where heteroskedasticity in the cross-section and time dimension and/or week serial correlations exist, the method still behaves very well, in particular when n is large (see results of DGP3-DGP4 in Tables 1 and 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

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

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

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.

Table 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 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.

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, independently of the number and locations of the jumps.

DGP7 ($S=0$)							
Method		post-SAW		LSDV	GLS	ILS	KSS
n	T	$\tilde{S}$	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 4: Simulation results of the Monte Carlo experiments for DGP7. The entries are the averages of 1000 replications.

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. Incidents such as the "flash crash" of 2010, 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 effects of AT on spreads 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 that detection of possible negative effects of AT on market quality have not been empirically documented. 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.

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:

$$MQ_t = \alpha + AT_t\overline{\beta} + e_t, \tag{57}$$

where the time index $t \in \{1, \ldots, T\}$. If the slope parameter is time varying then $\overline{\beta}$ in (57) presents only the time average of the true parameter, say β_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, as we explain below.

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 the 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 in which 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. Thus, 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 terms involving the covariances 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 liquidity dynamics 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 and thus increased AT can potentially decrease firm investment (relative to a market with no AT) through its effects on liquidity dynamics.

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.

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, Vol_{it}) over time period t divided by total order activity over time period t. We define order activity as the sum of trades (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 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 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 the latter 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.

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 (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 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 (PES_{it}) is quite similar to (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, 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 PQS_{it}, PES_{it} is also in terms of percentage points. An increase in 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 (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

$$\text{MQ}_{it} = \alpha_i + \theta_t + \text{AT}_{it}\beta_t + W_{it}^{'}\vartheta + e_{it}, \tag{58}$$

where MQ_{it} is the market quality measure under consideration, α_i and θ_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 MQ_{it}, which we assume to be independent 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 (58), $E(\text{AT}_{it}e_{it}) \neq 0$.

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 $\text{AT}_{-IND,it}$. 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,

$$\mathrm{AT}_{it} = a_i + g_t + b\mathrm{AT}_{-IND,it} + dW_{it} + \epsilon_{it} \tag{59}$$

to obtain an instrument, Z_{it}, for AT_{it} given by the fitted values from (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 (58) as described in Section 3. For comparison purposes, we additionally apply the conventional panel data model assuming constant slope parameter, i.e., $\beta_1 = \beta_2 = \cdots = \beta_T$.

7.3 Results

Table 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 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.

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 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

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

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

Dependent Variable		$\hat{AT}_{it}$	$\ln(ME)_{it-1}$	T/O_{it-1}	$1/P_{it-1}$	$H\text{-}L_{it-1}$	RV_{it-1}
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	
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	
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	
$H\text{-}L_{it}$	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 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.

in liquidity this is likely 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 6 through 9 present the results when we allow the parameter to jump discretely over time. The coefficient estimates in Tables 6 through 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 paper, the above results clearly point to a time varying relationship between the effects of AT on various measures of market quality.

8 Conclusion

This paper 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.

	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 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.

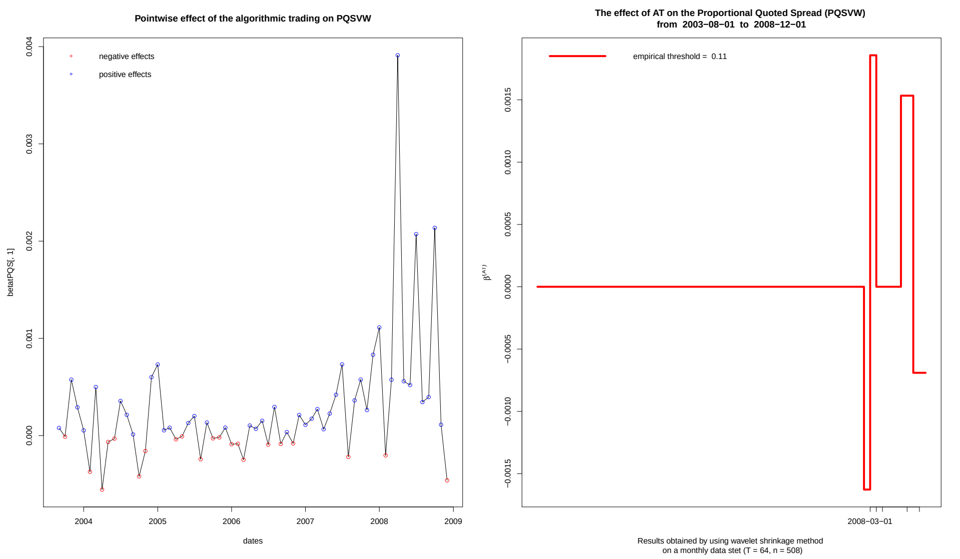

Figure 3: Time varying effect of algorithmic trading on the proportional quoted spread.

	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 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.

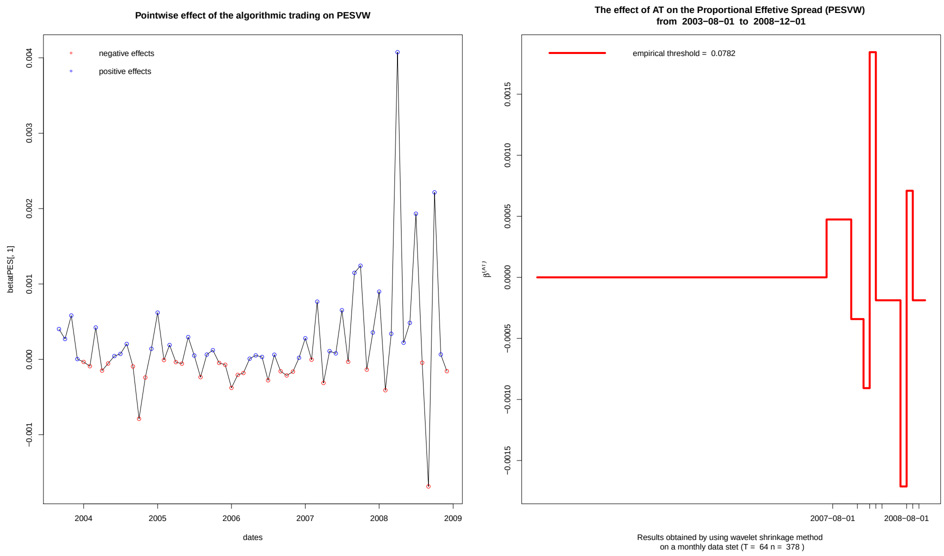

Figure 4: Time varying effect of algorithmic trading on the proportional effective spread.

	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 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.

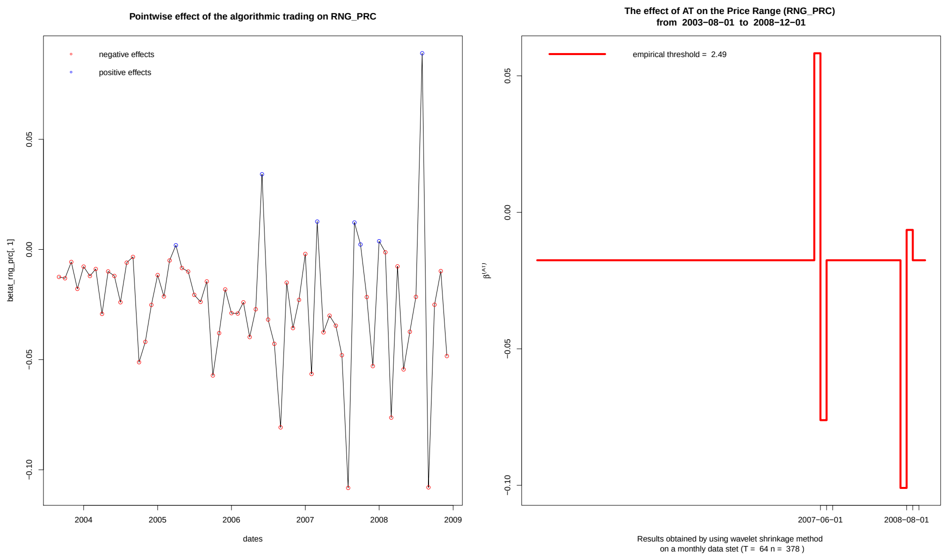

Figure 5: Time varying effect of algorithmic trading on the daily high-low price range.

	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 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.

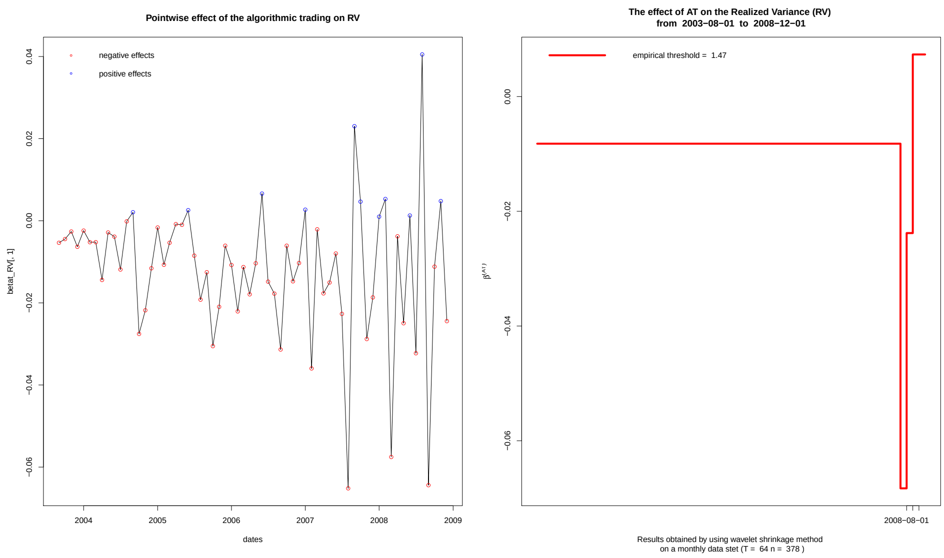

Figure 6: Time varying effect of algorithmic trading on the realized variance.

9 Theoretical Results and Proofs

10 Proofs of Section 2

Lemma 3 *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 (8) *and $h_{lk}(t)$ as* (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_τ 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 3: To prove the proposition, we show that β_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 (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

$$\bigcup_{l=1}^{l_\tau} \mathcal{I}_l = \{1,\dots,T\}$$

and

$$\mathcal{I}_{l_\tau} \subset \mathcal{I}_{l_\tau-1} \subset \cdots \subset \mathcal{I}_2 \subseteq \mathcal{I}_1 = \{1,\dots,T\}.$$

We now begin with the thinnest interval $\mathcal{I}_{l\tau}$ that contains the jump. Define

$$\beta_t^{(l_\tau)} = \begin{cases} \beta_t = \beta_\tau & \text{if} \quad t \le \tau \text{ and } t \in \mathcal{I}_{l\tau} \cap \{t|t \le \tau\} \\ \beta_t = \beta_{\tau+1} & \text{if} \quad t > \tau \text{ and } t \in \mathcal{I}_{l\tau} \cap \{t|t > \tau\} \\ 0 & \text{else.} \end{cases}$$

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

$$\beta_t^{(l_\tau)} = \begin{cases} \beta_\tau = \beta^{(l_\tau)} + a^*_{l_\tau,2k_{l_\tau}-1} b_{l_\tau,k_{l_\tau}} & \text{if} \quad t \le \tau \text{ and } t \in \mathcal{I}_{l\tau} \\ \beta_{\tau+1} = \beta^{(l_\tau)} - a^*_{l_\tau,2k_{l_\tau}} b_{l_\tau,k_{l_\tau}} & \text{if} \quad t > \text{ and } t \in \mathcal{I}_{l\tau} \\ 0 & \text{else.} \end{cases} \tag{60}$$

Using the definition of $w_{lk}(t)$, we can rewrite (60) as

$$\beta_t^{(l_\tau)} = \begin{cases} \beta_t = \beta^{(l_\tau)} + w_{l_\tau,k_{l_\tau}}(t) b_{l_\tau,k_{l_\tau}} & \text{if} \quad t \in \mathcal{I}_{l\tau} \\ 0 & \text{else.} \end{cases} \tag{61}$$

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 β_t is constant over $\mathcal{I}_{l_{\tau-1}} \setminus \mathcal{I}_{l_\tau}$; it can be either β_τ 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 $l \in \{2,\dots,l_\tau\}$, we end up with

$$\beta_t^{(l_\tau-l)} + w_{l_\tau-l+1,k_{l_\tau-l+1}}(t)b_{l_\tau-l+1,k_{l_\tau-l+1}} = \begin{cases} \beta^{(l_\tau-l)} + \sum_{s=l_\tau-l}^{l_\tau} w_{s,k_s}(t)b_{s,k_s} = \beta_t & \text{if} \quad t \in \mathcal{I}_{l_\tau-l} \\ 0 & \text{else.} \end{cases} \tag{62}$$

where $\beta^{(l_\tau-l)}$ is constant over $\mathcal{I}_{l_\tau-l}$. Finally, from (62), we can infer that, for all $t \in \mathcal{I}_1 = \{1,\dots,T\}$,

$$\beta_t = \beta^{(1)} + \sum_{l=2}^{l_\tau} w_{l,k_l}(t)b_{l,k_l} \qquad \forall t \in \{1,\dots,T\}.$$

Because $\beta^{(1)}$ is a constant and $w_{11}(t) = a_{11} \neq 0$, $\forall t \in \{1,\dots,T\}$, we can express β_t in terms of $l_\tau \leq L$ basis such that

$$\beta_t = \sum_{l=1}^{l_\tau} w_{l,k_l}(t)b_{l,k_l} \qquad \forall t \in \{1,\dots,T\}.$$

This completes the proof. □

Proof of Proposition 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 3. Let β 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 β 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 3, 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 3 *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 $l, l^{'} \in \{1,\dots,L\}$, $k, \in \{1,\dots,K_l\}$, *and* $k^{'}, \in \{1,\dots,K_{l'} | l \neq l^{'}\}$, *or* $k,k^{'}, \in \{1,\dots,K_l | k \neq k^{'}; l = l^{'}\}$.

Proof of Proposition 3: 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 $(l, k) = (l^{'}, k^{'})$,

$$\begin{aligned}
\mathcal{Z}_{l,k,it}\mathcal{X}_{l,k,it} &= X_{it}Z_{it}\left(w_{lk}(t)\right)^2 \\
&= X_{it}Z_{it}\left(a_{l,2k-1}h_{l,2k-1}(t) - a_{l,2k}h_{l,2k}(t)\right)^2 \\
&= X_{it}Z_{it}\left(a_{l,2k-1}^2h_{l,2k-1}^2(t) + a_{l,2k}^2h_{l,2k}^2(t) - 2a_{l,2k-1}a_{l,2k}\underbrace{h_{l,2k-1}(t)h_{l,2k}(t)}_{0}\right) \\
&= X_{it}Z_{it}\left(a_{l,2k-1}^2h_{l,2k-1}^2(t) + a_{l,2k}^2h_{l,2k}^2(t)\right),
\end{aligned} \tag{63}$$

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 (63), we can easily verify that (a) is a direct result of (i) for all $l \in \{2, \ldots, L\}$ and $k \in \{1, \ldots, K_l\}$.$\square$

11 Proofs of Section 3

Proof of Lemma 1: 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

$l > 1$,

$$\begin{aligned}
\sqrt{n(T-1)}(\tilde{b}_{l,k,p} - b_{l,k,p}) &= \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{P} \mathcal{Z}_{lk,it,q} \Delta e_{it}, \\
&= \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{P} W_{lk,pq}(t) Z_{it,q} \Delta e_{it}, \\
&= \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{P} A_{l,2k,pq} h_{l,2k}(t) Z_{it,q} \Delta e_{it} \\
&\quad - \frac{1}{\sqrt{n(T-1)}} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{q=1}^{P} A_{l,2k-1,pq} h_{l,2k-1}(t) Z_{it,q} \Delta e_{it}, \\
&= \frac{1}{\sqrt{n(2^{L-l-1}-1)}} \sum_{q=1}^{P} A_{l,2k,pq} \sum_{i=1}^{n} \sum_{t \in \{h_{l,2k}(t) \neq 0\}} Z_{it,q} \Delta e_{it} \\
&\quad - \frac{1}{\sqrt{n(2^{L-l-1}-1)}} \sum_{q=1}^{P} A_{l,2k-1,pq} \sum_{i=1}^{n} \sum_{t \in \{h_{l,2k-1}(t) \neq 0\}} Z_{it,q} \Delta e_{it},
\end{aligned}$$

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 $l = 1$,

$$\sqrt{n(T-1)}(\tilde{b}_{1,1,p} - b_{1,1,p}) = \frac{1}{\sqrt{n(2^{L}-1)}} \sum_{q=1}^{P} A_{1,1,pq} \sum_{i=1}^{n} \sum_{t=2}^{T} Z_{it,q} \Delta e_{it}.$$

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

$$\begin{aligned}
\Sigma_{l,k,p} &= E\big((\sqrt{n(T-1)}(\tilde{b}_{l,k,p} - b_{l,k,p}))^2\big), \\
&= E\left(\frac{1}{n(2^{L-l-1}-1)} \sum_{q,r=1}^{P} A_{l,2k,pq} A_{l,2k,pr} \sum_{i,j=1}^{n} \sum_{t,s \in H} Z_{it,q} Z_{js,r} E_c\big(\Delta e_{it} \Delta e_{js}\big) \right) \\
&\quad + E\left(\frac{1}{n(2^{L-l-1}-1)} \sum_{q,r=1}^{P} A_{l,2k-1,pq} A_{l,2k-1,pr} \sum_{i,j=1}^{n} \sum_{t,s \in H} Z_{it,q} Z_{js,r} E_c\big(\Delta e_{it} \Delta e_{js}\big) \right), \\
&= \Pi_{l,k,1} + \Pi_{l,k,2},
\end{aligned}$$

where $\sum_{q,r=1}^{P}$, $\sum_{i,j=1}^{n}$ and $\sum_{t,s\in H}$ denote the double summations $\sum_{q=1}^{P}\sum_{q=1}^{P}$ $\sum_{i=1}^{n}\sum_{j=1}^{n}$ and $\sum_{t\in\{H_{l,2k}(t)\neq 0\}}\sum_{s\in\{H_{l,2k}(s)\neq 0\}}$, respectively.

For $l=1$,

$$\begin{aligned}\Sigma_{1,1,p} &:= E\big((\sqrt{n(T-1)}(\tilde{b}_{1,1,p}-b_{1,1,p}))^2\big)\\ &= E\left(\sum_{q,r=1}^{P}\frac{1}{n(2^L-1)}A_{1,1,pq}A_{1,1,pr}\sum_{i,j=1}^{n}\sum_{t,s=2}^{T}Z_{it,q}Z_{js,r}E_c\big(\Delta e_{it}\Delta e_{js}\big)\right).\end{aligned}$$

By using Assumption C, we can infer

$$\begin{aligned}\Pi_{l,k,1} &= E\left(\frac{1}{n(2^{L-l-1}-1)}\sum_{q,r=1}^{P}A_{l,2k,pq}A_{l,2k,pr}\sum_{i,j=1}^{n}\sum_{t,s\in H}Z_{it,q}Z_{js,r}\sigma_{ij,ts}\right),\\ &\leq E\left(\frac{1}{n(2^{L-l-1}-1)}\sum_{q,r=1}^{P}A_{l,2k,pq}A_{l,2k,pr}\sum_{i,j=1}^{n}\sum_{t,s\in H}Z_{it,q}Z_{js,r}|\sigma_{ij,ts}|\right),\\ \Pi_{l,k,2} &\leq E\left(\frac{1}{n(2^{L-l-1}-1)}\sum_{q,r=1}^{P}A_{l,2k-1,pq}A_{l,2k-1,pr}\sum_{i,j=1}^{n}\sum_{t,s\in H}Z_{it,q}Z_{js,r}|\sigma_{ij,ts}|\right),\ \text{and}\\ \Sigma_{1,1,p} &\leq E\left(\frac{1}{n(2^{L}-1)}\sum_{q,r=1}^{P}A_{1,1,pq}A_{1,1,pr}\sum_{i,j=1}^{n}\sum_{t,s=2}^{T}Z_{it,q}Z_{js,r}|\sigma_{ij,ts}|\right).\end{aligned}$$

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

$$\begin{aligned}P\left(\left|\tilde{b}_{l,k,p}-b_{l,k,p}\right|>M^{\frac{1}{2}}\frac{c}{\sqrt{n(T-1)}}\right) &\leq P\left(\Sigma_{l,k,p}^{-\frac{1}{2}}\sqrt{n(T-1)}\left|\tilde{b}_{l,k,p}-b_{l,k,p}\right|>c\right),\\ &\leq \frac{1}{c}\exp(-\frac{c^2}{2}). \qquad (64)\end{aligned}$$

Using Boole's inequality and (64), 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 $\sum_{l=1}^{L}\sum_{k=1}^{K_l}\sum_{p=1}^{P}$. The assertion of the theorem follows by replacing c with $\sqrt{2\log((T-1)\underline{P})}c^*$ for any

$c^* > 0$. $\square$

Proof of Theorem 1: 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,

$$\tilde{\gamma}_{t,p} - \gamma_{t,p} = \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{lk,pq}(t)\hat{b}_{l,k,q} - \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{lk,pq}(t)b_{l,k,q}, \tag{65}$$

where

$$\hat{b}_{l,k,q} = \tilde{b}_{l,k,q} - \tilde{b}_{l,k,q}(|\tilde{b}_{l,k,q}| < \lambda_{n,T}). \tag{66}$$

and

$$\begin{aligned} W_{lk,pq}(t) &= A_{l,2k,pq}(t)H_{l,2k}(t) - A_{l,2k-1,pq}(t)H_{l,2k-1}(t), \qquad (67)\\ &= \sqrt{2^{l-2}}A_{l,2k,pq}\mathbf{I}(H_{l,2k}(t) \neq 0) - \sqrt{2^{l-2}}A_{l,2k-1,pq}\mathbf{I}(H_{l,2k-1}(t) \neq 0). \end{aligned}$$

Plugging (66) and (67) in (65) and using the absolute value inequality, we get

$$\begin{aligned} |\tilde{\gamma}_{t,p} - \gamma_{t,p}| \leq & \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} \sqrt{2^{l-2}}|A_{l,2k,pq}\mathbf{I}(H_{l,2k}(t) \neq 0)(\tilde{b}_{l,k,q} - b_{l,k,q})| \\ &+ \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} \sqrt{2^{l-2}}|A_{l,2k,pq}\mathbf{I}(H_{l,2k}(t) \neq 0)\tilde{b}_{l,k,q}\mathbf{I}(|\tilde{b}_{l,k,q}| < \lambda_{n,T})| \\ &+ \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} \sqrt{2^{l-2}}|A_{l,2k-1,pq}\mathbf{I}(H_{l,2k-1}(t) \neq 0)(\underline{\tilde{b}}_{l,k,q} - \underline{b}_{l,k,q})| \\ &+ \sum_{q=1}^{\underline{P}}\sum_{l=1}^{L}\sum_{k=1}^{K_l} \sqrt{2^{l-2}}|A_{l,2k,pq}\mathbf{I}(H_{l,2k-1}(t) \neq 0)\tilde{b}_{l,k,q}\mathbf{I}(|\tilde{b}_{l,k,q}| < \lambda_{n,T})|, \\ = & \; a + b + c + d. \end{aligned}$$

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 1, we can infer that

$$\begin{aligned}
\sup_{t,p} |\tilde{\gamma}_{t,p} - \gamma_{t,p}| &= \sup_{l,k,p} |\tilde{b}_{l,k,p} - b_{l,k,p}| \times O_p(\sqrt{T-1}) + \lambda_{n,T} \times O_p(\sqrt{T-1}), \\
&= O_p\Big(\sqrt{\frac{\log(T-1)}{n}} + \sqrt{T-1}\lambda_{n,T}\Big). \qquad (68)
\end{aligned}$$

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

$$\frac{1}{T-1}\sum_{t=2}^{T-2} f_t^2 \quad \leq \quad (\sup_{(l,k)\in\mathcal{L}_p,p} |\hat{b}_{l,k,p}|)^2 O_p(T-1).$$

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 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

$$\frac{1}{T-1}\sum_{t=2}^{T-1} e_t^2 \quad \leq \quad O_p\big(\frac{\log(T-1)}{n(T-1)} + \lambda_{n,T}^2\big)\left(\min\{(S^*+1)\log(T-1), (T-1)\}\right).$$

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. □

12 Proofs of Section 4

Proof of Lemma 2: We have to show that

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

for $m = s, u$.

For $p \in \{1,\dots,P\}$ and $m = s$, we have by construction

$$\begin{aligned}
\tilde{c}_{L,k,p}^{(s)} - c_{L,k,p}^{(s)} &= \frac{1}{T-1}\sum_{t=2}^{T}\psi_{L,k}(t-1)(\tilde{\gamma}_{t,p} - \gamma_{t,p}), \\
&= \frac{1}{T-1}\sum_{t=2}^{T}\psi_{L,k}(t-1)\sum_{l,m,q} W_{l,m,p,q}(t)(\tilde{b}_{l,m,q} - b_{l,m,q}), \\
&= \frac{1}{T-1}\sum_{t\in\{\psi_{L,k}(t-1)\neq 0\}}\psi_{L,k}(t-1)\sum_{l,m,q} W_{l,m,p,q}(t)(\tilde{b}_{l,m,q} - b_{l,m,q}),
\end{aligned}$$

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 1, we can easily verify that the term in the last inequality is $O_p(1)$. By Lemma 1, 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 2: 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 $p \in \{1, \dots, \underline{P}\}$:

$$\begin{aligned}
\mathcal{J}_p &:= \{\tau_{1,p}, \dots, \tau_{S_p,p}\}, \\
\mathcal{J}_p^c &:= \{1, \dots, T\} \setminus \mathcal{J}_p, \\
\overline{\mathcal{J}}_p &:= \{2, 4, \dots, T-1\} \cap \mathcal{J}_p, \\
\underline{\mathcal{J}}_p &:= \{3, 5, \dots, T\} \cap \mathcal{J}_p, \\
\overline{\mathcal{J}}_p^c &:= \{2, 4, \dots, T-1\} \setminus \overline{\mathcal{J}}_p, \text{ and} \\
\underline{\mathcal{J}}_p^c &:= \{3, 5, \dots, T\} \setminus \underline{\mathcal{J}}_p.
\end{aligned}$$

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

$$\omega_{n,T} := \{ \sup_{t \in \mathcal{J}_p^c,\, p \in \{1, \dots, P\}} \{ |\Delta \tilde{\beta}^{(u)}_{t,p}| \mathbf{I}_{\overline{\mathcal{J}}_p^c} + |\Delta \tilde{\beta}^{(s)}_{t,p}| \mathbf{I}_{\underline{\mathcal{J}}_p^c} \} = 0 \},$$

where $\mathbf{I}_{\overline{\mathcal{J}}_p^c} = \mathbf{I}(t \in \overline{\mathcal{J}}^c_p)$, $\mathbf{I}_{\underline{\mathcal{J}}_p^c} = \mathbf{I}(t \in \underline{\mathcal{J}}^c_p)$ 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 $\overline{\mathcal{J}}^c_p$ and $\underline{\mathcal{J}}^c_p$ are adjacent.

Let's now start with the no-jump case in $\overline{\mathcal{J}}^c_p$. By construction, we have, for all $t \in \{2, 4, \dots, T-1\}$,

$$\Delta\tilde{\beta}^{(u)}_{t,p} = \sum_{k=1}^{K_L} \Delta\psi_{L,k}(t)\hat{c}^{(u)}_{L,k,p}$$

Recall that at $l = L$, the construction of the wavelets basis implies that at each $t \in \{2, 4, \dots, T-1\}$ there is only one differenced basis $\Delta\psi_{L,k}(t)$ that is not zero. Let $\mathcal{K}^c_p = \{k|\Delta\psi_{L,k}(t) \neq 0,\ t \in \overline{\mathcal{J}}^c_p\} = \{k|\Delta\psi_{L,k}(t-1) \neq 0,\ t \in \underline{\mathcal{J}}^c_p\}$. We can infer that $\{\sup_{t\in\overline{\mathcal{J}}^c_p} |\sum_{k=1}^{K_L} \Delta\psi_{L,k}(t)\hat{c}^{(u)}_{L,k,p}| = 0\}$ occurs only if $\{\sup_{k\in\mathcal{K}^c_p} |c^{(u)}_{L,k,p}| = 0\}$ occurs.

By analogy, we can show the same assertion for the complement set $\underline{\mathcal{J}}^c_p$, i.e., $\{\sup_{t\in\underline{\mathcal{J}}^c_p} |\Delta\tilde{\beta}^{(s)}_{t,p}| = 0\}$ occurs only if $\{\sup_{k\in\mathcal{K}^c_p} |\hat{c}^{(s)}_{L,k,p}| = 0\}$ occurs.

To study $P(\omega_{n,T})$, it is hence sufficient to study

$$P(\sup_{k\in\mathcal{K}^c_p, m, p\in\{1,\dots,P\}} |\hat{c}^{(m)}_{L,k,p}| = 0) \quad = \quad P(\sup_{k\in\mathcal{K}^c_p, m, p\in\{1,\dots,P\}} |\tilde{c}^{(m)}_{L,k,p}| < \lambda_{n,T}).$$

By Lemma 2, $\sup_{k\in\mathcal{K}^c_p, m, p\in\{1,\dots,P\}} |\tilde{c}^{(m)}_{L,k,p}| = O_p\big(\sqrt{\log(T-1)/n(T-1)}\big)$, since $c^{(m)}_{L,k,p} = 0$, for all $k \in \mathcal{K}^c_p$, and $p \in \{1, \dots, P\}$. 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 $p \in \{1, \dots, P\}$ that is not detected and show the contradiction. If $\tau_{j,p} \in \overline{\mathcal{J}}_p$, then

$$|\Delta\tilde{\beta}^{(u)}_{\tau_{j,p},p}|\mathbf{I}_{\overline{\mathcal{J}}_p} + |\Delta\tilde{\beta}^{(s)}_{\tau_{j,p},p}|\mathbf{I}_{\underline{\mathcal{J}}_p} = |\Delta\tilde{\beta}^{(u)}_{\tau_{j,p},p}|.$$

Adding and subtracting $\Delta\beta^{(u)}_{\tau_{j,p},p}$, we get

$$\begin{aligned}
\Delta\tilde{\beta}^{(u)}_{\tau_{j,p},p} &= \sum_{k=1}^{K_L} \Delta\psi_{L,k}(\tau_{j,p})(\tilde{c}^{(u)}_{L,k,p} - c^{(u)}_{L,k,p}) - \sum_{k=1}^{K_L} \Delta\psi_{L,k}(\tau_{j,p})\tilde{c}^{(u)}_{L,k,p}\mathbf{I}(\tilde{c}^{(u)}_{L,k,p} < \lambda_{n,T}) \\
&\quad + \sum_{k=1}^{K_L} \Delta\psi_{L,k}(\tau_{j,p})c^{(u)}_{L,k,p}, \\
&= I + II + III.
\end{aligned}$$

By Lemma 2, $I = o_p(1)$, $II = o_p(1)$ 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: 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 2, 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 $Q_{(\tau)}^{\circ}$ and the second term converges in distribution to $N(0, V_{(\tau)}^{\circ})$. Slutsky's rule implies

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

It follows

$$\sqrt{n}\mathcal{T}_{(\tilde{\tau})}^{\frac{1}{2}}\big(\hat{\beta}_{(\tilde{\tau})} - \beta_{(\tau)}\big) = \sqrt{n}\mathcal{T}_{(\tau)}^{\frac{1}{2}}\big(\hat{\beta}_{(\tau)} - \beta_{(\tau)}\big) + o_p(1) \xrightarrow{d} N(0, (Q_{(\tau)}^{\circ})^{-1}\big(V_{(\tau)}^{\circ}\big)(Q_{(\tau)}^{\circ})^{-1}).$$

This completes the Proof. □

Proof of Proposition 2 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 $V^{(1)}_{(\tau)} = (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}Z'_{it,(\tau)}\sigma^2$, with $\sigma^2 = E_c(\Delta\dot{e}_{it})$.

$$\begin{aligned}
\hat{V}^{(1)}_{(\tilde{\tau})} - V^{(1)}_{(\tau)} &= \Big(\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\hat{e}^2_{it} - \sigma^2\Big)(n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}Z'_{it,(\tilde{\tau})}, \\
&= +\sigma^2\left((n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tilde{\tau})}Z'_{it,(\tilde{\tau})} - (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} Z_{it,(\tau)}Z'_{it,(\tau)}\right), \\
&= a + b.
\end{aligned}$$

From Assumption B(ii), we can infer

$$\begin{aligned}
||a|| &\leq \Big(\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\hat{e}^2_{it} - \sigma^2\Big)\frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}||(\mathcal{T}_{(\tilde{\tau})})^{-1/2}Z_{it,(\tilde{\tau})}||^2, \\
&= \left(\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}((\Delta\hat{e}^2_{it} - \Delta e^2_{it}) + (\Delta e^2_{it} - \sigma^2))\right)\frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}||(\mathcal{T}_{(\tilde{\tau})})^{-1/2}Z_{it,(\tilde{\tau})}||^2,
\end{aligned}$$

From

$$\begin{aligned}
\Delta\hat{\dot{e}}_{it} &= \Delta\dot{Y}_{it} - \Delta\dot{X}'_{it,(\tilde{\tau})}\hat{\beta}_{(\tilde{\tau})}, \\
&= \Delta\dot{e}_{it} + \Delta\dot{X}'_{it,(\tilde{\tau})}(\beta_{(\tilde{\tau})} - \hat{\beta}_{(\tilde{\tau})}), \qquad (69)
\end{aligned}$$

and by using Theorem 3 together with Assumption B(ii), we can show that

$$\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\hat{\dot{e}}_{it} - \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\dot{e}_{it} = o_p(1). \qquad (70)$$

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 2, we can infer that, given $\tilde{S}_1=S_1,\dots,\tilde{S}_P=S_P$,

$$(n\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tilde{\tau})}Z_{it,(\tilde{\tau})} \quad = \quad (n\mathcal{T}_{(\tau)})^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}Z_{it,(\tau)}Z_{it,(\tau)}+o_p(1).$$

Thus,

$$\hat{V}^{(1)}_{(\tilde{\tau})}-V^{(1)}_{(\tau)}=o_p(1).$$

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 $\sigma_i^2=E_c(\Delta\dot{e}_{it})$.

$$\begin{aligned}\hat{V}^{(2)}_{(\tilde{\tau})}-V^{(2)}_{(\tau)} &= \frac{1}{n}\sum_{i=1}^{n}(\hat{\sigma}_i^2-\sigma_i^2)(\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{t=2}^{T}Z_{it,(\tilde{\tau})}Z'_{it,(\tilde{\tau})},\\ &= +\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\left((\mathcal{T}_{(\tilde{\tau})})^{-1}\sum_{t=2}^{T}Z_{it,(\tilde{\tau})}Z'_{it,(\tilde{\tau})}-(\mathcal{T}_{(\tau)})^{-1}\sum_{t=2}^{T}Z_{it,(\tau)}Z'_{it,(\tau)}\right),\\ &= d+e.\end{aligned}$$

$$\begin{aligned}||d|| &\le \frac{1}{n}\sum_{i=1}^{n}(\hat{\sigma}_i^2-\sigma_i^2)\sum_{t=2}^{T}||(\mathcal{T}_{(\tilde{\tau})})^{-1/2}Z_{it,(\tilde{\tau})}||^2,\\ &= \frac{1}{n}\sum_{i=1}^{n}((\hat{\sigma}_i^2-\frac{1}{(T-1)}\sum_{t=2}^{T}\Delta\dot{e}_{it})-(\sigma_i^2-\frac{1}{(T-1)}\sum_{t=2}^{T}\Delta\dot{e}_{it})\frac{1}{n}\sum_{t=2}^{T}||(\mathcal{T}_{(\tilde{\tau})})^{-1/2}Z_{it,(\tilde{\tau})}||^2.\end{aligned}$$

From Equation (69), and Theorem 3, we can infer

$$\frac{1}{(T-1)}\sum_{t=2}^{T}\Delta\hat{\dot{e}}_{it}-\frac{1}{(T-1)}\sum_{t=2}^{T}\Delta\dot{e}_{it}=o_p(1)\nu_i, \tag{71}$$

where $\frac{1}{n}\sum_{i=1}^{n}|\nu_i|=O_p(1)$. Moreover,

$$\sigma_i^2-\frac{1}{(T-1)}\sum_{t=2}^{T}\Delta\dot{e}_{it}=o_p(1)\mu_i, \tag{72}$$

where $\frac{1}{n}\sum_{i=1}^{n}|\mu_i| = O_p(1)$. Note that the first terms in (71) and (72) 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 σ_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, 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. □

References

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.

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.

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.

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

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.

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.

Carr, P. and Wu, L. (2003). What Type of Process Underlies Options ? A Simple Robust Test. *The Journal of Finance*, LVIII(6):2581–2610.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3).

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.

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.

Inoue, A. and Rossi, B. (2011). Identifying the Sources of Instabilities in Macroeconomic Fluctuations. *The Review of Economics and Statistics*, 93(4):1186–1204.

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.

Lucas, R. E. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

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

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).

Qu, Z. and Perron, P. (2007). Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica*, 75(2):459–502.

Schmidt, P. and Sickles, R. (1977). Some Further Evidence on the Use of the Chow Test Under Heteroskedasticity. *Journal of the American Statistical Association*, 46:l293–l298.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.