# Unstable Diffusion Indexes: With an Application to Bond Risk Premia*

Daniele Massacci

Bank of England

February 14, 2019

**Abstract**

This paper studies the empirically relevant problem of estimation and inference in diffusion index forecasting models with structural instability. Factor model and factor augmented regression both experience a structural change with different unknown break dates. In the factor model, we estimate factors and loadings by principal components. We consider least squares estimation of the factor augmented regression and propose a break test. The empirical application uncovers instabilities in the linkages between bond risk premia and macroeconomic factors.

**JEL classification:** C12, C13, C38, C52, G12.

**Keywords:** Large Factor Model, Factor Augmented Regression, Structural Break, Bond Risk Premia.

*This work was started when the author was Franco Modigliani Research Fellow in Economics and Finance at the Einaudi Institute for Economics and Finance (EIEF); it was then completed after the author joined the Bank of England. The views in this paper are those of the author and do not necessarily reflect those of the Bank of England or its policy committees. The author really is highly indebted to Marco Lippi for introducing him to factor models and for several enlightening conversations. This paper benefits from comments from participants at the International Association for Applied Econometrics (IAAE) 2016 Annual Conference, at the Symposium on Statistical Penalization Methods and Dimension Reduction Methods for Economic and Financial Analysis at the University of York, and at the Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance 2017; and from conversations with Laura Coroneo, Pasquale Della Corte, Domenico Giannone, Christian Gourieroux, Refet Gürkaynak, George Kapetanios, Andrew Meldrum and Hashem Pesaran. Errors and omissions are the author's responsibility. Financial support from the Associazione Borsisti Marco Fanno and from UniCredit and Universities Foundation is gratefully acknowledged. Address correspondence to Daniele Massacci, Bank of England, Threadneedle Street, EC2R 8AH, United Kingdom. E-mail: daniele.massacci@gmail.com .

# 1 Introduction

The diffusion index forecasting model of Stock and Watson (1998, 2002), and Bai and Ng (2006) is a regression model with observable and latent covariates, where the latter are the common factors in the variables of a large scale dataset. The model is usually estimated under the assumption of structural stability in the loadings of the factor model and in the slope coefficients of the factor augmented regression: Connor and Korajczyk (1986, 1988, 1993), Bai and Ng (2002), Stock and Watson (2002), and Bai (2003) deal with linear static factor models; Forni *et al.* (2000, 2004), Forni and Lippi (2001), and Forni *et al.* (2015) study the linear generalized dynamic factor model; Stock and Watson (2002), and Bai and Ng (2006) focus on stable factor augmented regressions. Violation of the stability assumption may affect empirical results. For example, out-of-sample forecasts may not be accurate due to instabilities in the factor augmented regression (Stock and Watson (2002, 2009)) or in the factor model (Giannone (2007), and Banerjee *et al.* (2008)). In the latter, structural breaks enlarge the factor space (Breitung and Eickmeier (2011), and Chen *et al.* (2014)) without conveying more information: including additional factors in the factor augmented regression increases estimation noise and deteriorates forecast performance (Han and Inoue (2015)).

A number of contributions analyzes large dimensional factor models with structural instabilities. Bates *et al.* (2013) study the robustness of the principal components estimator as applied to factor models under neglected instability. Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015), Yamamoto and Tanaka (2015), and Barigozzi and Trapani (2017) develop statistical tools to detect breaks. Chen (2015), Cheng *et al.* (2016), Bai *et al.* (2017), and Baltagi *et al.* (2017) focus on estimation under the single break assumption. Baltagi *et al.* (2016), Ma and Su (2016), and Barigozzi *et al.* (2018) allow for multiple breaks. Massacci (2017) studies large dimensional threshold factor models with regime shifts in the loadings driven by a covariate. These contributions do not consider the whole diffusion index forecasting model. Corradi and Swanson (2014) propose a test for the joint null hypothesis of stable factor model and factor-augmented regression. Wang *et al.* (2015) estimate unstable factor augmented regressions with factors extracted from a linear model. To the very best of our knowledge, estimation and inference in the set up with instabilities in the factor model and in the factor augmented regression has not been studied: we aim at filling this gap.

We start from the single break factor model. Let $N$ and $T$ denote the cross-sectional and time series dimensions, respectively. We estimate the model by least squares by minimizing the sum of squared residuals (Baltagi *et al.* (2017), and Massacci (2017)): the resulting principal components estimator for factors and loadings has the same convergence rate $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$ as in the linear case (Bai and Ng (2002)). We then turn to the single break factor augmented regression. We estimate the parameters by least squares by replacing the latent factors with their estimates (Bai (1997), and Bai and Ng (2006)): despite the structural instability, the least squares estimator for the slope coefficients is $\sqrt{T}$ consistent and asymptotically normal provided that $\sqrt{T}/N \to 0$ as $N, T \to \infty$. We then propose a Lagrange multiplier test for the null hypothesis of stability: we show that the critical values provided in Andrews (1993) remain valid also in the presence of latent factors estimated from an unstable large dimensional model.

Finally, we apply our methodology to bond risk premia (Fama and Bliss (1987), Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009)). Common factors extracted from a large set of macroeconomic series help predicting bond excess returns (Ludvigson and Ng (2009)); however, the loadings are not stable over time (Breitung and Eickmeier (2011), Chen *et al.* (2014), and Cheng *et al.* (2016)). Pricing equations for bond risk premia experience structural breaks (Bikbov and Chernov (2010), and Smith and Taylor (2009)). We thus use our model to uncover instabilities in the linkages between bond market risk premia and macroeconomic fundamentals. Depending on the maturity of the bond, we show that: predictive regressions for bond risk premia are stable for most of the Great Moderation (Joslin *et al.* (2014)); a break occurred in the early 1980s (Smith and Taylor (2009)).

The paper is organized as follows. Section 2 describes the unstable diffusion index forecasting model. Section 3 gives relevant information on estimation and model selection in large dimensional factor models subject to structural break. Section 4 looks at estimation and inference in the unstable factor augmented regression. Section 5 performs a Monte Carlo analysis. Section 6 provides the application to bond risk premia. Section 7 outlines possible extensions and modifications. Section 8 concludes. Appendix A collects technical proofs.

Concerning notation, $\mathbb{I}(\cdot)$ denotes the indicator function; given a square matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$; the norm of a generic matrix $\mathbf{A}$ is $\|\mathbf{A}\| = [\mathrm{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$; for a given scalar $A$, $|A|$, $\mathbf{I}_A$ and $\mathbf{0}_A$ are the absolute value of $A$, the $A \times A$ identity matrix, and the $A \times A$ zero matrix, respectively; $\xrightarrow{p}$ denotes convergence in probability; $\overset{p}{\not\to}$ is the negation of $\xrightarrow{p}$; $\xrightarrow{d}$ denotes convergence in distribution; $\Rightarrow$ denotes weak convergence. Without round brackets inside, $[\cdot]$ is the integer part of the argument.

## 2 The Unstable Diffusion Index Forecasting Model

We consider

$$\mathbf{x}_t = \mathbb{I}(t/T \leq \pi_{\mathbf{x}})\, \mathbf{\Lambda}_1 \mathbf{f}_t + \mathbb{I}(t/T > \pi_{\mathbf{x}})\, \mathbf{\Lambda}_2 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \ldots, T, \tag{1}$$

$$y_{t+h} = \mathbb{I}(t/T \leq \pi_y) \left(\boldsymbol{\gamma}_1' \mathbf{f}_t + \boldsymbol{\beta}_1' \mathbf{w}_t\right) + \mathbb{I}(t/T > \pi_y) \left(\boldsymbol{\gamma}_2' \mathbf{f}_t + \boldsymbol{\beta}_2' \mathbf{w}_t\right) + \varepsilon_{t+h}, \quad t = 1, \ldots, T, \quad h \geq 0, \tag{2}$$

where $T$ is the time series dimension of the sample. Starting from (1), $\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})' \in \Re^N$ is the $N \times 1$ vector of observable dependent variables; $\mathbf{f}_t = (f_{1t}, \ldots, f_{Rt})' \in \Re^R$ is the $R \times 1$ vector of latent factors; $\mathbf{e}_t = (e_{1t}, \ldots, e_{Nt})' \in \Re^N$ is the $N \times 1$ vector of idiosyncratic errors; $\pi_{\mathbf{x}}$ is the break fraction; $\mathbf{\Lambda}_j = (\boldsymbol{\lambda}_{j1}, \ldots, \boldsymbol{\lambda}_{jN})'$ is the $N \times R$ matrix of factor loadings, with $i-th$ row defined as $\boldsymbol{\lambda}_{ji}' = (\lambda_{ji1}, \ldots, \lambda_{jiR})$, for $j = 1, 2$ and $i = 1, \ldots, N$. Moving to (2), $y_{t+h} \in \Re$ is the scalar dependent variable; $\mathbf{f}_t$ is the same vector of latent factors as in (1); $\mathbf{w}_t \in \Re^K$ is a

$K \times 1$ vector of observable variables; $\varepsilon_{t+h} \in \Re$ is the error term; $\pi_y$ is the break fraction; $\boldsymbol{\gamma}_j$ and $\boldsymbol{\beta}_j$ are $R \times 1$ and $K \times 1$ vectors of slope coefficients, respectively, for $j = 1, 2$.

The model in (1) is a large dimensional factor model with a break in the loadings. Given Assumption C3 in Section 3.1.2 below, we follow Chamberlain and Rothschild (1983) and allow for some degree of correlation in the idiosyncratic components on each side of the breakpoint: (1) then is an *approximate breakpoint factor model*; it is more general than an exact breakpoint factor model, which would extend the arbitrage pricing theory of Ross (1976) and would not allow for any correlation in the idiosyncratic components on any side of the breakpoint. Equation (2) is a factor augmented regression with a break in the slope coefficients of $\mathbf{f}_t$ or $\mathbf{w}_t$ (or both). Together with (1), equation (2) forms an *unstable diffusion index forecasting model*: it extends the linear set up of Stock and Watson (1998, 2002), and Bai and Ng (2006) by introducing structural instability in the factor model and in the factor augmented regression. Notice that the break fractions $\pi_{\mathbf{x}}$ and $\pi_y$ in (1) and (2), respectively, are not constrained to be equal.

# 3 The Breakpoint Factor Model

The paper studies diffusion index forecasting models with instability in the factor model *and* in the factor augmented regression: estimation and model selection in (1) is functional to Section 4. We adapt the methodology of Massacci (2017) for threshold factor models: Section 3.1 states the assumptions; Section 3.2 deals with estimation and model selection.

## 3.1 Assumptions

Let $\mathbb{I}_{1t}(\pi_{\mathbf{x}}) = \mathbb{I}(t/T \leq \pi_{\mathbf{x}})$ and $\mathbb{I}_{2t}(\pi_{\mathbf{x}}) = \mathbb{I}(t/T > \pi_{\mathbf{x}})$. Denote $R^0$, $\boldsymbol{\Lambda}_j^0 = \left(\boldsymbol{\lambda}_{j1}^0, \ldots, \boldsymbol{\lambda}_{jN}^0\right)'$, $\pi_{\mathbf{x}}^0$ and $\mathbf{f}_t^0$ the true values of $R$, $\boldsymbol{\Lambda}_j$, $\pi_{\mathbf{x}}$ and $\mathbf{f}_t$, respectively, for $j = 1, 2$. Define $\mathbf{f}_{jt}^0(\pi_{\mathbf{x}}) = \mathbb{I}_{jt}(\pi_{\mathbf{x}})\,\mathbf{f}_t^0$, for $j = 1, 2$ and $t = 1, \ldots, T$, and let $\boldsymbol{\delta}_{\mathbf{x}i}^0 = \boldsymbol{\lambda}_{2i}^0 - \boldsymbol{\lambda}_{1i}^0$, for $i = 1, \ldots, N$. We collect the assumptions into three sets: Section 3.1.1 states the assumption needed to identify (1) from a linear factor model; Section 3.1.2 lists the assumptions that are sufficient to ensure consistency of the estimator for factors, loadings and break fraction; Section 3.1.3 states the additional assumption required to obtain the convergence rates of the estimators.

#### 3.1.1 Identification

**Assumption I - Breakpoint Factor Model.** For $0.5 < \alpha^0 \leq 1$, $\boldsymbol{\delta}^0_{\mathbf{x}i} \neq 0$ for $i = 1, \ldots, \left[N^{\alpha^0}\right]$, and $\sum_{i=\left[N^{\alpha^0}\right]+1}^{N} \left\|\boldsymbol{\delta}^0_{\mathbf{x}i}\right\| = O\left(1\right)$.

According to Assumption I, *at least* a fraction $O\left(N^{\alpha^0}\right)$ of the $N$ series experiences a break, for $0.5 < \alpha^0 \leq 1$. Bates *et al.* (2013) show that if at most $O\left(N^{0.5}\right)$ series undergo a break, the principal components estimator applied to the misspecified linear model achieves the Bai and Ng (2002) convergence rate $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$: Assumption I ensures that enough series experience a break so that (1) is identified from a linear factor model when factors and loadings are estimated by principal components. Assumption I relates to Assumption 1 in Chen *et al.* (2014): at least $O\left(N^{\alpha^0}\right)$ breaks are *big*, as defined in Assumption 1(a); the remaining breaks are *small*, according to Assumption 1(b) with $N$ and $T$ of the same order.

#### 3.1.2 Consistency

**Assumption C1 - Factors.** $\mathrm{E}\left\|\mathbf{f}^0_t\right\|^4 < \infty$; for $j = 1, 2$, $T^{-1}\sum_{t=1}^{T} \mathbf{f}^0_{jt}\left(\pi_{\mathbf{x}}\right) \mathbf{f}^0_{jt}\left(\pi^0_{\mathbf{x}}\right)' \xrightarrow{p} \boldsymbol{\Sigma}^0_{j\mathbf{f}}\left(\pi_{\mathbf{x}}, \pi^0_{\mathbf{x}}\right)$ as $T \to \infty$ for all $\pi_{\mathbf{x}}$ and some positive definite matrix $\boldsymbol{\Sigma}^0_{j\mathbf{f}}\left(\pi_{\mathbf{x}}, \pi^0_{\mathbf{x}}\right)$.

**Assumption C2 - Factor Loadings.** For $j = 1, 2$ and $i = 1, \ldots, N$, $\left\|\boldsymbol{\lambda}^0_{ji}\right\| \leq \bar{\lambda} < \infty$, and $\left\|\boldsymbol{\Lambda}^{0\prime}_j \boldsymbol{\Lambda}^0_j / N - \mathbf{D}^0_{\boldsymbol{\Lambda}_j}\right\| \to 0$ as $N \to \infty$ for some $R^0 \times R^0$ positive definite matrix $\mathbf{D}^0_{\boldsymbol{\Lambda}_j}$.

**Assumption C3 - Time and Cross-Section Dependence and Heteroskedasticity.** There exists a positive constant $M_{\mathbf{x}} < \infty$ such that for $j = 1, 2$, for all $\pi_{\mathbf{x}}$ and for all $(N, T)$,

**(a)** $\mathrm{E}\left(e_{it}\right) = 0$ and $\mathrm{E}\left|e_{it}\right|^8 \leq M_{\mathbf{x}}$;

**(b)** $\mathrm{E}\left[\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right)\mathbb{I}_{jv}\left(\pi_{\mathbf{x}}\right) e_{it} e_{iv}\right] = \tau_{jitv}\left(\pi_{\mathbf{x}}\right)$ with $\left|\tau_{jitv}\left(\pi_{\mathbf{x}}\right)\right| \leq \left|\tau_{jtv}\right|$ for some $\tau_{jtv}$ and for all $i$, and $T^{-1}\sum_{t=1}^{T}\sum_{v=1}^{T}\left|\tau_{jtv}\right| \leq M_{\mathbf{x}}$;

**(c)** $\mathrm{E}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right)e_{it}e_{lt}\right]=\sigma_{jil}\left(\pi_{\mathbf{x}}\right)$, $\left|\sigma_{jll}\left(\pi_{\mathbf{x}}\right)\right|\leq M_{\mathbf{x}}$ for all $l$, and $N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\left|\sigma_{jil}\left(\pi_{\mathbf{x}}\right)\right|\leq M_{\mathbf{x}}$;

**(d)** $\mathrm{E}\left|T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right)e_{it}e_{lt}-\mathrm{E}\left[\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right)e_{it}e_{lt}\right]\right|^{4}\leq M_{\mathbf{x}}$ for every $(i,l)$.

**Assumption C4 - Weak Dependence between $\mathbf{f}_t^0$ and $e_{it}$.** There exists some positive constant $M_{\mathbf{x}}<\infty$ such that for all $\pi_{\mathbf{x}}$ and for all $(N,T)$,

$$\mathrm{E}\left\{N^{-1}\sum_{i=1}^{N}\left\|T^{-1/2}\left[\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right)\mathbf{f}_t^0 e_{it}\right]\right\|^2\right\}\leq M_{\mathbf{x}},\quad j=1,2.$$

Assumptions C1 to C4 are the natural extensions of Assumptions A to D imposed on linear factor models in Bai and Ng (2002), and accommodate the presence of the breakpoint. Assumption C1 restricts the sequence $\left\{\mathbf{f}_t^0\right\}_{t=1}^{T}$ so that appropriate second moments exist; it also imposes full rank conditions that exclude multicollinearity in the factors. According to Assumption C2, factor loadings are nonstochastic and factors have a nonnegligible effect on the variance of $\mathbf{x}_t$ on each side of the breakpoint. Under Assumption C3, limited degrees of time-series and cross-section dependence in the idiosyncratic components as well as heteroskedasticity are allowed: this makes (1) an approximate breakpoint factor model; in particular, Assumption C3(b) is aligned to Assumption C2 in Bai and Ng (2002). Assumption C4 provides an upper bound to the degree of dependence between the factors and the idiosyncratic components: if $\mathbf{f}_t^0$ and $e_{it}$ are independent, as in Assumption 5(i) in Barigozzi *et al.* (2018), then Assumption C4 is implied by Assumptions C1 and C3(a). Assumptions C1 to C4 (including $\mathrm{E}\left|e_{it}\right|^8\leq M_{\mathbf{x}}$ in C3(a)) ensure that $\pi_{\mathbf{x}}^0$ can be consistently estimated, so that consistency of the estimators for factors and loadings can also be achieved.

### 3.1.3 Convergence Rates

**Assumption CR - Mixing Condition and Moment Bounds:** For $i=1,\ldots,N$, and $t=1,\ldots,T$,

**(a)** $\left\{\mathbf{f}_t^0, \mathbf{e}_t\right\}_{t=1}^T$ is $\rho-$mixing, with $\rho-$mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$;

**(b)** $\mathrm{E}\left(\left\|\mathbf{f}_t^0 e_{it}\right\|^4\right) \leq C_{\mathbf{x}}$ for some $C_{\mathbf{x}} < \infty$.

Assumption CR is analogous to Assumption 1 in Hansen (2000). Assumption CR(a) allows for nonstationarity and suitably restricts the memory of the process $\left\{\mathbf{f}_t^0, \mathbf{e}_t\right\}_{t=1}^T$, and thus of $\{\mathbf{x}_t\}_{t=1}^T$, so that Lemma 3.4 in Peligrad (1982) can be used: structural breaks are allowed, whereas unit roots are ruled out. Assumption CR(b) imposes an unconditional moment bound.

## 3.2 Estimation and Model Selection

We study estimation and model selection in Sections 3.2.1 and 3.2.2, respectively.

### 3.2.1 Principal Components Estimation

We estimate factors and loadings by principal components, and $\pi_{\mathbf{x}}^0$ by concentrated least squares: the latter requires minimizing an objective function that depends only on $\pi_{\mathbf{x}}$. Define the $N \times 2R$ matrix of loadings $\mathbf{\Lambda}^R = \left(\mathbf{\Lambda}_1^R, \mathbf{\Lambda}_2^R\right)$ and the $R \times T$ matrix of factors $\mathbf{F}^R = \left(\mathbf{f}_1^R, \ldots, \mathbf{f}_T^R\right)$, where the superscript $R$ denotes dependence on the number of factors. Let the $N \times 2R^0$ matrix $\mathbf{\Lambda}^0 = \left(\mathbf{\Lambda}_1^0, \mathbf{\Lambda}_2^0\right)$ and the $R^0 \times T$ matrix $\mathbf{F}^0 = \left(\mathbf{f}_1^0, \ldots, \mathbf{f}_T^0\right)$ be the true value of $\mathbf{\Lambda}^R$ and $\mathbf{F}^R$, respectively. For given $R$ in (1), the objective function in terms of $\mathbf{\Lambda}^R$, $\mathbf{F}^R$ and $\pi_{\mathbf{x}}$ is

$$S\left(\mathbf{\Lambda}^R, \mathbf{F}^R, \pi_{\mathbf{x}}\right) = (NT)^{-1} \sum_{t=1}^T \left\{ \begin{array}{c} \left[\mathbf{x}_t - \mathbb{I}_{1t}\left(\pi_{\mathbf{x}}\right) \mathbf{\Lambda}_1^R \mathbf{f}_t^R - \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}\right) \mathbf{\Lambda}_2^R \mathbf{f}_t^R\right]' \\ \times \left[\mathbf{x}_t - \mathbb{I}_{1t}\left(\pi_{\mathbf{x}}\right) \mathbf{\Lambda}_1^R \mathbf{f}_t^R - \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}\right) \mathbf{\Lambda}_2^R \mathbf{f}_t^R\right] \end{array} \right\}. \quad (3)$$

For given $\pi_{\mathbf{x}}$, define

$$\hat{\mathbf{\Sigma}}_{j\mathbf{x}}\left(\pi_{\mathbf{x}}\right) = \left[(NT)^{-1} \sum_{t=1}^T \mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right) \mathbf{x}_t \mathbf{x}_t'\right], \quad j = 1, 2. \quad (4)$$

Given the constraints $N^{-1}\left(\mathbf{\Lambda}_1^{R\prime}\mathbf{\Lambda}_1^R\right) = N^{-1}\left(\mathbf{\Lambda}_2^{R\prime}\mathbf{\Lambda}_2^R\right) = \mathbf{I}_R$, from (3) the estimator $\hat{\pi}_{\mathbf{x}}^R$ for $\pi_{\mathbf{x}}^0$ is

$$\hat{\pi}_{\mathbf{x}}^R = \arg\min_{\pi_{\mathbf{x}}} (NT)^{-1} \sum_{t=1}^{T} \mathbf{x}_t' \left\{ \mathbf{I}_N - N^{-1} \left[ \begin{array}{c} \mathbb{I}_{1t}(\pi_{\mathbf{x}})\,\hat{\mathbf{\Lambda}}_1^R(\pi_{\mathbf{x}})\,\hat{\mathbf{\Lambda}}_1^R(\pi_{\mathbf{x}})' \\ +\mathbb{I}_{2t}(\pi_{\mathbf{x}})\,\hat{\mathbf{\Lambda}}_2^R(\pi_{\mathbf{x}})\,\hat{\mathbf{\Lambda}}_2^R(\pi_{\mathbf{x}})' \end{array} \right] \right\} \mathbf{x}_t, \qquad (5)$$

where $\hat{\mathbf{\Lambda}}_j^R(\pi_{\mathbf{x}}) = \left[\hat{\boldsymbol{\lambda}}_{j1}^R(\pi_{\mathbf{x}}), \ldots, \hat{\boldsymbol{\lambda}}_{jN}^R(\pi_{\mathbf{x}})\right]'$ is the estimator for $\mathbf{\Lambda}_j^0$ given $\pi_{\mathbf{x}}$, for $j = 1, 2$: $\hat{\mathbf{\Lambda}}_j^R(\pi_{\mathbf{x}})$ is $\sqrt{N}$ times the $N \times R$ matrix of eigenvectors of $\hat{\mathbf{\Sigma}}_{j\mathbf{x}}(\pi_{\mathbf{x}})$ corresponding to its $R$ largest eigenvalues. The estimator $\hat{\mathbf{F}}^R(\pi_{\mathbf{x}}) = \left[\hat{\mathbf{f}}_1^R(\pi_{\mathbf{x}}), \ldots, \hat{\mathbf{f}}_T^R(\pi_{\mathbf{x}})\right]$ for $\mathbf{F}^0$ for given $\pi_{\mathbf{x}}$ is

$$\hat{\mathbf{f}}_t^R(\pi_{\mathbf{x}}) = N^{-1}\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}})\,\hat{\mathbf{\Lambda}}_1^R(\pi_{\mathbf{x}}) + \mathbb{I}_{2t}(\pi_{\mathbf{x}})\,\hat{\mathbf{\Lambda}}_2^R(\pi_{\mathbf{x}})\right]' \mathbf{x}_t, \quad t = 1, \ldots, T.$$

Given $\hat{\pi}_{\mathbf{x}}^R$, the estimators for $\mathbf{\Lambda}^0$ and $\mathbf{F}^0$ are $\hat{\mathbf{\Lambda}}^R\left(\hat{\pi}_{\mathbf{x}}^R\right) = \left[\hat{\mathbf{\Lambda}}_1^R\left(\hat{\pi}_{\mathbf{x}}^R\right), \hat{\mathbf{\Lambda}}_2^R\left(\hat{\pi}_{\mathbf{x}}^R\right)\right]$ and $\hat{\mathbf{F}}^R\left(\hat{\pi}_{\mathbf{x}}^R\right)$, respectively: if $R^0$ is known, the estimators for $\mathbf{\Lambda}^0$, $\mathbf{F}^0$ and $\pi_{\mathbf{x}}^0$ are $\hat{\mathbf{\Lambda}} = \hat{\mathbf{\Lambda}}^{R^0}\left(\hat{\pi}_{\mathbf{x}}^{R^0}\right)$, $\hat{\mathbf{F}} = \hat{\mathbf{F}}^{R^0}\left(\hat{\pi}_{\mathbf{x}}^{R^0}\right)$ and $\hat{\pi}_{\mathbf{x}} = \hat{\pi}_{\mathbf{x}}^{R^0}$, respectively. If $R^0$ is unknown we proceed as in Section 3.2.2 below.

#### 3.2.2 Selecting the Number of Factors

There exist several procedures to determine the unknown number of factors $R^0$. Bai and Ng (2002), Alessi *et al.* (2010), Kapetanios (2010), Onatski (2010), Ahn and Horenstein (2013), and Caner and Han (2014) consider the static factor model. Amengual and Watson (2007) look at the restricted dynamic case. Hallin and Liška (2007), and Onatski (2009) focus upon the generalized dynamic factor model. Breitung and Eickmeier (2011) show that neglecting structural breaks in the loadings leads to overestimation of the number of factors. We robustify Bai and Ng (2002) selection criteria to account for the unknown break fraction.

Let $\bar{R}$ be any *a priori* number of factors $R = \bar{R}$ such that $\bar{R} \geq R^0$. From (5), the estimator for $\pi_{\mathbf{x}}^{0}$ is $\hat{\pi}_{\mathbf{x}}^{\bar{R}}$. Given (3), the breakpoint robust Bai and Ng (2002) information criteria are

$$IC_{p1}(R,R) = \ln S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right), \hat{\pi}_{\mathbf{x}}^{\bar{R}}\right] + (R+R)\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

$$IC_{p2}(R,R) = \ln S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right), \hat{\pi}_{\mathbf{x}}^{\bar{R}}\right] + (R+R)\left(\frac{N+T}{NT}\right)\ln\left(C_{NT}^{2}\right), \tag{6}$$

$$IC_{p3}(R,R) = \ln S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right), \hat{\pi}_{\mathbf{x}}^{\bar{R}}\right] + (R+R)\left[\frac{\ln\left(C_{NT}^{2}\right)}{C_{NT}^{2}}\right].$$

The following theorem states the validity of the proposed information criteria.

**Theorem 3.1** *Under Assumptions I, C1-C4 and CR, the criteria* $IC_{p1}(R,R)$*,* $IC_{p2}(R,R)$ *and* $IC_{p3}(R,R)$ *defined in* (6) *consistently estimate the number of factors* $R^0$*.*

The estimator $\hat{R}$ for $R^0$ is obtained by minimizing the information criteria in (6). Theorem 3.1 implies that $\hat{R}$ involves a two-step estimation strategy: in the first step, for any *a priori* specified number of factors $R = \bar{R}$ such that $\bar{R} \geq R^0$, the estimator $\hat{\pi}_{\mathbf{x}}^{\bar{R}}$ may be obtained from (5); in the second step, the criteria in (6) may be computed by plugging $\hat{\mathbf{\Lambda}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right)$, $\hat{\mathbf{F}}^{R}\left(\hat{\pi}_{\mathbf{x}}^{\bar{R}}\right)$ and $\hat{\pi}_{\mathbf{x}}^{\bar{R}}$ in (3). In practice, given a bounded integer $R^{\max} \geq R^0$, one may set $\bar{R} = R^{\max}$ in (5).

# 4 The Unstable Factor Augmented Regression

We study estimation and inference in (2) under the assumption that the true number of factors $R^0$ is known: if not, the selection criteria in (6) may be used to estimate $R^0$. Define $\boldsymbol{\theta}_l = \left(\boldsymbol{\gamma}_l', \boldsymbol{\beta}_l'\right)'$, and let $\boldsymbol{\theta}_l^0 = \left(\boldsymbol{\gamma}_l^{0\prime}, \boldsymbol{\beta}_l^{0\prime}\right)'$ and $\pi_y^0$ be the true values of $\boldsymbol{\theta}_l$ and $\pi_y$, respectively, for $l = 1, 2$; given $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, let $\boldsymbol{\theta}^0 = \left(\boldsymbol{\theta}_1^{0\prime}, \boldsymbol{\theta}_2^{0\prime}\right)'$ be the true value of $\boldsymbol{\theta}$. In Section 4.1 we consider least squares estimation of the $\left[\left(2R^0 + 2K\right) + 1\right] \times 1$ vector of coefficients $\left(\boldsymbol{\theta}^{0\prime}, \pi_y^0\right)'$: this goes beyond Wang *et al.* (2015), who estimate unstable factor augmented regressions when factors are extracted from linear models. In Section 4.2 we propose a test for the null hypothesis of structural stability $\mathbb{H}_0 : \left(\boldsymbol{\theta}_1^0 = \boldsymbol{\theta}_2^0\right)$ in (2): this complements Corradi and Swanson (2014)

Hausman-type test, which does not identify the source of instability between factor model and factor augmented regression, but it allows for the multiple break scenario under the alternative.

### 4.1 Estimation and Inference

Define $\mathbb{I}_{1t}(\pi_y) = \mathbb{I}(t/T \leq \pi_y)$ and $\mathbb{I}_{2t}(\pi_y) = \mathbb{I}(t/T > \pi_y)$; let $\mathcal{W}_{R^0+K}(\cdot)$ be a $\left(R^0 + K\right) \times 1$ vector of standard Brownian motions defined on $[0, 1]$.

**Assumption FR.** Let $\mathbf{z}_t^0 = \left(\mathbf{f}_t^{0\prime}, \mathbf{w}_t'\right)'$. Then:

**(a)** $T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y)\,\mathbf{z}_t^0\mathbf{z}_t^{0\prime} \xrightarrow{p} \pi_y\boldsymbol{\Sigma}_{\mathbf{z}}^0$ for all $\pi_y$, where $\boldsymbol{\Sigma}_{\mathbf{z}}^0 = \mathrm{E}\left(\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\right)$ is a positive definite matrix;

**(b)** For $l = 1, 2$, $T^{-1}\sum_{t=1}^{T}\mathbb{I}_{lt}(\pi_y)\,\mathbf{z}_t^0\varepsilon_{t+h} \xrightarrow{p} \mathbf{0}$ for all $\pi_y$ and for any $h \geq 0$;

**(c)** $\left\{\mathbf{z}_t^0, \varepsilon_{t+h}\right\}_{t=1}^{T}$ is $\rho-$mixing, with $\rho-$mixing coefficients satisfying $\sum_{m=1}^{\infty}\rho_m^{1/2} < \infty$, for any $h \geq 0$;

**(d)** $\mathrm{E}\left(\left\|\mathbf{z}_t^0\varepsilon_{t+h}\right\|^4\right) \leq C_y$ and $\mathrm{E}\left(\left\|\mathbf{z}_t^0\right\|^4\right) \leq C_y$ for some $C_y < \infty$, for $t = 1, \ldots, T$;

**(e)** $T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y)\,\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{z}_t^0\varepsilon_{t+h} \Rightarrow \left(\pi_{\mathbf{x}}^0\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right)^{1/2}\mathcal{W}_{R^0+K}(\pi_y)$ for all $\pi_y$ and any $h \geq 0$, where $\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0$ is a positive definite matrix defined as $\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0 = \lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\sum_{v=1}^{T}\mathrm{E}\left(\varepsilon_{t+h}\varepsilon_{v+h}\mathbf{z}_t^0\mathbf{z}_v^{0\prime}\right)$.

Assumption FR extends Assumption Y1 in Stock and Watson (2002), and Assumption E in Bai and Ng (2006). Assumption FR(a) is a full rank condition. Assumption FR(b) is required to obtain consistent estimators for $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$. Assumptions FR(c) and FR(d) are equivalent to Assumptions CR(a) and CR(b), respectively, as discussed in Section 3.1.3. Assumption FR(e) is needed to derive the asymptotic distribution of the estimators for $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$ and of the test statistic for the null hypothesis of stability: it allows for serial correlation and heteroskedasticity in the sequence of error terms $\{\varepsilon_{t+h}\}_{t=1}^{T}$.

The feasible objective function is the sum of squared residuals (divided by $T$)

$$L\left(\hat{\mathbf{F}},\boldsymbol{\theta},\pi_y\right)=T^{-1}\sum_{t=1}^{T}\left[y_{t+h}-\mathbb{I}_{1t}\left(\pi_y\right)\left(\boldsymbol{\gamma}_1'\hat{\mathbf{f}}_t+\boldsymbol{\beta}_1'\mathbf{w}_t\right)-\mathbb{I}_{2t}\left(\pi_y\right)\left(\boldsymbol{\gamma}_2'\hat{\mathbf{f}}_t+\boldsymbol{\beta}_2'\mathbf{w}_t\right)\right]^2.$$

Let $\hat{\mathbf{z}}_t=\left(\hat{\mathbf{f}}_t',\mathbf{w}_t'\right)'$. For given $\pi_y$, the least squares estimator for $\boldsymbol{\theta}^0$ is $\hat{\boldsymbol{\theta}}\left(\pi_y\right)=\left[\hat{\boldsymbol{\theta}}_1\left(\pi_y\right)',\hat{\boldsymbol{\theta}}_2\left(\pi_y\right)'\right]'$, where

$$\hat{\boldsymbol{\theta}}_l\left(\pi_y\right)=\left[\hat{\boldsymbol{\gamma}}_l\left(\pi_y\right)',\hat{\boldsymbol{\beta}}_l\left(\pi_y\right)'\right]'=\left[\sum_{t=1}^{T}\mathbb{I}_{lt}\left(\pi_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^{T}\mathbb{I}_{lt}\left(\pi_y\right)\hat{\mathbf{z}}_t y_{t+h}\right],\quad l=1,2,$$

is the estimator for $\boldsymbol{\theta}_l^0$, and $\hat{\boldsymbol{\gamma}}_l\left(\pi_y\right)$ and $\hat{\boldsymbol{\beta}}_l\left(\pi_y\right)$ are the estimators for $\boldsymbol{\gamma}_l^0$ and $\boldsymbol{\beta}_l^0$, respectively: the concentrated loss function is

$$L_{\boldsymbol{\theta}}\left(\hat{\mathbf{F}},\pi_y\right)=T^{-1}\sum_{t=1}^{T}\left[y_{t+h}-\mathbb{I}_{1t}\left(\pi_y\right)\hat{\boldsymbol{\theta}}_1\left(\pi_y\right)'\hat{\mathbf{z}}_t-\mathbb{I}_{2t}\left(\pi_y\right)\hat{\boldsymbol{\theta}}_2\left(\pi_y\right)'\hat{\mathbf{z}}_t\right]^2.$$

The estimator $\hat{\pi}_y$ and $\hat{\boldsymbol{\theta}}$ for $\pi_y^0$ and $\boldsymbol{\theta}^0$, respectively, then are

$$\hat{\pi}_y=\arg\min_{\pi_y}L_{\boldsymbol{\theta}}\left(\hat{\mathbf{F}},\pi_y\right),\quad\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}\left(\hat{\pi}_y\right)=\left[\hat{\boldsymbol{\theta}}_1\left(\hat{\pi}_y\right)',\hat{\boldsymbol{\theta}}_2\left(\hat{\pi}_y\right)'\right]':$$

it follows that $\hat{\boldsymbol{\theta}}_l=\hat{\boldsymbol{\theta}}_l\left(\hat{\pi}_y\right)=\left[\hat{\boldsymbol{\gamma}}_l\left(\hat{\pi}_y\right)',\hat{\boldsymbol{\beta}}_l\left(\hat{\pi}_y\right)'\right]'$ is the estimator for $\boldsymbol{\theta}_l^0$, and $\hat{\boldsymbol{\gamma}}_l=\hat{\boldsymbol{\gamma}}_l\left(\hat{\pi}_y\right)$ and $\hat{\boldsymbol{\beta}}_l=\hat{\boldsymbol{\beta}}_l\left(\hat{\pi}_y\right)$ are the estimators for $\boldsymbol{\gamma}_l^0$ and $\boldsymbol{\beta}_l^0$, respectively, for $l=1,2$.

### 4.1.1 Consistency

Define the $R^0\times T$ matrices of state-specific factors $\mathbf{F}_j^0\left(\pi_{\mathbf{x}}\right)=\left[\mathbf{f}_{j1}^0\left(\pi_{\mathbf{x}}\right),\ldots,\mathbf{f}_{jT}^0\left(\pi_{\mathbf{x}}\right)\right]$, for $j=1,2$, such that $\mathbf{F}_1^0\left(\pi_{\mathbf{x}}\right)+\mathbf{F}_2^0\left(\pi_{\mathbf{x}}\right)=\left(\mathbf{f}_1^0,\ldots,\mathbf{f}_T^0\right)=\mathbf{F}^0$ and $\mathbf{F}_1^0\left(\pi_{\mathbf{x}}\right)\mathbf{F}_2^0\left(\pi_{\mathbf{x}}\right)'=\mathbf{0}_{R^0}$. Let $\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}\right)$ be the rotation matrix

$$\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}\right)=\frac{\mathbf{F}_j^0\left(\pi_{\mathbf{x}}^0\right)\mathbf{F}_j^0\left(\pi_{\mathbf{x}}\right)'}{T}\frac{\boldsymbol{\Lambda}_j^{0\prime}\hat{\boldsymbol{\Lambda}}_j\left(\pi_{\mathbf{x}}\right)}{N}\hat{\mathbf{V}}_j\left(\pi_{\mathbf{x}}\right)^{-1},\quad j=1,2,\tag{7}$$

where $\hat{\mathbf{V}}_j(\pi_{\mathbf{x}})$ is the $R^0 \times R^0$ diagonal matrix of the first $R^0$ largest eigenvalues of $\hat{\mathbf{\Sigma}}_{j\mathbf{x}}(\pi_{\mathbf{x}})$ in (4) in decreasing order. Let $\hat{\mathbf{\Phi}}_{jj}(\pi_{\mathbf{x}}^0) = \text{diag}\left[\hat{\mathbf{H}}_{jj}(\pi_{\mathbf{x}}^0), \mathbf{I}_K\right]$ be a $(R^0 + K) \times (R^0 + K)$ block diagonal matrix, for $j = 1, 2$. Define the $(R^0 + K) \times (R^0 + K)$ block diagonal matrices $\hat{\mathbf{\Phi}}_{12}(\pi_{\mathbf{x}}^0, \pi_y^0)$ and $\hat{\mathbf{\Phi}}_{21}(\pi_{\mathbf{x}}^0, \pi_y^0)$ as

$$\hat{\mathbf{\Phi}}_{12}(\pi_{\mathbf{x}}^0, \pi_y^0) = \frac{\min\{\pi_{\mathbf{x}}^0, \pi_y^0\}}{\pi_y^0} \hat{\mathbf{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}(\pi_{\mathbf{x}}^0 < \pi_y^0) \frac{\pi_y^0 - \pi_{\mathbf{x}}^0}{\pi_y^0} \hat{\mathbf{\Phi}}_{22}(\pi_{\mathbf{x}}^0)$$

and

$$\hat{\mathbf{\Phi}}_{21}(\pi_{\mathbf{x}}^0, \pi_y^0) = \frac{1 - \max\{\pi_{\mathbf{x}}^0, \pi_y^0\}}{1 - \pi_y^0} \hat{\mathbf{\Phi}}_{22}(\pi_{\mathbf{x}}^0) + \mathbb{I}(\pi_{\mathbf{x}}^0 > \pi_y^0) \frac{\pi_{\mathbf{x}}^0 - \pi_y^0}{1 - \pi_y^0} \hat{\mathbf{\Phi}}_{11}(\pi_{\mathbf{x}}^0),$$

respectively. The following theorem states the consistency of $\hat{\pi}_y$, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ as estimators for $\pi_y^0$, $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$, respectively.

**Theorem 4.1** *Under Assumptions I, C1-C4, FR(a) and FR(b),* $\hat{\pi}_y \xrightarrow{p} \pi_y^0$, $\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{\Phi}}_{12}(\pi_{\mathbf{x}}^0, \pi_y^0)' \boldsymbol{\theta}_1^0 \xrightarrow{p} 0$ *and* $\hat{\boldsymbol{\theta}}_2 - \hat{\mathbf{\Phi}}_{21}(\pi_{\mathbf{x}}^0, \pi_y^0)' \boldsymbol{\theta}_2^0 \xrightarrow{p} 0$, *as* $N, T \to \infty$.

Theorem 4.1 extends Theorem 2 in Stock and Watson (2002). The factor augmented regression in (2) is identified up to a rotation because the latent factors satisfy

$$\begin{aligned}\mathbb{I}_{1t}(\pi_y^0)\boldsymbol{\gamma}_1^{0\prime}\mathbf{f}_t^0 + \mathbb{I}_{2t}(\pi_y^0)\boldsymbol{\gamma}_2^{0\prime}\mathbf{f}_t^0 &= \mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\mathbb{I}_{1t}(\pi_y^0)\boldsymbol{\gamma}_1^{0\prime}\mathbf{L}_{11}\mathbf{L}_{11}^{-1}\mathbf{f}_t^0 + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\mathbb{I}_{1t}(\pi_y^0)\boldsymbol{\gamma}_1^{0\prime}\mathbf{L}_{21}\mathbf{L}_{21}^{-1}\mathbf{f}_t^0 \\ &\quad + \mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\mathbb{I}_{2t}(\pi_y^0)\boldsymbol{\gamma}_2^{0\prime}\mathbf{L}_{12}\mathbf{L}_{12}^{-1}\mathbf{f}_t^0 + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\mathbb{I}_{2t}(\pi_y^0)\boldsymbol{\gamma}_2^{0\prime}\mathbf{L}_{22}\mathbf{L}_{22}^{-1}\mathbf{f}_t^0\end{aligned},$$

for some positive definite matrices $\mathbf{L}_{11}$, $\mathbf{L}_{21}$, $\mathbf{L}_{12}$ and $\mathbf{L}_{22}$. Theorem 4.1 relates to the difference between $\hat{\boldsymbol{\theta}}_l$ and the space spanned by $\boldsymbol{\theta}_l^0$, for $l = 1, 2$. The rotation induced around $\boldsymbol{\theta}_l^0$ is a convex linear combination of the rotations induced by $\hat{\mathbf{\Phi}}_{11}(\pi_{\mathbf{x}}^0)$ and $\hat{\mathbf{\Phi}}_{22}(\pi_{\mathbf{x}}^0)$, which depend on the matrices $\hat{\mathbf{H}}_{11}(\pi_{\mathbf{x}})$ and $\hat{\mathbf{H}}_{22}(\pi_{\mathbf{x}})$, respectively, as defined in (7); these arise from the rotational indeterminacy stemming from the underlying breakpoint factor model in (1). The rotation around $\boldsymbol{\theta}_l^0$ depends on the relative position of the break fractions $\pi_{\mathbf{x}}^0$ and $\pi_y^0$, for $l = 1, 2$. Let us consider $\hat{\boldsymbol{\theta}}_1$ (analogous arguments apply to $\hat{\boldsymbol{\theta}}_2$). If $\pi_{\mathbf{x}}^0 < \pi_y^0$, the break in the factor loadings occurs before the break in the factor augmented regression: the rotation induced around $\boldsymbol{\theta}_1^0$ is affected by the rotations around $\boldsymbol{\Lambda}_1^0$ and $\boldsymbol{\Lambda}_2^0$, with weights equal to $\pi_{\mathbf{x}}^0 / \pi_y^0$ and $\left(\pi_y^0 - \pi_{\mathbf{x}}^0\right) / \pi_y^0$, respectively. If $\pi_{\mathbf{x}}^0 \geq \pi_y^0$, the rotation around $\boldsymbol{\theta}_1^0$ only depends on the rotation around $\boldsymbol{\Lambda}_1^0$.

### 4.1.2 Rates of Convergence

The following theorem states the convergence rate of the concentrated least squares estimator $\hat{\pi}_y$ for the break fraction $\pi_y^0$.

**Theorem 4.2** *Under Assumptions I, C1-C4 and FR(a)-FR(d),*

$$T\left(\hat{\pi}_y - \pi_y^0\right) = O_p(1).$$

The $T$ convergence rate relates to the estimator for the break fraction $\hat{\pi}_y$: Theorem 4.2 implies that with high probability the difference between the estimator $\hat{T}_y = [T\hat{\pi}_y]$ for the break date and its true value $T_y^0 = \left[T\pi_y^0\right]$ is bounded, namely $\left(\hat{T}_y - T_y^0\right) = O_p(1)$ (Bai and Perron (1998)). Let $\hat{\mathbf{k}}_{lt}(\pi_y) = \mathbb{I}_{lt}(\pi_y)\hat{\varepsilon}_{t+h}(\pi_y)\hat{\mathbf{z}}_t$, for $l = 1, 2$, where $\hat{\varepsilon}_{t+h}(\pi_y) = y_{t+h} - \mathbb{I}_{1t}(\pi_y)\hat{\boldsymbol{\theta}}_1(\pi_y)'\hat{\mathbf{z}}_t - \mathbb{I}_{2t}(\pi_y)\hat{\boldsymbol{\theta}}_2(\pi_y)'\hat{\mathbf{z}}_t$. From Newey and West (1987), define: $\hat{\mathbf{K}}_{ld}(\pi_y) = T^{-1}\sum_{t=d+1}^{T}\hat{\mathbf{k}}_{lt}(\pi_y)\hat{\mathbf{k}}_{l,t-d}(\pi_y)'$, for $d = 0, \ldots, D_T$, with $D_T = o\left(T^{1/4}\right)$; $\hat{\boldsymbol{\Omega}}_l(\pi_y) = \hat{\mathbf{K}}_{l0}(\pi_y) + \sum_{d=1}^{D_T} w(d, D_T)\left[\hat{\mathbf{K}}_{ld}(\pi_y) + \hat{\mathbf{K}}_{ld}(\pi_y)'\right]$, where $w(d, D_T) = [1 - d/(D_T + 1)]$ is the Bartlett kernel. Theorem 4.2 feeds into the following theorem, which states the asymptotic distribution of the estimators for $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$.

**Theorem 4.3** *Under Assumptions I, C1-C4, CR and FR, if* $\sqrt{T}/N \to 0$ *then*

$$\sqrt{T}\left[\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0, \pi_y^0\right)'\boldsymbol{\theta}_1^0\right] \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_1}^0\right)$$

*and*

$$\sqrt{T}\left[\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\Phi}}_{21}\left(\pi_{\mathbf{x}}^0, \pi_y^0\right)'\boldsymbol{\theta}_2^0\right] \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_2}^0\right),$$

*with*

$$\boldsymbol{\Sigma}^0_{\hat{\boldsymbol{\theta}}_1} = \frac{1}{\pi^0_y}\left[\begin{array}{c} \frac{\min\left\{\pi^0_{\mathbf{x}}, \pi^0_y\right\}}{\pi^0_y}\boldsymbol{\Phi}^0_{11}\left(\pi^0_{\mathbf{x}}\right)'\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Sigma}^0_{\mathbf{z}\varepsilon}\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Phi}^0_{11}\left(\pi^0_{\mathbf{x}}\right) \\ +\mathbb{I}\left(\pi^0_{\mathbf{x}} < \pi^0_y\right)\frac{\pi^0_y - \pi^0_{\mathbf{x}}}{\pi^0_y}\boldsymbol{\Phi}^0_{22}\left(\pi^0_{\mathbf{x}}\right)'\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Sigma}^0_{\mathbf{z}\varepsilon}\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Phi}^0_{22}\left(\pi^0_{\mathbf{x}}\right) \end{array}\right]$$

*and*

$$\boldsymbol{\Sigma}^0_{\hat{\boldsymbol{\theta}}_2} = \frac{1}{1-\pi^0_y}\left[\begin{array}{c} \frac{1-\max\left\{\pi^0_{\mathbf{x}}, \pi^0_y\right\}}{1-\pi^0_y}\boldsymbol{\Phi}^0_{22}\left(\pi^0_{\mathbf{x}}\right)'\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Sigma}^0_{\mathbf{z}\varepsilon}\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Phi}^0_{22}\left(\pi^0_{\mathbf{x}}\right) \\ +\mathbb{I}\left(\pi^0_{\mathbf{x}} > \pi^0_y\right)\frac{\pi^0_{\mathbf{x}} - \pi^0_y}{1-\pi^0_y}\boldsymbol{\Phi}^0_{11}\left(\pi^0_{\mathbf{x}}\right)'\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Sigma}^0_{\mathbf{z}\varepsilon}\left(\boldsymbol{\Sigma}^0_{\mathbf{z}}\right)^{-1}\boldsymbol{\Phi}^0_{11}\left(\pi^0_{\mathbf{x}}\right) \end{array}\right],$$

*where* $\boldsymbol{\Phi}^0_{jj}\left(\pi^0_{\mathbf{x}}\right) = \operatorname{diag}\left[\boldsymbol{\Sigma}^0_{j\mathbf{f}}\left(\pi^0_{\mathbf{x}}, \pi^0_{\mathbf{x}}\right)\mathbf{Q}^0_{\boldsymbol{\Lambda}_j}\left(\pi^0_{\mathbf{x}}\right)\mathbf{V}^0_j\left(\pi^0_{\mathbf{x}}\right)^{-1}, \mathbf{I}_K\right]$, $\mathbf{Q}^0_{\boldsymbol{\Lambda}_j}\left(\pi^0_{\mathbf{x}}\right) = \operatorname{p}\lim\left[\boldsymbol{\Lambda}^{0\prime}_j\hat{\boldsymbol{\Lambda}}_j\left(\pi^0_{\mathbf{x}}\right)\right]/N$ *and* $\mathbf{V}^0_j\left(\pi^0_{\mathbf{x}}\right) = \operatorname{p}\lim\hat{\mathbf{V}}_j\left(\pi^0_{\mathbf{x}}\right)$. *Consistent estimators for* $\boldsymbol{\Sigma}^0_{\hat{\boldsymbol{\theta}}_1}$ *and* $\boldsymbol{\Sigma}^0_{\hat{\boldsymbol{\theta}}_2}$*, denoted by* $\widehat{\operatorname{Avar}}\left(\hat{\boldsymbol{\theta}}_1\right)$ *and* $\widehat{\operatorname{Avar}}\left(\hat{\boldsymbol{\theta}}_2\right)$*, respectively, are*

$$\widehat{\operatorname{Avar}}\left(\hat{\boldsymbol{\theta}}_l\right) = \left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{lt}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\hat{\boldsymbol{\Omega}}_l\left(\hat{\pi}_y\right)\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{lt}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}, \quad l = 1, 2. \tag{8}$$

Theorem 4.3 extends Theorem 1 in Bai and Ng (2006). It establishes convergence rate and limiting distribution of the least squares estimator $\hat{\boldsymbol{\theta}}_l$ for $\boldsymbol{\theta}^0_l$, for $l = 1, 2$. The estimated covariance matrix in (8) is robust to heteroskedasticity and autocorrelation. Let $\hat{\sigma}^2_{\varepsilon} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}^2_{t+h}$, with $\hat{\varepsilon}_{t+h} = \hat{\varepsilon}_{t+h}\left(\hat{\pi}_y\right)$. With homoskedastic and uncorrelated disturbances the estimator in (8) simplifies to

$$\widehat{\operatorname{Avar}}\left(\hat{\boldsymbol{\theta}}_l\right) = \hat{\sigma}^2_{\varepsilon}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{lt}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}, \quad l = 1, 2. \tag{9}$$

## 4.2 Testing for Structural Change

### 4.2.1 Testing Strategy

We now propose a test for the null hypothesis $\mathbb{H}_0 : \left(\boldsymbol{\theta}^0_1 = \boldsymbol{\theta}^0_2\right)$ in (2): this extends the literature on testing for a single break to allow for estimated factors (Andrews (1993) and Perron (2006)). We build a Lagrange multiplier statistic. Under the null hypothesis of linearity, the true factor augmented model becomes $y_{t+h} = \boldsymbol{\theta}^{0\prime}_1\mathbf{z}^0_t + \varepsilon_{t+h}$: the least squares estimator for $\boldsymbol{\theta}^0_1$ is $\hat{\boldsymbol{\theta}}^{\mathbb{H}_0}_1 =$

$\left(\sum_{t=1}^{T}\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right)^{-1}\left(\sum_{t=1}^{T}\hat{\mathbf{z}}_t y_{t+h}\right)$. Let $\hat{\mathbf{k}}_t^{\mathbb{H}_0}=\hat{\varepsilon}_{t+h}^{\mathbb{H}_0}\hat{\mathbf{z}}_t$, where $\hat{\varepsilon}_{t+h}^{\mathbb{H}_0}=y_{t+h}-\hat{\boldsymbol{\theta}}_1^{\mathbb{H}_0\prime}\hat{\mathbf{z}}_t$. From Newey and West (1987), define: $\hat{\mathbf{K}}_d^{\mathbb{H}_0}=T^{-1}\sum_{t=d+1}^{T}\hat{\mathbf{k}}_t^{\mathbb{H}_0}\hat{\mathbf{k}}_{t-d}^{\mathbb{H}_0\prime}$, for $d=0,\ldots,D_T$, with $D_T$ as in Section 4.1.2; $\hat{\boldsymbol{\Omega}}^{\mathbb{H}_0}=\hat{\mathbf{K}}_0^{\mathbb{H}_0}+\sum_{d=1}^{D_T}w\left(d,D_T\right)\left(\hat{\mathbf{K}}_d^{\mathbb{H}_0}+\hat{\mathbf{K}}_d^{\mathbb{H}_0\prime}\right)$, where $w\left(d,D_T\right)$ is the Bartlett kernel as in Section 4.1.2. Following Andrews (1993), the heteroskedasticity and autocorrelation robust Lagrange multiplier test statistic is defined as

$$\mathcal{LM}\left(\pi_y\right)=\frac{1}{\pi_y\left(1-\pi_y\right)}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\hat{\mathbf{z}}_t\hat{\varepsilon}_{t+h}^{\mathbb{H}_0}\right]'\left(\hat{\boldsymbol{\Omega}}^{\mathbb{H}_0}\right)^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\hat{\mathbf{z}}_t\hat{\varepsilon}_{t+h}^{\mathbb{H}_0}\right].$$

For known $\pi_y=\pi_y^0$ and under the null hypothesis, $\mathcal{LM}\left(\pi_y^0\right)$ has a $\chi^2$ limiting distribution with $\left(R^0+K\right)$ degrees of freedom as $N,T\to\infty$. However, $\pi_y^0$ is generally unknown and not identified under the null hypothesis. As in Andrews (1993), we propose the statistic

$$\sup\mathcal{LM}=\sup_{\pi_y\in\left[\underline{\pi}_y,1-\underline{\pi}_y\right]}\mathcal{LM}\left(\pi_y\right),$$

with $\underline{\pi}_y\in(0.00,0.50]$. The theorem below states the asymptotic distribution of $\sup\mathcal{LM}$.

**Theorem 4.4** *Let Assumptions I, C1-C4, CR, FR(a), FR(b) and FR(e) hold. Further assume that $\sqrt{T}/N\to 0$ as $N,T\to\infty$. Then for any $\underline{\pi}_y\in(0.00,0.50]$, and under the null hypothesis $\mathbb{H}_0:\left(\boldsymbol{\theta}_1^0=\boldsymbol{\theta}_2^0\right)$,*

$$\sup\mathcal{LM}\xrightarrow{d}\sup_{\pi_y\in\left[\underline{\pi}_y,1-\underline{\pi}_y\right]}\frac{\left[\mathcal{W}_{R^0+K}\left(\pi_y\right)-\pi_y\mathcal{W}_{R^0+K}\left(1\right)\right]'\left[\mathcal{W}_{R^0+K}\left(\pi_y\right)-\pi_y\mathcal{W}_{R^0+K}\left(1\right)\right]}{\pi_y\left(1-\pi_y\right)}$$

*as $N,T\to\infty$, provided that $\hat{\boldsymbol{\Omega}}^{\mathbb{H}_0}\xrightarrow{p}\boldsymbol{\Omega}^{0,\mathbb{H}_0}$ under $\mathbb{H}_0$, where*

$$\begin{aligned}\boldsymbol{\Omega}^{0,\mathbb{H}_0} &= \left\{\left(\pi_{\mathbf{x}}^0\right)^{1/2}\left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}+\left(1-\pi_{\mathbf{x}}^0\right)^{1/2}\left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\right\}\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0 \\ &\quad\times\left\{\left(\pi_{\mathbf{x}}^0\right)^{1/2}\left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}+\left(1-\pi_{\mathbf{x}}^0\right)^{1/2}\left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\right\}'\end{aligned}.$$

Theorem 4.4 extends Theorem 3 (b) in Andrews (1993) and implies that the critical values in Andrews (1993) apply to the $\sup\mathcal{LM}$ statistic; these values are used also in Breitung and

Eickmeier (2011), and Chen *et al.* (2014). The convergence in probability of $\hat{\mathbf{\Omega}}^{\mathbb{H}_0}$ to $\mathbf{\Omega}^{0,\mathbb{H}_0}$ is not stringent: factors are consistently estimated; $\hat{\mathbf{\Omega}}^{\mathbb{H}_0}$ is a heteroskedasticity and autocorrelation robust estimator for $\mathbf{\Omega}^{0,\mathbb{H}_0}$. With homoskedastic and serially uncorrelated disturbances, $\hat{\mathbf{\Omega}}^{\mathbb{H}_0}$ simplifies to $\hat{\mathbf{\Omega}}^{\mathbb{H}_0} = \left(\hat{\sigma}_{\varepsilon}^{\mathbb{H}_0}\right)^2 \left[T^{-1} \sum_{t=1}^{T} \mathbb{I}_{lt}\left(\hat{\pi}_y\right) \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t'\right]$, with $\left(\hat{\sigma}_{\varepsilon}^{\mathbb{H}_0}\right)^2 = T^{-1} \sum_{t=1}^{T} \left(\hat{\varepsilon}_{t+h}^{\mathbb{H}_0}\right)^2$.

By Theorem 4.1, under $\mathbb{H}_0 : \left(\boldsymbol{\theta}_1^0 = \boldsymbol{\theta}_2^0\right)$ it follows that $\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0, \pi_y^0\right)' \boldsymbol{\theta}_1^0 \xrightarrow{p} 0$ and $\hat{\boldsymbol{\theta}}_2 - \hat{\mathbf{\Phi}}_{21}\left(\pi_{\mathbf{x}}^0, \pi_y^0\right)' \boldsymbol{\theta}_1^0 \xrightarrow{p} 0$: under the null hypothesis, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ have different probability limits due to the different rotations around $\boldsymbol{\theta}_1^0$ induced by the structural break in the underlying factor model. However, this feature does not cause any size distortions, as it is fully captured by the asymptotic covariance matrix $\mathbf{\Omega}^{0,\mathbb{H}_0}$ defined in Theorem 4.4: the $\sup \mathcal{LM}$ statistic thus has the same asymptotic distribution as the one stated in Theorem 3 (b) in Andrews (1993).

#### 4.2.2 Robustness to Factor Model Representation

The test in Section 4.2.1 requires estimating the factors from the breakpoint model in (1). A factor model with structural instability has an equivalent linear representation (Breitung and Eickmeier (2011)), which one could use to test for stability in the factor augmented regression. Formally, let $\mathbf{f}_t^* = \mathbf{f}_t^*\left(\pi_{\mathbf{x}}, \mathbf{f}_t\right) = \left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}\right) \mathbf{f}_t', \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}\right) \mathbf{f}_t'\right]'$: the equivalent representation for the model in (1) is $\mathbf{x}_t = \mathbf{\Lambda} \mathbf{f}_t^* + \mathbf{e}_t$. Let $\mathbf{f}_t^{0*} = \mathbf{f}_t^*\left(\pi_{\mathbf{x}}^0, \mathbf{f}_t^0\right) = \left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right) \mathbf{f}_t^{0\prime}, \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right) \mathbf{f}_t^{0\prime}\right]'$ be the true value of $\mathbf{f}_t^*$: by Proposition 1 in Chen *et al.* (2014), the principal components estimator $\hat{\mathbf{f}}_t^*$ for the $\left(R^0 + R^0\right) \times 1$ vector $\mathbf{f}_t^{0*}$ is consistent up to a $\left(R^0 + R^0\right) \times \left(R^0 + R^0\right)$ rotation matrix and has convergence rate $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$. A test for structural change in the factor augmented regression based on the equivalent linear factor representation imposes $\left(R^0 + R^0 + K\right) \times 1$ restrictions as $N, T \to \infty$. The test in Section 4.2.1 has $\left(R^0 + K\right) \times 1$ restrictions. Estimating factors from the equivalent linear representation requires imposing a higher number of restrictions and results in a test with lower power.

# 5 Monte Carlo Analysis

The experiment related to model selection in the unstable factor model is described in Section 5.1. Those focusing on estimation, inference and stability testing in the factor augmented regression are detailed in Section 5.2. Section 5.3 discusses the results.

## 5.1 Model Selection in the Breakpoint Factor Model

We simulate the data using the two-factor Data Generating Process (DGP)

$$x_{it}^{s} = \mathbb{I}\left(t/T \leq \pi_{\mathbf{x}}^{0}\right)\left(\lambda_{11i}^{0} f_{1t}^{0s} + \lambda_{12i}^{0} f_{2t}^{0s}\right) + \mathbb{I}\left(t/T > \pi_{\mathbf{x}}^{0}\right)\left(\lambda_{21i}^{0} f_{1t}^{0s} + \lambda_{22i}^{0} f_{2t}^{0s}\right) + e_{it}^{s}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $s = 1, \ldots, 2000$ refers to the replication. We set $N = 25, 50, 100$ and $T = 100, 200, 400$. We define $\delta_{\mathbf{x}i}^{0} = \lambda_{21i}^{0} - \lambda_{11i}^{0} = \lambda_{22i}^{0} - \lambda_{12i}^{0}$, for $i = 1, \ldots, N$: we set $\delta_{\mathbf{x}i}^{0} > 0$ for $i = 1, \ldots, \left[N^{\alpha^0}\right]$, and $\delta_{\mathbf{x}i}^{0} = 0$ for $i = \left[N^{\alpha^0}\right] + 1, \ldots, N$. We fix the factor loadings $\lambda_{11i}^{0}$, $\lambda_{12i}^{0}$, $\lambda_{21i}^{0}$ and $\lambda_{22i}^{0}$ throughout the replications, with $\lambda_{11i}^{0} \sim \mathcal{N}(1,1)$ and $\lambda_{12i}^{0} \sim \mathcal{N}(1,1)$, for $i = 1, \ldots, N$. We control for: $(i)$ the number of cross-sectional units $\left[N^{\alpha^0}\right]$ with structural instability by setting $\alpha^0 = 0.60, 1.00$; $(ii)$ the magnitude of the break by setting $\delta_{\mathbf{x}i}^{0} = 0.25, 1.00$ for $i = 1, \ldots, \left[N^{\alpha^0}\right]$. As in Breitung and Eickmeier (2011), Chen *et al.* (2014), and Han and Inoue (2015), we set $\pi_{\mathbf{x}}^{0} = 0.50$.

We generate the factors $f_{1t}^{0s}$ and $f_{2t}^{0s}$ as

$$f_{kt}^{0s} = \rho_f f_{k,t-1}^{0s} + \left(1 - \rho_f^2\right)^{1/2} \varpi_{kft}^{s} \epsilon_{kft}^{s}, \quad f_{k,-50}^{0s} = 0, \quad k = 1, 2, \quad t = -49, \ldots, 0, \ldots, T, \tag{10}$$

with $\rho_f \sim \mathcal{U}(0.05, 0.95)$ fixed in repeated samples, $\mathrm{E}\left(\varpi_{kft}^{s}\right)^2 = 1$ and $\epsilon_{kft}^{s} \sim \mathrm{IID}\mathcal{N}(0,1)$, so $\mathrm{E}\left(f_{kt}^{0s}\right) = 0$ and $\mathrm{Var}\left(f_{kt}^{0s}\right) = 1$. We allow for conditional heteroskedasticity in $f_{kt}^{0s}$ through the $\mathrm{GARCH}(1,1)$ process $\left(\varpi_{kft}^{s}\right)^2 = \beta_{f1} + \beta_{f2}\left(\varpi_{kf,t-1}^{s}\right)^2 + \beta_{f3}\left(\varpi_{kf,t-1}^{s}\epsilon_{kf,t-1}^{s}\right)^2$, with $\left(\varpi_{kf,-50}^{s}\right)^2 = \mathrm{E}\left(\varpi_{kft}^{s}\right)^2 = 1$.

We generate the idiosyncratic components $e_{it}^s$ as

$$e_{it}^s = \rho_e e_{i,t-1}^s + \sigma_{ii}^{1/2} \left(1-\rho_e^2\right)^{1/2} \varpi_{e_it}^s \epsilon_{e_it}^s, \quad e_{i,-50}^s = 0, \quad i = 1, \dots, N, \quad t = -49, \dots, 0, \dots, T,$$

with $\rho_e \sim \mathcal{U}(0.05, 0.95)$ and $\sigma_{ii} \sim \chi(1)$ fixed in repeated samples. Let $\boldsymbol{\epsilon}_{et}^s = \left(\epsilon_{e_1t}^s, \dots, \epsilon_{e_Nt}^s\right)'$. We allow for cross-sectional dependence through the first order spatial autoregressive process $\boldsymbol{\epsilon}_{et}^s = \bar{\mathbf{Q}}\boldsymbol{\varrho}_{et}$, where $\bar{\mathbf{Q}} = \mathbf{Q}\left[N \Big/ \mathrm{tr}\left(\boldsymbol{\Sigma}_{e,\mathrm{diag}}^{1/2}\boldsymbol{\Omega}_{e,\mathrm{diag}}^{1/2}\mathbf{Q}\mathbf{Q}'\boldsymbol{\Omega}_{e,\mathrm{diag}}^{1/2}\boldsymbol{\Sigma}_{e,\mathrm{diag}}^{1/2}\right)\right]^{1/2}$, $\boldsymbol{\Sigma}_{e,\mathrm{diag}}^{1/2} = \mathrm{diag}\left[\left(\sigma_{11}^{1/2}, \dots, \sigma_{NN}^{1/2}\right)'\right]$, $\boldsymbol{\Omega}_{e,\mathrm{diag}}^{1/2} = \mathrm{diag}\left\{\left\{\left[\mathrm{E}\left(\varpi_{e_1t}^s\right)^2\right]^{1/2}, \dots, \left[\mathrm{E}\left(\varpi_{e_Nt}^s\right)^2\right]^{1/2}\right\}'\right\}$, $\boldsymbol{\varrho}_{et} \sim \mathrm{IID}\mathcal{N}(0, \mathbf{I}_N)$, and $\mathbf{Q} = \left(\mathbf{I}_N - \rho_{\mathbf{Q}}\mathbf{W}\right)^{-1}$: $\mathbf{W} = (w_{l_1l_2})$ is a rook-type matrix, namely all elements in $\mathbf{W}$ are zero except $w_{l_1+1,l_1} = w_{l_2-1,l_2} = 0.5$, for $l_1 = 1, \dots, N-2$ and $l_2 = 3, \dots, N$, with $w_{12} = w_{N,N-1} = 1$. In this way $\mathrm{Var}\left(e_{it}^s\right) = \sigma_{ii}\mathrm{E}\left(\varpi_{e_it}^s\right)^2 \Big/ \left[N^{-1}\sum_l^N \sigma_{ll}\mathrm{E}\left(\varpi_{e_lt}^s\right)^2\right]$ and $N^{-1}\sum_i^N \mathrm{Var}\left(e_{it}^s\right) = 1$. We model $\varpi_{e_it}^s$ as the GARCH(1,1) process $\left(\varpi_{e_it}^s\right)^2 = \beta_{f1} + \beta_{f2}\left(\varpi_{e_i,t-1}^s\right)^2 + \beta_{f3}\left(\varpi_{e_i,t-1}^s\epsilon_{e_i,t-1}^s\right)^2$, with $\left(\varpi_{e_i,-50}^s\right)^2 = \mathrm{E}\left(\varpi_{e_it}^s\right)^2 = 1$: it follows that $\mathrm{Var}\left(e_{it}^s\right) \to \sigma_{ii}$ as $N \to \infty$.

We study three cases: (*i*) time homoskedastic factors and idiosyncratic components, and cross-sectionally independent idiosyncratic components (CSI); (*ii*) time homoskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components (CSD); and (*iii*) time heteroskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components (CSDH). Under CSI, we set $\beta_{f1} = \beta_{e1} = 1$, $\beta_{f2} = \beta_{e2} = 0$, $\beta_{f3} = \beta_{e3} = 0$ and $\rho_{\mathbf{Q}} = 0$. We build CSD by imposing $\beta_{f1} = \beta_{e1} = 1$, $\beta_{f2} = \beta_{e2} = 0$, $\beta_{f3} = \beta_{e3} = 0$ and $\rho_{\mathbf{Q}} = 0.3$. We parameterize CSDH by setting $\beta_{f1} = \beta_{e1} = 0.1$, $\beta_{f2} = \beta_{e2} = 0.8$, $\beta_{f3} = \beta_{e3} = 0.1$ and $\rho_{\mathbf{Q}} = 0.3$.

To reduce the effect induced by the initial values $f_{k,-50}^{0s} = 0$, $\varpi_{kf,-50}^s = 1$, $e_{i,-50}^s = 0$ and $\varpi_{e_i,-50}^s = 1$, we discard the first 50 observations in the DGPs for $f_{kt}^{0s}$, $\varpi_{kft}^s$, $e_{it}^s$ and $\varpi_{e_it}^s$. We estimate factors and loadings as detailed in Section 3.2.1. We estimate $\pi_{\mathbf{x}}^0$ by grid search over the set of 19 equally spaced break fractions $\{0.05, 0.10, 0.15, \dots, 0.85, 0.90, 0.95\}$, which include the true value $\pi_{\mathbf{x}}^0 = 0.50$. The true number of factors $R^0 = 2$ is estimated with $R^{\max} = 8$. We assess the model selection criteria in (6) by reporting the average number of estimated factors over the replications.

### 5.2 Factor Augmented Regression

We consider the DGP

$$y_t^s = \mathbb{I}\left(t/T \leq \pi_y^0\right)\left(\gamma_1^0 f_{1t}^{0s} + \beta_1^0 w_t^s\right) + \mathbb{I}\left(t/T > \pi_y^0\right)\left(\gamma_2^0 f_{1t}^{0s} + \beta_2^0 w_t^s\right) + \varepsilon_t^s, \quad t = 1, \ldots, T,$$

with $\pi_y^0 = 0.50$, $\gamma_1^0 = \beta_1^0 = 1$, $\gamma_2^0 = \gamma_1^0 + \delta_y^0$, $\beta_2^0 = \beta_1^0 + \delta_y^0$, and $\delta_y^0$ chosen as discussed below. The latent factor $f_{1t}^{0s}$ is generated as in (10), with $\delta_{\mathbf{x}i}^0 = 1.00$ for $i = 1, \ldots, \left[N^{\alpha^0}\right]$, and $\delta_{\mathbf{x}i}^0 = 0$ for $i = \left[N^{\alpha^0}\right] + 1, \ldots, N$, with $\alpha^0 = 0.60$. We assume that the true number of factors $R^0 = 1$ is known. We generate the observable covariate $w_t^s$ as

$$w_t^s = \mu_w\left(1 - \rho_w\right) + \rho_w w_{t-1}^s + \left(1 - \rho_w^2\right)^{1/2} \epsilon_{wt}^s, \quad w_{-50}^s = 0, \quad \epsilon_{wt}^s \sim \mathrm{IID}\mathcal{N}\left(0, 1\right), \quad t = -49, \ldots, 0, \ldots, T,$$

with $\mu_w \sim \mathcal{N}(0, 1)$ and $\rho_w \sim \mathcal{U}(0.05, 0.95)$ fixed in repeated samples. We generate $\varepsilon_t^s$ as

$$\varepsilon_t^s = \rho_\varepsilon \varepsilon_{t-1}^s + \left(1 - \rho_\varepsilon^2\right)^{1/2} \varpi_{\varepsilon t}^s \epsilon_{\varepsilon t}^s, \quad \varepsilon_{-50}^s = 0, \quad \epsilon_{\varepsilon t}^s \sim \mathrm{IID}\mathcal{N}\left(0, 1\right), \quad t = -49, \ldots, 0, \ldots, T,$$

where $\varpi_{\varepsilon t}^s$ is parameterized as the GARCH $(1, 1)$ process $\left(\varpi_{\varepsilon t}^s\right)^2 = \beta_{\varepsilon 1} + \beta_{\varepsilon 2}\left(\varpi_{\varepsilon, t-1}^s\right)^2 + \beta_{\varepsilon 3}\left(\epsilon_{\varepsilon, t-1}^s \varpi_{\varepsilon, t-1}^s\right)^2$, with $\left(\varpi_{\varepsilon, -50}^s\right)^2 = \mathrm{E}\left(\varpi_{\varepsilon t}^s\right)^2 = 1$: under conditional homoskedasticity and independence, we fix $\beta_{\varepsilon 1} = 1$, $\beta_{\varepsilon 2} = \beta_{\varepsilon 3} = 0$ and $\rho_\varepsilon = 0$; we allow for conditional heteroskedasticity and serial dependence by setting $\beta_{\varepsilon 1} = 0.1$, $\beta_{\varepsilon 2} = 0.8$, $\beta_{\varepsilon 3} = 0.1$ and $\rho_\varepsilon = 0.3$. We reduce the effect of the initial values $w_{-50}^s = 0$, $\varepsilon_{-50}^s = 0$ and $\varpi_{\varepsilon, -50}^s = 1$ by discarding the first 50 observations in the DGPs for $w_t^s$, $\varepsilon_t^s$ and $\varpi_{\varepsilon t}^s$.

For estimation we focus on $\beta_1^0$ and $\pi_y^0$. We control for the magnitude of the break by setting $\delta_y^0 = 0.25, 1.00$. We assess the performance of the estimators $\hat{\beta}_1^s$ and $\hat{\pi}_y^s$ for $\beta_1^0$ and

$\pi_y^0$, respectively, from replication $s$ through their bias and RMSE. Given Theorem 4.3, $\hat{\beta}_1^s$ is asymptotically normal and we calculate the size at 5% level for the null hypothesis $\beta_1^0 = 1$: the asymptotic variance of $\hat{\beta}_1^s$ is estimated through (9) and (8) under homoskedasticity and independence, and under heteroskedasticity and dependence, respectively; in the latter case, we set $D_T = 5$ in the Bartlett kernel.

Finally, we assess the finite sample properties of the stability test proposed in Section 4.2. We compute size and power by setting $\delta_y^0 = 0.00$ and $\delta_y^0 = 0.25, 1.00$, respectively. We construct the test statistic under homoskedasticity and serial uncorrelation, as well as under the more general set up with heteroskedastic and correlated errors: in the latter case, we set $D_T = 5$ in the Bartlett kernel.

### 5.3 Results

The results are collected in three tables: Table 1 focuses on model selection in the unstable large dimensional factor model; Tables 2 and 3 show results from estimation and stability testing in the factor augmented regression, respectively.

Table 1 about here

Table 2 about here

Table 3 about here

Table 1 collects results for the selection criteria $IC_{p1}(R, R)$, $IC_{p2}(R, R)$ and $IC_{p3}(R, R)$ in (6) (Panels A, B and C, respectively) when $\alpha^0 = 0.60, 1.00$. The $IC_{p2}(R, R)$ criterion outperforms the other criteria under CSI, CSD and CSDH, and for all $\alpha^0$, $N$, $T$ and $\delta_{\mathbf{x}i}^0 > 0$: it overestimates the number of factors for $N = 25, 50$, whereas it works well for $N = 100$; compared to the CSI scenario, its performance tends to deteriorate as cross-sectional dependence and conditional heteroskedasticity are added into the DGP. The $IC_{p2}(R, R)$ criterion is not strongly affected by $\alpha^0$.

Table 2 shows results from estimation of the factor augmented regression. With homoskedastic and independent errors, the RMSE of the estimators for $\beta_1^0$ and $\pi_y^0$ decreases in $T$ and $\delta_y^0$, while it does not strongly depend upon $N$ (Panel B). Similar qualitative results hold for the bias (Panel A). The test for the null hypothesis $\beta_1^0 = 1$ (Panel C) is almost always correctly sized for $\delta_y^0 = 1.00$, with the empirical size diminishing as $T$ grows; for $\delta_y^0 = 0.25$, the break induces a weaker effect on the slope coefficients and the test is oversized, although the size improves as $T$ increases. With heteroskedastic and serially correlated errors, the size of the test deteriorates (Panel F): it improves as $N$, $T$ and $\delta_y^0$ increase; it reaches the nominal value for $N = 100$, $T = 400$ and $\delta_y^0 = 1.00$.

Table 3 reports results for the Lagrange multiplier test for stability in the factor augmented regression. For homoskedastic and independent idiosyncratic components (Panel A), the test is correctly sized for $T = 400$ regardless of the value of $N$, whereas it is sometimes undersized or oversized for smaller values of $T$; the power of the test increases with $\delta_y^0$ and $T$ (although the effect of size distortions needs to be accounted for), whereas it seems to be unaffected by $N$. With heteroskedastic and serially correlated errors (Panel B), the test still is correctly sized for $T = 400$; the power is however affected by the Newey and West (1987) correction.

In conclusion, the Monte Carlo results support the main theoretical contributions of this paper: the information criterion $IC_{p2}(R, R)$ in (6) is a valid tool to select the number of factors; the estimator for the factor augmented regression performs well in finite samples; the related stability test has good size and power properties.

## 6 Application to Bond Risk Premia

Section 6.1 links the paper to the literature. Section 6.2 describes the data. Section 6.3 details the factor model. Section 6.4 presents the results from the factor augmented regression.

### 6.1 Literature and Contribution

We apply our methodology to uncover potential structural instabilities in the empirical linkages between bond market risk premia and macroeconomic fundamentals. Fama and Bliss (1987) show that the one-year excess return on the $n$-year bond is predicted by the spread between the $n$-year forward rate and the one-year yield. Campbell and Shiller (1991) find similar results using yield spreads as predictors for yield changes. Fama and Bliss (1987), and Campbell and Shiller (1991), uncover maturity-specific predictors. Cochrane and Piazzesi (2005) show that a single common factor obtained as a linear combination of forward spreads predicts the one-year excess return on the $n$-year bond: they thus document that a common factor predicts the excess returns of all bonds. Ludvigson and Ng (2009) find that common factors extracted from a large set of macroeconomic series have predictive power for the one-year excess return on the $n$-year bond beyond the Cochrane and Piazzesi (2005) single factor.

We take Ludvigson and Ng (2009) as a starting point and depart from them in two ways: we extract the latent factors from a large number of macroeconomic variables under the maintained assumption of structural instability in the loadings; we allow for a structural break in the factor augmented pricing equation for bond excess returns. Existing empirical evidence suggests that the degree of cross-sectional dependence among macroeconomic series is not stable over time (Breitung and Eickmeier (2011), Chen *et al.* (2014), and Cheng *et al.* (2016)): this motivates the inclusion of a structural break in the factor model. Allowing for a potentially unstable factor augmented regression is also desirable: bond risk premia are highly persistent and require long time series of data; the longer the data span, the higher the likelihood of model instability. Bikbov and Chernov (2010) conduct their analysis over $1970-2002$: they raise concerns about structural stability due to the $1979-1982$ monetary policy experiment. In studying $1960-2006$, Smith and Taylor (2009) acknowledge the presence of a break in the early 1980s and *a priori* split their sample around that period. Joslin *et al.* (2014) focus on $1985-2007$: they estimate a two-state Markov-switching affine term structure model over $1971:11-2007:12$ to show that the period $1985 - 2007$ is well approximated by a single regime. We complement Joslin *et al.* (2014) by applying to bond risk premia the unstable diffusion index forecasting model proposed in this paper: Markov-switching models assumes that "history repeats" (Timmermann (2008)); our unstable model does not. Gürkaynak and Wright (2012) discuss the consequences of structural instability on the learning process of investors.

### 6.2 Data and Implementation

We study the period $1965:01 - 2007:12$, which extends the samples studied in Fama and Bliss (1987), Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009): the endpoint of the sample coincides with the end of the Great Moderation and it is the same as in Joslin *et al.* (2014). Our chosen period allows us to avoid the zero lower bound regime started in December 2008, during which the Federal Reserve has relied on unconventional monetary policy tools: an assessment of the impact of these tools on the yield curve has been a challenging task (Wu and Xia (2016), and references therein). Monthly bond return data are from the Fama-Bliss dataset available from the Center for Research in Securities Prices (CRSP): the dataset has prices for zero-coupon U.S. Treasury bonds for maturities between one to five years; yields, forward rates and returns are constructed as detailed in Cochrane and Piazzesi (2005).

We estimate the factors from the FRED-MD balanced panel of monthly macroeconomic series described in McCracken and Ng (2016)[1]. The data belong to eight groups: (*i*) output and income; (*ii*) labor market; (*iii*) consumption and housing; (*iv*) orders and inventories; (*v*) money and credit; (*vi*) interest rate and exchange rates; (*vii*) prices; (*viii*) stock market. We use the $2016:06$ vintage. We apply to the original dataset described in McCracken and Ng (2016) the changes outlined in the document "Changes to FRED-MD"[2]. We make the resulting 123 series stationary by applying the transformations provided in McCracken and Ng (2016)[3].

[1] The dataset is available at https://research.stlouisfed.org/econ/mccracken/fred-databases/ .
[2] The document is available at https://research.stlouisfed.org/econ/mccracken/fred-databases/ .
[3] A file containing all macroeconomic series used to estimate the factors is available upon request.

We formally implement the factor augmented pricing equation

$$rx_{t+12}^{(n)} = \mathbb{I}\left(t/T \leq \pi_y^{(n)}\right) \boldsymbol{\gamma}_1^{(n)\prime} \mathbf{f}_t + \mathbb{I}\left(t/T > \pi_y^{(n)}\right) \boldsymbol{\gamma}_2^{(n)\prime} \mathbf{f}_t + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)}, \qquad (11)$$

for $t = 1964 : 01, \ldots, 2006 : 12$ and $n = 2, 3, 4, 5$: the superscript $(n)$ denotes dependence on the maturity of the bond. The variable $rx_{t+12}^{(n)}$ is the excess log return for the $n$-year discount bond defined as in Cochrane and Piazzesi (2005). The vector $\mathbf{f}_t$ contains $R$ latent factors: based on the Monte Carlo results in Section 5.3, we obtain the estimate $\hat{R}$ for $R^0$ using the criterion $IC_{p2}(R, R)$ in (6). The observable covariate $CP_t$ is the single forward Cochrane and Piazzesi (2005) factor (i.e., a linear combination of forward rates), estimated over the period $1964 : 01 - 2006 : 12$. We estimate the true break fraction $\pi_y^{(n),0}$, $R^0 \times 1$ vectors $\boldsymbol{\gamma}_1^{(n),0}$ and $\boldsymbol{\gamma}_2^{(n),0}$, intercept $\beta_1^{(n),0}$ and slope coefficient $\beta_2^{(n),0}$ as detailed in Section 4: confidence intervals and stability testing are performed using the Newey and West (1987) correction with 18 lags to account for the MA(12) structure due to the overlapping data (Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009)).

We assume that $\beta_1^{(n),0}$ and $\beta_2^{(n),0}$ are not subject to structural instability: the information content of forward rates is stable over time (Cochrane and Piazzesi (2005)). The $CP_t$ factor is a linear combination of forward rates, and it is obtained by assuming that returns at all maturities are linear functions of forward rates: this implies that $\beta_2^{(n),0}$ is stable over time. If we were to allow for a break in $\beta_2^{(n),0}$, we would have to extend the procedure to estimate the Cochrane and Piazzesi (2005) factor accordingly: this goes beyond the purpose of the paper. The stability of $\beta_2^{(n),0}$ across all maturities is supported by the results in Section 6.4 below. The null hypothesis of stability then is $\mathbb{H}_0 : \left(\boldsymbol{\gamma}_1^{(n),0} = \boldsymbol{\gamma}_2^{(n),0}\right)$. Under the alternative hypothesis $\mathbb{H}_1 : \left(\boldsymbol{\gamma}_1^{(n),0} \neq \boldsymbol{\gamma}_2^{(n),0}\right)$, instabilities arise in the linkages between bond risk premia and macroeconomic factors.

Table 4 about here

Summary statistics for excess log returns are shown in Table 4 for levels and absolute values (Panels A and B, respectively). Both mean and standard deviation of bond excess returns increase with maturity: the former ranges between 0.4% and 0.9%, the latter between 1.9% and 5.8%. Bond risk premia are highly persistent: the first order autocorrelation is above 0.9 over all maturities, which is consistent with the MA(12) structure due to the overlapping data previously discussed; the autocorrelation of order 12 is positive and monotonically declines with maturity, ranging between 0.215 for $n = 2$ and 0.071 for $n = 5$. Absolute values of bond excess returns proxy bond risk premia volatilities: they are highly persistent, with mean and standard deviation increasing with maturity.

### 6.3 Breakpoint Factor Model: Macro Factors

We set $R^{\max} = 8$; following the Monte Carlo results in Section 5.3, we apply the $IC_{p2}(R, R)$ criterion in (6) and estimate $\hat{R} = 2$ factors. The estimated break fraction $\hat{\pi}_{\mathbf{x}} = 0.455$ corresponds to the estimated break date $\hat{T}_{\mathbf{x}} = 1983 : 06$: this is just before the beginning of the Great Moderation in early 1985. Our estimated number of factors is in line with what suggested in studies using datasets similar to ours: Stock and Watson (2012) *a priori* select five factors without accounting for structural breaks, which augment the factor space (Breitung and Eickmeier (2011), and Chen *et al.* (2014)); Cheng *et al.* (2016) select one factor during the period ranging from 1985 to 2007, which coincides with our post-break sample.

Table 5 about here

We interpret the estimated factors by studying their predictive power (Ludvigson and Ng (2009)). For each of the eight groups listed in Section 6.2, we set $\hat{\pi}_{\mathbf{x}} = 0.455$ (i.e., the estimate for $\pi^0_{\mathbf{x}}$) and compute the average of the $R^2$ obtained from the breakpoint regression of each variable within the group on a given factor. The results in Table 5 show that the first estimated factor $\hat{f}_{1t}$ captures the real side of the economy and inflation: the average $R^2$ from regressions of macroeconomic variables in Groups 1, 2, 3 and 7 are equal to 0.298, 0.166, 0.211 and 0.249, respectively. The second estimated factor $\hat{f}_{2t}$ loads on variables in Groups 6 and 8: it describes the financial side of the economy; it is also correlated with the $CP_t$ factor, with correlation coefficient equal to $-0.248$. The two factors exhibit time series dependence: the first order autocorrelation for $\hat{f}_{1t}$ and $\hat{f}_{2t}$ are equal to 0.485 and 0.630, respectively.

### 6.4 Unstable Factor Augmented Regression: Bond Risk Premia

Table 6 displays results from four specifications nested in (11): ($a$) the Cochrane and Piazzesi (2005) single factor model, with $\boldsymbol{\gamma}_1^{(n),0} = \boldsymbol{\gamma}_2^{(n),0} = \mathbf{0}$; ($b$) a linear model with macroeconomic factors only, with $\boldsymbol{\gamma}_1^{(n),0} = \boldsymbol{\gamma}_2^{(n),0}$ and $\beta_2^{(n),0} = 0$; ($c$) a factor model under the null hypothesis of linearity $\mathbb{H}_0 : \left(\boldsymbol{\gamma}_1^{(n),0} = \boldsymbol{\gamma}_2^{(n),0}\right)$; ($d$) the unstable factor augmented regression in (11).

Table 6 about here

The estimates for $\beta_2^{(n),0}$ from model ($a$) are 0.455, 0.857, 1.235 and 1.453 for $n = 2, 3, 4, 5$, respectively: these resemble the estimates in Cochrane and Piazzesi (2005) and do not contradict the assumption of stability imposed in (11) on $\beta_2^{(n),0}$. Based on the adjusted $R^2$, the $CP_t$ factor explains between 34% and 38% of the variation in bond excess returns: these values are in line with those in Cochrane and Piazzesi (2005). Model ($b$) shows that macroeconomic factors have less explanatory power than the $CP_t$ factor across all maturities. Judging from the values of the adjusted $R^2$ from the *linear* model ($c$), macroeconomic factors seem to have limited predictive power in excess of the $CP_t$ factor during the period of interest.

We estimate model ($d$) over the interval $\left[\underline{\pi}_y, 1 - \underline{\pi}_y\right]$, with $\underline{\pi}_y = 0.15$. The adjusted $R^2$ shows that allowing for structural instability provides a great deal of additional explanatory power with respect to the linear factor augmented regression ($c$). The $\sup \mathcal{LM}$ test rejects the null hypothesis of linearity at 10% level for $n = 2, 4, 5$: this provides statistical evidence in favor of breaks in the factor loadings in the pricing equation for bond excess returns. The estimated break fraction for $n = 2, 4$ is $\hat{\pi}_y^{(n),} = 0.510$, with corresponding estimated break date $\hat{T}_y^{(n),} = \left[T\hat{\pi}_y^{(n),}\right] = 1986 : 11$; for $n = 5$, we have $\hat{\pi}_y^{(n),} = 0.355$ and $\hat{T}_y^{(n),} = 1980 : 03$. When structural breaks are allowed for, macroeconomic factors become statistically significant at least on one side of the breakpoint: their predictive ability thus varies over time.

According to our results, the location of the break depends on the maturity of the bond. Our findings relate to Joslin *et al.* (2014), who show that, with the exception of the first year and three isolated months, the Great Moderation period between 1985 and 2007 is well approximated by a single regime; they are also in line with Smith and Taylor (2009), who acknowledge the presence of a break in the early 1980s due to a shift in the monetary policy rule. More generally, we contribute to the literature by shedding further light on the issue of structural instability in bond risk premia dynamics. Our findings run parallel to existing evidence of instabilities in stock returns prediction models (Rapach and Zhou (2013)).

In conclusion, the proposed unstable diffusion index forecasting model has both statistical and economic advantages with respect to existing competing models for bond risk premia: it generates higher adjusted $R^2$; it captures the effects of events such as the Great Moderation and shifts in monetary policy rules on bond excess returns.

### 6.5 Alternative Estimation Strategy

The vector $\left(\mathbf{x}_t', rx_{t+12}^{(n)}\right)'$ admits a common factor representation, for $n = 1, 2, 3, 4$, where $\mathbf{x}_t$ collects the macroeconomic variables. The common components could be estimated by projecting $\left(\mathbf{x}_t', rx_{t+12}^{(n)}\right)'$ on the estimated factors: see Section 3.3 in Forni *et al.* (2005). Only $\pi_{\mathbf{x}}^0$ and not $\pi_y^0$ would be identified though, as the contribution of $rx_{t+12}^{(n)}$ to the estimator of the common components would be negligible as $N \to \infty$. The formulation for the factor augmented regression

$$rx_{t+12}^{(n)} = \mathbb{I}\left(t/T \leq \pi_{\mathbf{x}}\right)\boldsymbol{\gamma}_1^{(n)\prime}\mathbf{f}_t + \mathbb{I}\left(t/T > \pi_{\mathbf{x}}\right)\boldsymbol{\gamma}_2^{(n)\prime}\mathbf{f}_t + \beta_1^{(n)} + \beta_2^{(n)}CP_t + \varepsilon_{t+12}^{(n)}$$

could be implemented, where $\pi_{\mathbf{x}}^0$ and $\mathbf{f}_t^0$ would be estimated from the factor model.

Table 7 about here

Table 7 collects results from this estimation strategy. For maturities $n = 2, 3, 4$, the results are aligned with the homologous findings in Table 6, whereas for $n = 5$ the differences are more pronounced. Each adjusted $R^2$ in Table 7 is lower than the homologous value in Table 6. We can conclude that the estimator proposed in this paper is likely to provide more accurate information about the pricing equation for bond risk premia.

# 7 Extensions and Modifications

We now consider possible extensions and modifications to the model in (1) and (2): Section 7.1 introduces the diffusion index threshold forecasting model; Section 7.2 discusses the model with multiple breaks; Section 7.3 proposes an alternative strategy to allow for parameter instability in the diffusion index forecasting model.

## 7.1 Diffusion Index Threshold Forecasting Model

The specification in (1) and (2) allows for structural instability in the factor model and in the forecasting equation, respectively. Most results in this paper hold for the following *diffusion index threshold forecasting model*

$$\mathbf{x}_t = \mathbb{I}\left(g_t \leq \varphi_{\mathbf{x}}\right)\boldsymbol{\Lambda}_1\mathbf{f}_t + \mathbb{I}\left(g_t > \varphi_{\mathbf{x}}\right)\boldsymbol{\Lambda}_2\mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \ldots, T,$$

$$y_{t+h} = \mathbb{I}\left(g_t \leq \varphi_y\right)\left(\boldsymbol{\gamma}_1'\mathbf{f}_t + \boldsymbol{\beta}_1'\mathbf{w}_t\right) + \mathbb{I}\left(g_t > \varphi_y\right)\left(\boldsymbol{\gamma}_2'\mathbf{f}_t + \boldsymbol{\beta}_2'\mathbf{w}_t\right) + \varepsilon_{t+h}, \quad t = 1, \ldots, T, \quad h \geq 0,$$

where $g_t \in \Re$ is the common observable threshold random variable, and $\varphi_{\mathbf{x}}$ and $\varphi_y$ are the unknown threshold values. The model lets $y_{t+h}$ depend upon $\mathbf{f}_t$, the vector of latent factors driving the comovement in $\mathbf{x}_t$ according to the threshold factor model of Massacci (2017). Let $\hat{\varphi}_{\mathbf{x}}$ and $\hat{\varphi}_y$ be the least squares estimators for $\varphi^0_{\mathbf{x}}$ and $\varphi^0_y$, respectively, namely for the true values of $\varphi_{\mathbf{x}}$ and $\varphi_y$, respectively. The break fractions $\pi^0_{\mathbf{x}}$ and $\pi^0_y$ in (1) and (2), respectively, are threshold values as applied to $t/T$. Under suitable assumptions, it can be shown that the results in Theorems 4.1, 4.2 and 4.3 hold for the diffusion index threshold forecasting model with respect to $\hat{\varphi}_{\mathbf{x}}$ and $\hat{\varphi}_y$. However, Theorem 4.4 no longer is valid: $g_t$ generally is a random variable; a test for threshold effect in the forecasting equation can be implemented by suitably generalizing the results in Hansen (1996).

### 7.2 Diffusion Index Forecasting Model with Multiple Breaks

The theoretical results in this paper accommodate one break in the factor model and one break in the factor augmented regression. A more empirically plausible scenario would be

$$\mathbf{x}_t = \sum_{j_{\mathbf{x}}=1}^{J^0_{\mathbf{x}}+1} \mathbb{I}\left(\pi_{\mathbf{x},j_{\mathbf{x}}-1} < t/T \leq \pi_{\mathbf{x},j_{\mathbf{x}}}\right) \mathbf{\Lambda}_{j_{\mathbf{x}}} \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \ldots, T,$$
$$y_{t+h} = \sum_{j_y=1}^{J^0_y+1} \mathbb{I}\left(\pi_{y,j_y-1} < t/T \leq \pi_{y,j_y}\right)\left(\boldsymbol{\gamma}'_{j_y}\mathbf{f}_t + \boldsymbol{\beta}'_{j_y}\mathbf{w}_t\right) + \varepsilon_{t+h}, \quad t = 1, \ldots, T, \quad h \geq 0,$$

where $\pi_{\mathbf{x},0} = \pi_{y,0} = 0$ and $\pi_{\mathbf{x},J_{\mathbf{x}}+1} = \pi_{y,J_y+1} = 1$: this allows for $J^0_{\mathbf{x}} \geq 1$ breaks in the factor model and $J^0_y \geq 1$ breaks in the factor augmented regression, where for ease of exposition we assume that $J^0_{\mathbf{x}}$ and $J^0_y$ are both known. To aid to the understanding of dealing with this more general model, consider first $J^0_{\mathbf{x}} = 1$ and $J^0_y = 2$, in which case (2) generalizes to

$$\begin{aligned} y_{t+h} &= \mathbb{I}\left(t/T \leq \pi_{y,1}\right)\left(\boldsymbol{\gamma}'_1\mathbf{f}_t + \boldsymbol{\beta}'_1\mathbf{w}_t\right) \\ &\quad + \mathbb{I}\left(\pi_{y,1} < t/T \leq \pi_{y,2}\right)\left(\boldsymbol{\gamma}'_2\mathbf{f}_t + \boldsymbol{\beta}'_2\mathbf{w}_t\right) \quad , \quad t = 1, \ldots, T, \quad h \geq 0. \\ &\quad + \mathbb{I}\left(t/T > \pi_{y,2}\right)\left(\boldsymbol{\gamma}'_3\mathbf{f}_t + \boldsymbol{\beta}'_3\mathbf{w}_t\right) + \varepsilon_{t+h} \end{aligned}$$

Let $\boldsymbol{\theta}_{j_y}^0 = \left(\boldsymbol{\gamma}_{j_y}^{0\prime}, \boldsymbol{\beta}_{j_y}^{0\prime}\right)'$ be the true value of $\boldsymbol{\theta}_{j_y} = \left(\boldsymbol{\gamma}_{j_y}', \boldsymbol{\beta}_{j_y}'\right)'$, for $j_y = 1, 2, 3$. Define the estimator $\hat{\boldsymbol{\theta}}_{j_y} = \left(\hat{\boldsymbol{\gamma}}_{j_y}', \hat{\boldsymbol{\beta}}_{j_y}'\right)'$ for $\boldsymbol{\theta}_{j_y}^0 = \left(\boldsymbol{\gamma}_{j_y}^{0\prime}, \boldsymbol{\beta}_{j_y}^{0\prime}\right)'$, for $j_y = 1, 2, 3$. Let $\pi_{y,1}^0$, $\pi_{y,2}^0$ and $\pi_{y,3}^0$ be the true values of $\pi_{y,1}$, $\pi_{y,2}$ and $\pi_{y,3}$, respectively. The results in Theorem 4.1 generalize as

$$\hat{\boldsymbol{\theta}}_1 - \left[\frac{\min\left\{\pi_{\mathbf{x},1}^0, \pi_{y,1}^0\right\}}{\pi_{y,1}^0} \hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x},1}^0\right) + \mathbb{I}\left(\pi_{\mathbf{x},1}^0 < \pi_{y,1}^0\right) \frac{\pi_{y,1}^0 - \pi_{\mathbf{x},1}^0}{\pi_{y,1}^0} \hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x},1}^0\right)\right]' \boldsymbol{\theta}_1^0 = o_p(1),$$

$$\hat{\boldsymbol{\theta}}_2 - \left[\begin{array}{c} \mathbb{I}\left(\pi_{y,1}^0 < \pi_{\mathbf{x},1}^0\right) \dfrac{\min\left\{\pi_{\mathbf{x},1}^0, \pi_{y,2}^0\right\} - \pi_{y,1}^0}{\pi_{y,2}^0 - \pi_{y,1}^0} \hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x},1}^0\right) \\ + \mathbb{I}\left(\pi_{\mathbf{x},1}^0 \leq \pi_{y,2}^0\right) \dfrac{\pi_{y,2}^0 - \max\left\{\pi_{\mathbf{x},1}^0, \pi_{y,1}^0\right\}}{\pi_{y,2}^0 - \pi_{y,1}^0} \hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x},1}^0\right) \end{array}\right]' \boldsymbol{\theta}_2^0 = o_p(1),$$

$$\hat{\boldsymbol{\theta}}_3 - \left[\frac{1 - \max\left\{\pi_{\mathbf{x},1}^0, \pi_{y,2}^0\right\}}{1 - \pi_{y,2}^0} \hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x},1}^0\right) + \mathbb{I}\left(\pi_{\mathbf{x},1}^0 > \pi_{y,2}^0\right) \frac{\pi_{\mathbf{x},1}^0 - \pi_{y,2}^0}{1 - \pi_{y,2}^0} \hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x},1}^0\right)\right]' \boldsymbol{\theta}_3^0 = o_p(1).$$

If we then allowed for $J_{\mathbf{x}}^0 = 2$ breaks, the factor model in (1) would become

$$\mathbf{x}_t = \mathbb{I}\left(t/T \leq \pi_{\mathbf{x},1}\right) \boldsymbol{\Lambda}_1 \mathbf{f}_t + \mathbb{I}\left(\pi_{\mathbf{x},1} < t/T \leq \pi_{\mathbf{x},2}\right) \boldsymbol{\Lambda}_2 \mathbf{f}_t + \mathbb{I}\left(t/T > \pi_{\mathbf{x},2}\right) \boldsymbol{\Lambda}_3 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \ldots, T:$$

in this case, the results regarding $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_2$ and $\hat{\boldsymbol{\theta}}_3$ stated above would have to be extended to account for the additional degree of rotational indeterminacy in the factor model due to the presence of the additional break. This strategy would allow us to generalize Theorems 4.1 and 4.3 regarding $\hat{\boldsymbol{\theta}}_{j_y}$, for $j_y = 1, \ldots, J_y^0$, for any $J_{\mathbf{x}}^0 \geq 1$ and $J_y^0 \geq 1$. In addition, in the case where the number of breaks $J_y^0$ no longer is known, Theorem 4.4 is not valid: the seminal contribution by Bai and Perron (1998) would be a natural starting point to conduct inference on the number of factors in the forecasting equation.

### 7.3 Alternative Specification of Parameter Instability

The model in (1) and (2) allows for different break fractions $\pi_{\mathbf{x}}$ and $\pi_y$ in the factor model and in the factor augmented regression, respectively, where $\pi_y$ is the same for latent factors $\mathbf{f}_t$ and observable covariates $\mathbf{w}_t$. Alternatively, one could assume a common break fraction $\pi^{\mathbf{f}}$ for $\mathbf{f}_t$ in the factor model and in the factor augmented regression, and a break fraction $\pi^{\mathbf{w}}$ for $\mathbf{w}_t$: formally, the model would be

$$\mathbf{x}_t = \mathbb{I}\left(t/T \leq \pi^{\mathbf{f}}\right)\boldsymbol{\Lambda}_1\mathbf{f}_t + \mathbb{I}\left(t/T > \pi^{\mathbf{f}}\right)\boldsymbol{\Lambda}_2\mathbf{f}_t + \mathbf{e}_t, \quad t = 1,\ldots,T,$$

$$\begin{aligned} y_{t+h} &= \mathbb{I}\left(t/T \leq \pi^{\mathbf{f}}\right)\boldsymbol{\gamma}_1'\mathbf{f}_t + \mathbb{I}\left(t/T > \pi^{\mathbf{f}}\right)\boldsymbol{\gamma}_2'\mathbf{f}_t \\ &\quad + \mathbb{I}\left(t/T \leq \pi^{\mathbf{w}}\right)\boldsymbol{\beta}_1'\mathbf{w}_t + \mathbb{I}\left(t/T > \pi^{\mathbf{w}}\right)\boldsymbol{\beta}_2'\mathbf{w}_t + \varepsilon_{t+h}, \end{aligned} \qquad t = 1,\ldots,T, \quad h \geq 0.$$

Let $\pi^{\mathbf{f},0}$ and $\pi^{\mathbf{w},0}$ denote the true values of $\pi^{\mathbf{f}}$ and $\pi^{\mathbf{w}}$, respectively. Under the maintained assumption of structural instability in the factor model, $\pi^{\mathbf{f},0}$ can be estimated from the factor model only, as the contribution to the underlying loss function coming from the factor augmented regression would be infinitesimal as $N \to \infty$. This has implications for estimation and inference in the factor augmented regression: only $\pi^{\mathbf{w},0}$ needs to be estimated; the critical values of the test statistic for the null hypothesis $\boldsymbol{\gamma}_1^0 = \boldsymbol{\gamma}_2^0$ are obtained from the $\chi^2\left(\hat{R}\right)$ distribution; the null hypothesis $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta}_2^0$ still needs to be tested using the procedure detailed in Andrews (1993).

More generally, one could also allow for different break fractions for $\mathbf{f}_t$ in the factor model and in the factor augmented regression: formally, the model would be

$$\mathbf{x}_t = \mathbb{I}\left(t/T \leq \pi^{\mathbf{f}}_{\mathbf{x}}\right)\boldsymbol{\Lambda}_1\mathbf{f}_t + \mathbb{I}\left(t/T > \pi^{\mathbf{f}}_{\mathbf{x}}\right)\boldsymbol{\Lambda}_2\mathbf{f}_t + \mathbf{e}_t, \quad t = 1,\ldots,T,$$

$$\begin{aligned} y_{t+h} &= \mathbb{I}\left(t/T \leq \pi^{\mathbf{f}}_y\right)\boldsymbol{\gamma}_1'\mathbf{f}_t + \mathbb{I}\left(t/T > \pi^{\mathbf{f}}_y\right)\boldsymbol{\gamma}_2'\mathbf{f}_t \\ &\quad + \mathbb{I}\left(t/T \leq \pi^{\mathbf{w}}_y\right)\boldsymbol{\beta}_1'\mathbf{w}_t + \mathbb{I}\left(t/T > \pi^{\mathbf{w}}_y\right)\boldsymbol{\beta}_2'\mathbf{w}_t + \varepsilon_{t+h}, \end{aligned} \qquad t = 1,\ldots,T, \quad h \geq 0.$$

The null hypotheses $\boldsymbol{\gamma}_1^0 = \boldsymbol{\gamma}_2^0$ and $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta}_2^0$ could be tested separately by using a procedure analogous to that described in Section 7.2: the former and the latter would require testing $\hat{R}$ and $K$ restrictions, respectively. A test for the null hypothesis $\boldsymbol{\theta}_1^0 = \boldsymbol{\theta}_2^0$ would involve the two nuisance parameters $\pi^{\mathbf{f}}_y$ and $\pi^{\mathbf{w}}_y$: we leave this interesting problem to future research.

# 8 Conclusions

We study estimation and inference in unstable diffusion index forecasting models: the application uncovers breaks in the linkages between bond risk premia and macroeconomic factors.

The novelty of the problem and the related technical difficulties make us analyze the single break case: the multiple break scenario is on our list of priorities (Bai and Perron (1998)).

The paper focuses on in-sample analysis. Diffusion indexes are widely used for out-of-sample forecasting. They may perform poorly in practice due to instabilities in the factor model (Giannone (2007), and Banerjee *et al.* (2008)) or in the factor augmented regression (Stock and Watson (2002, 2009)). Forecasting techniques robust to breaks would be desirable, especially if a break is located close to the end of the sample: one possible solution is to extend Pesaran and Timmermann (2007) procedure to select the estimation windows to include estimated factors.

# A Proofs of Theorems

## A.1 Proof of Theorem 3.1

**Proof of Theorem 3.1.** The proof is similar to that of Theorem 4.1 in Massacci (2017). ■

## A.2 Proofs of Theorems 4.1, 4.2, 4.3 and 4.4

**Lemma A.1** *For $j = 1, 2$, and under Assumptions I and C1-C4, as $N, T \to \infty$,*

$$\hat{\mathbf{f}}_t \xrightarrow{p} \left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right) + \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0.$$

**Lemma A.2** *Under Assumptions I, C1-C4 and CR, as $N, T \to \infty$:*

**(a)** $C_{NT}^2\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\hat{\mathbf{f}}_t - \left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right) + \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\right\}\varepsilon_{t+h}\right\} = O_p(1)$;

**(b)** $C_{NT}^2\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_j\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbf{f}_t^0 - \hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{f}}_t\right]\hat{\mathbf{z}}_t'\right\} = O_p(1)$, *for $j = 1, 2$.*

**Lemma A.3** *For $\pi_y \in [0, 1]$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\hat{\mathbf{z}}_t\hat{\varepsilon}_{t+h}^{\mathbb{H}_0} \Rightarrow \left\{\left(\pi_{\mathbf{x}}^0\right)^{1/2}\left[\mathbf{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1} + \left(1-\pi_{\mathbf{x}}^0\right)^{1/2}\left[\mathbf{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\right\}\left(\mathbf{\Sigma}_{\mathbf{z}\varepsilon}^0\right)^{1/2}\left[\mathcal{W}_{R^0+K}\left(\pi_y\right) - \pi_y\mathcal{W}_{R^0+K}(1)\right].$$

**Proof of Theorem 4.1.** Let $\hat{\varepsilon}_{t+h}$ and $\hat{h}_{t+h}$ be the estimated residual and the difference between true error and residual, respectively, namely

$$\hat{\varepsilon}_{t+h} = y_{t+h} - \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\left(\hat{\boldsymbol{\theta}}_1'\hat{\mathbf{z}}_t\right) - \mathbb{I}_{2t}\left(\hat{\pi}\right)\left(\hat{\boldsymbol{\theta}}_2'\hat{\mathbf{z}}_t\right), \quad \hat{h}_{t+h} = \varepsilon_{t+h} - \hat{\varepsilon}_{t+h}.$$

In order to prove the consistency of $\hat{\pi}_y$, we follow the same strategy used to prove Proposition 1 in Bai and Perron (1998). From the properties of projections,

$$T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{t+h}^2 \leq T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2; \tag{12}$$

from the definition of $\hat{h}_{t+h}$,

$$T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{t+h}^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2 + T^{-1}\sum_{t=1}^{T}\hat{h}_{t+h}^2 - 2T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}\hat{h}_{t+h}. \tag{13}$$

By Lemma A.1,

$$\begin{aligned}
\hat{\mathbf{z}}_t &= \begin{Bmatrix} \left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right) + \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0 + o_p(1) \\ \mathbf{w}_t \end{Bmatrix} \\
&= \mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0 + \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\left[\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0 + o_p(1):
\end{aligned}$$

we can then write

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}\hat{h}_{t+h} \\
= \; & T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2 - T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}\hat{\varepsilon}_{t+h} \\
= \; & T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2 - T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}\left[y_{t+h} - \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\left(\hat{\mathbf{z}}_t'\hat{\boldsymbol{\theta}}_1\right) - \mathbb{I}_{2t}\left(\hat{\pi}_y\right)\left(\hat{\mathbf{z}}_t'\hat{\boldsymbol{\theta}}_2\right)\right] \\
= \; & T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2 - T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}\left[\mathbb{I}_{1t}\left(\pi_y^0\right)\left(\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_1^0\right) + \mathbb{I}_{2t}\left(\pi_y^0\right)\left(\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_1^0\right) + \varepsilon_{t+h} - \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\left(\hat{\mathbf{z}}_t'\hat{\boldsymbol{\theta}}_1\right) - \mathbb{I}_{2t}\left(\hat{\pi}_y\right)\left(\hat{\mathbf{z}}_t'\hat{\boldsymbol{\theta}}_2\right)\right] \\
= \; & T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\hat{\pi}_y\right)\varepsilon_{t+h}\hat{\mathbf{z}}_t'\hat{\boldsymbol{\theta}}_1 + T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\hat{\pi}_y\right)\varepsilon_{t+h}\hat{\mathbf{z}}_t'\hat{\boldsymbol{\theta}}_2 - T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\varepsilon_{t+h}\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_1^0 - T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_y^0\right)\varepsilon_{t+h}\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_2^0 \\
= \; & T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\hat{\pi}_y\right)\varepsilon_{t+h}\left\{\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1} + \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\left[\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\right\}'\mathbf{z}_t^{0\prime}\hat{\boldsymbol{\theta}}_1 \\
& + T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\hat{\pi}_y\right)\varepsilon_{t+h}\left\{\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1} + \mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\left[\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\right\}'\mathbf{z}_t^{0\prime}\hat{\boldsymbol{\theta}}_2 \\
& - T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\varepsilon_{t+h}\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_1^0 - T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_y^0\right)\varepsilon_{t+h}\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_2^0 + o_p(1),
\end{aligned}$$

which implies

$$T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}\hat{h}_{t+h} = o_p(1) \tag{14}$$

by Assumption FR(b), since $\mathbb{I}_{lt}\left(\pi_{2y}\right) = \mathbb{I}_{jt}\left(\pi_{\mathbf{x}}\right)\mathbb{I}_{lt}\left(\pi_{1y}\right)$ for given $\pi_{\mathbf{x}}$ and suitably chosen $\pi_{1y}$ and $\pi_{2y}$, for $j, l = 1, 2$. Together with (12) and (13), (14) implies that $T^{-1}\sum_{t=1}^{T}\hat{h}_{t+h}^2 = o_p(1)$. We then proceed by showing that this result implies that $\hat{\pi}_y \xrightarrow{p} \pi_y^0$. In particular, suppose that $\hat{\pi}_y \not\xrightarrow{p} \pi_y^0$: it can then be easily shown that $T^{-1}\sum_{t=1}^{T}\hat{h}_{t+h}^2 \xrightarrow{p} D_y$, for some $D_y > 0$. Together with (13) and (14), this implies that

$$T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{t+h}^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2 + D_y + o_p(1) > T^{-1}\sum_{t=1}^{T}\varepsilon_{t+h}^2 + o_p(1),$$

which contradicts (12): it follows that $\hat{\pi}_y \xrightarrow{p} \pi_y^0$. Consider now

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_1 &= \left[\sum_{t=1}^{T}\mathbb{I}_{1t}(\hat{\pi}_y)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^{T}\mathbb{I}_{1t}(\hat{\pi}_y)\hat{\mathbf{z}}_t y_{t+h}\right] \\
&= \left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\hat{\pi}_y)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)' + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\hat{\pi}_y)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\left[\mathbb{I}_{1t}(\pi_y^0)\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_1^0 + \mathbb{I}_{2t}(\pi_y^0)\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_2^0 + \varepsilon_{t+h}\right]\right\} + o_p(1) \\
&= \left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)' + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\mathbb{I}_{1t}(\pi_y^0)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_1^0\right\} \\
&\quad+\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)' + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\mathbb{I}_{2t}(\pi_y^0)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\boldsymbol{\theta}_2^0\right\} \\
&\quad+\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)' + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\left[\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0) + \mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}\right\} + o_p(1) \\
&= \mathbb{I}(\pi_{\mathbf{x}}^0 \geq \pi_y^0)\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_y^0)\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\right]^{-1}\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0\right\} \\
&\quad+\mathbb{I}(\pi_{\mathbf{x}}^0 < \pi_y^0)\frac{\pi_{\mathbf{x}}^0}{\pi_y^0}\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{1t}(\pi_{\mathbf{x}}^0)\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\right]^{-1}\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0\right\} \\
&\quad+\mathbb{I}(\pi_{\mathbf{x}}^0 < \pi_y^0)\frac{\pi_y^0 - \pi_{\mathbf{x}}^0}{\pi_y^0}\left\{\sum_{t=1}^{T}\mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\mathbb{I}_{1t}(\pi_y^0)\left[\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\right]^{-1}\right\}^{-1} \\
&\quad\times\left\{\sum_{t=1}^{T}\mathbb{I}_{2t}(\pi_{\mathbf{x}}^0)\mathbb{I}_{1t}(\pi_y^0)\left[\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\right]^{-1}\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0\right\} + o_p(1):
\end{aligned}$$

it follows that

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_1 &= \mathbb{I}(\pi_{\mathbf{x}}^0 \geq \pi_y^0)\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0 + \mathbb{I}(\pi_{\mathbf{x}}^0 < \pi_y^0)\frac{\pi_{\mathbf{x}}^0}{\pi_y^0}\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0 + \mathbb{I}(\pi_{\mathbf{x}}^0 < \pi_y^0)\frac{\pi_y^0 - \pi_{\mathbf{x}}^0}{\pi_y^0}\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0 + o_p(1) \\
&= \frac{\min\{\pi_{\mathbf{x}}^0, \pi_y^0\}}{\pi_y^0}\hat{\boldsymbol{\Phi}}_{11}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0 + \mathbb{I}(\pi_{\mathbf{x}}^0 < \pi_y^0)\frac{\pi_y^0 - \pi_{\mathbf{x}}^0}{\pi_y^0}\hat{\boldsymbol{\Phi}}_{22}(\pi_{\mathbf{x}}^0)'\boldsymbol{\theta}_1^0 + o_p(1).
\end{aligned}$$

The result for $\hat{\boldsymbol{\theta}}_2$ may be proved in a similar way. This completes the proof of Theorem 4.1. ■

**Proof of Theorem 4.2.** The proof of Theorem 4.2 is similar to that of Theorem 3.4 in Massacci (2017). ■

**Proof of Theorem 4.3.** Adding and subtracting terms, the model may be written as

$$
\begin{aligned}
y_{t+h} &= \mathbb{I}_{1t}\left(\pi_y^0\right)\left(\gamma_1^{0\prime}\mathbf{f}_t^0+\boldsymbol{\beta}_1^{0\prime}\mathbf{w}_t\right)+\mathbb{I}_{2t}\left(\pi_y^0\right)\left(\gamma_2^{0\prime}\mathbf{f}_t^0+\boldsymbol{\beta}_2^{0\prime}\mathbf{w}_t\right)+\varepsilon_{t+h} \\
&= \mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\hat{\mathbf{f}}_t+\boldsymbol{\beta}_1^{0\prime}\mathbf{w}_t\right\} \\
&\quad +\mathbb{I}_{2t}\left(\pi_y^0\right)\left\{\gamma_2^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\hat{\mathbf{f}}_t+\boldsymbol{\beta}_2^{0\prime}\mathbf{w}_t\right\}+\varepsilon_{t+h} \\
&\quad +\mathbb{I}_{1t}\left(\pi_y^0\right)\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\} \\
&\quad +\mathbb{I}_{2t}\left(\pi_y^0\right)\gamma_2^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}. \\
&= \mathbb{I}_{1t}\left(\pi_y^0\right)\boldsymbol{\theta}_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\hat{\mathbf{z}}_t \\
&\quad +\mathbb{I}_{2t}\left(\pi_y^0\right)\boldsymbol{\theta}_2^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\hat{\mathbf{z}}_t+\varepsilon_{t+h} \\
&\quad +\mathbb{I}_{1t}\left(\pi_y^0\right)\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\} \\
&\quad +\mathbb{I}_{2t}\left(\pi_y^0\right)\gamma_2^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}
\end{aligned}
$$

The least squares estimator for $\boldsymbol{\theta}_1^0$ then is

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_1 &= \left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t y_{t+h}\right] \\
&= \left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t'\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]'\boldsymbol{\theta}_1^0\right\} \\
&\quad +\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{2t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t'\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]'\boldsymbol{\theta}_2^0\right\} \\
&\quad +\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}\right] \\
&\quad +\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{1t}\left(\pi_y^0\right)\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\} \\
&\quad +\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{2t}\left(\pi_y^0\right)\gamma_2^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& T^{1/2}\left[\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0\right] \\
= \; & T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t'\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]'\boldsymbol{\theta}_1^0\right\} \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}\right] \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{1t}\left(\pi_y^0\right)\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\}, \\
& -T^{1/2}\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0 \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{2t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t'\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]'\boldsymbol{\theta}_2^0\right\} \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\hat{\pi}_y\right)\hat{\mathbf{z}}_t\mathbb{I}_{2t}\left(\pi_y^0\right)\gamma_2^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\}
\end{aligned}
$$

and by Theorem 4.2

$$
\begin{aligned}
& T^{1/2}\left[\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0\right] \\
= \; & T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]'\boldsymbol{\theta}_1^0\right\} \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}\right] \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\}, \\
& -T^{1/2}\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0+o_p(1) \\
= \; & T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}\right] \\
& +T^{1/2}\left[\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\mathbb{I}_{1t}\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\}+o_p(1)
\end{aligned}
$$

so that

$$
\begin{aligned}
& T^{1/2}\left[\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0\right] \\
= \; & \left[T^{-1}\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[T^{-1/2}\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}\right] \\
& +\left[T^{-1}\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{T^{-1/2}\sum_{t=1}^T\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\}\right\}. \\
& +o_p(1)
\end{aligned}
\tag{15}
$$

Consider

$$
T^{-1/2}\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}=T^{-1/2}\sum_{t=1}^T \mathbb{I}_{1t}\left(\pi_y^0\right)\left(\hat{\mathbf{f}}_t',\mathbf{w}_t'\right)'\varepsilon_{t+h}
$$

and notice that

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{f}}_t\varepsilon_{t+h} &= T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\varepsilon_{t+h} \\
&\quad +T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\hat{\mathbf{f}}_t-\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\right\}\varepsilon_{t+h} \\
&= T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\varepsilon_{t+h}+o_p(1)
\end{aligned}$$

by Lemma A.2(a) since

$$\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\hat{\mathbf{f}}_t-\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\right\}\varepsilon_{t+h} \\
= \; & T^{1/2}\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\hat{\mathbf{f}}_t-\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\right\}\varepsilon_{t+h} \\
= \; & O_p\left[\frac{T^{1/2}}{\min(N,T)}\right]=o_p(1)
\end{aligned}$$

when $\sqrt{T}/N\to 0$: it follows that

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\varepsilon_{t+h} &= T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\right\}',\mathbf{w}_t'\right\}'\varepsilon_{t+h}+o_p(1) \\
&= T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}+o_p(1).
\end{aligned} \tag{16}$$

When $\sqrt{T}/N\to 0$ we have

$$\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\} \\
= \; & T^{1/2}\frac{1}{T}\sum_{t=1}^{T}\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\gamma_1^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right] \\ \times\left\{\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0-\hat{\mathbf{f}}_t\right\}\end{array}\right\} \\
= \; & \sum_{j=1}^{2}T^{-1/2}\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\gamma_1^{0\prime}\left[\mathbf{f}_t^0-\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{f}}_t\right] \\
= \; & O_p\left[\frac{T^{1/2}}{\min(N,T)}\right]=o_p(1),
\end{aligned} \tag{17}$$

since Lemma A.2(b) implies that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left\{\hat{\mathbf{z}}_t\left[\gamma_1^{0\prime}\mathbf{f}_t^0-\gamma_1^{0\prime}\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{f}}_t\right]\right\}'=\gamma_1^{0\prime}\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}_{jt}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbf{f}_t^0-\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{f}}_t\right]\hat{\mathbf{z}}_t'\right\}=C_{NT}^{-2}.$$

From (15), (16) and (17) we thus have

$$
\begin{aligned}
& T^{1/2}\left[\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0\right] \\
= & \left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\varepsilon_{t+h}\right]+o_p(1) \\
= & \left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right]^{-1}\left\{T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}\right\}+o_p(1) \\
= & \left\{T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)'+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)'\right]^{-1}\right\}^{-1} \\
& \times\left\{T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}\right\}+o_p(1) \\
= & \mathbb{I}\left(\pi_{\mathbf{x}}^0\geq\pi_y^0\right)\left\{T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)'\right]^{-1}\right\}^{-1} \\
& \times\left\{T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}\right\} \\
& +\mathbb{I}\left(\pi_{\mathbf{x}}^0<\pi_y^0\right)\frac{\pi_{\mathbf{x}}^0}{\pi_y^0}\left\{T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)'\right]^{-1}\right\}^{-1} \\
& \times\left\{T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}\right\} \\
& +\mathbb{I}\left(\pi_{\mathbf{x}}^0<\pi_y^0\right)\frac{\pi_y^0-\pi_{\mathbf{x}}^0}{\pi_y^0}\left\{T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\left[\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)'\right]^{-1}\right\}^{-1} \\
& \times\left\{T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\left[\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}\right\} \\
& +o_p(1) \\
= & \frac{\min\left\{\pi_{\mathbf{x}}^0,\pi_y^0\right\}}{\pi_y^0}\hat{\boldsymbol{\Phi}}_{11}\left(\pi_{\mathbf{x}}^0\right)'\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\varepsilon_{t+h}\right] \\
& +\mathbb{I}\left(\pi_{\mathbf{x}}^0<\pi_y^0\right)\frac{\pi_y^0-\pi_{\mathbf{x}}^0}{\pi_y^0}\hat{\boldsymbol{\Phi}}_{22}\left(\pi_{\mathbf{x}}^0\right)'\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\varepsilon_{t+h}\right]+o_p(1).
\end{aligned}
$$

Assumptions FR(a) and FR(e) imply that

$$T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\xrightarrow{p}\min\left\{\pi_{\mathbf{x}}^0,\pi_y^0\right\}\boldsymbol{\Sigma}_{\mathbf{z}}^0$$

and

$$T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\varepsilon_{t+h}\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\min\left\{\pi_{\mathbf{x}}^0,\pi_y^0\right\}\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right),$$

respectively; for $\pi_{\mathbf{x}}^0<\pi_y^0$ they also imply that

$$T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\mathbf{z}_t^{0\prime}\xrightarrow{p}\left(\pi_y^0-\pi_{\mathbf{x}}^0\right)\boldsymbol{\Sigma}_{\mathbf{z}}^0$$

and

$$T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbb{I}_{1t}\left(\pi_y^0\right)\mathbf{z}_t^0\varepsilon_{t+h}\xrightarrow{d}\mathcal{N}\left[\mathbf{0},\left(\pi_y^0-\pi_{\mathbf{x}}^0\right)\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right].$$

It therefore follows that

$$T^{1/2}\left[\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\Phi}}_{12}\left(\pi_{\mathbf{x}}^0,\pi_y^0\right)'\boldsymbol{\theta}_1^0\right]\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_1}^0\right).$$

The asymptotic distribution of $\hat{\boldsymbol{\theta}}_2$ may be derived in a similar way. The remaining part of the proof easily follows:

this completes the proof of Theorem 4.3. ■

**Proof of Theorem 4.4.** Since

$$\hat{\mathbf{z}}_t\hat{\varepsilon}_{t+h}^{\mathbb{H}_0} \xrightarrow{p} \left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{z}_t^0\varepsilon_{t+h}$$

then $\hat{\mathbf{\Omega}}^{\mathbb{H}_0} \xrightarrow{p} \mathbf{\Omega}^{0,\mathbb{H}_0}$ since $\hat{\mathbf{\Omega}}^{\mathbb{H}_0}$ is a Newey and West (1987) estimator for $\mathbf{\Omega}^{0,\mathbb{H}_0}$. The proof then follows from Lemma A.3: this completes the proof of Theorem 4.4. ■

**Proof of Lemma A.1.** Let $C_{NT} = \min\left\{\sqrt{N},\sqrt{T}\right\}$. Recall the $R^0 \times R^0$ matrix $\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}\right)$ defined in (7). Following similar steps as those in the proofs of Theorems 3.2 and 3.3 in Massacci (2017), it can be proved that

$$C_{NT}^2\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{ji}\left(\pi_{\mathbf{x}}^0\right)-\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}^0\right)'\boldsymbol{\lambda}_{ji}^0\right\|^2\right]=O_p\left(1\right),\quad j=1,2,$$

and $\hat{\pi}_{\mathbf{x}} \xrightarrow{p} \pi_{\mathbf{x}}^0$ as $N,T\rightarrow\infty$, respectively. Lemma A.1 follows by noting that rank $\left[\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}\right)\right]=R^0$ for all $\pi_{\mathbf{x}}$. ■

**Proof of Lemma A.2.** Following similar steps as those in the proof of Theorem 3.4 in Massacci (2017),

$$N^{\alpha^0}T\left(\hat{\pi}_{\mathbf{x}}-\pi_{\mathbf{x}}^0\right)=O_p\left(1\right)$$

and

$$C_{NT}^2\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{ji}\left(\hat{\pi}_{\mathbf{x}}\right)-\hat{\mathbf{H}}_{jj}\left(\pi_{\mathbf{x}}^0\right)'\boldsymbol{\lambda}_{ji}^0\right\|^2\right]=O_p\left(1\right),\quad j=1,2.$$

As a direct consequence,

$$C_{NT}^2\left[\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\mathbf{f}}_t-\left[\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{11}\left(\pi_{\mathbf{x}}^0\right)+\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\hat{\mathbf{H}}_{22}\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\mathbf{f}_t^0\right\|^2\right]=O_p\left(1\right).$$

The proofs of Lemmas A.2(a) and A.2(b) are then similar to the proofs of Lemmas A.1.(iv) and A.1.(iii) in Bai and Ng (2006), respectively. ■

**Proof of Lemma A.3.** Define $\hat{\mathbf{Z}}=\left(\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_T\right)'$, $\mathbf{M}_{\hat{\mathbf{Z}}}=\mathbf{I}_T-\hat{\mathbf{Z}}\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'$, $\boldsymbol{\varepsilon}_h=\left(\varepsilon_{1+h},\ldots,\varepsilon_{T+h}\right)'$, $\hat{\boldsymbol{\varepsilon}}_h^{\mathbb{H}_0}=\left(\hat{\varepsilon}_{1+h}^{\mathbb{H}_0},\ldots,\hat{\varepsilon}_{T+h}^{\mathbb{H}_0}\right)'$ and $\mathbb{I}_1\left(\pi_y\right)=\text{diag}\left\{\left[\mathbb{I}_{11}\left(\pi_y\right),\ldots,\mathbb{I}_{1T}\left(\pi_y\right)\right]'\right\}$. Assumptions FR(a) and FR(e), and Lemma

A.2(a) imply that, under the null hypothesis $\mathbb{H}_0 : \left(\boldsymbol{\theta}_1^0 = \boldsymbol{\theta}_2^0\right)$,

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\hat{\mathbf{z}}_t\hat{\varepsilon}_{t+h}^{\mathbb{H}_0} &= T^{-1/2}\hat{\mathbf{Z}}'\mathbb{I}_1\left(\pi_y\right)\hat{\boldsymbol{\varepsilon}}_h^{\mathbb{H}_0} \\
&= T^{-1/2}\hat{\mathbf{Z}}'\mathbb{I}_1\left(\pi_y\right)\mathbf{M}_{\hat{\mathbf{Z}}}\boldsymbol{\varepsilon}_h \\
&= T^{-1/2}\hat{\mathbf{Z}}'\mathbb{I}_1\left(\pi_y\right)\left[\mathbf{I}_T - \hat{\mathbf{Z}}\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\right]\boldsymbol{\varepsilon}_h \\
&= T^{-1/2}\hat{\mathbf{Z}}'\mathbb{I}_1\left(\pi_y\right)\boldsymbol{\varepsilon}_h - T^{-1/2}\hat{\mathbf{Z}}'\mathbb{I}_1\left(\pi_y\right)\hat{\mathbf{Z}}\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\boldsymbol{\varepsilon}_h \\
&= T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\hat{\mathbf{z}}_t\varepsilon_{t+h} - \pi_y T^{-1/2}\sum_{t=1}^{T}\hat{\mathbf{z}}_t\varepsilon_{t+h} + o_p(1) \\
&= \left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{z}_t^0\varepsilon_{t+h} + \left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_y\right)\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{z}_t^0\varepsilon_{t+h} \\
&\quad -\pi_y\left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{z}_t^0\varepsilon_{t+h} - \pi_y\left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\pi_{\mathbf{x}}^0\right)\mathbf{z}_t^0\varepsilon_{t+h} + o_p(1) \\
&\Rightarrow \left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\left(\pi_{\mathbf{x}}^0\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right)^{1/2}\mathcal{W}_{R^0+K}\left(\pi_y\right) + \left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\left[\left(1-\pi_{\mathbf{x}}^0\right)\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right]^{1/2}\mathcal{W}_{R^0+K}\left(\pi_y\right) \\
&\quad -\pi_y\left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\left(\pi_{\mathbf{x}}^0\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right)^{1/2}\mathcal{W}_{R^0+K}(1) - \pi_y\left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\left[\left(1-\pi_{\mathbf{x}}^0\right)\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right]^{1/2}\mathcal{W}_{R^0+K}(1) \\
&= \left\{\left(\pi_{\mathbf{x}}^0\right)^{1/2}\left[\boldsymbol{\Phi}_{11}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1} + \left(1-\pi_{\mathbf{x}}^0\right)^{1/2}\left[\boldsymbol{\Phi}_{22}^0\left(\pi_{\mathbf{x}}^0\right)\right]^{-1}\right\}\left(\boldsymbol{\Sigma}_{\mathbf{z}\varepsilon}^0\right)^{1/2}\left[\mathcal{W}_{R^0+K}\left(\pi_y\right) - \pi_y\mathcal{W}_{R^0+K}(1)\right],
\end{aligned}
$$

which completes the proof of Lemma A.3. ■

# References


[1] Ahn, S. C., and A. R. Horenstein (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81 (3), 1203 − 1227.

[2] Alessi, L., M. Barigozzi, and M. Capasso (2010), "Improved Penalization for Determining the Number of Factors in Approximate Factor Models," *Statistics and Probability Letters*, 80 (23 − 24), 1806 − 1813.

[3] Amengual, D., and M. W. Watson (2007), "Consistent Estimation of the Number of Dynamic Factors in a Large $N$ and $T$ Panel," *Journal of Business and Economic Statistics*, 25 (1), 91 − 96.

[4] Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61 (4), 821 − 856.

[5] Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics*, 79 (4), 551 – 563.

[6] Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71 (1), 135 – 171.

[7] Bai, J., X. Han, and Y. Shi (2017), "Estimation and Inference of Change Points in High-Dimensional Factor Models," Working Paper, Columbia University, City University of Hong Kong and Texas A&M University.

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

[9] Bai, J., and S. Ng (2006), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74 (4), 1133 – 1150.

[10] Bai, J., and P. Perron (1998), "Estimation and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66 (1), 47 – 78.

[11] Baltagi, B. H., C. Kao, and F. Wang (2016), "Estimating and Testing High Dimensional Factor Models with Multiple Structural Changes," Working Paper, Syracuse University.

[12] Baltagi, B. H., C. Kao, and F. Wang (2017), "Identification and Estimation of a Large Factor Model with Structural Instability," *Journal of Econometrics*, 197 (1), 87 – 100.

[13] Banerjee, A., M. Marcellino, and I. Masten (2008), "Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change," EUI Working Paper, ECO 2008/17.

[14] Barigozzi, M., and L. Trapani (2017), "Sequential Testing for Structural Stability in Approximate Factor Models," Working Paper, London School of Economics and Cass Business School.

[15] Barigozzi, M., H. Cho, and P. Fryzlewicz (2018), "Simultaneous Multiple Change-Point and Factor Analysis for High-Dimensional Time Series," *Journal of Econometrics*, 206 (1), 187 – 225.

[16] Bates, B. J., M. Plagborg-Møller, J. H. Stock, and M. W. Watson (2013), "Consistent Factor Estimation in Dynamic Factor Models with Structural Instability," *Journal of Econometrics*, 177 (2), 289 – 304.

[17] Bikbov, R., and M. Chernov (2010), "No-Arbitrage Macroeconomic Determinants of the Yield Curve," *Journal of Econometrics*, 159 (1), 166 – 182.

[18] Breitung, J., and S. Eickmeier (2011), "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163 (1), 71 – 84.

[19] Campbell, J. Y., and R. J. Shiller (1991), "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, 58 (3), 495 – 514.

[20] Caner, M., and X. Han (2014), "Selecting the Correct Number of Factors in Approximate Factor Models: The Large Panel Case With Group Bridge Estimators," *Journal of Business & Economic Statistics*, 32 (3), 359 – 374.

[21] Chamberlain, G., and M. Rothschild (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51 (5), 1281 – 1304.

[22] Chen, L. (2015), "Estimating the Common Break Date in Large Factor Models," *Economics Letters*, 131, 70 – 74.

[23] Chen, L., J. J. Dolado, and J. Gonzalo (2014), "Detecting Big Structural Breaks in Large Factor Models," *Journal of Econometrics*, 180 (1), 30 – 48.

[24] Cheng, X., Z. Liao, and F. Schorfheide (2016), "Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities," *Review of Economic Studies*, 83 (4), 1511 – 1543.

[25] Cochrane, J. H., and M. Piazzesi (2005), "Bond Risk Premia," *American Economic Review*, 95 (1), 138 − 160.

[26] Connor, G., and R. A. Korajczyk (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15 (3), 373 − 394.

[27] Connor, G., and R. A. Korajczyk (1988), "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21 (2), 255 − 289.

[28] Connor, G., and R. A. Korajczyk (1993), "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48 (4), 1263 − 1288.

[29] Corradi, V., and N. R. Swanson (2014), "Testing for Structural Stability of Factor Augmented Forecasting Models," *Journal of Econometrics*, 182 (1), 100 − 118.

[30] Fama, E. F., and R. R. Bliss (1987), "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77 (4), 680 − 692.

[31] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82 (4), 540 − 554.

[32] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004), "The Generalized Dynamic Factor Model: Consistency and Rates," *Journal of Econometrics*, 119 (2), 231 − 255.

[33] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100 (471), 830 − 840.

[34] Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015), "Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations," *Journal of Econometrics*, 185 (2), 359 − 371.

[35] Forni, M., and M. Lippi (2001), "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory*, 17 (6), 1113 – 1141.

[36] Giannone, D. (2007), "Discussion on: "Forecasting in Dynamic Factor Models Subject to Structural Instability" by J. H. Stock and M. W. Watson," in: *New Developments in Dynamic Factor Modelling*, Bank of England, Centre for Central Banking Studies.

[37] Gürkaynak, R. S., and J. H. Wright (2012), "Macroeconomics and the Term Structure," *Journal of Economic Literature*, 50 (2), 331 – 367.

[38] Hallin, M., and R. Liška (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102 (48), 603 – 617.

[39] Han, X., and A. Inoue (2015), "Tests for Parameter Instability in Dynamic Factor Models," *Econometric Theory*, 31 (5), 1117 – 1152.

[40] Hansen, B. E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica*, 64 (2), 413 – 430.

[41] Hansen, B. E. (2000), "Sample Splitting and Threshold Estimation", *Econometrica*, 68 (3), 575 – 603.

[42] Joslin, S., M. Priebsch, and K. J. Singleton (2014), "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, 69 (3), 1197 – 1233.

[43] Kapetanios, G. (2010), "A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets," *Journal of Business and Economic Statistics*, 28 (3), 397 – 409.

[44] Ludvigson, S. C., and S. Ng (2009), "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22 (12), 5027 – 5067.

[45] Ma, S., and L. Su (2016), "Estimation of Large Dimensional Factor Models with an Unknown Number of Breaks," Working Paper, University of California, Riverside and Singapore Management University.

[46] Massacci, D. (2017), "Least Squares Estimation of Large Dimensional Threshold Factor Models," *Journal of Econometrics*, 197 (1), 101 – 129.

[47] McCracken, M. W., and S. Ng (2016), "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business and Economics Statistics*, 34 (4), 574 – 589.

[48] Newey, W. K., and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autororrelation Consistent Covariance Matrix," *Econometrica*, 55 (3), 703 – 708.

[49] Onatski, A. (2009), "Testing Hypotheses About the Number of Factors in Large Factor Models," *Econometrica*, 77 (5), 1447 – 1479.

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

[51] Peligrad, Magda (1982), "Invariance Principles for Mixing Sequences of Random Variables," *Annals of Probability*, 10 (4), 968 – 980.

[52] Perron, P. (2006), "Dealing with Structural Breaks," in K. Patterson and T.C. Mills (eds.), *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, Palgrave Macmillan, 278 – 352.

[53] Pesaran, M. H., and A. Timmermann (2007), "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics*, 137 (1), 134 – 161.

[54] Rapach, D. E., and G. Zhou (2013), "Forecasting Stock Returns," in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting, Volume* 2*, Part A*, North-Holland, 329 – 383.

[55] Ross, S. (1976), "The Arbitrage Theory of Capital Pricing," *Journal of Economic Theory*, 13 (3), 341 – 360.

[56] Smith, J. M., and J. B. Taylor (2009), "The Term Structure of Policy Rules," *Journal of Monetary Economics*, 56 (7), 907 – 917.

[57] Stock, J. H., and M. W. Watson (1998), "Diffusion Indexes," NBER Working Paper 6702.

[58] Stock, J. H., and M. W. Watson (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97 (460), 1167 – 1179.

[59] Stock, J. H., and M. W. Watson (2009), "Forecasting in Dynamic Factor Models Subject to Structural Instability," in J. Castle and N. Shephard (eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, Oxford: Oxford University Press.

[60] Stock, J. H., and M. W. Watson (2012), "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business and Economics Statistics*, 30 (4), 481 – 493.

[61] Timmermann, A. (2008). Elusive Return Predictability. *International Journal of Forecasting*, 24 (1), 1 – 18.

[62] Wang, S., G. Cui, and K. Li (2015), "Factor-Augmented Regression Models with Structural Change," *Economics Letters*, 130, 124 – 127.

[63] Wu, C. J., and F. D. Xia (2016), "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound," *Journal of Money, Credit and Banking*, 48 (2 – 3), 253 – 291.

[64] Yamamoto, Y., and S. Tanaka (2015), "Testing for Factor Loading Structural Change under Common Breaks," *Journal of Econometrics*, 189 (1), 187 – 206.

**Table 1: Factor Model, Model Selection, $R^0 = 2$**

**Panel A:** $IC_{p1}(R, R)$

| | | $\alpha^0 = 0.60$ | | | | | | $\alpha^0 = 1.00$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta^0_{\mathbf{x}i} > 0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| | CSI | 2.7885 | 2.7975 | 3.1255 | 3.1040 | 2.0400 | 2.0405 | 2.7880 | 2.7835 | 3.1095 | 3.0995 | 2.0395 | 2.0395 |
| 100 | CSD | 2.8735 | 2.8960 | 3.9080 | 3.8700 | 2.0795 | 2.0825 | 2.8765 | 2.8840 | 3.8830 | 3.8695 | 2.0835 | 2.0830 |
| | CSDH | 2.8985 | 2.9185 | 3.9595 | 3.9200 | 2.2760 | 2.2665 | 2.8995 | 2.9075 | 3.9410 | 3.9165 | 2.2780 | 2.2745 |
| | CSI | 2.9060 | 2.9160 | 3.1070 | 3.0955 | 2.0275 | 2.0220 | 2.9015 | 2.9070 | 3.0870 | 3.0935 | 2.0270 | 2.0210 |
| 200 | CSD | 2.9480 | 2.9630 | 4.0420 | 3.9855 | 2.0555 | 2.0520 | 2.9450 | 2.9555 | 4.0095 | 3.9935 | 2.0540 | 2.0520 |
| | CSDH | 2.8830 | 2.9085 | 4.1060 | 4.0690 | 2.2025 | 2.2025 | 2.8855 | 2.8995 | 4.0750 | 4.0695 | 2.2015 | 2.2020 |
| | CSI | 2.9690 | 2.9765 | 3.0695 | 3.0590 | 2.0070 | 2.0070 | 2.9675 | 2.9730 | 3.0645 | 3.0530 | 2.0070 | 2.0070 |
| 400 | CSD | 2.9880 | 2.9950 | 4.3180 | 4.2740 | 2.0255 | 2.0220 | 2.9875 | 2.9915 | 4.3015 | 4.2855 | 2.0245 | 2.0235 |
| | CSDH | 2.9325 | 2.9520 | 4.3405 | 4.3095 | 2.1415 | 2.1360 | 2.9305 | 2.9455 | 4.3200 | 4.3155 | 2.1370 | 2.1355 |

**Panel B:** $IC_{p2}(R, R)$

| | | $\alpha^0 = 0.60$ | | | | | | $\alpha^0 = 1.00$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta^0_{\mathbf{x}i} > 0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| | CSI | 2.6050 | 2.6070 | 2.4035 | 2.3865 | 2.0010 | 2.0005 | 2.5955 | 2.6005 | 2.3895 | 2.3875 | 2.0005 | 2.0005 |
| 100 | CSD | 2.6930 | 2.7245 | 3.0970 | 3.0815 | 2.0020 | 2.0005 | 2.6950 | 2.7180 | 3.0795 | 3.0725 | 2.0020 | 2.0010 |
| | CSDH | 2.6695 | 2.7045 | 3.2285 | 3.2010 | 2.0425 | 2.0410 | 2.6695 | 2.6785 | 3.2150 | 3.2000 | 2.0440 | 2.0405 |
| | CSI | 2.8255 | 2.8405 | 2.6210 | 2.6215 | 2.0005 | 2.0005 | 2.8225 | 2.8315 | 2.6275 | 2.6035 | 2.0005 | 2.0005 |
| 200 | CSD | 2.8945 | 2.9200 | 3.4270 | 3.3990 | 2.0020 | 2.0020 | 2.8900 | 2.9115 | 3.4110 | 3.4010 | 2.0025 | 2.0025 |
| | CSDH | 2.8125 | 2.8415 | 3.6220 | 3.6020 | 2.0695 | 2.0685 | 2.8100 | 2.8305 | 3.6225 | 3.6040 | 2.0695 | 2.0680 |
| | CSI | 2.9455 | 2.9535 | 2.7880 | 2.7940 | 2.0010 | 2.0005 | 2.9440 | 2.9485 | 2.7850 | 2.7810 | 2.0005 | 2.0005 |
| 400 | CSD | 2.9790 | 2.9885 | 3.7295 | 3.7045 | 2.0045 | 2.0035 | 2.9760 | 2.9845 | 3.7500 | 3.7160 | 2.0035 | 2.0035 |
| | CSDH | 2.9110 | 2.9340 | 3.9730 | 3.9515 | 2.0715 | 2.0660 | 2.9065 | 2.9230 | 3.9740 | 3.9610 | 2.0690 | 2.0655 |

**Panel C:** $IC_{p3}(R, R)$

| | | $\alpha^0 = 0.60$ | | | | | | $\alpha^0 = 1.00$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta^0_{\mathbf{x}i} > 0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| | CSI | 3.1320 | 3.1475 | 6.0565 | 5.9915 | 7.9895 | 7.9850 | 3.1400 | 3.1490 | 6.0305 | 5.9755 | 7.9880 | 7.9820 |
| 100 | CSD | 3.4475 | 3.4720 | 6.2625 | 6.2315 | 7.9995 | 7.9995 | 3.4515 | 3.4785 | 6.2430 | 6.1925 | 7.9995 | 7.9990 |
| | CSDH | 3.6980 | 3.7055 | 6.3870 | 6.3295 | 8.0000 | 7.9995 | 3.7010 | 3.7215 | 6.3670 | 6.2895 | 8.0000 | 7.9995 |
| | CSI | 2.9870 | 2.9905 | 5.2225 | 5.1615 | 3.5835 | 3.5200 | 2.9860 | 2.9890 | 5.1790 | 5.1590 | 3.5650 | 3.5225 |
| 200 | CSD | 3.0115 | 3.0120 | 5.6805 | 5.6565 | 4.3795 | 4.2520 | 3.0120 | 3.0140 | 5.6730 | 5.6530 | 4.3535 | 4.2625 |
| | CSDH | 3.0600 | 3.0755 | 5.4505 | 5.4050 | 4.6495 | 4.5505 | 3.0620 | 3.0685 | 5.4225 | 5.3980 | 4.6390 | 4.5480 |
| | CSI | 2.9940 | 2.9950 | 4.6365 | 4.5900 | 2.5395 | 2.5285 | 2.9935 | 2.9950 | 4.6195 | 4.5915 | 2.5285 | 2.5225 |
| 400 | CSD | 2.9985 | 2.9995 | 5.6245 | 5.6290 | 2.7365 | 2.7295 | 2.9990 | 2.9980 | 5.6205 | 5.6190 | 2.7375 | 2.7255 |
| | CSDH | 2.9715 | 2.9795 | 5.2770 | 5.2465 | 2.8445 | 2.8280 | 2.9690 | 2.9765 | 5.2475 | 5.2440 | 2.8400 | 2.8360 |

This table presents results for the model selection criteria in (6). The DGP is detailed in Section 5.1. CSI denotes time homoskedastic factors and idiosyncratic components, and cross-sectionally independent idiosyncratic components. CSD denotes time homoskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components. CSDH denotes time heteroskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components.

**Table 2: Factor Augmented Regression, Estimation**

| Panel A: Homoskedastic and Independent Case, Bias, with $\alpha^0 = 0.60$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1^0 = 1$ | | | | | | $\pi_y^0 = 0.50$ | | | | | |
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta_{\mathbf{x}i}^0 > 0$ | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| $\delta_y^0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| 100 | CSI | -0.0196 | -0.0033 | -0.0234 | -0.0017 | -0.0321 | -0.0095 | 0.0025 | -0.0002 | 0.0044 | -0.0010 | 0.0025 | -0.0021 |
| | CSD | -0.0233 | -0.0041 | -0.0194 | -0.0017 | -0.0339 | -0.0086 | -0.0020 | 0.0000 | 0.0093 | -0.0011 | 0.0018 | -0.0016 |
| | CSDH | -0.0223 | -0.0047 | -0.0215 | -0.0021 | -0.0267 | -0.0074 | -0.0058 | 0.0001 | 0.0080 | -0.0013 | 0.0039 | -0.0003 |
| 200 | CSI | -0.0220 | 0.0014 | -0.0209 | 0.0023 | -0.0214 | -0.0021 | -0.0008 | 0.0000 | -0.0001 | 0.0001 | 0.0016 | -0.0003 |
| | CSD | -0.0216 | 0.0016 | -0.0201 | 0.0023 | -0.0211 | -0.0022 | 0.0004 | 0.0001 | 0.0016 | 0.0001 | 0.0026 | -0.0003 |
| | CSDH | -0.0200 | 0.0016 | -0.0199 | 0.0023 | -0.0205 | -0.0022 | -0.0001 | 0.0002 | 0.0038 | 0.0001 | 0.0017 | -0.0006 |
| 400 | CSI | -0.0169 | -0.0003 | -0.0178 | -0.0008 | -0.0199 | 0.0019 | 0.0000 | 0.0001 | -0.0068 | -0.0001 | -0.0049 | -0.0002 |
| | CSD | -0.0176 | -0.0001 | -0.0180 | -0.0008 | -0.0194 | 0.0019 | -0.0014 | 0.0002 | -0.0054 | 0.0000 | -0.0032 | -0.0002 |
| | CSDH | -0.0171 | -0.0002 | -0.0175 | -0.0008 | -0.0194 | 0.0018 | 0.0015 | 0.0001 | -0.0069 | 0.0000 | -0.0025 | -0.0003 |

| Panel B: Homoskedastic and Independent Case, RMSE, with $\alpha^0 = 0.60$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1^0 = 1$ | | | | | | $\pi_y^0 = 0.50$ | | | | | |
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta_{\mathbf{x}i}^0 > 0$ | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| $\delta_y^0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| 100 | CSI | 0.4097 | 0.1498 | 0.4415 | 0.1476 | 0.4330 | 0.1531 | 0.2698 | 0.0471 | 0.2770 | 0.0432 | 0.2734 | 0.0450 |
| | CSD | 0.4300 | 0.1510 | 0.4302 | 0.1477 | 0.4209 | 0.1537 | 0.2763 | 0.0487 | 0.2779 | 0.0431 | 0.2645 | 0.0454 |
| | CSDH | 0.4266 | 0.1538 | 0.4313 | 0.1472 | 0.4107 | 0.1563 | 0.2753 | 0.0549 | 0.2774 | 0.0450 | 0.2633 | 0.0504 |
| 200 | CSI | 0.2151 | 0.1043 | 0.2193 | 0.0991 | 0.2172 | 0.1026 | 0.2369 | 0.0192 | 0.2462 | 0.0180 | 0.2378 | 0.0191 |
| | CSD | 0.2213 | 0.1050 | 0.2167 | 0.0993 | 0.2155 | 0.1028 | 0.2393 | 0.0196 | 0.2446 | 0.0176 | 0.2341 | 0.0184 |
| | CSDH | 0.2283 | 0.1052 | 0.2197 | 0.0992 | 0.2190 | 0.1026 | 0.2395 | 0.0208 | 0.2468 | 0.0190 | 0.2346 | 0.0189 |
| 400 | CSI | 0.1160 | 0.0693 | 0.1185 | 0.0693 | 0.1237 | 0.0694 | 0.1845 | 0.0054 | 0.1818 | 0.0055 | 0.1891 | 0.0065 |
| | CSD | 0.1163 | 0.0698 | 0.1186 | 0.0695 | 0.1236 | 0.0695 | 0.1859 | 0.0060 | 0.1813 | 0.0057 | 0.1881 | 0.0066 |
| | CSDH | 0.1164 | 0.0697 | 0.1190 | 0.0694 | 0.1230 | 0.0694 | 0.1887 | 0.0063 | 0.1823 | 0.0060 | 0.1897 | 0.0071 |

| Panel C: Homoskedastic and Independent Case, Size 5% Level, with $\alpha^0 = 0.60$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_1^0 = 1$ | | | | | | $\pi_y^0 = 0.50$ | | | | | |
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta_{\mathbf{x}i}^0 > 0$ | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| $\delta_y^0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| 100 | CSI | 0.1675 | 0.0600 | 0.1965 | 0.0605 | 0.1780 | 0.0655 | - | - | - | - | - | - |
| | CSD | 0.1715 | 0.0605 | 0.1955 | 0.0600 | 0.1680 | 0.0675 | - | - | - | - | - | - |
| | CSDH | 0.1680 | 0.0620 | 0.1960 | 0.0620 | 0.1680 | 0.0700 | - | - | - | - | - | - |
| 200 | CSI | 0.1655 | 0.0560 | 0.1545 | 0.0455 | 0.1680 | 0.0605 | - | - | - | - | - | - |
| | CSD | 0.1700 | 0.0520 | 0.1545 | 0.0450 | 0.1630 | 0.0605 | - | - | - | - | - | - |
| | CSDH | 0.1620 | 0.0515 | 0.1555 | 0.0460 | 0.1645 | 0.0600 | - | - | - | - | - | - |
| 400 | CSI | 0.1150 | 0.0415 | 0.1105 | 0.0495 | 0.1205 | 0.0465 | - | - | - | - | - | - |
| | CSD | 0.1170 | 0.0425 | 0.1095 | 0.0505 | 0.1245 | 0.0465 | - | - | - | - | - | - |
| | CSDH | 0.1210 | 0.0410 | 0.1130 | 0.0490 | 0.1260 | 0.0475 | - | - | - | - | - | - |

This table presents results from estimation of and inference on the factor augmented regression. The DGP is detailed in Section 5.2. CSI, CSD and CSDH denote the same scenarios as in Table 1.

**Table 2-*Continued*: Factor Augmented Regression, Estimation**

**Panel D: Heteroskedastic and Dependent Case, Bias, with $\alpha^0 = 0.60$**

| | | $\beta_1^0 = 1$ | | | | | | $\pi_y^0 = 0.50$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta_{\mathbf{x}i}^0 > 0$ | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| $\delta_y^0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| 100 | CSI | -0.0180 | -0.0101 | -0.0081 | -0.0047 | -0.0110 | -0.0028 | 0.0038 | 0.0001 | -0.0040 | -0.0027 | 0.0071 | 0.0003 |
| | CSD | -0.0237 | -0.0101 | -0.0060 | -0.0043 | -0.0092 | -0.0016 | 0.0051 | 0.0014 | -0.0039 | -0.0020 | 0.0092 | 0.0007 |
| | CSDH | -0.0150 | -0.0113 | -0.0092 | -0.0041 | -0.0071 | -0.0029 | 0.0071 | 0.0002 | -0.0020 | -0.0029 | 0.0073 | -0.0003 |
| 200 | CSI | -0.0178 | -0.0009 | -0.0229 | -0.0074 | -0.0136 | -0.0017 | 0.0032 | 0.0000 | 0.0076 | -0.0010 | 0.0126 | -0.0003 |
| | CSD | -0.0170 | -0.0008 | -0.0233 | -0.0072 | -0.0127 | -0.0013 | 0.0021 | 0.0000 | 0.0079 | -0.0008 | 0.0133 | 0.0001 |
| | CSDH | -0.0176 | -0.0011 | -0.0199 | -0.0067 | -0.0161 | -0.0014 | 0.0005 | 0.0001 | 0.0079 | -0.0007 | 0.0146 | -0.0003 |
| 400 | CSI | -0.0093 | 0.0008 | -0.0131 | 0.0003 | -0.0112 | -0.0015 | 0.0008 | 0.0000 | 0.0027 | 0.0000 | 0.0042 | -0.0005 |
| | CSD | -0.0102 | 0.0008 | -0.0128 | 0.0003 | -0.0107 | -0.0015 | -0.0005 | 0.0000 | 0.0036 | 0.0000 | 0.0040 | -0.0005 |
| | CSDH | -0.0055 | 0.0010 | -0.0138 | 0.0005 | -0.0091 | -0.0020 | -0.0005 | 0.0000 | 0.0049 | 0.0001 | 0.0033 | -0.0006 |

**Panel E: Heteroskedastic and Dependent Case, RMSE, with $\alpha^0 = 0.60$**

| | | $\beta_1^0 = 1$ | | | | | | $\pi_y^0 = 0.50$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta_{\mathbf{x}i}^0 > 0$ | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| $\delta_y^0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| 100 | CSI | 0.3827 | 0.1710 | 0.4633 | 0.1712 | 0.4469 | 0.1598 | 0.2731 | 0.0625 | 0.2750 | 0.0573 | 0.2715 | 0.0562 |
| | CSD | 0.3971 | 0.1733 | 0.4610 | 0.1722 | 0.4318 | 0.1606 | 0.2750 | 0.0660 | 0.2746 | 0.0562 | 0.2641 | 0.0548 |
| | CSDH | 0.4085 | 0.1675 | 0.4388 | 0.1703 | 0.4406 | 0.1587 | 0.2782 | 0.0636 | 0.2750 | 0.0588 | 0.2761 | 0.0571 |
| 200 | CSI | 0.2426 | 0.1289 | 0.2262 | 0.1094 | 0.2235 | 0.1080 | 0.2484 | 0.0268 | 0.2497 | 0.0278 | 0.2464 | 0.0236 |
| | CSD | 0.2490 | 0.1305 | 0.2259 | 0.1095 | 0.2180 | 0.1082 | 0.2482 | 0.0282 | 0.2504 | 0.0269 | 0.2418 | 0.0226 |
| | CSDH | 0.2393 | 0.1258 | 0.2202 | 0.1067 | 0.2121 | 0.1065 | 0.2510 | 0.0278 | 0.2515 | 0.0254 | 0.2437 | 0.0241 |
| 400 | CSI | 0.1423 | 0.0755 | 0.1468 | 0.0741 | 0.1380 | 0.0731 | 0.2073 | 0.0082 | 0.2075 | 0.0085 | 0.2007 | 0.0096 |
| | CSD | 0.1428 | 0.0760 | 0.1478 | 0.0744 | 0.1385 | 0.0731 | 0.2077 | 0.0085 | 0.2098 | 0.0087 | 0.2010 | 0.0094 |
| | CSDH | 0.1422 | 0.0747 | 0.1429 | 0.0730 | 0.1351 | 0.0719 | 0.2119 | 0.0081 | 0.2096 | 0.0084 | 0.1999 | 0.0091 |

**Panel F: Heteroskedastic and Dependent Case, Size 5% Level, with $\alpha^0 = 0.60$**

| | | $\beta_1^0 = 1$ | | | | | | $\pi_y^0 = 0.50$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | 50 | | 100 | | 25 | | 50 | | 100 | |
| $\delta_{\mathbf{x}i}^0 > 0$ | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| $\delta_y^0$ | | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | | | | |
| 100 | CSI | 0.2855 | 0.1315 | 0.3080 | 0.1340 | 0.2865 | 0.1215 | - | - | - | - | - | - |
| | CSD | 0.2890 | 0.1310 | 0.3060 | 0.1345 | 0.2730 | 0.1190 | - | - | - | - | - | - |
| | CSDH | 0.2845 | 0.1320 | 0.3120 | 0.1365 | 0.2820 | 0.1215 | - | - | - | - | - | - |
| 200 | CSI | 0.2140 | 0.0780 | 0.2170 | 0.0855 | 0.2135 | 0.0855 | - | - | - | - | - | - |
| | CSD | 0.2200 | 0.0750 | 0.2135 | 0.0840 | 0.2080 | 0.0850 | - | - | - | - | - | - |
| | CSDH | 0.2085 | 0.0720 | 0.2200 | 0.0840 | 0.2065 | 0.0845 | - | - | - | - | - | - |
| 400 | CSI | 0.1595 | 0.0650 | 0.1670 | 0.0640 | 0.1710 | 0.0595 | - | - | - | - | - | - |
| | CSD | 0.1625 | 0.0685 | 0.1740 | 0.0630 | 0.1695 | 0.0590 | - | - | - | - | - | - |
| | CSDH | 0.1680 | 0.0665 | 0.1750 | 0.0650 | 0.1685 | 0.0585 | - | - | - | - | - | - |

**Table 3: Factor Augmented Regression, Lagrange Multiplier Test for Stability**

| **Panel A: Homoskedastic and Independent Case, with $\alpha^0 = 0.60$** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | | 50 | | | 100 | | |
| $\delta^0_{\mathbf{x}i} > 0$ | | 1.00 | | | 1.00 | | | 1.00 | | |
| | | **Size** | **Power** | | **Size** | **Power** | | **Size** | **Power** | |
| $\delta^0_y$ | | 0.00 | 0.25 | 1.00 | 0.00 | 0.25 | 1.00 | 0.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | |
| 100 | CSI | 0.0575 | 0.0890 | 0.9895 | 0.0225 | 0.0570 | 0.9910 | 0.0540 | 0.0845 | 0.9920 |
| | CSD | 0.0220 | 0.0525 | 0.9875 | 0.0225 | 0.0555 | 0.9905 | 0.1215 | 0.1530 | 0.9920 |
| | CSDH | 0.0300 | 0.0585 | 0.9825 | 0.0200 | 0.0570 | 0.9900 | 0.1300 | 0.1590 | 0.9900 |
| 200 | CSI | 0.0340 | 0.1345 | 1.0000 | 0.0290 | 0.1170 | 1.0000 | 0.0395 | 0.1415 | 1.0000 |
| | CSD | 0.0260 | 0.1255 | 1.0000 | 0.0290 | 0.1180 | 1.0000 | 0.0790 | 0.1740 | 1.0000 |
| | CSDH | 0.0280 | 0.1230 | 1.0000 | 0.0290 | 0.1190 | 1.0000 | 0.0915 | 0.1840 | 1.0000 |
| 400 | CSI | 0.0440 | 0.3415 | 1.0000 | 0.0450 | 0.3360 | 1.0000 | 0.0455 | 0.3495 | 1.0000 |
| | CSD | 0.0430 | 0.3405 | 1.0000 | 0.0475 | 0.3325 | 1.0000 | 0.0505 | 0.3540 | 1.0000 |
| | CSDH | 0.0395 | 0.3355 | 1.0000 | 0.0480 | 0.3260 | 1.0000 | 0.0580 | 0.3550 | 1.0000 |

| **Panel B: Heteroskedastic and Dependent Case, with $\alpha^0 = 0.60$** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | | 25 | | | 50 | | | 100 | | |
| $\delta^0_{\mathbf{x}i} > 0$ | | 1.00 | | | 1.00 | | | 1.00 | | |
| | | **Size** | **Power** | | **Size** | **Power** | | **Size** | **Power** | |
| $\delta^0_y$ | | 0.00 | 0.25 | 1.00 | 0.00 | 0.25 | 1.00 | 0.00 | 0.25 | 1.00 |
| $T$ | **DGP** | | | | | | | | | |
| 100 | CSI | 0.0215 | 0.0205 | 0.0755 | 0.0200 | 0.0215 | 0.0735 | 0.0285 | 0.0305 | 0.1035 |
| | CSD | 0.0160 | 0.0145 | 0.0655 | 0.0190 | 0.0205 | 0.0690 | 0.0340 | 0.0370 | 0.1135 |
| | CSDH | 0.0185 | 0.0230 | 0.0790 | 0.0210 | 0.0225 | 0.0820 | 0.0295 | 0.0280 | 0.1010 |
| 200 | CSI | 0.0255 | 0.0570 | 0.9875 | 0.0335 | 0.0550 | 0.9850 | 0.0390 | 0.0585 | 0.9865 |
| | CSD | 0.0235 | 0.0535 | 0.9855 | 0.0345 | 0.0550 | 0.9845 | 0.0705 | 0.0890 | 0.9865 |
| | CSDH | 0.0255 | 0.0595 | 0.9875 | 0.0325 | 0.0560 | 0.9840 | 0.0470 | 0.0690 | 0.9855 |
| 400 | CSI | 0.0430 | 0.2095 | 1.0000 | 0.0390 | 0.2115 | 1.0000 | 0.0450 | 0.2210 | 1.0000 |
| | CSD | 0.0405 | 0.2070 | 1.0000 | 0.0390 | 0.2100 | 1.0000 | 0.0515 | 0.2270 | 1.0000 |
| | CSDH | 0.0550 | 0.2080 | 1.0000 | 0.0400 | 0.2135 | 1.0000 | 0.0555 | 0.2405 | 1.0000 |

This table presents results for the Lagrange multiplier test for stability in the factor augmented regression. The DGP is detailed in Section 5.2. CSI, CSD and CSDH denote the same scenarios as in Table 1.

**Table 4: Empirical Results, Bond Excess Returns, Summary Statistics, 1965:01 - 2007:12**

| **Panel A: Levels** | | | | |
|---|---|---|---|---|
| | $n=2$ | $n=3$ | $n=4$ | $n=5$ |
| Mean($rx_t^n$) | 0.004 | 0.007 | 0.009 | 0.009 |
| Std. Dev.($rx_t^n$) | 0.019 | 0.034 | 0.047 | 0.058 |
| Corr ($rx_t^n, rx_{t-1}^n$) | 0.931 | 0.933 | 0.932 | 0.922 |
| Corr ($rx_t^n, rx_{t-12}^n$) | 0.215 | 0.147 | 0.109 | 0.071 |
| **Panel B: Absolute Values** | | | | |
| | $n=2$ | $n=3$ | $n=4$ | $n=5$ |
| Mean($\lvert rx_t^n \rvert$) | 0.015 | 0.027 | 0.037 | 0.045 |
| Std. Dev.($\lvert rx_t^n \rvert$) | 0.012 | 0.022 | 0.031 | 0.037 |
| Corr ($\lvert rx_t^n \rvert, \lvert rx_{t-1}^n \rvert$) | 0.857 | 0.862 | 0.856 | 0.836 |
| Corr ($\lvert rx_t^n \rvert, \lvert rx_{t-12}^n \rvert$) | 0.070 | 0.129 | 0.150 | 0.157 |

This table presents summary statistics for yearly bond excess returns expressed in levels (Panel A) and absolute values (Panel B) computed over the maturities $n = 2, 3, 4, 5$. The data sample ranges from January 1965 to December 2007, for a sample size of 516 monthly observations. Mean and standard deviation are displayed in decimals per annum (i.e., 0.01 equals 1 percentage point).

**Table 5: Empirical Results, Estimated Factors, Average Marginal $R^2$, 1964:01 - 2006:12**

| | **Group 1: Output and Income** | **Group 2: Labor Market** |
|---|---|---|
| $\hat{f}_{1t}$ | 0.298 | 0.166 |
| $\hat{f}_{2t}$ | 0.107 | 0.078 |
| | **Group 3: Consumption and Orders** | **Group 4: Orders and Inventories** |
| $\hat{f}_{1t}$ | 0.211 | 0.100 |
| $\hat{f}_{2t}$ | 0.046 | 0.086 |
| | **Group 5: Money and Credit** | **Group 6: Interest Rate and Exchange Rates** |
| $\hat{f}_{1t}$ | 0.032 | 0.084 |
| $\hat{f}_{2t}$ | 0.025 | 0.263 |
| | **Group 7: Prices** | **Group 8: Stock Market** |
| $\hat{f}_{1t}$ | 0.249 | 0.002 |
| $\hat{f}_{2t}$ | 0.004 | 0.185 |

For each of the eight groups listed in Section 6.2, this table displays the average of the $R^2$ obtained from the breakpoint regression with $\hat{\pi}_{\mathbf{x}} = 0.455$ of each variable within the group on a given factor. The data sample ranges from January 1964 to December 2006, for a sample size of 516 monthly observations.

**Table 6: Empirical Results, Bond Excess Returns, Least Squares Estimation, 1965:01 - 2007:12**

| | $n=2$ | | | | $n=3$ | | | | $n=4$ | | | | $n=5$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| $\pi_y^{(n),0}$ | - | - | - | 0.510 | - | - | - | 0.510 | - | - | - | 0.510 | - | - | - | 0.355 |
| $\gamma_{11}^{(n),0}$ | - | 0.008 | 0.005 | 0.007 | - | 0.012 | 0.006 | 0.009 | - | 0.015 | 0.005 | 0.009 | - | 0.017 | 0.005 | 0.029** |
| | - | (0.006) | (0.004) | (0.006) | - | (0.011) | (0.008) | (0.010) | - | (0.014) | (0.010) | (0.013) | - | (0.017) | (0.012) | (0.015) |
| $\gamma_{12}^{(n),0}$ | - | -0.009* | -0.001 | 0.010*** | - | -0.021** | -0.006 | 0.013* | - | -0.028** | -0.007 | 0.016* | - | -0.032** | -0.007 | 0.003 |
| | - | (0.005) | (0.006) | (0.004) | - | (0.008) | (0.009) | (0.007) | - | (0.011) | (0.011) | (0.009) | - | (0.013) | (0.012) | (0.014) |
| $\gamma_{21}^{(n),0}$ | - | - | - | -0.003 | - | - | - | -0.005 | - | - | - | -0.006 | - | - | - | -0.016 |
| | - | - | - | (0.002) | - | - | - | (0.003) | - | - | - | (0.005) | - | - | - | (0.015) |
| $\gamma_{22}^{(n),0}$ | - | - | - | -0.021*** | - | - | - | -0.035*** | - | - | - | -0.043*** | - | - | - | -0.013 |
| | - | - | - | (0.005) | - | - | - | (0.009) | - | - | - | (0.012) | - | - | - | (0.014) |
| $\beta_1^{(n),0}$ | 0.001 | 0.004* | 0.001 | 0.000 | 0.001 | 0.008* | 0.001 | -0.001 | 0.000 | 0.010 | 0.000 | -0.002 | -0.002 | 0.010 | -0.001 | 0.000 |
| | (0.002) | (0.003) | (0.002) | (0.002) | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) | (0.006) | (0.005) | (0.005) | (0.006) | (0.007) | (0.006) | (0.005) |
| $\beta_2^{(n),0}$ | 0.455*** | - | 0.438*** | 0.467*** | 0.857*** | - | 0.823*** | 0.868*** | 1.235*** | - | 1.199*** | 1.256*** | 1.453*** | - | 1.417*** | 1.464*** |
| | (0.052) | - | (0.058) | (0.058) | (0.103) | - | (0.115) | (0.118) | (0.150) | - | (0.165) | (0.168) | (0.192) | - | (0.208) | (0.202) |
| $\bar{R}^2$ | 0.344 | 0.102 | 0.352 | 0.415 | 0.358 | 0.095 | 0.362 | 0.404 | 0.382 | 0.084 | 0.383 | 0.415 | 0.347 | 0.065 | 0.347 | 0.370 |
| $\sup \mathcal{LM}$ | - | - | - | 10.756* | - | - | - | 9.868 | - | - | - | 10.295* | - | - | - | 11.224* |

This table reports results from least squares estimation of the following models for $n$-year bond excess returns $rx_{t+12}^{(n)}$, for $t = 1964:01, \ldots, 2006:12$:

$$
\begin{aligned}
(a)\quad & rx_{t+12}^{(n)} = \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)};\\
(b)\quad & rx_{t+12}^{(n)} = \gamma_{11}^{(n)} \hat{f}_{1t} + \gamma_{12}^{(n)} \hat{f}_{2t} + \beta_1^{(n)} + \varepsilon_{t+12}^{(n)};\\
(c)\quad & rx_{t+12}^{(n)} = \gamma_{11}^{(n)} \hat{f}_{1t} + \gamma_{12}^{(n)} \hat{f}_{2t} + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)};\\
(d)\quad & rx_{t+12}^{(n)} = \mathbb{I}\left(t/T \le \pi_y^{(n)}\right)\left(\gamma_{11}^{(n)} \hat{f}_{1t} + \gamma_{12}^{(n)} \hat{f}_{2t}\right) + \mathbb{I}\left(t/T > \pi_y^{(n)}\right)\left(\gamma_{21}^{(n)} \hat{f}_{1t} + \gamma_{22}^{(n)} \hat{f}_{2t}\right) + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)}.
\end{aligned}
$$

The vector of factors $\hat{\mathbf{f}}_t = \left(\hat{f}_{1t}, \hat{f}_{2t}\right)'$ is estimated as detailed in Section 3.2.1; the number of factors $\hat{R} = 2$ (i.e., the dimension of $\hat{\mathbf{f}}_t$) is determined using the $IC_{p2}(R, R)$ criterion in (6). The factors are extracted from the 123 macroeconomic series described in Section 6.2 over the period $1964:01 - 2006:12$. $CP_t$ is the Cochrane and Piazzesi (2005) factor, namely a linear combination of five forward spreads estimated over the period $1964:01 - 2006:12$. Newey and West (1987) standard errors with 18 months lag order are reported in parentheses. $\bar{R}^2$ is the adjusted $R^2$. $^*$, $^{**}$ and $^{***}$ denote significance at 10%, 5% and 1% level, respectively. Critical values for the $\sup \mathcal{LM}$ statistic with $\underline{\pi}_y = 15\%$ trimming and $\hat{R} = 2$ restrictions are 10.01, 11.79 and 15.51 for 10%, 5% and 1% level, respectively (Andrews (1993)).

**Table 7: Alternative Estimation Strategy, Bond Excess Returns, 1965:01 - 2007:12**

| | $n=2$ | $n=3$ | $n=4$ | $n=5$ |
|---|---|---|---|---|
| $\gamma_{11}^{(n),0}$ | 0.007 | 0.009 | 0.017 | 0.009 |
| | (0.006) | (0.010) | (0.010) | (0.015) |
| $\gamma_{12}^{(n),0}$ | 0.010*** | 0.013* | 0.017* | 0.016 |
| | (0.004) | (0.007) | (0.010) | (0.012) |
| $\gamma_{21}^{(n),0}$ | -0.002 | -0.004 | -0.006 | -0.007 |
| | (0.002) | (0.003) | (0.004) | (0.005) |
| $\gamma_{22}^{(n),0}$ | -0.016*** | -0.028*** | -0.035*** | -0.035*** |
| | (0.006) | (0.009) | (0.012) | (0.014) |
| $\beta_{1}^{(n),0}$ | 0.0001 | 0.000 | -0.001 | -0.003 |
| | (0.002) | (0.004) | (0.005) | (0.006) |
| $\beta_{2}^{(n),0}$ | 0.482*** | 0.895*** | 1.290*** | 1.506*** |
| | (0.060) | (0.119) | (0.168) | (0.213) |
| $\bar{R}^2$ | 0.398 | 0.396 | 0.409 | 0.364 |

This table reports results from least squares estimation of the following model for $n$-year bond excess returns $rx_{t+12}^{(n)}$, for $t = 1964:01, \ldots, 2006:12$:

$$rx_{t+12}^{(n)} = \mathbb{I}\left(t/T \leq \hat{\pi}_{\mathbf{x}}\right)\left(\gamma_{11}^{(n)}\hat{f}_{1t} + \gamma_{12}^{(n)}\hat{f}_{2t}\right) + \mathbb{I}\left(t/T > \hat{\pi}_{\mathbf{x}}\right)\left(\gamma_{21}^{(n)}\hat{f}_{1t} + \gamma_{22}^{(n)}\hat{f}_{2t}\right) + \beta_1^{(n)} + \beta_2^{(n)}CP_t + \varepsilon_{t+12}^{(n)}.$$

The vector of factors $\hat{\mathbf{f}}_t = \left(\hat{f}_{1t}, \hat{f}_{2t}\right)'$ and the break fraction $\hat{\pi}_{\mathbf{x}}$ are estimated as detailed in Section 3.2.1; the number of factors $\hat{R} = 2$ (i.e., the dimension of $\hat{\mathbf{f}}_t$) is determined using the $IC_{p2}(R, R)$ criterion in (6). The factors are extracted from the 123 macroeconomic series described in Section 6.2 over the period $1964:01 - 2006:12$. $CP_t$ is the Cochrane and Piazzesi (2005) factor, namely a linear combination of five forward spreads estimated over the period $1964:01 - 2006:12$. Newey and West (1987) standard errors with 18 months lag order are reported in parentheses. $\bar{R}^2$ is the adjusted $R^2$. * and *** denote significance at 10% and 1% level, respectively.