May 2018

STUDIES ON HOUSING MARKET DYNAMICS AND COINTEGRATION ANALYSIS WITH LATENT FACTORS

Shulin Shen

Syracuse University

ABSTRACT

This dissertation consists of two essays on housing market dynamics and cointegration analysis with latent factors. The theme of this dissertation is housing market dynamics, with the first essay an application of advanced panel time series models to the studies of housing market dynamics, and the second essay a theoretic derivation of an econometric tool on cointegration analysis with latent factors that can be applied to the housing market analysis.

This first essay develops a parsimonious dynamic model to study the impact of a common demand shock in the housing market on housing prices and construction activities across a set of locations with heterogeneous supply side conditions. Embedded within the model is a lead-lag structure that allows one to identify from where shocks propagate while allowing for and yielding estimates of cross-sectional differences in housing supply elasticities. The findings indicate that local supply conditions may matter more than distance when modeling spatiotemporal dynamics in the housing market.

The second essay considers estimating and testing cointegration between an integrated series of interest and a vector of possibly cointegrated nonstationary latent factors. The fully modified least squares (FM-OLS) estimation is adopted to the estimation of the cointegration relation of interest. The asymptotic properties of the FM-OLS estimators are derived, and the residual-based cointegration tests are shown to work as usual even with latent factors. Based on the estimated cointegration relation, it is demonstrated that an error correction term added to the traditional diffusion index forecast model improves forecasting accuracy.

STUDIES ON HOUSING MARKET DYNAMICS AND COINTEGRATION ANALYSIS WITH LATENT FACTORS

by

Shulin Shen

B.S., Xi’an Jiaotong University, 2011
M.S., Xi’an Jiaotong University, 2013

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in *Economics*.

Syracuse University

May 2018

ACKNOWLEDGEMENTS

I thank all the people who helped me complete this dissertation and contributed to my intellectual progress at Syracuse University. In particular, my deepest gratitude goes to Professor Stuart Rosenthal, Professor Yoonseok Lee, Professor Badi Baltagi, and Professor Chihwa Kao. Their encouragement, support and insightful guidance have led me through each step of my progress. I thank Professors William Horrace, Jeffrey D. Kubik, Abdulaziz Shifa, Jerry S. Kelly, Alfonso Flores-Lagunes, Jan Ondrich, and Donald H. Dutkowsky for helping me broaden my view of economic research. I would also like to thank Professors Terry McConnell, Pinyuen Chen, and H. Hyune-Ju Kim for enhancing my mathematic skills. I appreciate the contribution of Professors Yulong Wang and Yingyi Ma to my dissertation defense.

My special thanks go to Yusen Kwoh, Chuntien Hu, Daigee Shaw, Yuan Zhao, and Weihua Yu. Their passion for academic research and encouragement have led me to pursue my PhD degree. I also benefit a lot from the suggestions and encouragement from Fa Wang and Bin Peng. I would also like to acknowledge the assistance from Peg Austin, Karen Cimilluca, Emily Henrich, Sue Lewis, Candi Patterson, Mary Santy, Laura Walsh, Katrina Wingle, Matthew O'Keefe, and other staff in the Economics Department and Center for Policy Research.

Lastly, I would like to thank my family and friends for their support. My parents, Xiaoqiu Li and Keming Shen, have devoted decades in helping me pursue my academic career. Last, but most important, my lifelong appreciation goes to my husband, Jindong Pang. His company and encouragement have carried me through all these difficulties and challenges in my academic career from the very beginning of my college training. My daughter, Evelyn, offers me lots of oy in the last year of my graduate study. Thanks to all my friends and classmates and I have enjoyed my time at Syracuse.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2: Unobserved Demand Shocks and Housing Market Dynamics in a Model with Spatial Variation in the Elasticity of Supply

LIST OF TABLES

Chapter 2: Unobserved Demand Shocks and Housing Market Dynamics in a Model with Spatial Variation in the Elasticity of Supply

Chapter 3: Fully Modified Least Squares Estimation of Factor-Augmented Cointegration Regressions

Chapter 1: Studies on Housing Market Dynamics and Cointegration Analysis with Latent Factors

The theme of this dissertation is housing market dynamics. Housing markets in the US as well as in many other countries exhibit huge volatilities during the past several decades, including the most recent Great Recession. The fluctuations in the housing market have very influential impacts on our economy, either through mortgage markets and the construction activities or through many other channels such as the consumption and saving behavior of households. Besides huge volatilities in the housing market, there are substantial heterogeneity in the dynamics of local housing markets. For these and many other reasons, a growing number of studies have attempted to model and forecast housing price dynamics. Given that housing prices are mostly nonstationary time series and are highly spatially correlated across local housing markets, being able to model the nonstationarities and spatial correlations in the housing markets is the key issue in the housing market analysis.

This dissertation consists of two essays on housing market dynamics and cointegration analysis with latent factors. The first essay is an application of advanced panel time series models to housing dynamics studies, in which the spatial correlations of housing markets are rooted in the spatially correlated demand shocks to the housing markets and a lead-lag diffusion pattern of the housing demand shocks in a regional housing market is identified and estimated. The second essay is on a theoretic derivation of an econometric tool on cointegration analysis with latent factors. Given the fact that housing prices are mostly nonstationary time series and housing markets in a given geographic region may subject to common shocks, the idea of taking advantage of large dimensional nonstationary data sets to the housing market analysis is appealing as well as challenging. The second essay provides a theoretic tool of estimating the cointegration relation between an integrated series of interest and a vector of possibly integrated factors. The vector of possibly integrated factors provides a method to summarize the co-movements in a large nonstationary panel data set. The theoretic results can be applied to the housing market to study the common cycles and long-run equilibrium relations in local housing markets.

To be more specific, the first essay in this dissertation extends a parsimonious error correction model to study the underlying unobservable spatially correlated demand shocks across a set of locations. Currently, most of these studies on housing market dynamics focus on price movements only and are not able to provide insights on the heterogeneity in the diffusion patterns in the local housing markets. More importantly, as documented in the business cycle literature, sometimes national or state level building permits may be a better leading indicator for economic activities than housing prices. So there may be quite different roles played by housing prices and construction activities. Being able to model these two closely related important components of the housing market will help us gain a much broader and more comprehensive view of the housing market dynamics.

In the first essay, we build our model on a simple supply and demand model of a local housing market. With the assumption that there is only a demand shock to the local housing market, the demand shock can be written as a function of the observable price change and new construction. By modeling the underlying demand shocks, we are able to derive two reduced form diffusion models, one for price movements, and the other for construction activities. In these derived diffusion models, the coefficient estimates depend on the local price elasticities of housing supply explicitly, which enable us to model the heterogeneous price response and construction response across locations. Another feature of this model is that it not only controls for local spillover effects in the housing market by adding spatial lag terms in the main equations, but also allows for the identification of a leading area from where the housing market shocks originate and spread out contemporaneously.

The data we use is the Federal Housing Finance Agency (FHFA) house price indices and building permits data from US Census for 22 largest MSAs in California from 1980 to 2016. Our estimation results first indicate that housing market in San Jose could be treated as a leading market in these 22 MSAs of California. Secondly, conditioning on the local spillover effect, the response to a common demand shock in a local housing market is quite different for locations with different local supply conditions. For coastal cities in California with less elastic housing supply, the price adjustments are much more substantial than the construction adjustments given a common demand shock to this market. In contrast, for most inland cities of California with more elastic housing supply, they adjust construction more than price when facing a common housing market shock. Another important finding is that the coefficient estimates from the two reduced form diffusion models, display correct positive correlations with the Saiz (2010) price elasticities measurements. This positive correlation provides support for our modeling of the underlying demand shocks and can work as an alternative method to get estimates of local supply conditions.

The second paper in this dissertation studies the estimation and testing of the co-integration relation between an integrated variable of interest and a vector of latent integrated factors. The latent factors are unobservable but can be estimated from a large panel of integrated series. One motivation of this study is dimension reduction. In macroeconomic literature, as the data are getting much easier to collect, the number of potential useful variables for analysis could be huge, and most the macroeconomic variables are nonstationary intrinsically. Also, there may not exist any economic theory guiding us how to model the long run relation among these integrated series. One way to take advantage of this large panel data set is through the factor analysis to extract the common stochastic trends, and study the possible long run relation between the common stochastic trends and an integrated series of interest. Since the factors are of a much lower dimension, the co-integration analysis will be much easier to conduct. Another motivation is the growing literature on the Factor-augmented error correction model (FECM). The FECM model focuses on the co-integration relation between a smaller subset of the series in the large panel set and the set of latent factors. The method has been used empirically by adding an error correction term to the forecasting of the first-differenced series. However, there is no theoretical evidence to support the cointegration analysis and the estimation of the FECM model using estimated factors. The estimation errors in the latent integrated factors could accumulate across time and may cause problems in the cointegration analysis.

The second paper in this dissertation tries to fill in this gap by studying the estimation and testing of the co-integration relation between the latent factors and another integrated series of interest. The nonstationary factor used in this paper is a more general one, which allows for nonstationary idiosyncratic error terms in the factor model. We also allow for possible endogeneity in the latent factors in the main cointegration equation. Following Phillips and Hansen (1990) fully modified least squares estimator, we show that under some restrictions on the sample sizes and the bandwidth expansion rates of the long run covariance matrices estimator, the fully modified least estimator of the cointegration coefficient using estimated factors have a mixed normal limiting distribution, which will help with hypothesis testing and statistical inference.

Another theoretical result the second paper verifies is that the conventional residual-based cointegration test work as usual as long as the factors are consistently estimated. At the end of the paper, we propose a possible application of the fully modified estimator of the cointegration coefficient to the traditional diffusion index forecasting literature. After testing and estimating for the cointegration relation, we could add an error correction term to the conventional forecasting equation of a differenced integrated variable if there exists any cointegration relation between the level of the variable and the level of the factors. Our empirical example shows that the proposed forecasting method may outperform existing methods under some cases.

Chapter 2: Unobserved Demand Shocks and Housing Market Dynamics in a Model with Spatial Variation in the Elasticity of Supply

1. Introduction

1.1 Overview

This paper extends a parsimonious dynamic model developed in Holly et al. (2011) (hereafter HPY) to estimate the influence of spatially correlated unobserved demand shocks on house price movements and construction across locations. The extended model has embedded within it cross-sectional differences in housing supply elasticities while also allowing for the possibility that shocks may propagate out over time from a dominant location. The emphasis on the heterogeneity in housing supply elasticities in our model is similar in spirit to papers like Glaeser and Gyourko (2005) and Glaeser et al (2008) who demonstrate that housing supply elasticities have important effects on house price volatility.

The original model of HPY offers a parsimonious structure for analyzing spatial and temporal diffusion of house price shocks in a dynamic system. HPY estimate separate house price diffusion models for different cities in the U.K. allowing for the possibility that price in a given city may be cointegrated with price movements in a “dominant” city (which is London in their case). This structure allows for possible lead-lag relationships by allowing demand shocks to hit the dominant location first and then propagate out over time to secondary locations. We extend HPY by explicitly modeling the unobserved demand shocks allowing for cross-sectional differences in housing supply elasticities as suggested above. Our model is then used to examine both house price dynamics and construction whereas HPY focus on price movements only. In this sense, the model in HPY is a restricted version of the model developed in this paper.

The need to do a better job of modeling housing market dynamics for the U.S. became especially obvious following the crash of 2007. Sharply falling housing prices prompted massive numbers of mortgage defaults, dramatic declines in new construction, and pushed the economy into the Great Recession (Leamer, 2007; Iacoviello, 2005). Nevertheless, despite the onset of an historic national recession, the recent boom and bust in housing prices and mortgage default did not hit all metropolitan areas similarly. Cities like Phoenix, Los Angeles and Sarasota saw prices more than double in the few years leading up to the 2006 peak only to fall precipitously in the following few years. Other large growing cities like Denver and Houston experienced comparatively little change in housing prices over the same period. For these and other reasons, a growing number of studies have attempted to model and forecast housing price dynamics in a manner that allows for spatial correlation and patterns across cities, but most often in a reduced form context.

Based on our extended diffusion model of unobserved demand shocks, we derive a price diffusion and a construction diffusion model for each individual metropolitan area, and illustrate their features using data on house prices and construction for 22 metropolitan areas in California from 1975 to present. Results indicate strong evidence that metro-level house prices are cointegrated in California, where cointegrating coefficients are positively correlated with local supply elasticities. These estimated cointegrating coefficients also allow us to infer estimates of the elasticity of supply for individual cities (up to a scale factor) as noted above. Those estimates correlate closely with elasticity measures obtained by Saiz (2010) using very different data on topography of land forms.

Based on cointegration and exogeneity tests, additional findings indicate that price changes in San Jose can be treated as a common factor for all other metropolitan area price changes. This is consistent with San Jose being the center of the high-tech industry, an industry that is both volatile and which generates enormous amounts of income and employment in the California economy. Besides the important role of the leader's price shocks, our results also highlight the importance of cross-sectional differences in the price elasticities of housing supply. The effect of the dominant area's price shocks tend to be inversely related to local supply elasticities. Inelastic locations also exhibit larger and faster price adjustments following shocks to the dominant area while elastic metro areas exhibit larger and faster changes in the level of new construction.

Additionally, several panel model specifications are estimated for groups of locations outside of San Jose (with San Jose treated as a separate dominant area). Impulse response functions are also used to highlight related dynamics.[1] The panel model estimates of the construction diffusion model further indicate that San Jose's contemporaneous effects are sizable and significant, and tend to be larger in annual and biannual data as compared to quarterly data. The effect of data frequency is consistent with the fact that supply elasticities tend to increase with the time horizon. Such a perspective emerges naturally out of our model with our explicit modeling of unobserved demand shocks and supply elasticities. That perspective, however, has been mostly overlooked in most previous papers on housing market dynamics which adopt a more reduced form specification.

The rest of the paper is set out as follows. The next subsection provides further background on related literature. In Section 2, we derive the demand diffusion model and then derive the price and construction diffusion models. We also show how panel estimation of the price and diffusion models take into account the supply side conditions in Section 2.3. The local projection method of spatial-temporal impulse responses is presented in Section 2.4. In Section 3, we report estimates of the price and construction diffusion model using quarterly, annual and biannual data for 22 metro areas in California over the period 1980Q1-2016Q4. In Section 4, we draw some conclusions.

[1] We use the local projection method of Jordà (2005) to study the high dimensional spatial-temporal impulse response functions. Without the need to invert a high dimensional matrix and allowing for estimating the impulse response functions of a different dependent variable, the local projection method of Jordà (2005) provides an easy-to-implement way of diffusion analysis. From the impulse response analysis, we find that a positive shock to San Jose house price spills over to other regions gradually regardless of the distance to San Jose and regardless of the supply side conditions. In addition, a positive San Jose's house price shock will have a significant and persistent effect on construction in metro areas with more elastic housing supply conditions.

1.2 **Previous literature**

Our paper builds off a number of different studies that have examined housing market dynamics from several different perspectives. The most relevant literature is the study of spatial correlations of housing market dynamics. One of the most important forms of cross section dependence arises from contemporaneous dependence across space by relating each cross section unit to its neighbors (Whittle, 1954; Cliff and Ord, 1973; Anselin, 2013; Kelejian and Robinson, 1995; Kelejian and Prucha, 1999, 2010; Lee, 2004; Brady, 2011). Another approach to dealing with cross sectional dependence is to make use of multifactor error processes where the cross section dependence is characterized by a finite number of unobserved common factors (Pesaran, 2006; Bai 2003, 2009). However, there is no clear guidance whether the spatial dependence is pervasive or attenuates across space empirically. Holly et al. (2010) model house prices at the level of US states and find there is evidence of significant spatial dependence even when the strong form of cross sectional dependence has been swept away by the use of cross sectional averages.

As compared to purely spatial or purely factor models analyzed in the literature, the spatial-temporal model developed in HPY uses London house prices as the common factor and then models the remaining dependencies conditional on London house prices. This paper extends the HPY model to study the diffusion patterns of the unobserved underlying demand shocks. This ensures an important role for local supply conditions that have the potential to dampen or amplify the impact of demand shocks on price and quantity responses but which are mostly ignored in HPY. Instead, HPY argued that the supply of housing is very inelastic in the UK, with a supply elasticity of 0.5 compared to an elasticity of 1.4 for the US. (Swank et al., 2002). Clearly, if the price elasticity of housing supply differed markedly across regions, then responses to both region specific and national demand shocks could generate very different house price dynamics (Glaeser and Gyourko, 2005; Glaeser et al., 2008).

Another highly relevant literature is the study of supply constraints and housing market dynamics. Since DiPasquale's (1999) review of the literature to that date, academic work on housing supply has expanded extensively. Several papers have made it clear that constraints on housing supply vary markedly across regions of the United States, and that these constraints can explain large differences in house prices and the level of construction (Mayer and Somerville, 2000; Glaeser et al., 2005; Gyourko and Saiz, 2006; Quigley and Raphael, 2005; Green et al., 2005; Ihlanfeldt, 2007; Glaeser and Ward, 2009; Paciorek, 2013). These and related papers, however, typically posit a relatively simple relationship between price and housing investment that ignores spatial spillovers and patterns that contribute to cross-sectional variation in housing market dynamics. This paper starts by building a diffusion model of the unobserved demand shocks across space, and then derives two reduced form diffusion models for price shocks and new construction, respectively. The relationship between price and investments and the impact of supply side conditions on this relationship are implicitly embedded in the construction diffusion model.

This paper is also closely related to the literature on housing market efficiency, housing bubbles and business cycles. Papers such as Hosios and Pesando (1991) and Case and Shiller (1989, AER) find evidence that quality-adjusted house prices are serially correlated on a quarterly basis, implying future house prices are forecastable. Capozza et al. (2004) finds that higher construction costs were associated with higher serial correlation and lower mean reversion in housing prices, presenting conditions for price overshooting. Even though these papers model house price dynamics, their main focus is to assess whether house prices were forecastable and thus test if there is a bubble in the housing market (Flood and Hodrick, 1990). The time series methods applied were relatively simple ignoring spatial correlations, underlying demand shocks, and lead-lag patterns.

In the housing bubble literature, Glaeser et al. (2008) find that the duration and magnitude of housing bubbles are sensitive to the housing supply elasticity, with larger price increases in supply-inelastic areas during booms. Complementing Glaeser et al. (2008), Huang and Tang (2012) also find a significant link between the supply inelasticity and price declines during a bust. These papers provide evidence that supply elasticities may amplify (or mute) housing market boom and bust patterns but do not formally model underlying supply and demand factors. More recently, Liu et al. (2016) document within-city heterogeneity in response to a bubble, and Landvoigt et al. (2015) find that cheaper credit for poor households was a major driver of prices during the 2000s boom, especially at the low end of the market. These two papers formally model supply (Liu et al., 2016) and demand factors (Landvoigt et al., 2015) and document within city heterogeneity during a housing boom and bust episode.[2]

The importance of modeling housing market dynamics has been reinforced by a growing number of macroeconomic studies that treat volatility in the housing market as a source and not simply a consequence of business cycle fluctuations. Bernanke (2008), Leamer (2007), and Davis and Heathcote (2005) argue that housing is a leading driver of business cycles and suggest that housing should be treated differently from other types of investments in macroeconomic models. More recently, Strauss (2013) finds that national and state-level building permits significantly lead economic activity in nearly all US states over the past three decades, while Ghent and Owyang (2010) find that national permits are a better leading indicator for a city's employment and that declines in house prices are often not followed by declines in employment. While the focus of our paper is not on links between the housing market and local business cycles per se, by formally modeling the manner in which unobserved demand shocks contribute to spatiotemporal patterns of home prices and housing

[2] In related work, Genesove and Han (2012) use commuting time as a proxy for within-city variation in supply elasticity and report evidence that during a housing crash prices fall more in the city center than at a city's edge.

construction our model provides a framework that can be used to help explain cross-sectional differences in boom-bust patterns.[3]

2. Demand Diffusion Model

2.1 A demand shock diffusion model

In this paper, we apply the dynamic system of HPY to the cumulative demand shocks derived below. To simplify notation, we use p_{it} (or lnP_{it}) to denote the log of house prices, and use lnQ_{it} to denote the log of house stocks over time for $t = 1,2,\dots,T$, and over areas $i = 0,1,2,\dots,N$. Given the assumption that there is only a demand shock to each local housing market, and under the premise that the supply and demand functions of housing follow a log linear form, the demand shock at time period t for location i, denoted by Δd_{it}, can be expressed as the vertical distance between the new demand curve and the old one. As illustrated in Figure 1, using simple algebra, we have

$$\Delta d_{it} = \left(1 + \frac{|\varepsilon_i^s|}{|\varepsilon_i^d|}\right) \Delta lnP_{it} = \left(\frac{1}{|\varepsilon_i^s|} + \frac{1}{|\varepsilon_i^d|}\right) \Delta lnQ_{it},$$

with ε^s being the price elasticity of supply of housing, ε^d being the price elasticity of demand for housing, and the symbol Δ signifies changes in relevant variables. The cumulative demand shock at time t for location i is given by

$$d_{it} = \sum \Delta d_{it} = \left(1 + \frac{|\varepsilon_i^s|}{|\varepsilon_i^d|}\right) lnP_{it}.$$

We assume that one of the areas, say area 0, is dominant in the sense that shocks to it propagate to other areas simultaneously and over time, whilst shocks to the remaining areas have little immediate impact on area 0. For the dominant area, the first order linear error correction specification is given by:

[3] Ghent and Owyang (2010), Del Negro and Otrok (2007), and Hernández-Murillo et al. (2015) all find that housing cycles may have both national and regional elements but that the more pervasive national cycle is dominated by cross-sectional heterogeneity upon disaggregating the data. The lead-lag diffusion model of unobserved demand shocks in this paper analogously allows for a common regional factor in addition to idiosyncratic city-specific drivers of housing market volatility.

$$\Delta\, d_{0t} = \phi_{0s}\left(d_{0,t-1} - \omega_0 \bar{d}^{\,s}_{0,t-1}\right) + a_0 + a_{01}\Delta d_{0,t-1} + b_{01}\Delta \bar{d}^{\,s}_{0,t-1} + \varepsilon_{0t}.$$

For the remaining areas, it is given by:

$$\Delta\, d_{it} = \phi_{is}\left(d_{i,t-1} - \omega_i \bar{d}^{\,s}_{i,t-1}\right) + \phi_{i0}\left(d_{0,t-1} - \delta_i d_{i,t-1} - \rho_i t\right) + a_i + a_{i1}\Delta d_{i,t-1} + b_{i1}\Delta \bar{d}^{\,s}_{i,t-1} + c_{io}\,\Delta d_{ot} + \varepsilon_{it},$$

for $i = 1,2, \dots, N$, where $\bar{d}^{\,s}_{it}$ denotes the spatial variable for area i defined by

$$\bar{d}^{\,s}_{it} = \sum_{j=0}^{N} s_{ij} d_{jt} \text{ with } \sum_{j=0}^{N} s_{ij} = 1.$$

In the empirical application, we use an inverse distance measure where s_{ij} is proportional to $1/D_{ij}$, with D_{ij} being the distance between location i and location j. In the above specification, we assume that cumulative demand shocks for other locations $i = 1,2, \dots, N$, are cointegrated with that of the dominant area with cointegration relation given by $d_{ot} - \delta_i d_{it} - \rho_i$t. The size of δ_i and ρ_i depend on the relative income and population growth in location i relative to the leading area. These two parameters measure the long-run relation among fundamental driving forces of the demand for housing across different locations.

In practice, it is hard to estimate the above model of demand shocks since there is no accurate measure of the level of the demand shocks. However, from the simple linear algebra, we can express the demand shock as a function of the housing prices. Substituting relevant expressions into the above system and normalizing the coefficients of the left-hand side (LHS) variables, we get[4]

$$\Delta p_{0t} = \tilde{a}_0 + a_{01}\Delta p_{0,t-1} + \tilde{b}_{01}\Delta \bar{p}^{\,s}_{0,t-1} + \tilde{\varepsilon}_{0t},$$

[4] To simplify the illustration, we first ignore the error correction term involving the spatial average of neighbor's demand shocks. The model reduces into

$$\Delta\, d_{0t} = a_0 + a_{01}\Delta d_{0,t-1} + b_{01}\Delta \bar{d}^{\,s}_{0,t-1} + \varepsilon_{0t},$$

And for the remaining areas

$$\Delta\, d_{it} = \phi_{i0}\left(d_{0,t-1} - \delta_i d_{i,t-1} - \rho_i t\right) + a_i + a_{i1}\Delta d_{i,t-1} + b_{i1}\Delta \bar{d}^{\,s}_{i,t-1} + c_{io}\,\Delta d_{ot} + \varepsilon_{it}.$$

After substituting relevant expressions into above equations, we have for the dominant area

$$\left(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\right)\Delta p_{0t} = a_0 + a_{01}\left(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\right)\Delta p_{0,t-1} + b_{01}\Delta \bar{d}^{\,s}_{0,t-1} + \varepsilon_{0t},$$

and for the remaining areas:

$$\left(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\right)\Delta\, p_{it} = \phi_{i0}\left(\left(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\right)p_{0,t-1} - \delta_i\left(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\right)p_{i,t-1} - \rho_i t\right) + a_i + a_{i1}\left(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\right)\Delta\, p_{it-1} + b_{i1}\Delta \bar{d}^{\,s}_{i,t-1} + c_{io}\left(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\right)\Delta p_{0t} + \varepsilon_{it}.$$

$$\Delta\, p_{it} = \tilde{\phi}_{i0}\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right) + \tilde{a}_i + a_{i1}\Delta\, p_{it-1} + \tilde{b}_{i1}\Delta\bar{p}\,^{s}_{i,t-1} + \tilde{c}_{io}\,\Delta p_{0t} + \tilde{\varepsilon}_{it}\,,$$

where $\bar{p}\,^{s}_{it}$ denotes the spatial variable for area i defined by

$$\bar{p}\,^{s}_{it} = \sum_{j=0}^{N} s_{ij} p_{jt} \text{ with } \sum_{j=0}^{N} s_{ij} = 1.$$

Thus, we can derive the error correction specification for the log of house prices (housing price diffusion model) from the cointegrating relation between the demand shocks across locations. For location i, the coefficient on the error correction term $\tilde{\phi}_{i0}$, the coefficient on the contemporaneous effect of the leading area $\tilde{c}_{i0}$, and the cointegrating coefficients are as follows:

$$\tilde{\phi}_{i0} = \phi_{i0}\left(1 + \frac{|\varepsilon_0^s|}{|\varepsilon_0^d|}\right) / \left(1 + \frac{|\varepsilon_i^s|}{|\varepsilon_i^d|}\right),\ \tilde{c}_{i0} = c_{i0}\left(1 + \frac{|\varepsilon_0^s|}{|\varepsilon_0^d|}\right) / \left(1 + \frac{|\varepsilon_i^s|}{|\varepsilon_i^d|}\right),$$

$$\beta_i = \delta_i \frac{\left(1+|\varepsilon_i^s|/\left|\varepsilon_i^d\right|\right)}{(1+|\varepsilon_0^s|/|\varepsilon_0^d|)} \text{ and } \gamma_i = \frac{\rho_i}{(1+|\varepsilon_0^s|/|\varepsilon_0^d|)}.$$

For the leading area, $\tilde{a}_0 = \frac{a_0}{(1+|\varepsilon_0^s|/|\varepsilon_0^d|)}$, and $\tilde{b}_{01} = \frac{b_{01}}{(1+|\varepsilon_0^s|/|\varepsilon_0^d)}$.

Compare the above specification with the price diffusion model in HPY:

$$\Delta\, p_{0t} = \phi_{0s}\left(p_{0,t-1} - \bar{p}\,^{s}_{0,t-1}\right) + a_0 + a_{01}\Delta p_{0,t-1} + b_{01}\Delta\bar{p}\,^{s}_{0,t-1} + \varepsilon_{0t},$$

$$\Delta\, p_{it} = \phi_{is}\left(p_{i,t-1} - \bar{p}\,^{s}_{i,t-1}\right) + \phi_{i0}\left(p_{i,t-1} - p_{0,t-1}\right) + a_i + a_{i1}\Delta p_{i,t-1}$$
$$+ b_{i1}\Delta\bar{p}\,^{s}_{i,t-1} + c_{io}\,\Delta p_{ot} + \varepsilon_{it}.$$

There are several differences. First of all, this paper tries to model the cumulative demand shocks and argues that the modelling of HPY is implicitly built on the modelling of the demand shocks. As the above derivation shows, starting from the error correction model of the cumulative demand shocks, we get the error correction model of the log of housing prices. Secondly, in our derived error correction model of housing prices, convergence is not necessary. Log of real housing prices could diverge across locations. The error correction term $\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right)$ allows for different trending pattern and cointegrating vector other than (1, -1). Of course, the form of the cointegrating relation is an empirical issue. As shown later, in the metro areas of California, some areas' HPIs have quite different trending pattern and most of them have a cointegrating vectors other than (1, -1).[5] Thirdly, coefficients in the derived price diffusion model contain useful information about local supply elasticities. As shown in the expressions for the error correction coefficient $\tilde{\phi}_{i0}$ and the coefficient on the contemporaneous effect of the leading area $\tilde{c}_{i0}$, areas with more elastic housing supply (larger $|\varepsilon_i^s|$) will adjust prices to a less extent than areas with more inelastic housing supply in response to a common demand shock. As verified in the empirical exercise, the leading area's contemporaneous effect $\tilde{c}_{i0}$ indeed has a negative relation with supply elasticities estimates from Saiz (2010).

Starting with the error correction model of the demand shocks, we can also derive an error correction model for ΔlnQ_{it}. However, the measurement of the change in housing stock involves a frequency issue. Also new construction exhibits very obvious seasonality patterns. Hence the signal embedded in a change in the housing stock involves lots of irrelevant noise. In order to study the demand side shocks embedded in the quantity response, we can use the signal imbedded in prices to study the diffusion of the demand shocks onto housing stocks. Substituting the relevant expressions on the right-hand side (RHS) of the error correction model of the demand shocks with expressions of housing prices, and substituting the dependent variable with the expressions for the housing stocks,[6] one can get the following equations after normalizing the coefficients on the LHS variables:

$$\Delta lnQ_{0t} = \hat{a}_0 + \hat{a}_{01}\Delta p_{0,t-1} + \hat{b}_{01}\Delta\bar{p}\,^s_{0,t-1} + \hat{\varepsilon}_{0t},$$

$$\Delta\, lnQ_{it} = \hat{\phi}_{i0}\big(\, p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\big) + \hat{a}_i + \hat{a}_{i1}\Delta\, p_{it-1} + \hat{b}_{i1}\Delta\bar{p}\,^s_{i,t-1} + \hat{c}_{io}\,\Delta p_{0t} + \hat{\varepsilon}_{it}.$$

[5] The error correction term involving spatial averages also possess a coefficient different from 1. The error correction term $d_{i,t-1} - \omega_i \bar{d}\,^s_{i,t-1}$ indicates that each metro area's cumulative demand shocks shares a common trend with its neighbor's cumulative demand shocks, with a cointegrating coefficient given by $(1, -\omega_i)$. As verified later, this cointegrating relation among cumulative demand shocks implies a similar cointegrating relation among the log of housing prices. Thus if we include the error correction term in the log of housing prices, the error correction term will take the form $p_{i,t-1} - \widetilde{\omega}_i \bar{p}\,^s_{i,t-1}$.

[6] For the dominant area, $\big(1/|\varepsilon_0^s| + 1/|\varepsilon_0^d|\big)\Delta lnQ_{0t} = a_0 + a_{01}\big(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\big)\Delta p_{0,t-1} + b_{01}\Delta\bar{d}\,^s_{0,t-1} + \varepsilon_{0t}$. For the remaining areas, $\big(1/|\varepsilon_i^s| + 1/|\varepsilon_i^d|\big)\Delta lnQ_{it} = \phi_{i0}\big(\,\big(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\big)p_{0,t-1} - \delta_i\big(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\big)p_{i,t-1} - \rho_i t\big) + a_i + a_{i1}\big(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\big)\Delta\, p_{it-1} + b_{i1}\Delta\bar{d}\,^s_{i,t-1} + c_{io}\,\big(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\big)\Delta p_{0t} + \varepsilon_{it}$.

We call the above simplified model the construction diffusion model. For the leading area 0, $\hat{a}_{01} = a_{01}|\varepsilon_0^s|$, and $\hat{a}_0 = a_0/\left(1/|\varepsilon_0^s| + 1/|\varepsilon_0^d|\right)$. For location i, the coefficient on the error correction term $\hat{\phi}_{i0}$ and the coefficient on the contemporaneous effect of the leading area $\hat{c}_{io}$ are given by

$$\hat{\phi}_{i0} = \phi_{i0}\frac{(1+|\varepsilon_0^s|/|\varepsilon_0^d|)}{(1/|\varepsilon_i^s|+1/|\varepsilon_i^d|)} = \tilde{\phi}_{i0}|\varepsilon_i^s|, \text{ and } \hat{c}_{io} = c_{i0}\frac{(1+|\varepsilon_0^s|/|\varepsilon_0^d|)}{(1/|\varepsilon_i^s|+1/|\varepsilon_i^d|)} = \tilde{c}_{i0}|\varepsilon_i^s|.$$

And $\hat{a}_{i1} = a_{i1}\frac{\left(1+|\varepsilon_i^s|/\left|\varepsilon_i^d\right|\right)}{(1/|\varepsilon_i^s|+1/|\varepsilon_i^d|)}$=$a_{i1}|\varepsilon_i^s|$, and $\hat{a}_i = \frac{a_i}{(1/|\varepsilon_i^s|+1/|\varepsilon_i^d|)}$.

Expressing the change in the housing stock as a function of the price error correction terms and price changes, we could examine the different diffusion patterns of the demand shocks. As demand side shocks originated from a leading area diffuse across surrounding areas, responses of prices and quantities could be very different. All of these responses hinge on the relative price elasticities of supply and demand with respect to the leading area. For areas with smaller price elasticities of supply of housing and with more restrictive regulations on construction, we would expect a larger contemporaneous response in price movements (measured by $\tilde{c}_{i0}$) and a smaller contemporaneous response in construction activities (measured by $\hat{c}_{io}$). On the contrary, for places with more open land and less restrictive zonings on construction, we would expect more construction activity other than price movements.

Given the above argument, this paper is able to study the two sides of the housing markets, i.e., price and quantity. By studying price movements as well as construction activity, we are able to capture a more complete picture of the diffusion of the demand side shocks. As we will see later in the empirical evidence, indeed, quantity response behaves quite differently from price response.

2.2 **Panel time series model**

Estimation of the demand shock diffusion model could be done for each individual location using time series analysis. Given the short time period for the construction permits data described below, the OLS estimation of the construction diffusion model is not reliable. Hence, we resort to panel estimations for the analysis of the construction diffusion model.[7] When using panel regression to analyze the diffusion model, we first assume that housing market shocks have the same diffusion pattern, and then divide locations under consideration into two groups based on price elasticities of housing supply from Saiz (2010) and estimate separate panel regressions for each group.

Under the assumption that all the following locations within a group have the same diffusion pattern, we can write the following panel data model for metro areas' log house prices:

$$\Delta\, p_{it} = \tilde{\phi}_s(p_{i,t-1} - \widetilde{\omega}_i \bar{p}^{\,s}_{i,t-1}) + \tilde{\phi}_0\big(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\big) + \tilde{a}_i + a_1 \Delta\, p_{it-1} + \tilde{b}_1 \Delta \bar{p}^{\,s}_{i,t-1} + \tilde{c}_o\, \Delta p_{0t} + \tilde{\varepsilon}_{it}, \quad i = 1,2,\dots,N; t = 1,2,\dots,T.$$

And for the construction diffusion model, we have the following panel data regression model:

$$\Delta\, lnQ_{it} = \hat{\phi}_s\big(p_{i,t-1} - \widetilde{\omega}_i \bar{p}^{\,s}_{i,t-1}\big) + \hat{\phi}_0\big(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\big) + \hat{a}_i + \hat{a}_{i1} \Delta\, p_{it-1} + \hat{b}_{i1} \Delta \bar{p}^{\,s}_{i,t-1} + \hat{c}_{io}\, \Delta p_{0t} + \hat{\varepsilon}_{it}, \quad i = 1,2,\dots,N; t = 1,2,\dots,T.$$

In the above two specifications, the diffusion coefficients are the same across locations except the area fixed effects $\tilde{a}_i$ and $\hat{a}_i$. Notice that in both of the panel regressions, we exclude the dominant area and focus on the diffusion analysis of the following areas. The error correction terms, $p_{i,t-1} - \widetilde{\omega}_i \bar{p}^{\,s}_{i,t-1}$, and $p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t$, are estimated from the bivariate VAR(4) models of each location's house price and its neighbor's local averages,

[7] Applying panel data techniques to the housing market dynamics, we should pay special attention to the heterogeneity and cross sectional dependence issues, since housing markets are quite localized. In the individual OLS estimation of the diffusion models, cross sectional dependence has been taken into account by the inclusion of spatial averages of the neighbors' shocks, and heterogeneity is assumed automatically since each individual location has its own regression equation.

and the bivariate VAR(4) models of each location's house price and the dominant area's house price, respectively.

2.3 Spatial-Temporal Impulse Response Functions

Starting from the price diffusion model, HPY rewrite the system of equations into a vector autoregression model (VAR), in which some coefficient matrices reflect temporal dependence of house prices while other matrices reflect spatial dependence. Based on the estimates of these coefficient matrices, the VAR model can be used for forecasting or impulse response analysis. This approach involves inverting an $(N+1)\times(N+1)$ matrix. Hence, this impulse response analysis is computationally intensive for large N. Moreover, it cannot generate the impulse response analysis for the construction diffusion model since the dependent variable is different from the explanatory variables. Instead of following HPY's impulse response analysis, this paper uses Jordà's (2005) location projection method, which allows one to estimate the dynamics of regional housing prices as well as construction controlling for spatial correlation across regions. As shown in Jordà (2005), the impulse response function for an individual variable in a vector of endogenous variables can be estimated consistently from a regression of this variable on the lags in the system for each horizon, h. (See Jordà (2005) for a complete explanation of the local projection method and Jordà (2007) and Jordà and Kozicki (2007) for additional explanation).

By Jordà's (2005) location projection method, for the housing price diffusion model, the impulse responses of a unit shock to house prices in the dominant area on the following area $i = 1,2,\ldots,N,$ at horizon h periods ahead is given by $\tilde{c}_{io}^{h}$ in the following equation:

$$\Delta p_{it+h} = \tilde{\phi}_{is}^{h}\left(p_{i,t-1} - \widetilde{\omega}_i \bar{p}_{i,t-1}^{\,s}\right) + \tilde{\phi}_{i0}^{h}\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right) + \tilde{a}_i^{h}$$
$$+ a_{i1}^{h}\Delta p_{it-1} + \tilde{b}_{i1}^{h}\Delta \bar{p}_{i,t-1}^{\,s} + \tilde{c}_{io}^{h}\Delta p_{0t} + \tilde{\varepsilon}_{it+h}.$$

For the construction diffusion model, the impulse responses of a unit shock to house prices in the dominant area on the following area i at horizon h periods is given by $\hat{c}_{io}^{h}$ in the following equation:

$$\Delta \, lnQ_{it+h} = \hat{\phi}_{is}^{h}(p_{i,t-1} - \widetilde{\omega}_i \bar{p}_{i,t-1}^{\;s}) + \hat{\phi}_{i0}^{h}(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t) + \hat{a}_i^{h} +$$
$$\hat{a}_{i1}^{h} \Delta \, p_{it-1} + \hat{b}_{i1}^{h} \Delta \bar{p}_{i,t-1}^{\;s} + \hat{c}_{io}^{h} \, \Delta p_{0t} + \hat{\varepsilon}_{it+h}.$$

The impulse responses analysis for the panel time series model can be derived similarly.

3. Empirical Results

3.1 Metro areas and the leading area

We apply the methodology described in HPY to quarterly All-Transactions (Estimated using Sales Prices and Appraisal Data) House Price Index (hereafter HPI) from the Federal Housing Finance Agency (FHFA) for Metropolitan areas in California.[8] The nominal HPI series are deflated using Consumer Price Index-All Urban Consumers: Less Shelter for US.[9] Definitions of Metropolitan areas are based on the Office of Management and Budget (OMB) 2013 delineations. Since there are more missing observations for HPI for smaller metropolitan areas, this paper selects metropolitan areas with population larger than 250,000 (based on 2010 Census Population and Housing Tables). The final housing price data include quarterly All-Transactions HPI series of FHFA over the period 1980Q1-2016Q4 for 22 metro areas listed in Table 1.

To construct the variable ΔlnQ_{it}, notice that

$$\Delta lnQ_{it} = lnQ_{it} - lnQ_{i,t-1} = ln(Q_{i,t-1} + \Delta Q_{it}) - lnQ_{i,t-1} \approx \frac{\Delta Q_{it}}{Q_{i,t-1}}.$$

The approximation of the above equation is valid since the change in housing stock is quite small relative to the existing housing stock. In this paper, we use housing permits as our

[8] The FHFA HPI series can be downloaded at the following URL: https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#qat.
[9] Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/CUUR0000SA0L2.

measure of new quantity ΔQ_{it}, and we scale up ΔlnQ_{it} by 100. In other words, we use the percent change (100* ΔlnQ_{it}) in the housing stock as our measure of the quantity response of demand shocks. Monthly county level permits data are obtained from the SOCDS Building Permits Database of U.S. Department of Housing and Urban Development.[10] The county level permits data cover the period 1997Q1-2016Q4. We aggregate across counties and months to create quarterly metropolitan area level aggregates using the 2013 definitions provided by the census. County housing stock estimates are from the Census 2000 housing units counts.[11] We first aggregate across counties to create metropolitan area level housing units counts in 2000. To form quarterly estimates of housing units counts for quarters after 2000, we add cumulative building permits for total units from 2000 on to the 2000 housing units counts. Similarly, to from quarterly estimates of housing units counts for quarters before 2000, we subtract the reverse cumulative building permits for total units from 2000 backwards from the 2000 housing units counts.

HPY pick London as the leader for the argument that London is the largest city in Europe but more significantly is a major world financial center. As the largest places for economic activity, it is highly possible that economic shocks will first arrive at London and then propagate out to the surrounding regions in UK. In this paper, we find that it is not necessary that the largest metropolitan area lead other areas in the housing market. In terms of the 2010 Census population, Los Angeles-Long Beach-Glendale (hereafter LA), is the most populous area. However, in testing for cointegration among housing prices, only 5 out of 21 areas show a significant cointegration relation with LA at the 5% significance level. In contrast, 20 out of 21 areas show significant cointegration relation with San Jose-Sunnyvale-Santa Clara (hereafter San Jose) at the 5% significance level. Theoretically, if house prices of

[10] The building permit database contains data on permits for residential construction issued by about 21,000 jurisdictions collected in the Census Bureau's Building Permits Survey. (https://socds.huduser.gov/permits/summary.odb)
[11] The Census 2000 housing units counts are available at American FactFinder website https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF1_H001&prodType=table.

all other areas are cointegrated with a leading area's house price, we should expect that any pair of locations' house prices are cointegrated. However, because of the finite sample properties of the cointegration rank test (hereafter CI), the pairwise CI tests indicate quite different cointegration patterns when choosing different leading areas. By the CI test based on the bivariate vector error correction model, choosing San Jose as a leading area yields the most meaningful results. Hence, in this paper, we pick our leading area as the one that shows the most cointegrated relations with other areas and further confirm the exogeneity of the leading area's price shocks using the Wu-Hausman test later on in the estimation of the price diffusion model.

3.2 Convergence of house price indexes in California

The logarithm of real HPI and their quarterly rates of change cross the 22 regions are displayed in Figure 2. There is a clear upward trend for most of California metro areas over the 1975-2016 period, with prices in San Francisco and San Jose rising faster than other metro areas. Even though all of these metro areas' HPI indices move downward or upward together most of the time, there are obvious diverging behaviors in these HPI indices for the post-2006 period. As all of these metro areas' housing market recover from the crisis, there are persistent gaps in the HPI indices, and it seems that these gaps will continue to exist for a while.

Using San Jose as the dominant region, in the left panel of Table 2, we present trace statistics for testing cointegration between San Jose and metro area i house price indexes, computed based on a bivariate VAR(4) specification in p_{0t} and p_{it} for $i = 1,2,\dots,21$. The null hypothesis that the log of real house price index in San Jose is not cointegrated with that in other metro areas is rejected at the 10% significance level or less in all cases. As stated in HPY, cointegration whilst necessary for long-run convergence of house prices is not sufficient. We further test for the cotrending and the cointegrating vector corresponding to (p_{it}, p_{ot}) is (1, -1). The joint hypothesis that p_{it} and p_{0t} are cotrending and their cointegrating vector can be represented by (1, -1) is tested using the log-likelihood ratio statistic with an asymptotic chi-squared distribution with degree of freedom 2. In this paper, we follow the algorithm of Cavaliere, Nielsen, and Rahbek (2015) to calculate the 95% and 90% bootstrapped critical values of the joint test, in which the null hypothesis is imposed on the bootstrap sample. Cavaliere, Nielsen, and Rahbek (2015) show that the bootstrap test constructed this way is asymptotically valid and it outperforms other existing methods.[12]

As shown in the right panel of Table 2, the null of the joint test under consideration is rejected at the 10% level for all the cases, except San Francisco and Visalia. Most of these rejections are not marginal. For 11 out of these 21 metro areas, the null is rejected at the 5% significance level. Thus, it seems that in California, HPI for metro areas are not converging in the long-run. To understand the divergence of HPI in California, we run two separate marginal tests for the cotrending hypothesis and for the CI vector being (1, -1), based on a bivariate VAR(4) with unrestricted intercepts and restricted trend coefficients using the log-likelihood ratio statistic. These two individual test statistics have a χ_1^2 limiting distribution. Again, the critical values are based on the bootstrapping algorithm of Cavaliere, Nielsen, and Rahbek (2015).

As shown in the left panel of Table 3, the null hypothesis of cotrending is rejected at the 5% level for 8 metro areas, including Anaheim, LA, Salinas, San Diego, San Luis Obispo, San Rafael, Santa Cruz, and Santa Rosa. For the test of cointegrating vector being (1, -1) with the leader's HPI based on the bivariate VAR(4) model with unrestricted intercepts and

[12] It is well known that the finite-sample properties of tests of hypotheses on the cointegrating vectors in vector autoregressive models can be quite poor, and that current solutions based on Bartlett-type corrections or bootstrap based on unrestricted parameter estimators are unsatisfactory, in particular in those cases where also asymptotic χ_2^2 tests fail most severely.

restricted trend (middle panel of Table 3), the null is rejected at the 10% level or less for the same set of 8 areas for which the cotrending hypothesis is rejected.

One can conclude that except for these 8 metro areas, other metro areas in CA show evidence of long-run convergence of log of real HPI with log of San Jose's real HPI. However, it should be pointed out that the base VAR model for those who do share a common trend with the leading area is misspecified. The testing of CI vector being (1, -1) for these area sharing a common trend with the leading area should be based on a bivariate VAR(4) model with unrestricted intercepts only. Thus we run another log-likelihood test of the CI vector being (1, -1), based on the bivariate VAR(4) model with unrestricted intercepts and restricted trend coefficients if the cotrending test is rejected, otherwise based on a bivariate VAR(4) with unrestricted intercepts. The last three columns of Table 3 show the test results. Again, the critical values are based on the bootstrapping algorithm of Cavaliere, Nielsen, and Rahbek (2015). For all of these 8 areas for which the cotrending test with San Jose's HPI is rejected, the null hypothesis that log of real HPI of these areas is cointegrated with that of San Jose with CI vector (1, -1) is rejected at the 10% level or less. For the remaining 13 areas that show cotrending evidence with San Jose, the null of CI vector being (1, -1) is rejected for 9 of them. In total, the null of the CI vector being (1, -1) is rejected for 17 metro areas in CA.

From the above testing of over-identifying restrictions in bivariate VAR(4) models, there is little evidence that the HPIs of metro areas in CA are converging in the long run. Even though the HPIs of these metro areas are co-integrated with that of San Jose, 8 of them tend to have different trending patterns than San Jose, and almost all of them have quite different cointegrating coefficients than (1, -1). [13]

[13] We also study the long run converging relation between each metro area's log of HPI and the log of HPI of the local average of its neighbors. The empirical results show that p_{it} share a common linear trend with $\bar{p}\,_{it}^{s}$, but the cointegrating vector differs from (1,-1).

3.3 Price elasticity of supply of housing and convergence of house prices

We estimate error correction coefficients of log of real HPI of San Jose and other CA metro areas in a cointegrating bivariate VAR(4) with unrestricted intercepts and restricted trend coefficients if the cotrending test is rejected. Otherwise, the error correction term is estimated based on a bivariate VAR(4) with unrestricted intercepts. From the simple demand shock model, we find that the cointegration coefficient depends on the relative magnitude of the supply elasticities through $\beta_i = \delta_i\left(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\right)/\left(1 + |\varepsilon_0^s|/|\varepsilon_0^d|\right)$, which suggests a positive relation between the CI coefficient and the price elasticity of supply of housing.

The primary measure of supply side conditions is taken from Saiz (2010), as shown in the third column of Table 1. Such supply elasticity estimates are simple nonlinear combinations of the available data on physical and regulatory constraints, and predetermined population levels in 2000. Because the definitions of metro area differ in this paper, only 19 metro areas (18 following areas and 1 leading area) have the supply elasticity measures. To test the empirical application of the CI coefficient, we run the following regression using the supply elasticity estimates from Saiz (2010):

$\beta_i = c + b * \varepsilon_i^s + v_i$, for i=1, 2,…,N.

Excluding Bakersfield for which the CI coefficient (17.13) is an outlier, we are left with 17 following metro areas (with such small sample size, standard errors are from bootstrapping with 1000 replications). As shown in Table 4, the first column shows the regression result of the CI coefficient on the estimated elasticities of Saiz (2010). The coefficient on the estimated price elasticity of supply is positive and significant at the 1% significance level. The positive significant coefficient on elasticity verifies the positive relation between the CI coefficient and the price elasticity of supply, which is further depicted in Figure 3.

In the second regression of Table 4, the explanatory variable is the share of unavailable land for development (unaval). The results show that the higher the share of unavailable land, the smaller the CI coefficient. From the logic that for severely land-constrained places housing supply is highly inelastic as in Saiz (2010), this negative and significant coefficient on the share of unavailable land is consistent with the derivation that the CI coefficient is positively correlated with the supply elasticity of housing. However, we find little evidence of a significant correlation between the CI coefficient and the WRLURI index (a measure of the strictness of the local regulatory environment based on results from a 2005 survey of over 2000 localities across the country from Gyourko, Saiz and Summers, 2008). Also, the population size in 2000 and the percent change in population from 2000 to 2010 show little impact on the CI coefficient.

3.4 **Estimates of house price diffusion models**

The regression results for the price diffusion model in which San Jose acts as the dominant metro area are summarized in Table 5. Estimates of the error correction coefficients, $\tilde{\phi}_{i0}$ and $\tilde{\phi}_{is}$, are provided in columns 2 and 3 of Table 5. The estimates, $\tilde{\phi}_{i0}$, the coefficient on the error correction term $\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right)$, captures the effect of deviations of area i's log of HPI from that of San Jose, and $\tilde{\phi}_{is}$ is associated with $\left(p_{i,t-1} - \omega_i \bar{p}^{\,s}_{i,t-1}\right)$, which measures the effect of deviations of area i's log of HPI from its neighbors.

For the error correction term measured relative to San Jose, we find that it is only statistically significant in five coastal areas (San Francisco, San Luis Obispo, San Rafael, Santa Cruz, and Santa Rosa). In other words, only these five coastal metro areas show significant adjustments to price deviations from the dominant region's price level. The error correction term measured relative to neighboring areas is statistically significant in seven areas (San Jose, Merced, Sacramento, Salinas, San Diego, Stockton, and Vallejo).

The remaining 10 areas, with none of these two error correction terms significant, include the Los Angeles-Long Beach Combined Statistical Area (composed of LA, Anaheim, Riverside, and Oxnard), the Fresno-Madera Combined Statistical Area (composed of Fresno and Madera), Bakersfield, Modesto, Oakland, and Santa Maria. The non-significance of these two error correction terms for these 10 areas are hard to explain. As stated in HPY, this insignificance may be due to the fact that the sample period might not be sufficiently informative in this regard, or these areas might have different error correcting properties that the parsimonious specification can fully take into account.

Next let us turn to the short-term dynamics and spatial effects. As in HPY, we report the sum of lagged coefficients, with the associated t-ratios provided in brackets (by the delta method). Different from HPY, the own lag effects in this paper are quite significant with moderate magnitudes for most of the areas, excluding only five areas, namely, Riverside, San Diego, Stockton, Vallejo, and Visalia. Likewise, the lagged HPI changes from neighboring areas are statistically significant for most of the areas, with the exception of San Jose, Anaheim, LA, San Francisco, and Santa Maria. This significant evidence of the own lag effects and of the lagged neighbors' HPI changes, clearly highlight the importance of dynamic spill-over effects from the neighboring areas as well as the persistence of the housing prices movements.

The contemporaneous effect of San Jose HPI are sizeable and statistically significant in all areas. There is no clear relation between the size of this contemporaneous effect and the commuting distance of the area to San Jose. For most of the areas considered, the coefficients on the San Jose lag effects offset a significant part of the San Jose contemporaneous effects. We combine the San Jose contemporaneous effects and the lagged San Jose effects for each area by summing the two estimates. Still we find no clear relation between the size of the combined coefficients and the commuting distance to San Jose. In Figure 4, we plot the sum of the contemporaneous effect and lag effect of the leader's HPI on each metro areas against the supply elasticity estimates from Saiz (2010). As shown in the figure, metro areas with more inelastic housing supply will be affected by the leader's house price changes to a larger extent than areas with more elastic housing supply. This negative relation between the supply elasticity and the combined coefficient on leader's price changes is consistent with the derivation of $\tilde{c}_{i0} = c_{i0}\left(1 + |\varepsilon_0^s|/\left|\varepsilon_0^d\right|\right)/\left(1 + |\varepsilon_i^s|/|\varepsilon_i^d|\right)$ from the diffusion model of the demand shock.

The Wu-Hausman statistics, which test the hypothesis that HPI changes in San Jose are exogenous to the evolution of house prices in other areas, show that the null cannot be rejected for all of the metro areas at the 1% significance level. Only for Oakland and Santa Cruz, the null is rejected at the 5% level, and for Stockon the null is rejected at the 10% significance level. By the Wu-Hausman test results, we verify the assumption that housing price changes in San Jose are exogenous to all other metro areas' price changes and hence confirm the assertion that San Jose leads the housing markets in all of the metro areas of CA.[14]

3.5 **Panel model estimation**

In this section, we pool all of the individual estimations into panel regressions with metro area fixed effects[15], and use quarterly data, annual data, and then biannual data to explore how the frequency of the data affects the demand shock diffusion patterns. In order to allow for heterogeneous diffusion patterns implied by varying local supply side conditions, we run these panel regressions for three groups of metro areas. The first group consists of all of the 21 following metro areas, and the second group includes 6 metro areas with the most

[14] We also study the time series estimation of each individual metro area's construction diffusion model. Because of the small sample size (76), the results are quite noisy and there is no clear pattern on the effect of the leader's contemporaneous price changes and the construction adjustments to short-run price deviations from its long-run equilibrium level. We resort to the panel analysis of the construction diffusion model taking advantage of more estimation power.

[15] In the panel regression, we set all of the lag orders to the maximum number 4 and include both of these two error correction terms whether they are significant or not.

inelastic housing supply (LA, Oakland, Oxnard, San Diego, San Francisco, and Santa Maria). The last group is made up of the remaining 15 metro areas with more elastic housing supply.

The estimation results for the price diffusion model are summarized in Table 6. The first three columns use the quarterly housing price data, with the first column for all of the 21 metro areas, and the second for the 6 least elastic areas, and the third for the remaining 15 relative elastic areas. From the first panel regression (Column 1 of Table 6), we find sizable and significant leader contemporaneous effect (0.74 with standard error 0.022), and this estimate is comparable to that from the individual time series estimates. This difference in the estimates for these two groups of areas with different supply elasticities are not significant.

The coefficients on these two error correction terms are significant with the correct signs, indicating housing prices in following areas will adjust upwards if they are below their long-run equilibrium with the dominant area's house price or with their neighbors' house prices. The error correction coefficients differ substantially for these two groups of areas. For the inelastic metro areas, the coefficient on EC1 is 0.014, compared to 0.00088 for the elastic metro areas. This result indicates that metro areas with more inelastic supply conditions will adjust prices faster to any deviation from their long-run equilibrium with the dominant area's price level. The coefficient on EC2 is only significant in the elastic metro group, indicating that only elastic areas' housing prices will respond to short-run deviation from its long-run equilibrium with its neighbors' housing prices.

The leader lag effects are significant to the 4^{th} lag in the full sample and are similar in magnitude for these two groups with different supply side conditions. Neighbor lag effects are also significant, with a slightly larger magnitude for the elastic group. Own lag effects are also significant, with similar magnitudes for both groups. Again, in the panel analysis, we see that dynamic spillover effects from the neighboring areas are important in the diffusion analysis through the error correction terms as well as through the spatial lag terms, and it is more important for areas with more unrestricted housing supply conditions.

Comparing the estimates of the first three columns with those of the middle three columns of Table 6, we can see how the frequency affects the diffusion patterns of price shocks. [16] As we change from quarterly data to annual data, the error correction coefficients are significantly larger. This indicates that in a longer time horizon, local housing markets will adjust more thoroughly to the short-run price deviations from their long-run equilibrium. The difference in the leader contemporaneous effect between the inelastic group and the elastic group is still not statistically significant.

The last three columns of Table 6 show the results using the biannual data. Again, the error correction coefficients become even larger with only EC1 significant. These two error correction terms are not significant for the inelastic group. This is consistent with the interpretation that the error correction coefficients measure the adjustment speeds of the prices to their short-run deviation from their long-run equilibrium. As the data frequency become lower, i.e., a longer time gap between observations, we may not be able to estimate the short-run adjustment speeds. However, under the biannual estimation, the difference in the leader contemporaneous effect between the inelastic group and the elastic group is much larger and significant (0.80 with standard error 0.032 for inelastic and 0.59 with standard error 0.029 for elastic).

To summarize for the panel regression of the price diffusion model, San Jose's contemporaneous effects are sizable and significant, and tend to be larger in metro areas with more inelastic housing supply conditions. We also find strong evidence on price responses to price deviations from their long-run equilibrium, with inelastic places adjusting prices faster to the deviations from the dominant area's price level. Moreover, there exist significant

[16] Notice that in the annual and biannual regression, we only include 2 lag terms to save on observations lost due to lagging.

spillover effects and own lag effects in the diffusion of price shocks. Also, longer horizons allow for more thorough adjustments to price deviations.

The estimation results for the construction diffusion model are summarized in Table 7. From the first panel regression (Column 1 of Table 7), we find sizable and significant leader contemporaneous effect (0.44 with standard error 0.21). This estimate seems to be larger for the group of metro areas with elastic housing supply (0.57 with standard error 0.28 in Column 3 of Table 7) than that for the group with inelastic housing supply (0.39 with standard error 0.22 in Column 2 of Table 7). However, this difference is not statistically significant. The own lag effects (coefficient on LD.lnHPI) is larger and more significant than the leader contemporaneous effects.

The coefficients on these two error correction terms are significant with opposite signs as in the price equations, indicating that following areas will depress construction if their housing prices are below their long-run equilibrium with the dominant area's house price or with their neighbors' house prices. The error correction coefficients differ substantially for these two groups. For the inelastic metro areas, the coefficient on EC1 is 0.02 with standard error 0.022 (not significantly different from 0), compared to -0.042 with standard error 0.0045 for the elastic metro areas. This result indicates that metro areas with more elastic supply conditions will adjust construction faster in response to any price deviation from their long-run equilibrium with the dominant area's price level. The coefficient on EC2 is -0.066 with standard error 0.033 for the inelastic group, compared to 0.058 with standard error 0.025 for the elastic group. The negative sign on EC2 for the inelastic group is hard to explain in that it implies that these inelastic areas will boost construction in the short-run even when their housing prices are below their long-run equilibrium with their neighbor's house prices.

The leader lag effects and the neighbor lag effects are not significant. In contrast, own lag effects are significant, with similar magnitudes for both groups. Thus, in the panel analysis of the construction diffusion model, we see that dynamic price spillover effects from the neighboring areas are important in the diffusion analysis only through the error correction terms.

As we change from quarterly to annual data, the coefficient on EC1 is significantly larger, indicating that in a longer time horizon, local housing markets will adjust construction more to the short-run price deviations from their long-run equilibrium. However, the coefficient on EC2 becomes insignificant for the annual regression. This insignificance of EC2 in the annual data indicates that short-run spillover effects from neighbors are only observed in higher frequency data. The difference in the leader contemporaneous effect between the inelastic group and the elastic group is larger (0.92 with standard error 0.47 for inelastic and 1.41 with standard error 0.59 for elastic) even though the difference is still not statistically significant.

The last three columns of Table 7 show the results using the biannual data. As the data frequency becomes lower, the short-run adjustments are not significant any more. The own lag effects also tend to be insignificant. However, under the biannual estimation, the leader contemporaneous effect become even larger, even though the difference between the inelastic group and the elastic group is still not significant (1.58 with standard error 0.2 for the whole group, 1.19 with standard error 0.25 for the inelastic group, and 0.59 with standard error 0.029 for the elastic group).

To summarize for the panel regression of the construction diffusion model, San Jose's contemporaneous effects are sizable and significant, and tend to be larger in lower frequency data. We find strong evidence on construction response to price deviations from their long-run equilibrium in higher frequency data. Elastic places tend to adjust construction faster and to a larger extent to the deviations from the dominant area's price level. Moreover, spillover effects on construction work only through the error correction terms, and own lag effect is larger than the leader contemporaneous effect in high frequency data but are dominated by the leader contemporaneous effect in low frequency data.

Comparing the panel regression of the price diffusion model and that of the construction diffusion model, we find quite different diffusion patterns of the demand shocks. The leader contemporaneous effects are sizable and significant in both estimations, but the difference in the leader contemporaneous effects for metro areas with different local supply conditions is more significant for the price diffusion model. Short-run adjustment of prices to price deviation from the equilibrium with the leader's price level are faster for the inelastic metro areas, while the elastic metro areas will adjust prices more rapidly to price deviation from the equilibrium with their neighbor's price level. In contrast, elastic areas will adjust construction faster to price deviation from the equilibrium with the leader's price level. Spillover effects of neighbors' demand shocks are transmitted in the price equations through the neighbor lag effect as well as the error correction term, while neighbors' spillover effects impact construction only through the error correction term. Own lag effects are more important in the construction diffusion model for higher frequency data, and leader contemporaneous effect become more important for the construction diffusion model for lower frequency data.

3.6 **Spatial-temporal impulse response**

In this section, we use the local projection method of Jordà (2005) outlined in Section 2.4 to study the impulse response of the effects of a positive unit shock to San Jose house prices over time and across space. We study not only the effects of a leader's price shock on the other areas' price changes, but also the effects of a leader's price shock on the other areas' construction response. We first apply the local projection method to the price and construction diffusion models for each individual metro areas, and then to the panel regressions for both diffusion models.

In Figure 5, we plot the impulse response of the effects of a positive unit shock to San Jose house price changes on the house price changes in other areas for the individual price diffusion estimations. The left panel of Figure 5 shows the effects of the shock on house price changes in 6 metro areas with the least elastic supply conditions (LA, Oakland, Oxnard, San Diego, San Francisco, and Santa Maria), whilst the right panel shows the effects on house price changes in the other 6 metro areas with the most elastic supply conditions (Bakersfield, Fresno, Merced, Modesto, Stockton, and Visalia). As we can see from these impulse response functions (hereafter IRFs) of the house price changes, the spontaneous responses are of the same magnitude for most of these 12 metro areas regardless of the supply side conditions and most of these IRFs go to zero in less than 5 quarters (except LA and Oakland). Thus, we do not find very significant differences in the transmissions of the leader's price shocks to areas with different supply conditions.

Figure 6 illustrates the IRFs of a positive unit shock to San Jose's price changes on price changes (left panel) and housing stock changes (right panel) in other areas estimated from the panel regressions. In each panel, "Elastic" stands for estimates from the panel regression with 15 metro areas with relative elastic housing supply conditions, and "Inelastic" stands from estimates from the panel regression with the 6 metro areas with the least elastic housing supply conditions, and "All MSA" stands for estimates from the panel regression with all of the 21 following areas. The left panel shows that IRFs of price changes do not exhibit significant differences between elastic areas and inelastic areas, and the responses of price changes to a unit shock of the leader's price changes decrease to zero gradually within 10 quarters. In contrast, the right panel indicates that elastic areas exhibit significant larger construction responses to a unit shock to the leader's price changes, and this response grows larger after 5 quarters and remain above zero for more than 18 quarters. For inelastic metro areas, these IRFs are not significantly different from zero.

To summarize the impulse response analysis, we find that a positive shock to San Jose house price changes spills over to other regions' price changes gradually regardless of the distance to San Jose and regardless of the supply side conditions. However, a positive San Jose's house price shock will have a significant and persistent effect on construction in metro areas with more elastic housing supply conditions.

4. Conclusion

This paper incorporates supply side conditions into the spatial and temporal dispersion of shocks in a non-stationary dynamic system. Using California metro area house prices we establish that San Jose is a dominant area in the sense of Pesaran and Chudik (2010). House prices within each metro area respond directly to a shock to San Jose and the overall effect of the dominant area's shock is negatively correlated with the local supply elasticities. Construction within each metro area also responds directly to a shock to San Jose, and the overall effect of the dominant area's shock is positively correlated with the local supply elasticities. Impulse response analysis indicates that the construction response is more persistent than the price response for metro areas with more elastic housing supply.

An important finding in this paper relative to Holly et al. (2011) is that local supply conditions have greater impact on the diffusion patterns of a common demand shock in the housing market than physical distance. When San Jose experiences a price shock, the effects on price and construction in other areas tend not to attenuate with distance to San Jose. On the other hand, impulse response functions (IRF) and other results indicate that local supply conditions have important impacts on the responses to shocks in the dominant area. These findings complement the cross sectional dependence literature and reinforce the view that local supply conditions may matter more than distance when modeling spatiotemporal dynamics in the housing market.

Acknowledgements

Chapter 2 is based on the working paper Baltagi, Rosenthal, and Shen (2017).

Figure 1: Housing market with demand side shocks

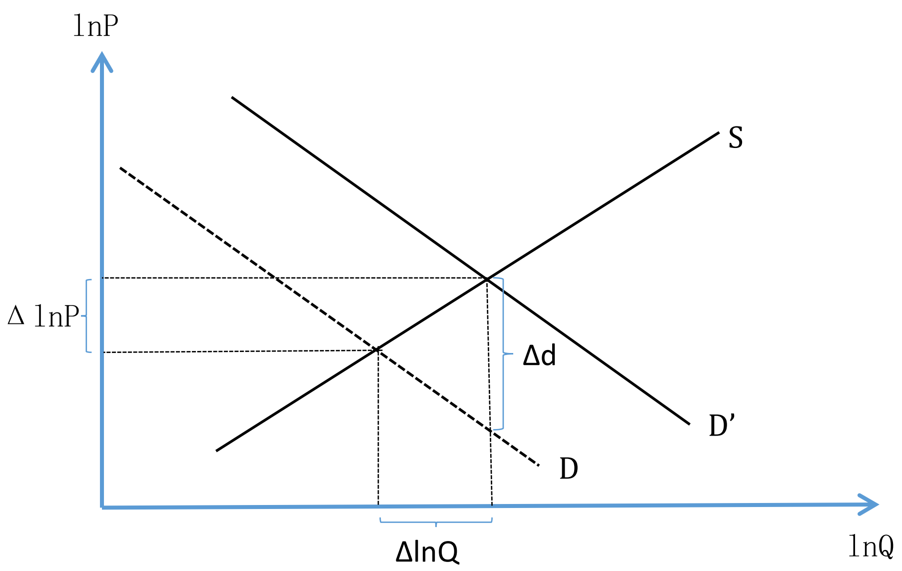


Figure 2: California (CA) Real House Price Indices by Metro Areas

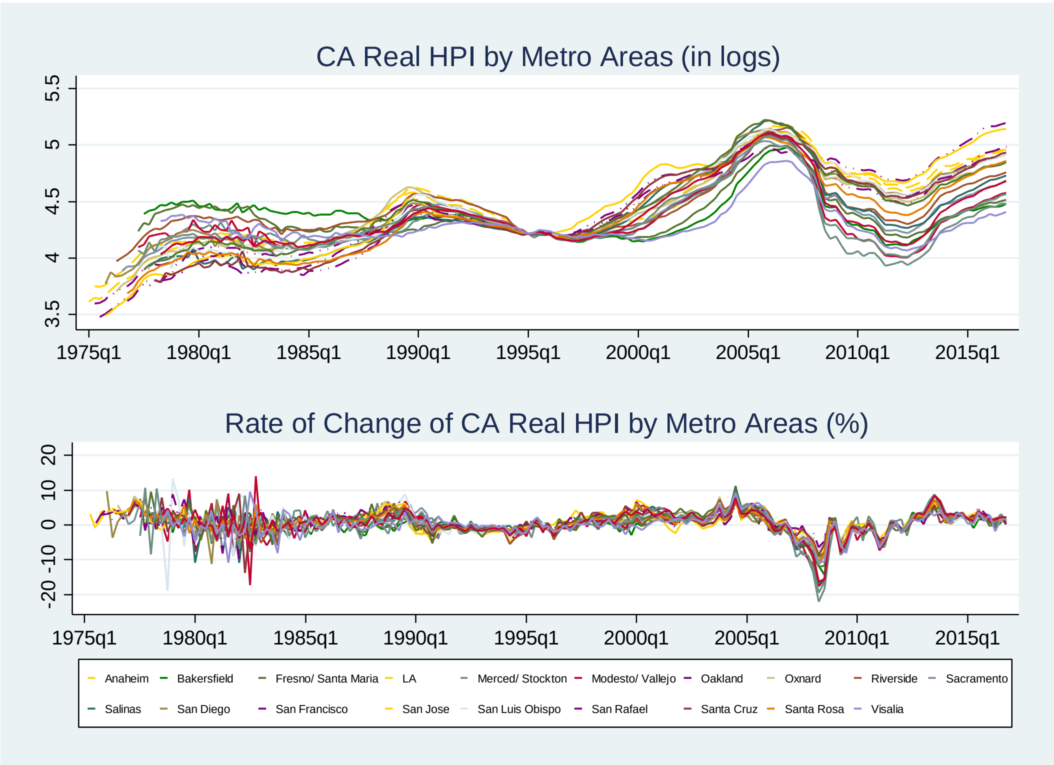
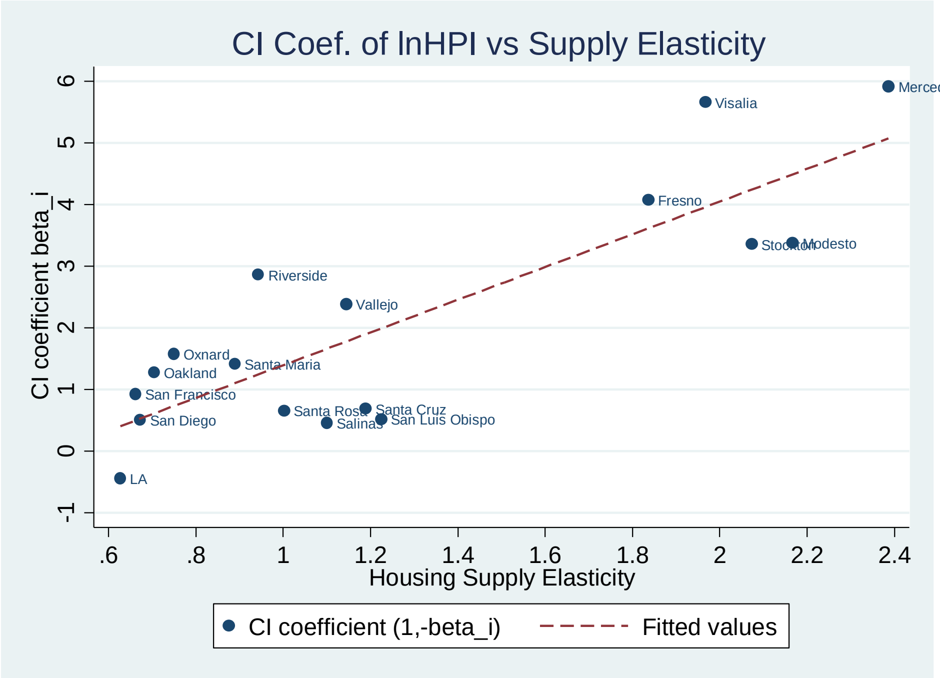
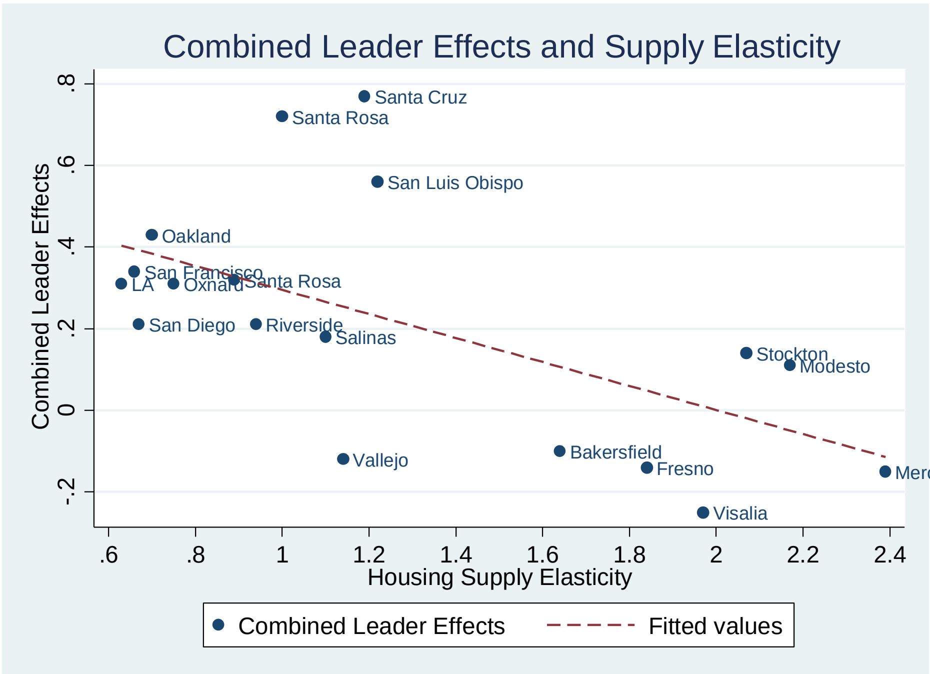


Figure 3: CI coefficients of log HPI versus supply elasticities



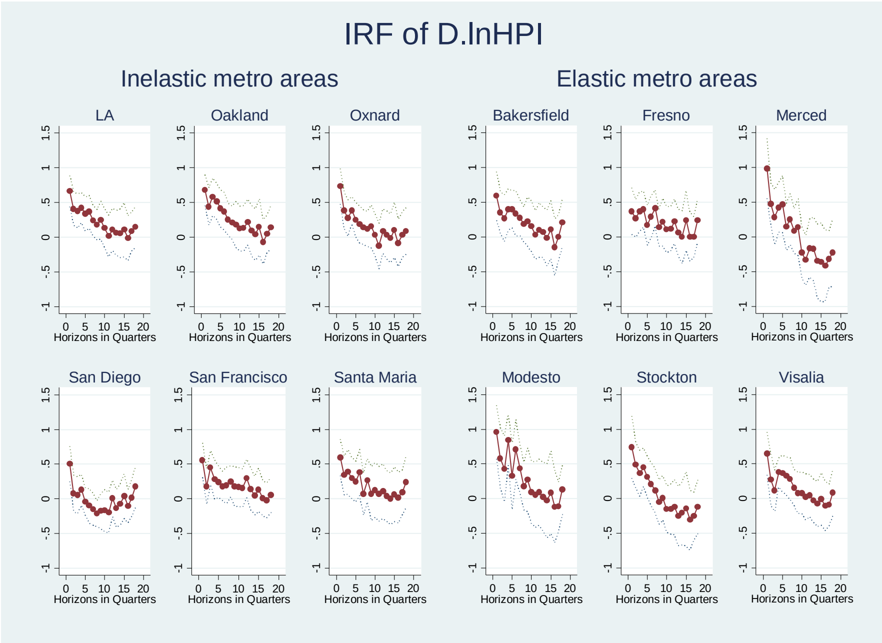
Notes: On the vertical axis is the CI coefficient β_i in the CI relation $p_{0t} - \beta_i p_{it}$, which is estimated from a bivariate VAR(4) specification of the log real HPI in San Jose (p_{0t}) and the other metro area (p_{it}) with unrestricted intercepts and restricted trend coefficients if rejecting the cotrending test, otherwise from a bivariate VAR(4) specification with a unrestricted intercepts only. On the horizontal axis is the price elasticities of housing supply ε_i^s from Saiz (2010). Each dot stands for a following area and the red dotted line stands for regression $\beta_i = c + b * \varepsilon_i^s + v_i$.

Figure 4: Combined leader effects versus supply elasticities



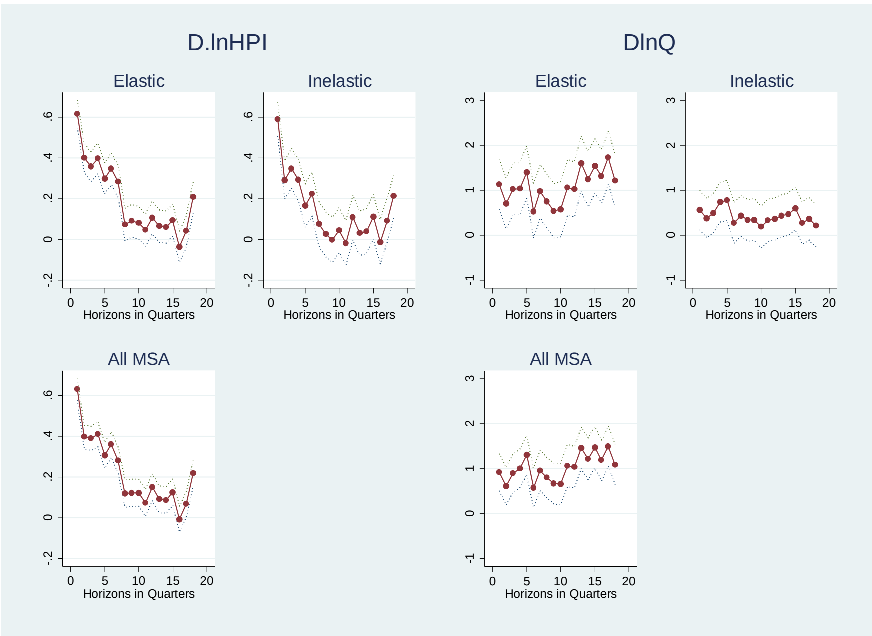
Notes: On the vertical axis is the sum of the contemporaneous and lag effect of the leader's HPI on each metro areas, i.e., $\sum_{l=0}^{k_{ic}} \tilde{c}_{il}$ from $\Delta p_{it} = \tilde{\phi}_{io}\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right) + \tilde{\phi}_{is}\left(p_{i,t-1} - \omega_i \bar{p}^{\,s}_{i,t-1}\right) + \tilde{a}_i + \sum_{l=1}^{k_{ia}} a_{il} \Delta p_{it-l} + \sum_{l=1}^{k_{ib}} \tilde{b}_{il} \Delta \bar{p}^{\,s}_{i,t-l} + \sum_{l=0}^{k_{ic}} \tilde{c}_{il}\, \Delta p_{0,t-l} + \tilde{\varepsilon}_{it},\ for\ i = 1,2,\ldots,N$. On the horizontal axis is the price elasticities of housing supply from Saiz (2010). Each dot stands for a following area.

Figure 5: Impulse Response Functions of one unit shock to San Jose house price changes over time from Individual OLS Regressions of the price diffusion model



Notes: The plotted IRFs are $\tilde{c}_{io}^h$ estimates in $\Delta p_{it+h} = \tilde{\phi}_{is}^h (p_{i,t-1} - \tilde{\omega}_i \bar{p}^{\,s}_{i,t-1}) + \tilde{\phi}_{i0}^h (p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t) + \tilde{a}_i^h + a_{i1}^h \Delta p_{it-1} + \tilde{b}_{i1}^h \Delta \bar{p}^{\,s}_{i,t-1} + \tilde{c}_{io}^h \Delta p_{0t} + \tilde{\varepsilon}_{it+h}$ for each horizon h. Each graph stands for an individual metro area.

Figure 6: Impulse Response Functions of one unit shock to San Jose house price changes over time from panel regression of the price and the construction diffusion model



Notes: The plotted IRFs in the left panel are $\tilde{c}_o^h$ estimates in panel regression $\Delta p_{it+h} = \tilde{\phi}_s^h (p_{i,t-1} - \tilde{\omega}_i \bar{p}^{\,s}_{i,t-1}) + \tilde{\phi}_0^h (p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t) + \tilde{a}_i^h + a_1^h \Delta p_{it-1} + \tilde{b}_1^h \Delta \bar{p}^{\,s}_{i,t-1} + \tilde{c}_0^h \Delta p_{0t} + \tilde{\varepsilon}_{it+h}$ for each group of metro areas at each horizon h. The IRFs in the right panel are $\hat{c}_o^h$ estimates in $\Delta\, lnQ_{it+h} = \hat{\phi}_s^h (p_{i,t-1} - \tilde{\omega}_i \bar{p}^{\,s}_{i,t-1}) + \hat{\phi}_0^h (p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t) + \hat{a}_i^h + \hat{a}_1^h \Delta p_{it-1} + \hat{b}_1^h \Delta \bar{p}^{\,s}_{i,t-1} + \hat{c}_0^h \Delta p_{0t} + \hat{\varepsilon}_{it+h}$ for each group of metro areas at each horizon h. Group "ALL MSA" includes all of the 21 following areas, while group "Elastic" includes 6 metro areas with the most inelastic housing supply (LA, Oakland, Oxnard, San Diego, San Francisco, and Santa Maria), and group "Inelastic" is made up of the remaining 15 metro areas with more elastic housing supply.

Table 1: Metro areas, abbreviations, and data

Metro Areas	Abbrev.	Elasticity	pop_2010
Anaheim-Santa Ana-Irvine, CA	Anaheim		3010232
Bakersfield, CA	Bakersfield	1.64	839631
Fresno, CA	Fresno	1.84	930450
Los Angeles-Long Beach-Glendale, CA	LA	0.63	9818605
Merced, CA	Merced	2.39	255793
Modesto, CA	Modesto	2.17	514453
Oakland-Hayward-Berkeley, CA	Oakland	0.70	2559296
Oxnard-Thousand Oaks-Ventura, CA	Oxnard	0.75	823318
Riverside-San Bernardino-Ontario, CA	Riverside	0.94	4224851
Sacramento--Roseville--Arden-Arcade, CA	Sacramento		2149127
Salinas, CA	Salinas	1.10	415057
San Diego-Carlsbad, CA	San Diego	0.67	3095313
San Francisco-Redwood City-South San Francisco, CA	San Francisco	0.66	1523686
San Jose-Sunnyvale-Santa Clara, CA	San Jose	0.76	1836911
San Luis Obispo-Paso Robles-Arroyo Grande, CA	San Luis Obispo	1.22	269637
San Rafael, CA	San Rafael		252409
Santa Cruz-Watsonville, CA	Santa Cruz	1.19	262382
Santa Maria-Santa Barbara, CA	Santa Maria	0.89	423895
Santa Rosa, CA	Santa Rosa	1.00	483878
Stockton-Lodi, CA	Stockton	2.07	685306
Vallejo-Fairfield, CA	Vallejo	1.14	413344
Visalia-Porterville, CA	Visalia	1.97	442179

Notes: Definitions of Metropolitan areas are based on the Office of Management and Budget (OMB) 2013 delineations. Column Elasticity is the supply elasticity estimates from Saiz (2010). Such supply elasticity estimates are based on economic fundamentals related to natural and man-made land constraints. Column pop_2010 is the population counts of 2010 Census.

Table 2:

Trace cointegration tests with unrestricted intercepts and restricted trend coefficients, and test of over-identifying restrictions in bivariate VAR(4) models of log HPI of CA Metro Areas (1980Q1-2016Q4) Trace cointegration tests with unrestricted intercepts and restricted trend coefficients, and test of over-identifying restrictions in bivariate VAR(4) models of log HPI of CA Metro Areas (1980Q1-2016Q4)

CBSA	Areas	Trace Statistics		H_0: Cotrending and Cointegrating vector is (1,-1) with San Jose		
		H_0: r=0 vs. H_1: r >=1	H_0: r<=1 vs. H_1: r >=2	LR statistics	95% BCV	90% BCV
11244	Anaheim	35.16***	10.96*	17.13**	15.06	13.15
12540	Bakersfield	29.89**	14.06**	13.34*	13.49	12.21
23420	Fresno	28.8**	11.89*	12.7*	12.93	11.43
31084	LA	32.42***	9.50	16.92**	15.98	14.24
32900	Merced	29.8**	13.04**	14.26**	13.76	12.61
33700	Modesto	28.13**	11.94*	12.93*	14.63	12.77
36084	Oakland	30.46***	12.7**	12.62*	13.54	12.00
37100	Oxnard	34.29***	12.61**	17.41**	15.70	13.73
40140	Riverside	26.23**	10.52*	12.93*	14.16	12.71
40900	Sacramento	30.41**	11.92*	15.84*	17.16	15.73
41500	Salinas	27.49**	6.45	18.34**	15.70	14.08
41740	San Diego	27.47**	5.62	19.51**	13.80	11.57
41884	San Francisco	23.08*	7.82	9.16	14.20	12.10
42020	San Luis Obispo	26.52**	5.21	18.95**	15.48	12.98
42034	San Rafael	33.56***	6.98	17.52**	14.84	12.20
42100	Santa Cruz	27.79**	2.75	23.96**	15.46	13.70
42200	Santa Maria	32.49***	11.18*	16.13*	16.69	14.69
42220	Santa Rosa	31.38***	9.35	18.04**	15.48	13.89
44700	Stockton	28.38**	12.78**	12.76*	13.41	12.32
46700	Vallejo	30.69***	10.76*	17**	14.93	13.64
47300	Visalia	26.79**	11.27*	12.42	14.94	13.06

[1]The trace statistics reported are based on the bivariate VAR(4) specification of log of real HPI of San Jose and other metro areas in CA, with unrestricted intercepts and restricted trend coefficients.

[2]The trace statistic is the cointegration test statistic of Johansen (1991). The log likelihood ratio (LR) statistic reported is for testing the cotrending restriction with the cointegration vector given by (1,-1) for the log real HPI in San Jose and the other metro area.

[3]For the trace test, the 99%, 95%, and 90% critical values of the test for H_0: r=0 are 30.45, 25.32, and 22.76. For the trace test, the 99%, 95%, and 90% critical values of the test for H_0: r<=1 are 16.26, 12.25, and 10.49.

[4]BCV stands for bootstrap critical values, based on 1000 bootstrap replications. Bootstrapping algorithm is from Cavaliere, Nielsen, and Rahbek (2015).

[5]* signifies that test rejects the null at the 10% level; ** signifies test rejects the null at the 5% level; **** signifies that test rejects the null at the 1% level.

Table 3: Test of over-identifying restrictions in bivariate VAR(4) models of log HPI of CA Metro Areas (1980Q1-2016Q4)

Areas	H_0: Cotrending with Leader			H_0: Cointegrating vector is (1,-1) with Leader			H_0: Cointegrating vector is (1,-1) with Leader based on cotrending test		
	LR stat.	95% BCV	90% BCV	LR stat.	95% BCV	90% BCV	LR stat.	95% BCV	90% BCV
Anaheim	10.3**	9.98	8.81	13.06*	13.27	11.41	13.06*	13.27	11.41
Bakersfield	0.02	6.48	5.36	1.22	8.62	6.98	13.32**	9.99	8.92
Fresno	4.75	8.45	7.16	3.89	8.08	6.57	7.95*	8.03	6.89
LA	9.55**	9.36	7.66	12.92*	13.06	11.73	12.92*	13.06	11.73
Merced	3.62	8.18	6.46	3.24	8.19	7.05	10.64*	10.85	9.55
Modesto	4.16	8.64	6.94	2.64	9.60	7.95	8.77	11.04	9.76
Oakland	3.74	9.16	7.56	5.00	9.74	8.57	8.87**	7.79	6.62
Oxnard	6.88	9.79	8.16	9.08	12.13	10.21	10.53**	8.28	7.21
Riverside	1.82	6.76	5.25	4.72	8.92	7.73	11.11**	9.86	9.05
Sacramento	6.54	12.82	10.98	4.68	10.71	9.12	9.29**	8.97	8.03
Salinas	14.1**	12.32	10.48	14.37**	11.85	10.26	14.37**	11.85	10.26
San Diego	16.09**	13.33	11.49	13.91**	13.53	11.74	13.91**	13.53	11.74
San Francisco	6.88	9.81	8.56	7.44	10.91	9.63	2.28	10.53	8.77
San Luis Obispo	16.01**	12.96	11.20	14.39**	13.42	11.78	14.39**	13.42	11.78
San Rafael	15.38**	11.46	9.80	9.01*	9.43	7.79	9.01*	9.43	7.79
Santa Cruz	22.23**	9.39	7.87	20.52**	11.81	9.53	20.52**	11.81	9.53
Santa Maria	9.89	13.85	11.96	9.73	13.88	12.31	6.23*	7.02	6.07
Santa Rosa	12.68**	12.67	10.63	10.47*	11.20	9.70	10.47*	11.20	9.70
Stockton	2.70	8.45	6.91	2.57	8.26	7.09	10.06**	10.00	9.07
Vallejo	9.14	11.11	9.35	7.73	10.84	9.13	7.86	9.08	8.15
Visalia	4.12	8.80	6.84	2.12	9.08	7.51	8.30	11.29	9.81

[1] The first log likelihood ratio (LR) statistic reported is for testing the cotrending restriction for the log real HPI in San Jose and the other metro area, based on the bivariate VAR(4) specification with unrestricted intercepts and restricted trend coefficients.
[2] The second log likelihood ratio (LR) statistic reported is for testing the cointegration vector given by (1,-1) for the log real HPI in San Jose and the other metro area, based on the bivariate VAR(4) specification with unrestricted intercepts and restricted trend
[3] The third log likelihood ratio (LR) statistic reported is for testing the cointegration vector given by (1,-1) for the log real HPI in San Jose and the other metro area, based on the bivariate VAR(4) specification with unrestricted intercepts and restricted trend coefficients if rejecting the cotrending test, otherwise the base bivariate VAR(4) specification only has a unrestricted intercepts.
[4] BCV stands for bootstrap critical values, based on 1000 bootstrap replications. Bootstrapping algorithm is from Cavaliere, Nielsen, and Rahbek (2015).
[5] * signifies that test rejects the null at the 10% level; ** signifies test rejects the null at the 5% level; **** signifies that test rejects the null at the 1% level.

Table 4: CI coefficients of lnHPI versus supply side conditions

	(1)	(2)	(3)	(4)	(5)
VARIABLES	beta_i	beta_i	beta_i	beta_i	beta_i
elasticity	2.65***				
	(0.48)				
unaval		-5.88***	-6.29***	-5.47***	-6.30***
		(1.14)	(1.24)	(1.20)	(1.46)
WRLURI			0.87	0.12	0.87
			(0.97)	(1.04)	(0.90)
c.population_2010				-4.7e-07	
#c.unaval				(4.3e-07)	
c.percent_change_80_10					0.049
#c.unaval					(2.17)
Constant	-1.26**	4.97***	4.55***	5.07***	4.54***
	(0.57)	(0.67)	(0.86)	(0.91)	(0.92)
Observations	17	17	17	17	17
R-squared	0.704	0.674	0.691	0.777	0.691

Notes: beta_i stands for the CI coefficient β_i in the CI relation $p_{0t} - \beta_i p_{it}$, from a bivariate VAR(4) of log of real HPI of San Jose (p_{0t}) and other CA metro area (p_{it}) with unrestricted intercepts and restricted trend coefficients if the cotrending test is rejected, otherwise with unrestricted intercepts only. Variable elasticity is the supply elasticity estimates from Saiz (2010), which are simple nonlinear combinations of the available data on physical and regulatory constraints. Variable unaval is the share of unavailable land for development from Saiz (2010). Variable WRLURI is from the 2005 Wharton Regulation Survey of Gyourko, Saizm and Summers (2008) on the elasticity of supply. Variable c.population_2010#c.unaval is an interaction term of the 2010 Census population counts with the variable unaval, while variable c.percent_change_80_10#c.unaval is an interaction term of the percent change of population from 1980 Census to 2010 Census with the variable unaval. Because the definitions of metro area differ from Saiz (2010), only 18 following metro areas have the supply elasticity measures (with Bakersfield for which the CI coefficient (17.13) is an outlier, we are left with 17 metro areas). Standard errors in parentheses are bootstrapped from 1000 repetitions. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Estimation of Region Specific House Price Diffusion Equation with San Jose as a Dominant Region (1980Q1-2016Q4)

Areas	EC1	EC2	Own Lag Effects	Neighbor Lag Effects	Leader Lag Effects	Leader Contemporaneous Effects	Wu-Hausman Statistics	k_a	k_b	k_c
San Jose		0.02***	0.75***	0.08				1	3	
		(2.97)	(8.12)	(0.95)						
Anaheim			0.49***	0.16	-0.44***	0.73***	-1.35	1	1	1
			(4.1)	(1.33)	(-5.35)	(12.02)				
Bakersfield			0.47***	0.60***	-0.74***	0.64***	0.56	3	2	1
			(3.71)	(3.25)	(-5.84)	(7.10)				
Fresno			0.31**	0.76***	-0.74***	0.6***	-0.61	4	1	1
			(2.32)	(3.95)	(-5.10)	(6.18)				
LA			0.64***	0.04	-0.41***	0.72***	-1.64	1	1	1
			(4.84)	(0.26)	(-4.79)	(11.74)				
Merced		-0.03***	0.17*	1.26***	-0.83***	0.68***	-1.09	2	1	4
		(-3.13)	(1.67)	(6.62)	(-4.50)	(5.83)				
Modesto			-0.44***	1.67***	-0.52**	0.62***	-0.44	3	1	1
			(-2.7)	(5.94)	(-2.41)	(4.64)				
Oakland			0.25*	0.34***	-0.43***	0.86***	-2.07**	1	3	1
			(1.88)	(3.30)	(-5.7)	(19.77)				
Oxnard			0.27**	0.37***	-0.5***	0.8***	-1.08	1	1	2
			(2.5)	(2.98)	(-4.71)	(11.10)				
Riverside			0.12	0.85***	-0.54***	0.75***	-1.28	1	1	1
			(0.91)	(4.25)	(-4.46)	(8.80)				
Sacramento		-0.09***	0.8***	0.05***	-0.66***	0.81***	-0.2	2	4	1
		(-3.09)	(6.08)	(0.27)	(-5.77)	(10.58)				
Salinas		0.00**	-0.2**	1.22***	-0.79***	0.97***	-0.8	1	2	1
		(-2.00)	(-1.98)	(7.26)	(-5.99)	(11.19)				
San Diego		0.06***	0.14	0.54***	-0.43***	0.64***	0.52	4	2	2
		(2.85)	(0.73)	(2.91)	(-3.92)	(7.81)				

San Francisco	0.04*		0.47***	0.11	-0.47***	0.81***	-0.84	4	1	2
	(1.91)		(4.41)	(1.62)	(-4.53)	(15.58)				
San Luis Obispo	0.08***		-0.03	0.22**		0.56***	-0.6	1	1	0
	(3.72)		(-0.30)	(2.1)		(6.58)				
San Rafael	0.08***		-0.5***	0.35***		0.8***	0.64	2	1	0
	(3.54)		(-5.01)	(4.22)		(12.42)				
Santa Cruz	0.14***		-0.55***	0.5***		0.77***	-2.13**	2	1	0
	(4.75)		(-5.22)	(5.46)		(10.87)				
Santa Maria			0.55***	0.1	-0.39***	0.71***	-0.6	3	2	1
			(3.28)	(0.57)	(-3.47)	(8.38)				
Santa Rosa	0.06***		-0.22**	0.46***		0.72***	0.93	1	1	0
	(3.07)		(2.12)	(4.19)		(13.00)				
Stockton		-0.03**	0.07	1.04***	-0.84***	0.97***	-1.77*	1	1	1
		(-2.52)	(0.63)	(5.42)	(-6.37)	(10.56)				
Vallejo		-0.04***	-0.05	1.42***	-0.8***	0.68***	-0.41	4	1	1
		(-3.16)	(-0.43)	(6.88)	(-5.85)	(7.48)				
Visalia			0.06	0.96***	-0.81***	0.55***	0.05	1	1	2
			(0.63)	(6.44)	(-5.15)	(5.21)				

Notes: This table reports estimates based on the price equations $\Delta p_{it} = \tilde{\phi}_{i0}\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right) + \tilde{\phi}_{is}\left(p_{i,t-1} - \omega_i \bar{p}_{i,t-1}^{\,s}\right) + \tilde{a}_i + \sum_{l=1}^{k_{ia}} a_{il} \Delta p_{it-l} + \sum_{l=1}^{k_{ib}} \tilde{b}_{il} \Delta \bar{p}_{i,t-l}^{\,s} + \sum_{l=0}^{k_{ic}} \tilde{c}_{il} \Delta p_{0,t-l} + \tilde{\varepsilon}_{it},\ for\ i = 1,2,\dots,N$. For $i = 0$ denoting the San Jose equation, we put a priori restriction, $\tilde{\phi}_{00} = \tilde{c}_{00} = 0$. “EC1”, “EC2”, “Own lag effects”, “Neighbor lag effects”, “Leader lag effects”, “Leader contemporaneous effects” relate to estimates of $\tilde{\phi}_{i0}$, $\tilde{\phi}_{is}$, $\sum_{l=1}^{k_{ia}} a_{il}$, $\sum_{l=1}^{k_{ib}} \tilde{b}_{il}$, $\sum_{l=1}^{k_{ic}} \tilde{c}_{il}$, and $\tilde{c}_{i0}$, respectively. T-ratios are in the parenthesis. *** signifies that the test rejects the null at the 1% level, ** at the 5% level, and * at the 10% level. The error correction coefficients are restricted such that at most one of them are statistically significant at the 5% level. Wu-Hausman is the t-ratio for testing $H_0\colon \mu_i = 0$ in the augmented regression $\Delta p_{it} = \tilde{\phi}_{i0}\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right) + \tilde{\phi}_{is}\left(p_{i,t-1} - \omega_i \bar{p}_{i,t-1}^{\,s}\right) + \tilde{a}_i + \sum_{l=1}^{k_{ia}} a_{il} \Delta p_{it-l} + \sum_{l=1}^{k_{ib}} \tilde{b}_{il} \Delta \bar{p}_{i,t-l}^{\,s} + \sum_{l=0}^{k_{ic}} \tilde{c}_{il} \Delta p_{0,t-l} + \mu_i \tilde{\varepsilon}_{0t} + \tilde{\varepsilon}_{it}$, where $\tilde{\varepsilon}_{0t}$ is the residual of the San Jose house price equation. In selecting the lag orders, k_{ia}, k_{ib}, and k_{ic}, the maximum lag-order is set to 4 and the lag orders are selected by Schwarz Bayesian criterion. All the regressions include an intercept term.

Table 6: Panel Regression of House Price Diffusion Equation with San Jose as a Dominant Region (1980Q1-2016Q4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Quarterly All Metro	Quarterly Inelastic Metro	Quarterly Elastic Metro	Annual All Metro	Annual Inelastic Metro	Annual Elastic Metro	Biannual All Metro	Biannual Inelastic Metro	Biannual Elastic Metro
EC1	0.0012**	0.014***	0.00088*	0.011***	0.081***	0.010***	0.035***	0.12	0.035**
	(0.00045)	(0.0032)	(0.00050)	(0.0029)	(0.021)	(0.0032)	(0.013)	(0.088)	(0.014)
EC2	-0.0060***	0.0036	-0.0074***	-0.028**	0.060*	-0.039***	-0.058	0.13	-0.11*
	(0.0019)	(0.0048)	(0.0022)	(0.012)	(0.036)	(0.014)	(0.057)	(0.20)	(0.066)
D.price_leader	0.74***	0.77***	0.72***	0.66***	0.70***	0.62***	0.66***	0.80***	0.59***
	(0.022)	(0.030)	(0.028)	(0.028)	(0.043)	(0.036)	(0.023)	(0.032)	(0.029)
LD.price_leader	-0.43***	-0.33***	-0.48***	-0.24***	-0.32***	-0.23***	0.37***	0.46***	0.34***
	(0.036)	(0.049)	(0.046)	(0.045)	(0.065)	(0.059)	(0.066)	(0.10)	(0.087)
L2D.price_leader	-0.053	-0.13**	-0.024	0.029	0.059	-0.019	0.14**	0.048	0.12
	(0.037)	(0.050)	(0.048)	(0.046)	(0.068)	(0.059)	(0.062)	(0.098)	(0.082)
L3D.price_leader	-0.14***	-0.092*	-0.16***						
	(0.037)	(0.051)	(0.048)						
L4D.price_leader	0.13***	0.021	0.15***						
	(0.032)	(0.044)	(0.041)						
LD.Spatial_lnHPI	0.68***	0.43***	0.77***	0.23***	0.096	0.32***	0.048	0.22	-0.030
	(0.038)	(0.055)	(0.048)	(0.081)	(0.12)	(0.11)	(0.12)	(0.22)	(0.16)
L2D.Spatial_lnHPI	-0.22***	-0.14**	-0.23***	-0.099	-0.0056	-0.055	0.10	0.015	0.13
	(0.042)	(0.061)	(0.053)	(0.077)	(0.12)	(0.10)	(0.12)	(0.18)	(0.16)
L3D.Spatial_lnHPI	-0.080*	0.0014	-0.084						
	(0.044)	(0.063)	(0.056)						
L4D.Spatial_lnHPI	-0.059	-0.12**	-0.037						
	(0.038)	(0.055)	(0.049)						
LD.lnHPI	0.064***	0.12***	0.042*	0.42***	0.49***	0.37***	-0.65***	-0.78***	-0.53***
	(0.021)	(0.041)	(0.025)	(0.057)	(0.11)	(0.069)	(0.079)	(0.18)	(0.10)
L2D.lnHPI	0.14***	0.14***	0.12***	-0.046	-0.099	-0.055	-0.33***	-0.10	-0.35***

	(0.020)	(0.039)	(0.024)	(0.057)	(0.11)	(0.068)	(0.079)	(0.17)	(0.10)
L3D.lnHPI	0.18***	0.097**	0.19***						
	(0.020)	(0.039)	(0.024)						
L4D.lnHPI	0.074***	0.18***	0.054**						
	(0.021)	(0.039)	(0.024)						
Constant	-0.00090**	-0.000058	-0.0013***	-0.0060**	-0.0027	-0.0071**	-0.0032	-0.011	-0.00089
	(0.00038)	(0.00051)	(0.00048)	(0.0025)	(0.0035)	(0.0032)	(0.010)	(0.013)	(0.013)
Observations	3,003	858	2,145	714	204	510	336	96	240
R-squared	0.659	0.770	0.636	0.717	0.810	0.701	0.828	0.949	0.769
Number of cbsa_md	21	6	15	21	6	15	21	6	15
Metro FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table reports estimates based on the price equations $\Delta p_{it} = \tilde{\phi}_0\big(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\big) + \tilde{\phi}_s\big(p_{i,t-1} - \omega_i \bar{p}\,^s_{i,t-1}\big) + \tilde{a}_i + \sum_{l=1}^4 a_l \Delta\, p_{it-l} + \sum_{l=1}^4 \tilde{b}_l \Delta\bar{p}\,^s_{i,t-l} + \sum_{l=0}^4 \tilde{c}_l\, \Delta p_{0,t-l} + \tilde{\varepsilon}_{it},\ for\ i = 1,2,\ldots,N$. The dominant area is excluded in this panel regression. Variable D. price_leader ($\Delta p_{0,t}$) is the contemporaneous price changes in San Jose, and LlD.price_leader ($\Delta p_{0,t-l}$) is the lagged price changes of order l in San Jose for l=1, 2, 3, 4. Variable LlD.Spatial_lnHPI ($\Delta\bar{p}\,^s_{i,t-l}$) is the lagged price changes of order l of the neighbor of metro area i for l=1, 2, 3, 4. Variable LlD.lnHPI ($\Delta p_{i,t-l}$) is the lagged price changes of order l in metro area i for l=1, 2, 3, 4. “EC1”, “EC2”, “D.price_leader” (Leader contemporaneous effects), “LD.price_leader”—“ L4D.price_leader” (Leader lag effects) , “LD.Saptial_lnHPI”—“ L4D. Saptial_lnHPI” (Neighbor lag effects), “LD. lnHPI”—“ L4D. lnHPI” (Own lag effects), relate to estimates of $\tilde{\phi}_0$, $\tilde{\phi}_s$, $\tilde{c}_0$, $\tilde{c}_1 - \tilde{c}_4$, $\tilde{b}_1 - \tilde{b}_4$, and $\tilde{a}_1 - \tilde{a}_4$, respectively. Standard errors are in the parenthesis. *** signifies that the test rejects the null at the 1% level, ** at the 5% level, and * at the 10% level. The first three columns use quarterly HPI from 1980Q1 to 2016Q4; the first regression includes all of the 21 following areas; the second regression is for metro areas with supply elasticity less than 0.9, more specifically including LA, Oakland, Oxnard, San Diego, San Francisco, and Santa Maria; and the third regression is for the remaining 15 metro areas. Column 4 to Column 6 use annual HPI from 1980 to 2016, and the last three columns use biannual data from 1980 to 2016.

Table 7: Panel Regression of Construction Diffusion Equation with San Jose as a Dominant Region (1997Q1-2015Q4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Quarterly All Metro	Quarterly Inelastic Metro	Quarterly Elastic Metro	Annual All Metro	Annual Inelastic Metro	Annual Elastic Metro	Biannual All Metro	Biannual Inelastic Metro	Biannual Elastic Metro
EC1	-0.041***	0.020	-0.042***	-0.10***	0.15	-0.11***	-0.028	-0.26	-0.024
	(0.0039)	(0.022)	(0.0045)	(0.034)	(0.19)	(0.039)	(0.090)	(0.52)	(0.10)
EC2	0.044**	-0.066**	0.058**	0.15	-0.16	0.15	0.91*	0.77	1.11*
	(0.020)	(0.033)	(0.025)	(0.18)	(0.33)	(0.22)	(0.51)	(1.59)	(0.66)
D.price_leader	0.44**	0.39*	0.57**	1.23***	0.92*	1.41**	1.58***	1.19***	1.71***
	(0.21)	(0.22)	(0.28)	(0.43)	(0.47)	(0.59)	(0.20)	(0.25)	(0.27)
LD.price_leader	-0.49	0.041	-0.80	-2.08***	-0.57	-2.81***	0.039	0.58	-0.052
	(0.39)	(0.40)	(0.52)	(0.65)	(0.65)	(0.88)	(0.55)	(0.76)	(0.77)
L2D.price_leader	-0.63	-0.57	-0.68	-0.062	0.27	-0.32	0.24	0.51	0.024
	(0.43)	(0.44)	(0.57)	(0.67)	(0.68)	(0.91)	(0.50)	(0.68)	(0.74)
L3D.price_leader	-0.25	0.47	-0.60						
	(0.42)	(0.43)	(0.56)						
L4D.price_leader	0.22	0.23	0.21						
	(0.33)	(0.35)	(0.45)						
LD.Spatial_lnHPI	-0.063	-0.77	0.57	0.80	-1.44	2.15	0.92	1.55	1.18
	(0.51)	(0.58)	(0.67)	(1.11)	(1.25)	(1.51)	(1.07)	(2.27)	(1.52)
L2D.Spatial_lnHPI	0.46	0.028	0.66	1.31	-1.00	2.00	1.96*	1.40	2.52
	(0.60)	(0.67)	(0.79)	(1.12)	(1.33)	(1.49)	(1.05)	(1.54)	(1.62)
L3D.Spatial_lnHPI	0.39	-0.88	0.93						
	(0.61)	(0.66)	(0.81)						
L4D.Spatial_lnHPI	0.68	-0.47	0.93						
	(0.50)	(0.55)	(0.67)						
LD.lnHPI	1.16***	0.88*	1.00**	1.94**	2.80**	1.26	-0.022	-1.24	-0.29

	(0.36)	(0.52)	(0.44)	(0.82)	(1.21)	(1.03)	(0.83)	(1.84)	(1.12)
L2D.lnHPI	0.61	1.19**	0.38	-0.11	1.73	-0.47	-1.56**	-1.52	-1.84*
	(0.39)	(0.55)	(0.48)	(0.88)	(1.23)	(1.09)	(0.71)	(1.25)	(1.04)
L3D.lnHPI	0.20	0.57	0.063						
	(0.39)	(0.54)	(0.48)						
L4D.lnHPI	-0.078	1.32***	-0.40						
	(0.36)	(0.50)	(0.45)						
Constant	0.22***	0.14***	0.25***	0.84***	0.51***	0.95***	1.10***	0.73***	1.24***
	(0.0043)	(0.0046)	(0.0059)	(0.039)	(0.043)	(0.054)	(0.11)	(0.13)	(0.14)
Observations	1,575	450	1,125	399	114	285	189	52	137
R-squared	0.351	0.387	0.366	0.324	0.386	0.338	0.322	0.449	0.318
Number of cbsa_md	21	6	15	21	6	15	21	6	15
Metro FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table reports estimates based on the price equations $100 * \Delta lnQ_{it} = \tilde{\phi}_0\left(p_{0,t-1} - \beta_i p_{i,t-1} - \gamma_i t\right) + \tilde{\phi}_s\left(p_{i,t-1} - \omega_i \bar{p}_{i,t-1}^{\,s}\right) + \tilde{a}_i + \sum_{l=1}^{4} a_l \Delta\, p_{it-l} + \sum_{l=1}^{4} \tilde{b}_l \Delta \bar{p}_{i,t-l}^{\,s} + \sum_{l=0}^{4} \tilde{c}_l\, \Delta p_{0,t-l} + \tilde{\varepsilon}_{it},\ for\ i = 1,2,\dots,N$. The dominant area is excluded in this panel regression. Variable D. price_leader ($\Delta p_{0,t}$) is the contemporaneous price changes in San Jose, and LlD.price_leader ($\Delta p_{0,t-l}$) is the lagged price changes of order l in San Jose for l=1, 2, 3, 4. Variable LlD.Spatial_lnHPI ($\Delta \bar{p}_{i,t-l}^{\,s}$) is the lagged price changes of order l of the neighbor of metro area i for l=1, 2, 3, 4. Variable LlD.lnHPI ($\Delta p_{i,t-l}$) is the lagged price changes of order l in metro area i for l=1, 2, 3, 4. "EC1", "EC2", "D.price_leader" (Leader contemporaneous effects), "LD.price_leader"—" L4D.price_leader" (Leader lag effects) , "LD.Saptial_lnHPI"—" L4D. Saptial_lnHPI" (Neighbor lag effects), "LD. lnHPI"—" L4D. lnHPI" (Own lag effects), relate to estimates of $\tilde{\phi}_0$, $\tilde{\phi}_s$, $\tilde{c}_0$, $\tilde{c}_1 - \tilde{c}_4$, $\tilde{b}_1 - \tilde{b}_4$, and $\tilde{a}_1 - \tilde{a}_4$, respectively. Standard errors are in the parenthesis. *** signifies that the test rejects the null at the 1% level, ** at the 5% level, and * at the 10% level. The first three columns use quarterly HPI from 1997Q1 to 2015Q4; the first regression includes all of the 21 following areas; the second regression is for metro areas with supply elasticity less than 0.9, more specifically including LA, Oakland, Oxnard, San Diego, San Francisco, and Santa Maria; and the third regression is for the remaining 15 metro areas. Column 4 to Column 6 use annual HPI from 1997 to 2015, and the last three columns use biannual data from 1997 to 2015.

References

Anselin, Luc. *Spatial econometrics: methods and models*. Vol. 4. Springer Science & Business Media, 2013.

Bai, Jushan. "Inferential theory for factor models of large dimensions." *Econometrica* 71.1 (2003): 135-171.

Bai, Jushan. "Panel data models with interactive fixed effects." *Econometrica* 77.4 (2009): 1229-1279.

Bernanke, Ben S. "Housing, mortgage markets, and foreclosures." *Remarks at The Federal Reserve System Conference on Housing and Mortgage Markets, Washington, DC*. 2008.

Brady, Ryan R. "Measuring the diffusion of housing prices across space and over time." *Journal of Applied Econometrics* 26.2 (2011): 213-231.

Capozza, Dennis R., Patric H. Hendershott, and Charlotte Mack. "An anatomy of price dynamics in illiquid markets: analysis and evidence from local housing markets." *Real Estate Economics* 32.1 (2004): 1-32.

Case, Karl E., and Robert J. Shiller. "The efficiency of the market for single-family homes." *American Economic Review* 79 (1989): 125-137.

Cavaliere, Giuseppe, Heino Bohn Nielsen, and Anders Rahbek. "Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models." *Econometrica* 83.2 (2015): 813-831.

Cliff, Andrew David, and J. Keith Ord. *Spatial autocorrelation*. Vol. 5. London: Pion, 1973.

Davis, Morris A., and Jonathan Heathcote. "Housing and the business cycle." *International Economic Review* 46.3 (2005): 751-784.

Del Negro, Marco, and Christopher Otrok. "99 Luftballons: Monetary policy and the house price boom across US states." *Journal of Monetary Economics* 54.7 (2007): 1962-1985.

DiPasquale, Denise. "Why don't we know more about housing supply?" *The Journal of Real Estate Finance and Economics* 18.1 (1999): 9-23.

Flood, Robert P., and Robert J. Hodrick. "On testing for speculative bubbles." *The Journal of Economic Perspectives* 4.2 (1990): 85-101.

Genesove, David, and Lu Han. "A spatial look at housing boom and bust cycles." *Housing and the Financial Crisis*. University of Chicago Press, 2012. 105-141.

Ghent, Andra C., and Michael T. Owyang. "Is housing the business cycle? Evidence from US cities." *Journal of Urban Economics* 67.3 (2010): 336-351.

Glaeser, Edward L., and Joseph Gyourko. "Urban decline and durable housing." *Journal of Political Economy* 113.2 (2005): 345-375.

Glaeser, Edward L., Joseph Gyourko, and Albert Saiz. "Housing supply and housing bubbles." *Journal of Urban Economics* 64.2 (2008): 198-217.

Glaeser, Edward L., Joseph Gyourko, and Raven Saks. “Why have housing prices gone up?” *American Economic Review Papers and Proceedings* 95.2 (2005): 329-333.

Glaeser, Edward L., and Bryce A. Ward. "The causes and consequences of land use regulation: Evidence from Greater Boston." *Journal of Urban Economics* 65.3 (2009): 265-278.

Green, Richard K., Stephen Malpezzi, and Stephen K. Mayo. "Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources." *The American Economic Review* 95.2 (2005): 334-339.

Gyourko, Joseph, and Albert Saiz. "Construction costs and the supply of housing structure." *Journal of Regional Science* 46.4 (2006): 661-680.

Hernández-Murillo, Rubén, Michael Owyang, and Margarita Rubio. "Clustered housing cycles." (2015).

Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata. "A spatio-temporal model of house prices in the USA." *Journal of Econometrics* 158.1 (2010): 160-173.

Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata. "The spatial and temporal diffusion of house prices in the UK." *Journal of Urban Economics* 69.1 (2011): 2-23.

Hosios, Arthur J., and James E. Pesando. "Measuring prices in resale housing markets in Canada: evidence and implications." *Journal of Housing Economics* 1.4 (1991): 303-317.

Huang, Haifang, and Yao Tang. "Residential land use regulation and the US housing price cycle between 2000 and 2009." *Journal of Urban Economics* 71.1 (2012): 93-99.

Iacoviello, Matteo. "House prices, borrowing constraints, and monetary policy in the business cycle." *The American Economic Review* 95.3 (2005): 739-764.

Ihlanfeldt, Keith R. "The effect of land use regulation on housing and land prices." *Journal of Urban Economics* 61.3 (2007): 420-435.

Jordà, Òscar. "Estimation and inference of impulse responses by local projections." *The American Economic Review* 95.1 (2005): 161-182.

Kelejian, Harry H., and Ingmar R. Prucha. "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review* 40.2 (1999): 509-533.

Kelejian, Harry H., and Ingmar R. Prucha. "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances." *Journal of Econometrics* 157.1 (2010): 53-67.

Kelejian, Harry H., and Dennis P. Robinson. "Spatial correlation: a suggested alternative to the autoregressive model." *New directions in spatial econometrics*. Springer Berlin Heidelberg, 1995. 75-95.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider. "The housing market (s) of San Diego." *The American Economic Review* 105.4 (2015): 1371-1407.

Leamer, Edward E. *Housing is the business cycle*. No. w13428. National Bureau of Economic Research, 2007.

Lee, Lung-Fei. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models." *Econometrica* 72.6 (2004): 1899-1925.

Liu, Crocker H., Adam Nowak, and Stuart S. Rosenthal. "Housing price bubbles, new supply, and within-city dynamics." *Journal of Urban Economics* 96 (2016): 55-72.

Mayer, Christopher J., and C. Tsuriel Somerville. "Land use regulation and new construction." *Regional Science and Urban Economics* 30.6 (2000): 639-662.

Paciorek, Andrew. "Supply constraints and housing market dynamics." *Journal of Urban Economics* 77 (2013): 11-26.

Pesaran, M. Hashem. "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica* 74.4 (2006): 967-1012.

Quigley, John M., and Steven Raphael. "Regulation and the high cost of housing in California." *The American Economic Review* 95.2 (2005): 323-328.

Saiz, Albert. "The geographic determinants of housing supply." *The Quarterly Journal of Economics* 125.3 (2010): 1253-1296.

Strauss, Jack. "Does housing drive state-level job growth? Building permits and consumer expectations forecast a state's economic activity." *Journal of Urban Economics* 73.1 (2013): 77-93.

Swank, Job, Jan Kakes, and Alexander F. Tieman. "The housing ladder, taxation, and borrowing constraints." (2003). DNB Report No.9, Amsterdam.

Whittle, Peter. "On stationary processes in the plane." *Biometrika* (1954): 434-449.

Chapter 3: Fully Modified Least Squares Estimation of Factor-Augmented Cointegration Regressions

1 Introduction

In this paper, we study estimation and a test of cointegration relations between an observed integrated variable and some latent integrated factors. Usually, cointegration analysis is on observable integrated series to explore possible long run equilibrium relations. Cointegration relations between an integrated variable and some latent unobserved integrated factors have been understudied, but the need of this study is highlighted in the recent development in the literature of forecasting under the nonstationary setting with cointegration and large dynamic factor model involved.

One motivation of considering cointegration relations with latent integrated factors is to find the most relevant long run equilibrium information through dimension reduction. Under the case when the number of integrated series is large and there is no clear economic theory on the long run equilibrium relation between the series of interest and the large panel of integrated series, latent factors can work as an efficient way to summarize the pervasive source of nonstationarity in the large panel, which may help to explain the series of interest better in the long run. Also, cointegration relations between the series of interest and the latent factors of this large panel could be estimated much more easily because of the much smaller number of series involved. Another motivating examples is the diffusion index forecasts with integrated (or I(1)) variables, where the forecasting equation is in the form of an error correction model (ECM) and there is a need to estimate the error correction (hereafter EC) term. Estimating the EC term is basically estimating the cointegration regression between the variable of interest and the latent diffusion index.

This idea of diffusion index forecasts in which covariability in a large number of economic variables can be modeled by a relatively few number of unobserved latent variables (the latter also known as diffusion indexes) is appealing and has proved to be useful in dealing with this high-dimensional problem (see Stock and Watson (1998), (2002a), (2002b)). Most of the diffusion index forecasts have been done in a stationary setting by transforming integrated series into stationary series, but most economic time series frequently exhibit characteristics that are widely believed to be intrinsically nonstationary. Cointegration among some integrated macroeconomic variables may help with forecasts by adding long-run information into the model. Transforming integrated series into stationary series may throw useful long run information away and result in over-differenced equations.

The Factor-augmented Error Correction Model (FECM) introduced by Banerjee and Marcellino (2009) is an extension of the diffusion index forecasts to I(1) variables with possible cointegration relation taken into account. By adding an cointegration relation to the dynamic factor models and modeling the factors jointly with a limited set of economic variables of interest from the large dataset, the FECM method have been shown to improve over both the Error Correction Model (ECM), by relaxing the dependence of cointegration analysis on a small set of variables, and the Factor-augmented Vector Autoregression (FAVAR, Bernanke, Boivin, and Bernanke et al., 2005), by allowing for the inclusion of error correction terms in the equations for the key variables under analysis. Further studies in Banerjee, Marcellino, and Masten (2014a, 2014b) show that the FECM generally offers a higher forecasting precision relative to the FAVAR.

However, in the above studies of FECM, the authors outline their underlying data generating process (DGP) using the true latent factors, while their estimation processes are based on estimated factors. It is well known that estimated factors involve estimation errors even under a stationary setting (Stock and Watson, 1998; Bai and Ng, 2002, Bai, 2003). In models with weak stationary factors, estimated factors may be very noisy and may fail to provide useful information for the purpose of forecasting. However, as long as the latent factors embed strong signals in the large panel of data and could be consistently estimated, the estimation errors in the factors are negligible and inference for factor-augmented regressions could be conducted as usual as shown in Bai and Ng (2006). Based on results in Bai (2003), Bai and Ng (2006) show that the least squares estimators obtained from factor-augmented regressions are consistent with usual converging speeds and are asymptotically normal, given that signals embedded in factors are strong and could be consistently estimated.

For large panels of integrated series, estimation errors in latent integrated factors could be substantial given the fact that estimators of the integrated factors are usually constructed as partial sums of the principal component estimators to a first-differenced panel. No theoretical examination has been undertaken to show that the estimation errors in the estimated integrated factors are negligible and thus to show that the usage of estimated integrated factors for the cointegration estimation and the factor-augmented error correction model estimation are valid. In this paper, we try to fill this gap by developing asymptotic theories for estimators of the cointegration regression between an integrated variable and some latent factors. Given that the latent integrated factors are strong and could be consistently estimated, our results indicate that the direct least squares estimator of the cointegration relation based estimated factors are consistent. This will provide theoretical justification for the usage of estimated factors in the estimation of FECM. We also show that given the factors are consistently estimated the traditional residual-based cointegration tests between the integrated variable and these latent factors also work as usual.

As stated above, the cointegration estimation considered in this paper involves a generated regressor issue. Pagan (1984) provides extensive discussions on situations when regressions involve generated regressors from another regression, and provides results on the consistency and the efficiency of two-step estimators as compared to joint estimators of the two regressions. The analysis in Pagan (1984) is quite classic in the sense that regressors are all stationary and the first-step estimations of the two-step estimators are usually least squares estimations. In the cointegration regression considered in this paper, we also use a two-step procedure, with estimating the latent factors in the first step and estimating the cointegration regression using the estimated factors in the second step. The main difference from Pagan (1984) is that the main regression we focus on is a cointegration regression with integrated regressors and the factor analysis in our first step can not be treated as a least squared regression given the factor model we are considering. And given the nature of the large dimensional factor model, a joint estimation of the factor model and the cointegration regression seems impossible and thus we do not have a benchmark to infer the efficiency of our two-step estimators. Hence, in this paper, we focus on the consistency and inference of the cointegration relation estimator using generated factors from a large panel of integrated series.

The factor model this paper assumes is a more realistic nonstationary large-dimension factor model which allows for possible I(1) idiosyncratic components as in Bai and Ng (2004). The factor model in the current FECM literature, such as in Banerjee and Marcellino (2009), only allows for stationary idiosyncratic components, imposing a large number of cointegration relations in the large-dimension factor model. This corresponds to the factor model considered in Bai (2004), which seems unrealistic in the real world given the fact that many macroeconomic variables are not cointegrated. Hence, this paper adopts the factor model in Bai and Ng (2004), and try to estimate and test the cointegration relation between these pervasive sources of nonstationarity in this large panel of integrated series and another integrated variable of interest. The latent nonstationary factors are allowed to cointegrate to some extent, which is equivalent to allowing for stationary common factors in the factor model. Also, the integrated variable of interest could be one series outside of the large panel dataset from which the factors are extracted.

Given the large-dimension nonstationary factor model and the estimates of the latent integrated factors in Bai and Ng (2004), the next step is to explore the asymptotic properties of the estimates of the cointegration relation between the integrated variable of interest and the latent factors using estimated factors. Another extension this paper highlights is that we allow for the correlation between the latent regressors and the error term in the cointegration equation of interest, which implies endogeneity in the latent regressors but is often assumed missing in previous literature. To account for potential serial correlation and endogeneity in the cointegration regression of interest, we adopt the fully modified least squares (FM-OLS) estimation of the cointegration equation developed in Phillips and Hansen (1990) and Phillips (1995).

As shown in Phillips and Durlauf (1986), for regressions with integrated processes, the asymptotic theory for conventional tests and estimates involves major departures from classical theory and raises new issues of the presence of nuisance parameters in the limiting distribution theory. To get nuisance parameter-free asymptotic distributions of estimates for regressions with integrated processes, Phillips and Hansen (1990) and Phillips (1995) propose fully modified least squares (FM-OLS) regression, based on which the asymptotic distribution of Wald test statistic is shown to involve chi-squared distributions. These FM-OLS estimates account for serial correlation and endogeneity in the regressors. We follow the FM-OLS regression of Phillips and Hansen (1990) and Phillips (1995) to get estimates of the cointegration coefficients with asymptotic distributions free of nuisance parameters, which in turn facilitate hypothesis testing. Nonstationarity in the latent regressors does not affect the consistency of estimates even when the latent regressors are correlated with error terms.

In some sense, our setting up is similar to the cointegrating regressions with messy regressors considered in Miller (2010). In Miller (2010), the integrated regressors are messy in the sense that the data may be mismeasured, missing, observed at mixed frequencies, or may have mildly nonstationary noise. It is shown in Miller (2010) that canonical cointegrating regression (CCR) is valid even when the error term is not covariance stationary. Just like FM-OLS, CCR is also a covariance-based technique used to estimate the cointegrating vector of a prototypical cointegrating regression (Park 1992). In the cointegrating regression considered in our paper, we can think of the integrated factors as the messy regressors with measurement errors. The measurement errors (or the estimation errors) of the latent factors are shown to be covariance-stationary in Bai and Ng (2004). Thus there is no need to resort to the CCR and we can get by using the FM-OLS, which requires covariance stationary errors.

In short, our estimation and testing of the cointegration relation between an observed nonstationary series and some latent factors works under a two-step process. The first step is to estimate nonstationary factors from the large nonstationary panel dataset consistently following the method in Bai and Ng (2004). The second step is to get the FM-OLS estimates of the cointegration relation between the integrated variable of interest and the latent integrated factors using the estimated factors from the first step. We derive the asymptotic properties of the FM-OLS estimates of the cointegration coefficients, which allows for possible hypothesis testing and inferences. Traditional residual-based cointegration tests with estimated factors are shown to have usual limiting distributions given factors are estimated consistently and thus could be used in empirical work without doubt. In the Application section, we propose the Factor-Augmented Diffusion Index (FADI) forecasting method by adding an error correction term into the traditional diffusion index forecasts of Stock and Watson (2002a). In the last section, we use a large panel data set of US macroeconomic variables from Stock and Watson (2005) to study possible cointegration relations among the series in the large panel and the factors, and show that the FADI method with consistently estimated factors could improve over the FECM method in Banerjee and Marcellino (2009) for certain variables under study in short forecasting horizons.

The paper proceeds as follows. Section 2 introduces the model and states the underlying assumptions. Section 3 derives the properties of the FM-OLS estimates and their asymptotic distributions. Section 4 discusses the cointegration test among an observable nonstationary series and a set of possibly cointegrated nonstationary latent factors. The Factor-Augmented Diffusion Index (FADI) forecasting method is discussed in Section 5, and an empirical example on the nonstationary panel of Stock and Watson (2005) is discussed in Section 6. Section 7 concludes the paper and summarizes its main results. Derivations and proofs are given in the Appendix.

The notation and terminology that we use in the paper are taken from Phillips (1995) and Bai ang Ng (2004). We define the matrix $\Omega = \sum_{k=-\infty}^{\infty} E(u_k u_0^{'})$ as the long-run variance matrix of the covariance stationary time series u_t and write lrvar$(u_t) = \Omega$. Similarly, we designate long-run covariance matrices as lrcov$(\cdot)$, and we use lrcov$_+(\cdot)$ to signify one-sided sums of covariance matrices, e.g., $\Delta = \sum_{k=0}^{\infty} E(u_k u_0^{'})$, which is called the one-sided long-run covariance. $BM(\Omega)$ denotes a vector Brownian motion with covariance matrix Ω, and we usually write integrals like $\int_0^1 B(s)ds$ as $\int_0^1 B$ or simply $\int B$ when there is no ambiguity over limits. The notation $y_t \equiv I(1)$ signifies that the time series y_t is integrated of order one, so that $\Delta y_t \equiv I(0)$. In addition, the inequality " >0" denotes positive definite when applied to matrices, and the symbols " $\xrightarrow{d}$ ", " $\xrightarrow{p}$ ", "$a.s.$", " $\equiv$ " and " $:=$ " signify convergence in distribution, convergence in probability, almost surely, equality in distribution, and notational definition, respectively. We use $||A||$ to signify the matrix norm $(tr(A^{'}A))^{1/2}$, $|A|$ to denote the determinant of A, vec$(\cdot)$ to stack the rows of a matrix into a column vector, $[x]$ to denote the largest integer $\leq$ x, and all limits in the paper are taken as the sample size $(n, T) \to \infty$, except where otherwise noted.

2 Model and Assumptions

In this paper we are interested in estimating and testing the cointegration relation between an observed I(1) variable and latent I(1) factors illustrated in the following equation:

$$y_t = \alpha^{'} F_t + \varepsilon_t, \tag{1}$$

where y_t is an integrated scalar series, F_t is an r-dimensional vector of integrated latent factors, and ε_t is a stationary scalar. The motivating example for the above cointegration analysis is the diffusion index forecasts with I(1) variables, where the forecasting equation is in the form of an ECM model as follows:

$$\Delta y_t = \gamma\beta^{'} \begin{pmatrix} y_{t-1} \\ F_{t-1} \end{pmatrix} + A_1 \begin{pmatrix} \Delta y_{t-1} \\ \Delta F_{t-1} \end{pmatrix} + ... + A_q \begin{pmatrix} \Delta y_{t-q} \\ \Delta F_{t-q} \end{pmatrix} + \epsilon_t. \tag{2}$$

In the above forecasting equation, the key component is the EC term, $\beta^{'}(y_{t-1}, F_{t-1}^{'})^{'}$. Since the factors are unobserved, estimated factors are used to form forecasts in empirical applications. However, there is no theoretical work to justify the usual estimation of cointegration regressions and the above factor-augmented error correction model using estimated factors. This paper tries to fill this gap by studying the direct estimation of the cointegration relation in equation (1) and discuss the cointegration test between the integrated variable of interest y_t and the latent vector of integrated factors F_t.

The vector F_t is unobservable, but could be estimated from the following factor model as in Bai and Ng (2004):

$$X_{it} = c_i + \beta_i t + \lambda_i^{'} F_t + e_{it}, \tag{3}$$

$$(I - L)F_t = C(L)\eta_t, \tag{4}$$

$$(1 - \rho_i L)e_{it} = D_i(L)\epsilon_{it} \tag{5}$$

where X_{it} $(i = 1, 2, ..., n;\ t = 1, 2, ..., T)$ is a large set of integrated observable variables, $C(L) = \sum_{j=0}^{\infty} C_j L^j$ and $D_i(L) = \sum_{j=0}^{\infty} D_{ij} L^j$. The factor, F_t, is an r dimensional vector of random walks. We assume that there are r_0 cointegration relations and r_1 common trends among these I(1) factors, with $r = r_0 + r_1$. In the above factor model, the idiosyncratic components are allowed to be nonstationary. If $\rho_i < 1$, the idiosyncratic error e_{it} is stationary, while if $\rho_i = 1$, the idiosyncratic error e_{it} is I(1). The possibility of nonstationary idiosyncratic components in the above model allows us to model difference sources of nonstationarity in X_{it}. If F_t is nonstationary but e_{it} is stationary, the nonstationarity of X_{it} is due to a pervasive source. On the other hand, if F_t is stationary but e_{it} is nonstationary, then the nonstationarity of X_{it} is from a series-specific source. The PANIC method-Panel Analysis of Nonstationary in Idiosyncratic and Common components developed in Bai and Ng (2004) can detect whether the nonstationarity in a series is pervasive, or variable-specific, or both. Also, Bai and Ng (2004) have shown how to estimate the latent factors by the method of principal components and determine the number of common trends r_1 when neither F_t nor e_{it} is observed.

Let $M < \infty$ be a generic positive number, not depending on T or n. The factor model satisfies the following assumptions as in Bai and Ng (2004):

Assumption 1 *(i) For nonrandom* λ_i, $\|\lambda_i\| \leq M$*; for random* λ_i, $E\|\lambda_i\|^4 \leq M$*; (ii)* $\frac{1}{n}\sum_{i=1}^{n} \lambda_i \lambda_i' \xrightarrow{p} \Sigma_\Lambda > 0$ *as* $n \to \infty$ *for some* $(r \times r)$ *positive definite non-random matrix* Σ_Λ.

Assumption 2 *(i)* $\eta_t \sim iid(0, \Sigma_\eta)$, $E\|\eta_t\|^4 \leq M$*; (ii)* $var(\Delta F_t) = \sum_{j=0}^{\infty} C_j \Sigma_\eta C_j' > 0$*; (iii)* $\sum_{j=0}^{\infty} j\|C_j\| < M$*; and (iv) C(1) has rank* r_1, $0 \leq r_1 \leq r$.

Assumption 3 *(i) For each i,* $\epsilon_{it} \sim iid(0, \sigma_{\epsilon i}^2)$, $E|\epsilon_{it}|^8 \leq M$, $\sum_{j=0}^{\infty} j|D_{ij}| < M$, $\omega_{\epsilon i}^2 = D_i(1)^2 \sigma_{\epsilon i}^2 > 0$*; (ii)* $E(\epsilon_{it}\epsilon_{jt}) = \pi_{ij}$ *with* $\sum_{i=1}^{N} |\pi_{ij}| \leq M$ *for all j; (iii)* $E|N^{-1/2} \sum_{i=1}^{N} [\epsilon_{is}\epsilon_{it} - E(\epsilon_{is}\epsilon_{it})]|^4 \leq M$*, for every* (t, s).

Assumption 4 *The errors* ϵ_{it}, $\{\eta_t\}$*, and the loadings* $\{\lambda_i\}$ *are three groups of mutually independent groups.*

Assumption 5 $E\|F_0\| \leq M$*, and for every* $i = 1, 2, ..., n$, $E|e_{i0}| \leq M$.

Assumption 1 on the factor loadings is to guarantee that the factor structure is identifiable. Assumption 2 assumes that the short run variance of ΔF_t is positive definite, which guarantees that the principal component analysis of the first-differenced factor model work. However, the long-run covariance of ΔF_t can be reduced rank to permit linear combinations of I(1) factors to be stationary. When there are no stochastic trends,

$r_1 = 0$ and C(1) is null because ΔF_t is over-differenced. On the other hand, when $r_1 > 0$, we can rotate the original F_t space by an orthogonal matrix A such that the first r_1 elements of AF_t are integrated, while the final r_0 elements are stationary. We can denote this rotation by $A = [A_1, A_2]'$, where A_1 is $r \times r_1$ satisfying $A_1'A_1 = I_{r_1}$, and $A_1'A_2 = 0$. Under Assumption 3, $(1 - \rho_i L)e_{it}$ (with ρ_i possibly different across i) is allowed to be weakly serially and cross-sectionally correlated. Assumption 4 assumes ϵ_{it}, $\{\eta_t\}$, and $\{\lambda_i\}$ are mutually independent across i and t, while Assumption 5 is an initial condition assumption imposed commonly in unit root analysis.

The factor estimates are based on the application of principal component analysis to the first-differenced data as in Bai and Ng (2004). Normally, the principal component method is applied to data in level. When the idiosyncratic term e_{it} is stationary, the principal components estimators for F_t and λ_i have been shown to be consistent when all the factors are I(0) (Bai and Ng, 2002) and when some or all of them are I(1) (Bai, 2004). But when e_{it} has a unit root, a regression of X_{it} on F_t is spurious, and the estimates of F_t and λ_i based on data in level will not be consistent. The method of principal components to the first-differenced data in Bai and Ng (2004) could obtain estimates of F_t and e_{it} that preserve their orders of integration, both when e_{it} is I(1) and when it is I(0).

To be precise, suppose the data in level is denoted by X, a data matrix with T time-series observations and n cross-section units. Taking the first difference of X to yield x, a set of $(T - 1) \times n$ stationary variables, we could get the first-differenced factor model:

$$x_{it} = \lambda_i' f_t + z_{it}, \tag{6}$$

where $x_{it} = \Delta X_{it}$, $f_t = \Delta F_t$, and $z_{it} = \Delta e_{it}$. Let $f = (f_2, f_3, ..., f_T)'$ and $\Lambda = (\lambda_1, ..., \lambda_N)'$. The principal component estimator of f, denoted $\hat{f}$, is $\sqrt{T-1}$ times the r eigenvectors corresponding to the first r largest eigenvalues of the $(T-1)\times(T-1)$ matrix xx'. Under the normalization $\hat{f}'\hat{f}/(T-1) = I_r$, the estimated loading matrix is $\hat{\Lambda} = x'\hat{f}/(T-1)$. Define for $t = 2, ..., T$,

$$\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s. \tag{7}$$

According to Bai and Ng (2004), under Assumptions 1-5, there exists a matrix H with rank r such that as $(n, T) \to \infty$,

$$\max_{1\le t\le T} \|\hat{F}_t - HF_t + HF_1\| = O_p(T^{1/2}n^{-1/2}) + O_p(T^{-1/4}).$$

Without loss of generality, we assume that at $t = 1$, $F_1 = 0$. Then we have $\max_{1\le t\le T} \|\hat{F}_t - HF_t\| = O_p(T^{1/2}N^{-1/2}) + O_p(T^{-1/4})$. This result implies that $\hat{F}_t$ is uniformly consistent for HF_t (up to a shift factor HF_1) provided $T/n \to 0$ as $(n, T) \to \infty$.

Since the factor estimator is estimating a rotation of the original factors, we assume that there exist an orthogonal matrix A such that the first r_1 elements of AHF_t are integrated, while the final r_0 elements are stationary. One such rotation is given by $A = [A_1, A_2]'$, where A_1 is $r \times r_1$ satisfying $A_1'A_1 = I_{r_1}$, and $A_1'A_2 = 0$. We define $F_{1t} = A_1'HF_t$ to be the r_1 common stochastic trends and $F_{2t} = A_2'HF_t$ to be the r_0 stationary elements resulting from such a rotation.

In this paper, we consider the possibility that nonstationary regressors, the unobservable regressors F_t, may be endogenous in the regression equation (1). As in Phillips (1995), which studies the fully modified least squares estimates to account for serial correlation effects and for the endogeneity in the regressors, we allow for the innovations of F_t to be serially correlated and possibly correlated with the idiosyncratic terms in the regression equation (1). Recall from the factor model (3)-(5), we have

$$\Delta F_{1t} = (I - L)A_1'HF_t = A_1'HC(L)\eta_t := u_{1t}, \tag{8}$$

$$F_{2t} = A_2'HF_t = A_2'H(F_0 + \sum_{s=1}^{t} C(L)\eta_s) := u_{2t}, \tag{9}$$

Let $u_t = (u_{1t}^{'}, u_{2t}^{'})^{'}$, $v_t = (\varepsilon_t, u_t^{'})^{'} = (\varepsilon_t, u_{1t}^{'}, u_{2t}^{'})^{'}$, and $\psi_t = \varepsilon_t \otimes u_{2t}$. As in Phillips (1995), we assume that v_t is a linear process that satisfies the following assumption.

Assumption 6 *(EC–Error Condition)*

(a) $v_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} C_j \epsilon_{t-j}$, $\sum_{j=0}^{\infty} j^a ||C_j|| < \infty$, $|C(1)| \neq 0$ *for some* $a > 1$.

(b) ϵ_t *is i.i.d. with zero mean, variance matrix* $\Sigma_\epsilon > 0$ *and finite fourth order cumulants.*

(c) $E(\psi_{t,j}) = E(\varepsilon_{t+j} \otimes u_{2t}) = 0$ *for all* $j \geq 0$.

Assumption 6 (EC) ensures the following functional central limit theorem (FCLT) for v_t to hold:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} v_t \xrightarrow{d} B(r) \equiv BM(\Omega), \text{ for } r \in [0, 1],$$

where $\Omega = C(1)\Sigma_\epsilon C(1)^{'}$ is the long-run variance matrix of v_t. We use $\Sigma = E\left(v_0 v_0^{'}\right)$ to denote the variance matrix of v_t. The variance matrix Σ and long-run variance matrix Ω of v_t are partitioned into cell submatrices Σ_{ij} and Ω_{ij} (i, j=0, 1, 2) conformably with v_t. The Brownian motion B(r) can be partitioned into cell vectors $B_i(r)$ (i=0, 1, 2) similarly. We also have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_{t,0} \xrightarrow{d} N(0, \Omega_{\psi\psi}), \quad \Omega_{\psi\psi} = \sum_{j=-\infty}^{\infty} E(\varepsilon_t \varepsilon_{t+j}^{'} \otimes u_{2t} u_{2t+j}^{'}).$$

The one-sided long-run covariances are defined as

$$\Lambda = \sum_{k=1}^{\infty} E\left(v_k v_0^{'}\right),$$

and

$$\Delta = \Sigma + \Lambda = \sum_{k=0}^{\infty} E\left(v_k v_0^{'}\right),$$

which can also be partitioned into cell submatrices conformably with v_t.

The approach we are following requires the estimation of both Ω and Δ, which is typically achieved by kernel smoothing of the component sample autocovariances. Since factors are unobservable, the sample autocovariances depend on estimated factors. Kernel estimates of Ω and Δ take the following general form (see, e.g., Priestley (1981))

$$\hat{\Omega} = \sum_{j=-T+1}^{T-1} \omega(j/K)\hat{\Gamma}(j), \text{ and } \hat{\Delta} = \sum_{j=0}^{T-1} \omega(j/K)\hat{\Gamma}(j), \tag{10}$$

where $\omega(\cdot)$ is a kernel function and K is a bandwidth parameter, with truncation in the sums given above occurs when $\omega(j/K) = 0$ for $|j| \geq K$. The sample covariances in (10) are given by

$$\hat{\Gamma}(j) = T^{-1} \sum_{1 \leq t, t+j \leq T} \hat{v}_{t+j}\hat{v}_t^{'},$$

where $\hat{v}_t = \left(\hat{\varepsilon}_t, \hat{u}_{1t}^{'}, \hat{u}_{2t}^{'}\right)^{'}$, $\hat{u}_{1t} = \hat{F}_{1t} - \hat{F}_{1,t-1} = A_1^{'}\Delta\hat{F}_t$, $\hat{u}_{2t} = \hat{F}_{2t} = A_2^{'}\hat{F}_t$, and $\hat{\varepsilon}_t$ is the residual from a preliminary least squares regression of y_t on $\hat{F}_t$. Again, $\hat{\Omega}$ and $\hat{\Delta}$ can be partitioned into cell submatrices conformably with v_t.

We also define $u_{at} = (u_{1t}^{'}, \Delta u_{2t}^{'})^{'} = AHf_t$, where the subscript "a" is denoting the elements corresponding to u_{1t} and Δu_{2t}, which occur after the rotation A is taken. Similarly, the long-run covariance matrices Ω_{0a}, Ω_{aa}, Δ_{0a}, Δ_{aa} and their kernel estimates are defined in terms of the autocovariances and sample autocovariances of u_{at}. As pointed out in Phillips (1995), the submatrix of Ω_{aa} corresponding to the difference Δu_{2t}, i.e. $\Omega_{\Delta u_2 \Delta u_2}$, is a zero matrix, since Δu_2 is an I(-1) process and therefore has zero long-run variance. By the same reasoning, the submatrix of Ω_{0a}, viz. $\Omega_{0\Delta u_2}$, is also a zero matrix. The presence of some stationary components (viz. F_{2t}) in the regression equation (1) leads to these degeneracies in the long-run covariance matrices Ω_{0a} and Ω_{aa}. One thing to keep in mind is that since we assume that the rotation matrix A is unknown beforehand as in Phillips (1995), the kernel estimates that the Fully-Modified approach relies on, $\hat{\Omega}_{0f}$ and $\hat{\Omega}_{ff}$, are kernel estimates of the long-run covariances $\Omega_{0f} = lrcov(\varepsilon_t, \Delta HF_t)$ and $\Omega_{ff} = lrcov(\Delta HF_t, \Delta HF_t)$. These kernel estimates and long-run covariances are the same as those of Ω_{0a} and Ω_{aa} after transformation by A. Because of the degeneracies in the long-run covariance matrices, the limit behavior of the kernel estimates of these matrices needs to be handled carefully. (In the proof, we borrow some results from Lemma 8.1 in the Appendix of Phillips (1995).)

We use the same class of admissible kernels as in Phillips (1995).

Assumption 7 *(KL–Kernel Condition) The kernel function* $\omega(\cdot)$*:* $\mathbb{R} \to [-1, 1]$ *is a twice continuously differentiable even function with*

(a) $\omega(0) = 1$, $\omega^{'}(0) = 0$, $\omega^{''}(0) \neq 0$*; and either*

(b) $\omega(x) = 0$ *for* $|x| \geq 1$*, with* $lim_{|x|\to 1}\omega(x)/(1-|x|)^2$*=constant, or*

(b') $\omega(x) = O(x^{-2})$*, as* $|x| \to 1$*.*

Under Assumption 7 (KL) we have

$$\lim_{x\to 0}(1-\omega(x))/x^2 = -(1/2)\omega^{''}(0),$$

and thus the characteristic exponent (r) of the kernel $\omega(x)$ as defined in Parzen (1957) is $r = 2$. Under Assumption 7 (KL) with (a) and (b) come the commonly used Parzen and Tukey-Hanning kernels, and under Assumption 7 (KL) with (a) and $(b^{'})$ comes the Bartlett-Priestley or quadratic spectral kernel (Priestley 1981, p.463).

The bandwidth expansion rate of $K = K(T)$ as $T \to \infty$ are defined according to Phillips (1995):

Definition 1 *(expansion rate order symbol* O_e*): For some* $k > 0$ *and for* K *monotone* *increasing in T we write*

$$K = O_e(T^k) \text{ if } K \sim c_T(T^k) \text{ as } T \to \infty,$$

where c_T is slowly varying at infinity (i.e., $c_{Tx}/c_T \to 1$ as $T \to \infty$ for $x > 0$).

Using this notation we outline a set of conditions on the bandwidth expansion rate as $T \to \infty$.

Assumption 8 *(BW–Bandwidth Expansion Rate). The bandwidth parameter K in the kernel estimates (10) has an expansion rate of the form*

$$BW(i). \quad K = O_e(T^k) \text{ for some } k \in (1/4, 2/3);$$

i.e., $K \sim c_T(T^k)$ for some slowly varying function c_T and thus $K/T^{2/3} + T^{1/4}/K \to 0$ and $K^4/T \to \infty$ as $T \to \infty$. Some of our results require other bandwidth expansion rates which we designate as

$$BW(ii). \quad K = O_e(T^k) \text{ for some } k \in (0, 2/3),$$

$$BW(iii). \quad K = O_e(T^k) \text{ for some } k \in (1/4, 1),$$

$$BW(iv). \quad K = O_e(T^k) \text{ for some } k \in (0, 1).$$

As will be shown in Theorem 1 of this paper, Assumption 8 (BW) is not enough to guarantee the consistency of the kernel estimates when the regressors involve estimation errors. In the estimated factor context, an extra condition requiring that the estimation errors in the factors do not accumulate at a rate faster than the expansion rate of the bandwidth K should be imposed.

3 Inference with Estimated Factors

3.1 OLS estimation

Recall the regression equation given in (1):

$$y_t = \alpha' F_t + \varepsilon_t.$$

Let $\hat{\delta}$ be the least squares estimates of the regression of y_t on $\hat{F}_t$ (given in equation (7)) for $t = 1, ..., T$. The OLS estimates can be written as $\hat{\delta} = (\hat{F}'\hat{F})^{-1}\hat{F}'Y$ in which $Y = (y_1, ..., y_T)'$ and $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)'$. Define $\delta = H^{-1'}\alpha$. Denote $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$, $F_1 = FH'A_1$, $F_2 = FH'A_2$, $\hat{F}_1 = \hat{F}A_1$, and $\hat{F}_2 = \hat{F}A_2$.

Lemma 1 *Suppose Assumptions 1-5 and 6 (EC) hold. As $(n, T) \to \infty$, if $T/\sqrt{n} \to 0$,*

(a) $TA_1'(\hat{\delta} - \delta) \xrightarrow{d} (\int B_1 B_1')^{-1}(\int_0^1 B_1 dB_0 + \Delta_{10})$,

(b) $\sqrt{T}A_2'(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1})$.

This lemma establishes the consistency of the feasible OLS estimator using estimated factors and the different converging speeds of the nonstationary coefficient estimator and stationary coefficient estimator. As observed in Vogelsang and Wagner (2014), when ε_t is uncorrelated with u_{1t} and hence uncorrelated with F_{1t}, we have (i) $\Delta_{10} = 0$, and (ii) $B_0(r)$ is independent of $B_1(r)$. Because of the independence between $B_0(r)$ and $B_1(r)$ in this case, the limiting distribution of $TA_1'(\hat{\delta} - \delta)$ is a zero mean Gaussian conditioning on $B_1(r)$. Therefore, the t and Wald statistics for testing hypotheses about $A_1'\delta$ have the usual $N(0, 1)$ and chi-squared limits when consistent robust standard errors are used to handle the serial correlation in ε_t.

When the factors are endogenous, the limiting distribution of $TA_1'(\hat{\delta} - \delta)$ is non-standard given the correlation between $B_0(r)$ and $B_1(r)$ and the presence of the nuisance parameters in the vector Δ_{10}. No asymptotic normal result can be obtained conditioning on $B_1(r)$, and the asymptotic bias introduced by Δ_{10} make this limiting distribution more complicated. Inference is difficult in this situation because nuisance parameters cannot be removed by simple scaling methods.

Phillips and Hansen (1990) and Phillips (1995) develop the FM-OLS estimator to remove Δ_{10} and to deal with the correlation between $B_0(r)$ and $B_1(r)$ in the above limiting distribution. The key component in this FM-OLS estimator is to construct a stochastic process independent of $B_1(r)$ as follows:

$$B_{0\cdot 1} = B_0 - \Omega_{01}\Omega_{11}^{-1}B_1 \equiv BM(\sigma_{00\cdot 1}^2),$$

where $\sigma_{00\cdot 1}^2 = \Omega_{00} - \Omega_{01}\Omega_{11}^{-1}\Omega_{10}$. This stochastic process is independent of $B_1(r)$ by construction. Using $B_{0\cdot 1}(r)$, we can write

$$\int_0^1 B_1(r)dB_0(r) + \Delta_{10} = \int_0^1 B_1(r)dB_{0\cdot 1}(r) + \int_0^1 B_1(r)dB_1'(r)\Omega_{11}^{-1}\Omega_{10} + \Delta_{10}.$$

Because $B_1(r)$ and $B_{0\cdot 1}(r)$ are independent, we can show that $\int_0^1 B_1(r)dB_{0\cdot 1}(r)$ is a zero mean Gaussian mixture conditioning on $B_1(r)$. As is clear from the above expression, the FM-OLS estimator rests upon two transformations, with one transformation removing the term $\int_0^1 B_1(r)dB_1'(r)\Omega_{11}^{-1}\Omega_{10}$ and the other removing Δ_{10}. Because these terms depend on Ω and Δ, the two transformations require estimates of Ω and Δ_{10}. As shown in the next section, when factors are latent and are estimated from the large panel of integrated dataset, the consistency of the estimates of Ω and Δ_{10} require extra conditions on the bandwidth expansion rate and the sample sizes T and n.

3.2 The FM-OLS estimation

As in Phillips (1995), the FM-OLS estimator given below is constructed by making corrections for endogeneity and for serial correlation to the least squares estimator $\hat{\delta} = (\hat{F}'\hat{F})^{-1}\hat{F}'Y$. For the endogeneity correction, the variable y_t is modified with the transformation

$$y_t^+ = y_t - \hat{\Omega}_{0\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\Delta\hat{F}_t = y_t - \hat{\Omega}_{0\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{f}_t.$$

In this transformation, $\hat{\Omega}_{0\hat{f}}$ and $\hat{\Omega}_{\hat{f}\hat{f}}$ are kernel estimates of the long-run covariances $\Omega_{0f} = lrcov(\varepsilon_t, \Delta HF_t) = lrcov(\varepsilon_t, Hf_t)$ and $\Omega_{ff} = lrcov(\Delta HF_t, \Delta HF_t) = lrcov(Hf_t, Hf_t)$ taking forms

$$\hat{\Omega}_{0\hat{f}} = \sum_{j=-T+1}^{T-1} \omega(j/K)\hat{\Gamma}_{0\hat{f}}(j), \text{ and } \hat{\Omega}_{\hat{f}\hat{f}} = \sum_{j=-T+1}^{T-1} \omega(j/K)\hat{\Gamma}_{\hat{f}\hat{f}}(j),$$

where $\hat{\Gamma}_{0\hat{f}}(j) = T^{-1}\sum_{1\leq t,t+j\leq T}\hat{\varepsilon}_{t+j}\hat{f}_t^{'}$, and $\hat{\Gamma}_{\hat{f}\hat{f}}(j) = T^{-1}\sum_{1\leq t,t+j\leq T}\hat{f}_{t+j}\hat{f}_t^{'}$, where $\hat{f}_t$ are the principal component estimates of factors of the first-differenced factor model and $\hat{\varepsilon}_{t+j}$ is the residual from a preliminary least squares regression of y_t on $\hat{F}_t$. Recalling that $A\Delta HF_t = (u_{1t}^{'}, \Delta u_{2t}^{'})^{'} = u_{at}$, we have

$$\hat{\Omega}_{0\hat{a}} = \hat{\Omega}_{0\hat{f}}A^{'}, \text{and } \hat{\Omega}_{\hat{a}\hat{a}} = A\hat{\Omega}_{\hat{f}\hat{f}}A^{'},$$

where $\hat{\Omega}_{0\hat{a}} = \sum_{j=-T+1}^{T-1}\omega(j/K)\hat{\Gamma}_{0\hat{a}}(j)$, $\hat{\Omega}_{\hat{a}\hat{a}} = \sum_{j=-T+1}^{T-1}\omega(j/K)\hat{\Gamma}_{\hat{a}\hat{a}}(j)$, with $\hat{\Gamma}_{0\hat{a}}(j) = T^{-1}\sum_{1\leq t,t+j\leq T}\hat{\varepsilon}_{t+j}\hat{u}_{at}^{'}$, and $\hat{\Gamma}_{\hat{a}\hat{a}}(j) = T^{-1}\sum_{1\leq t,t+j\leq T}\hat{u}_{a,t+j}\hat{u}_{at}^{'}$, where $\hat{u}_{at} = A\hat{f}_t$ and $\hat{\varepsilon}_{t+j}$ is the residual from a preliminary least squares regression of y_t on $\hat{F}_t$.

The purpose of the endogeneity correction is to deal with endogeneity in the regressors F_t associated with any cointegrating links between y_t and F_t. Since factors F_t are unobservable, we use principal component estimates of the factors, $\hat{F}_t$, and estimated $\hat{u}_t$ to form this transformation. This highlights the major difference from the FM-OLS considered in Phillips (1995). As will be shown shortly, errors introduced by the estimation of factors are negligible in the asymptotic distribution of the FM-OLS estimator provided that the cross-sectional sample size n is large enough relative to the time series sample size T and the bandwidth K asymptotically.

The serial correlation correction term takes the form

$$\hat{\Delta}^{+}_{\hat{f}0} = \hat{\Delta}_{\hat{f}0} - \hat{\Delta}_{\hat{f}\hat{f}}\hat{\Omega}^{-1}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}0}, \tag{11}$$

where $\hat{\Delta}_{\hat{f}0}$ and $\hat{\Delta}_{\hat{f}\hat{f}}$ are kernel estimates of the one-sided long-run covariances $\Delta_{f0} = lrcov_{+}(\Delta HF_t, \varepsilon_t) = lrcov_{+}(Hf_t, \varepsilon_t)$ and $\Delta_{ff} = lrcov_{+}(\Delta HF_t, \Delta HF_t) = lrcov_{+}(Hf_t, Hf_t)$ taking forms

$$\hat{\Delta}_{\hat{f}0} = \sum_{j=0}^{T-1} \omega(j/K)\hat{\Gamma}_{\hat{f}0}(j), \text{ and } \hat{\Delta}_{\hat{f}\hat{f}} = \sum_{j=0}^{T-1} \omega(j/K)\hat{\Gamma}_{\hat{f}\hat{f}}(j).$$

This correction is to deal with the effects of serial covariance in the shocks u_{1t} that drive the nonstationary regressors $F_{1t} = A_1{}'HF_t$ and any serial covariance between the equation error ε_t and the past history of u_{1t}. By the same taken as above, we have

$$\hat{\Delta}_{0\hat{a}} = \hat{\Delta}_{0f}A', \text{and } \hat{\Delta}_{\hat{a}\hat{a}} = A\hat{\Delta}_{\hat{f}\hat{f}}A'.$$

Combining the endogeneity and serial correlation corrections we have the FM-OLS regression formula

$$\hat{\delta}_{FM} = (\hat{F}'\hat{F})^{-1}(\hat{F}'Y^{+} - T\hat{\Delta}^{+}_{\hat{f}0}).$$

As pointed out in Phillips (1995), the sample moment matrices of the data and their orders of magnitude (which depend on the directions of stationarity and nonstationarity in the regressors) are the keys in deriving a limit theory $\hat{\delta}_{FM}$. Meanwhile, the behavior of the kernel estimates $\hat{\Delta}_{\hat{f}0}$, $\hat{\Delta}_{\hat{f}\hat{f}}$, $\hat{\Omega}_{0\hat{f}}$, and $\hat{\Omega}_{\hat{f}\hat{f}}$ that appear in the correction terms of $\hat{\delta}_{FM}$ are also in a need of special attention. The latter is especially important because the kernel estimator $\hat{\Omega}_{\hat{f}\hat{f}}$ tends to a singular limit due to the fact that $\Omega_{f_2f_2} = A_1'\Omega_{ff}A_1 = 0$ (because of the presence of stationary components (viz., $\hat{F}_{2t}$) in the regressors $\hat{F}_t$). The technical Lemmas A.4, A.5, and A.6 in the Appendix enable us to take this singularity into account in the asymptotic analysis and determine what impact it has on the asymptotic behavior of the estimator $\hat{\delta}_{FM}$ in both stationary and nonstationary directions. In this regard, the bandwidth expansion rate of K turns out to be very important.

One more complexity this paper involves is that the regressors F_t are unobservable and estimated from a factor model. The estimated factors innovations, $\hat{u}_t$, and the residuals from a preliminary least squares regression of y_t on $\hat{F}_t$, $\hat{\varepsilon}_t$, involve errors from the estimation of the factors. So the fully modified transformations and the kernel estimates of the long-run variance-covariance matrices involve estimation errors from the estimated factors. To guarantee the estimation errors in the factors do not contaminate the asymptotic properties of the FM-OLS estimator, we need more strict restrictions on the bandwidth expansion rate of K than in Phillips (1995), and more strict restrictions on the relative expansion rate of the cross sectional and time series sample sizes n and T than in Bai and Ng (2004).

The following theorem outlines our main results of the FM-OLS estimators when the regressors are latent and estimated factors are used for estimation.

Theorem 1 *Under Assumptions 1-5, 6 (EC), 7 (KL), and 8 (BW),*

(a) under the assumption that $K = O_e(T^k)$ *for some* $k \in (0, 2/3)$, $K\sqrt{T/n} \to 0$, *and* $T/\sqrt{n} \to 0$ *as* $(n, T) \to \infty$, *we have*

$$TA_1'(\hat{\delta}_{FM} - \delta) \xrightarrow{d} (\int B_1 B_1')^{-1} \int_0^1 B_1 dB_{0\cdot 1};$$

(b) under the assumption that $K = O_e(T^k)$ *for some* $k \in (1/4, 2/3)$, $K^{3/2}\sqrt{T/n} \to 0$, *and* $T/\sqrt{n} \to 0$ *as* $(n, T) \to \infty$,

$$\sqrt{T}A_2'(\hat{\delta}_{FM} - \delta) \xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1}),$$

where $B_{0\cdot 1} = B_0 - \Omega_{01}\Omega_{11}^{-1}B_1 \equiv BM(\sigma^2_{00\cdot 1})$ *in which* $\sigma^2_{00\cdot 1} = \Omega_{00} - \Omega_{01}\Omega_{11}^{-1}\Omega_{10}$.

Notice that the assumption $K = O_e(T^k)$ for some $k \in (0, 2/3)$ as $(n, T) \to \infty$ is the same as in Phillips (1995) for the nonstationary coefficient estimates. However, we require the extra condition that $K\sqrt{T/n} \to 0$ in addition to the condition that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$. In Lemma 8.1 of Phillips (1995, p.1058), which shows the consistency of the kernel estimates with observable regressors, the only requirement on the bandwidth expansion rate is the one stated in Assumption 8 (BW). But with estimation errors in the factors (converge at rate $O_p(\sqrt{T/n})$), the induced errors in the kernel estimates will accumulate at rate $O_p(K\sqrt{T/n})$. Thus in order to guarantee the consistency of the kernel estimates, the extra restriction $K\sqrt{T/n} \to 0$ should be imposed. In another words, using estimated factors does not affect the consistency of the kernel estimates as long as the estimation errors of the factors converge to zero fast enough relative to the bandwidth expansion rate.

For the stationary coefficient estimates, the assumption $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$ as $(n,T) \to \infty$ is tighter than that $K = O_e(T^k)$ for some $k \in (1/4, 1)$ as $T \to \infty$ in Phillips (1995). This tighter bandwidth expansion rate comes from the accumulation of estimation errors in the factors across the summation of K sample covariances. Lemma A.6 (b) gives the stationary coefficient correction more explicitly (and when it is scaled by $T^{1/2}$), with the correction term in this case having magnitude $O_p(T^{1/2}/K^2) + O_p(1/\sqrt{K}) + O_p(T/\sqrt{n}) + O_p(K^{3/2}/T) + O_p(K^{3/2}\sqrt{T/n})$. The correction term is $o_p(1)$ when the bandwidth expansion rate $K = O_e(T^k)$ satisfies $1/4 < k < 2/3$ and $K^{3/2}\sqrt{T/n} \to 0$. To guarantee the estimation error in the factors does not contaminate the limiting behavior of the long-run covariance estimates, we do not allow the Bandwidth expansion rate to be too large.

We also impose the more strict relative expansion rate $K^{3/2}\sqrt{T/n} \to 0$ for the stationary FM estimates than for the nonstationary FM estimates (which only requires $K\sqrt{T/n} \to 0$, which is needed in the consistency of the long-run covariance estimates $\hat{\Omega}_{0\hat{a}}$). This condition $K^{3/2}\sqrt{T/n} \to 0$ could be written as $\sqrt{T^3/n} \to 0$ since $O_p(K^{3/2}/T) = o_p(1)$ under the assumption that $K = O_e(T^k)$ satisfies $1/4 < k < 2/3$. This bandwidth expansion rate along with the extra requirement that $K^{3/2}\sqrt{T/n} \to 0$ is different than that in Phillips (1995) because of the extra error terms $O_p(T/\sqrt{n}) + O_p(K^{3/2}/T) + O_p(K^{3/2}\sqrt{T/n})$ in the correction expression. These terms are the results of the estimation error in the factors. In order to guarantee that the estimation error in the factors does not contaminate the limiting behavior of the FM estimates, we need more strict requirement on the relative rate of the bandwidth expansion rate, the cross sectional and time series sample sizes, i.e., $K^{3/2}\sqrt{T/n} \to 0$ as $(n,T) \to \infty$.

Assumptions 1-5 are concerned with the consistency of the principal component estimates of the factors while Assumptions 6 (EC), 7 (KL), and 8 (BW) are concerned with the consistency of kernel estimates of Ω and Δ. Bai and Ng (2004) shows that under Assumptions 1-5, the principal component estimates of the factors, $\hat{F}_t$, are consistent for the true factors F_t up to a rotation H, and the time average of the squared estimation errors converges to zero when as $T/n \to 0$ as $(n,T) \to \infty$. When regressors are observable, Assumptions 6 (EC), 7 (KL), and 8 (BW) guarantee that the kernel estimates $\hat{\Omega}$ and $\hat{\Delta}$ are consistent (Andrews 1991; Phillips 1995). When regressors are unobservable and estimated factors are used to form the kernel estimates, additional restrictions should be imposed to guarantee that estimation errors from the regressors do not impact the consistency of the kernel estimates. The two extra conditions $K\sqrt{T}/\sqrt{n} \to 0$ and $K^{3/2}\sqrt{T}/\sqrt{n} \to 0$ serve this purpose since it indicates that estimation errors from the factors should converge to zero fast enough relative to the expansion rate of the bandwidth K.

A consistent estimator for the asymptotic variance of $\hat{\delta}_{FM}$ is

$$A\hat{v}ar(\hat{\delta}_{FM}) = (D_T^{-1}\hat{F}'\hat{F}D_T^{-1})^{-1} \begin{pmatrix} \frac{\hat{\sigma}^2_{00\cdot 1}}{T^2}\sum_{t=1}^{T-h}\hat{F}_{1t}\hat{F}'_{1t} & 0 \\ 0 & \frac{1}{T}\sum_{t=1}^{T-h}\hat{\varepsilon}_t^2\hat{F}_{2t}\hat{F}'_{2t} \end{pmatrix} \cdot (D_T^{-1}\hat{F}'\hat{F}D_T^{-1})^{-1} \tag{12}$$

where $D_T = diag(TI_{r_1}, \sqrt{T}I_{r_0})$, and $\hat{\sigma}^2_{00\cdot 1} = \hat{\Omega}_{00} - \hat{\Omega}_{01}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10}$. Asymptotically pivotal t and Wald statistics with $N(0,1)$ and chi-squared limiting distributions can then be constructed.

As discussed in Bai and Ng (2004), the factor model is unidentified because $\alpha^{'}LL^{-1}F_t = \alpha^{'}F_t$ for any invertible matrix L. The above theorem is a result pertaining to the difference between $\hat{\delta}_{FM}$ and the space spanned by δ. Consistency of the FM estimators follows from the fact that the averaged squared deviations between $\hat{F}_t$ and $H\hat{F}_t$ vanish as n and T both tend to infinity. Furthermore, having estimated endogenous I(1) factors as regressors does not affect the consistency of the FM parameter estimates.

4 Cointegration Tests

Before running the regression equation in (1), it is always desirable to test for the cointegration between the observable nonstationary series y_t and the set of possibly cointegrated latent factors F_t in the first place. In this section, we discuss how to test for cointegration between y_t and F_t and establish the asymptotic properties of the residual-based cointegration test statistics. To test for cointegration relation between y_t and F_t, we can simply run the unit root test on the residuals from the OLS regression of y_t on F_t. Since the factors F_t are unobserved, we use the estimated factors $\hat{F}_t$ instead.

Let $\hat{\delta}$ be the least squares estimates of the regression of y_t on $\hat{F}_t$ for $t = 1, ..., T$. The OLS estimates can be written as $\hat{\delta} = (\hat{F}'\hat{F})^{-1}\hat{F}'Y$ in which $Y = (y_1, ..., y_T)^{'}$ and $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)^{'}$. Define $\delta = H^{-1'}\alpha$. Let $\hat{\varepsilon}_t$ denote the residuals from the OLS regression of y_t on $\hat{F}_t$ for $t = 1, ..., T$. Let $\hat{\rho}_T$ be the least squares estimates of the regression of $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}$ for $t = 2, ..., T$, which could be written as $\hat{\rho}_T = \frac{\sum_{t=2}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2}$. To derive the asymptotic property of the unit root test under the null hypothesis that there exists a unit root in ε_t, we have to modify the Assumption 6 (EC). Now define $u_t = (u_{1t}^{'}, u_{2t}^{'})^{'}$, $v_t = (\Delta\varepsilon_t, u_{1t}^{'}, u_{2t}^{'})^{'} = (\Delta\varepsilon_t, u_t^{'})^{'}$ and $\psi_t = \Delta\varepsilon_t \otimes u_{2t}$, in which $u_{1t} = \Delta F_{1t} = (I - L)A_1'HF_t = A_1'HC(L)\eta_t$ and $u_{2t} = F_{2t} = A_2'HF_t = A_2'H(F_0 + \sum_{s=1}^{t} C(L)\eta_s)$. We assume that v_t is a linear process that satisfies the following assumption.

Assumption 9 *(EC'–Error Condition)*

(a) $v_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} C_j \epsilon_{t-j}$, $\sum_{j=0}^{\infty} j^a ||C_j|| < \infty$, $|C(1)| \neq 0$ *for some* $a > 1$.

(b) ϵ_t *is i.i.d. with zero mean, variance matrix* $\Sigma_\epsilon > 0$ *and finite fourth order cumulants.*

(c) $E(\psi_{t,j}) = E(\Delta\varepsilon_{t+j} \otimes u_{2t}) = 0$ *for all* $j \geq 0$.

Again, Assumption 9 (EC') ensures the validity of functional central limit theorem (FCLT) for v_t. Like the case in Assumption 6 (EC), we can partition the corresponding Brownian motion B(r) into cell vectors $B_i(r)$ (i=0, 1, 2). Under the null hypothesis that there exists a unit root in ε_t, the OLS regression of y_t on F_t is spurious and we have the following lemma:

Lemma 2 *Suppose Assumptions 1-5 and Assumption 9 (EC') hold. As* $(n, T) \to \infty$, *if* $T/n \to 0$,

(a) $A_1'(\hat{\delta} - \delta) \xrightarrow{d} (\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0)$,

(b) $A_2'(\hat{\delta} - \delta) \xrightarrow{d} \Sigma_{22}^{-1}(\int_0^1 dB_2 B_0 + \Delta_{20}) - \Sigma_{22}^{-1}(\int_0^1 dB_2 B_1 + \Delta_{21})(\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0)$.

As we can see from the above lemma, the OLS estimates are no longer consistent. In the following, we follow Hamilton (1994, Chapter 19) closely to construct the cointegration test. Notice that the main difference of the test in this paper is that the cointegration regression involves estimation errors in F_t and thus we need to derive the limiting distribution of the cointegration test statistics under the existence of these estimation errors.

Let s_T^2 be the OLS estimate of the variance of the residual $\varkappa_t$ for the regression

$$\hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + \varkappa_t, \text{ for t= 2, 3, ..., T},$$

yielding

$$s_T^2 = (T-2)^{-1} \sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\rho}_T \hat{\varepsilon}_{t-1})^2.$$

Let $\hat{\sigma}_{\hat{\rho}_T}$ be the standard error of $\hat{\rho}_T$ from the above regression:

$$\hat{\sigma}_{\hat{\rho}_T}^2 = s_T^2 \div \{\sum_{t=2}^{T} \hat{\varepsilon}_{t-1}^2\}.$$

Finally, let $\hat{c}_{j,T}$ be the jth sample autocovariance of the estimated residuals:

$$\hat{c}_{j,T} = (T-1)^{-1} \sum_{t=j+2}^{T} \hat{\varkappa}_t \hat{\varkappa}_{t-j} \text{ for j=0, 1, 2,..., T-2}$$

for $\hat{\varkappa}_t = \hat{\varepsilon}_t - \hat{\rho}_T \hat{\varepsilon}_{t-1}$; and let the square of $\hat{\lambda}_T$ be given by

$$\hat{\lambda}_T^2 = \hat{c}_{0,T} + 2 \sum_{j=1}^{q} [1 - j/(q+1)] \hat{c}_{j,T},$$

where q is the number of autocovariances to be used. The Phillips-Ouliaris Z_ρ statistic (1987) can be calculated as:

$$Z_{\rho,T} = T(\hat{\rho}_T - 1) - 1/2\{(T-1)^2 \hat{\sigma}_{\hat{\rho}_T}^2 \div s_T^2\}\{\hat{\lambda}_T^2 - \hat{c}_{0,T}\}.$$

If y_t and F_t are not cointegrated, then the regression of y_t and F_t is a spurious regression and $\hat{\rho}_T$ should be close to 1. On the other hand, if $\hat{\rho}_T$ is quite below 1, and the calculation of $Z_{\rho,T}$ yields a negative value with a large absolute value, then the null hypothesis that y_t and F_t are not cointegrated should be rejected. The following theorem provides a formal statement of the asymptotic distributions of the above test statistics.

Theorem 2 *Suppose Assumptions 1-5 and Assumption 9 (EC') hold. As $(n,T)\to\infty$, if $T/n\to 0$,*

(a) $T(\hat{\rho}_T-1)\xrightarrow{d}\frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0+\Lambda_{\tilde{0}\tilde{0}}}{\int_0^1 \tilde{B}_0\tilde{B}_0}$, *where* $\tilde{B}_0=B_0-(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}B_1$. *The Brownian motion* $\tilde{B}_0$ *has long-run covariance matrix*

$$\Omega_{\tilde{0}\tilde{0}}=\Omega_{00}-(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Omega_{10}-\Omega_{01}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0)$$
$$+(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Omega_{11}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0),$$

and one-sided long-run covariance

$$\Lambda_{\tilde{0}\tilde{0}}=\Lambda_{00}-(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{10}-\Lambda_{01}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0)$$
$$+(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{11}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0).$$

(b) If $q\to\infty$ *as* $T\to\infty$ *but* $q/T\to 0$, *then the statistic* $Z_{\rho,T}$ *satisfies* $Z_{\rho,T}\xrightarrow{d}Z_n$, *where*

$$Z_n=\frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0}{\int_0^1 \tilde{B}_0\tilde{B}_0}=\frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)W(r)dr},$$

in which $W(r)$ *is a one dimensional standard Brownian motion.*

Result (a) implies that $\hat{\rho}_T\xrightarrow{p}1$. When the regression of y_t on F_t is spurious, the estimated residuals will behave like a unit root process. The above results are similar to Proposition 19.4 of Hamilton (1994, Chapter 19) except the fact that we allow for the cointegration among the regressors F_t and we use the estimated factors $\hat{F}_t$ to estimate the cointegration regression and construct the cointegration test. Like the case in Proposition 19.4 of Hamilton (1994, Chapter 19), the limiting distribution of $T(\hat{\rho}_T-1)$ and $Z_{\rho,T}$ depend only on the number of stochastic explanatory factors in the cointegration regression (r_1). The above limiting distributions are derived under the case that there is no constant term appearing in the cointegration of y_t on F_t. So the critical values for the Phillips Z_ρ statistic can be found in Case 1 of Table B.8 in Hamilton (1994).

5 Application: Factor-Augmented Diffusion Index Forecasts

So far, we have constructed the FM-OLS estimates of the cointegration regression of an I(1) process y_t and some latent possibly cointegrated nonstatioanry factors F_t, and showed that the usual cointegration test works even under the case when the factors are estimated. In this section, we discuss possible applications of the FM-OLS estimates derived above to the Diffusion Index Forecasts literature and compare with forecasting with factor-augmented error correction models literature.

Usually, macroeconomic forecasting with a large set of possible predictors is done through adding factors to an otherwise standard forecasting model, such as "diffusion index forecast model" (DI) of Stock and Watson (2002a) and factor-augmented vector autoregressive (FAVAR) models of Bernanke, Boivin, and Bernanke et al. (2005). Under these methods, the large panel of data are transformed in the first place to get estimates of a much smaller number of stationary factors, and these estimated stationary factors are added to the forecasting equation of a properly transformed variable of interest. The estimation of the factors and the forecasting of the variable of interest are done in a stationary setting with all of the nonstationarity has been taken care of by taking logarithms, first-differencing or even twice differencing.

However, most macroeconomic variables are nonstationary in nature. To explore the nonstationarity and cointegration relations in the forecasting of a small number of nonstationary variables using a large panel of possibly nonstationary predictors, Banerjee and Marcellino (2009) suggested using factors extracted from large nonstationary panels in small-scale error correction models to control for long-run cointegration relations.

To facilitate the comparison of different forecasting methods, we take the FECM from Banerjee and Marcellino (2009) and repeat it here. Banerjee and Marcellino (2009) assume that there are n I(1) variables which can be partitioned into the n_A of major interest, x_{At}, and the $n_B = n - n_A$ remaining ones, x_{Bt}. The common trend specification of the factor model could be written as:

$$\begin{pmatrix} x_{At} \\ x_{Bt} \end{pmatrix} = \begin{pmatrix} \Psi_A \\ \Psi_B \end{pmatrix} f_t + \begin{pmatrix} u_{At} \\ u_{Bt} \end{pmatrix}, \tag{13}$$

where $u_t = (u_{At}^{'}, u_{Bt}^{'})^{'}$ is an n-dimensional vector of stationary errors, and f_t is a r-dimensional vector of uncorrelated I(1) common stochastic trends. From the above specification, all of the series in the panel, $x_t = (x_{At}^{'}, x_{Bt}^{'})^{'}$, are cointegrated with f_t. Especially, x_{At} and f_t are cointegrated. By the Granger representation theorem, we have the following error correction specification with added lagged terms to take care of series correlations in the errors:

$$\begin{pmatrix} \Delta x_{At} \\ \Delta x_{Bt} \end{pmatrix} = \begin{pmatrix} \gamma_A \\ \gamma \end{pmatrix} \delta^{'} \begin{pmatrix} x_{At-1} \\ f_{t-1} \end{pmatrix} + A_1 \begin{pmatrix} \Delta x_{At-1} \\ \Delta f_{t-1} \end{pmatrix} + ... + A_q \begin{pmatrix} \Delta x_{At-q} \\ \Delta f_{t-q} \end{pmatrix} + \begin{pmatrix} \epsilon_{At} \\ \epsilon_t \end{pmatrix}. \tag{14}$$

The above model is referred by Banerjee and Marcellino (2009) as the Factor-augmented Error Correction Model (FECM). Banerjee, Marcellino, and Masten (2014a) showed that FECM generally offers a better forecasting performance relative to both FAVARs and standard small-scale ECMs, in that FECMS nest both FAVARs and ECMs.

In this paper, we propose a forecasting method in which the cointegration could be taken into account by augmenting the DI forecasting of a first-differenced series with the error-correction terms estimated by the FM-OLS methods. To be more precise, consider the h-step-ahead DI forecast of Stock and Watson (2002a),

$$\Delta \hat{y}^h_{T+h|T} = \hat{c}_h + \sum_{j=1}^{k_1} \hat{\alpha}^{'}_{hj} \Delta y_{T-j+1} + \sum_{j=1}^{k_2} \hat{\beta}^{'}_{hj} \Delta \hat{F}_{T-j+1}. \tag{15}$$

We augment the above DI forecast with the error correction term $y_T - \hat{\delta}^{'}_{FM} \hat{F}_T$:

$$\Delta \hat{y}^h_{T+h|T} = \hat{c}_h + \hat{\gamma}_h (y_T - \hat{\delta}^{'}_{FM} \hat{F}_T) + \sum_{j=1}^{k_1} \hat{\alpha}^{'}_{hj} \Delta y_{T-j+1} + \sum_{j=1}^{k_2} \hat{\beta}^{'}_{hj} \Delta \hat{F}_{T-j+1}. \tag{16}$$

We call the above forecasting method Factor-augmented Diffusion Index Forecasts (FADI). The EC term, $y_T - \hat{\delta}^{'}_{FM} \hat{F}_T$, is included in the above forecasting equation only if there exists cointegration relation between the series of interest y_t and the vector of latent factors F_t. So in the implementation of the above forecasting method, we first test for cointegration relation between the series of interest y_t and the vector of latent factors F_t, and then form forecasts based on the above equation if there exists cointegration relation. Otherwise, forecasts are based on the usual DI forecasting method.

The proposed FADI method may look like the FECM method proposed in Banerjee and Marcellino (2009) at the first glance. However, there are several main differences between our FADI method and the FECM. Firstly, the factor model based on which the nonstationary factors are estimated are different. In our paper, we allow for the nonstationarity in the idiosyncratic components of the factor model (i.e., some of u_{Bt} in (13) can be I(1)), while in Banerjee and Marcellino (2009), the idiosyncratic components are all assumed to be stationary. To be more specific, the factor model we adopt here is from Bai and Ng (2004), while Banerjee and Marcellino (2009) assume the factor model in Bai (2004). Given the number of nonstationary series in the large panel (n) is large, the assumption that all of the idiosyncratic components are stationary is not realistic. It is more pragmatic to allow for idiosyncratic source of nonstationarity in the large panel.

The second main difference between our FADI method and the FECM is that we allow for possible cointegration among the factors. In the FECM of Banerjee and Marcellino (2009), all of the nonstationary factors are assumed to be uncorrelated random walks. On the contrary, in our FADI method, the nonstationary factors could be cointegrated to some extent. The allowance of cointegration among I(1) factors is equivalent to allow for the existence of nonstationary as well as stationary factors in the factor model. In the empirical applications of FECM, Banerjee, Marcellino, and Masten (2014a) consider a modification of the FECM, denoted FECMc, with FECM augmented with common factors extracted from the stationary component of x_t after the I(1) factors f_t and their corresponding loadings have been estimated. Their consideration of the possible stationary component of x_t and the evidence of this extra stationary factor in their example highlight the necessity of allowing cointegration among the factors. In this sense, our method could nest both FECM amd FECMc considered in Banerjee, Marcellino, and Masten (2014a) from a theoretical framework.

Finally, in the FADI method we use the FM-OLS estimator of the cointegration vector among y_t and the latent factors f_t in the EC term. The FM-OLS estimator corrects the second-order bias resulting from the existence of serial correlation and correlations of the innovations in the variable of interest, ε_t, and the innovations in the latent factors, u_t. More importantly, we test for the existence of the cointegration relation between the series y_t and the vector of latent factors F_t in the first place and include this extra EC term in the forecasting equation if there exists such cointegration relation. If there appears no cointegration between the series y_t and the vector of latent factors F_t, we do not include this EC term in the forecasting equation. In the FECM model of Banerjee and Marcellino (2009), since they assume stationary idiosyncratic components in the factor model and they are considering forecasting of some set of variables from this large integrated panel, the cointegration relation between the series of interest and the vector of latent factors are assumed implicitly. However, as pointed out above, it is more realistic to allow for nonstationary idiosyncratic components in the large integrated panel, and hence it is necessary to test for cointegration relation between the series of interest and the vector of latent factors in the first step. We will demonstrate these points using US macroeconomic data in the following section.

6 Empirical Example: Testing Cointegration of Stock and Watson (2005)

In this section, we take a large panel of monthly US macroeconomic variables from Stock and Watson (2005) to analyze the source of nonstationarity in the panel and study possible cointegration relations among the series in the large panel and the factors. The data set in Stock and Watson (2005) records monthly observations on 132 U.S. macroeconomic time series from 1959:1 through 2003:12, with 14 categories' predictors ranging from real output and income to price indexes and miscellaneous. Banerjee, Marcellino, and Masten (2014a) use this data set to simulate real-time forecasting using the FECM model as discussed in the previous section. For their FECM model to hold, they have to assume that the idiosyncratic error terms of the large panel must be stationary and they do not allow for cointegration among the nonstationary factors themselves. However, in their empirical example, they do not provide any testing result verifying all of these assumptions. In this section, we are going to test the stationarity assumptions of the idiosyncratic error terms and study the cointegration relation among the factors themselves as well as the cointegration relation among the factors and some variable of interests in the large panel.

As in Banerjee, Marcellino, and Masten (2014a), we retain only 104 series that were considered as I(1) by Stock and Watson and focus on the sample period of 1960:1 to 1998:12. Instead of transforming all of these series into approximate stationary series as in

Stock and Watson (2005), we follow appropriate steps to transform all of these series to I(1) series. In general, logarithms are used for real quantity variables, levels are used for nominal interest rates, and first differences of logarithms (growth rates) for price series. Specific transformations and the list of series are given in Table 17 of Banerjee, Marcellino, and Masten (2014a).

From the dataset of I(1) variables, we estimate the I(1) factors using the method in Bai and Ng (2004), and test for unit roots in the idiosyncratic errors and the estimated factors. To be more specific, we take the first difference of the nonstationary panel X, and apply the principal component method to the first differenced panel ΔX to get the factor estimates $\hat{f}_t$ and the loading estimates $\hat{\lambda}_i$, and then construct the factor estimates for the nonstationary panel as $\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s$ and estimate the idiosyncratic errors as $\hat{e}_{it} = X_{it} - \hat{\lambda}_i^{'} \hat{f}_t$. As in Bai and Ng (2004), let $ADF_{\hat{e}}^{c}(i)$ be the t statistic for testing $d_{i0} = 0$ in the univariate augmented autoregression (with no deterministic terms)

$$\Delta \hat{e}_{it} = d_{i0}\hat{e}_{it-1} + d_{i1}\Delta \hat{e}_{it-1} + ... + d_{ip}\Delta \hat{e}_{it-p} + error.$$

According to Bai and Ng (2004), the asymptotic distribution of $ADF_{\hat{e}}^{c}(i)$ is the same with the DF test developed by Dickey and Fuller (1979) for the case of no constant with -1.95 as the critical value at the 5% significance level. In the testing, the number of lagged differences, i.e., p in the above equation, is chosen by BIC criteria. The right panel of Table 1 summarizes the unit root testing results. For the 104 I(1) series in the large panel, 44 of them have unit roots in their idiosyncratic error terms. (Notice that even though the panel X is constructed to include only I(1) series, the unit root tests of X show that there are only I(1) 61 series in the panel.) The unit root testings of each factor estimate indicate that the estimated 5 factors are stationary. The individual unit root testings on the estimated factors may be size distorted and thus the number of nonstationary factors may be understated.

Under the assumption that all of the idiosyncratic errors in the large panel are stationary, we could also use the principal component-based estimator to the level of the data to estimate the I(1) factors as suggested in Bai (2004) and test for unit roots in the estimated idiosyncratic errors. The ADF test results in the left panel of Table 1 indicate that among all of the 104 series, the estimated idiosyncratic errors are nonstationary for 5 series. It seems that the assumption of stationary idiosyncratic errors are reasonable given the test results. However, if we use the method suggested in Bai and Ng (2004) of using differenced data to estimate the factors, the test results suggest 44 nonstationary idiosyncratic errors in the large panel. As discussed in Bai and Ng (2004), the estimation method of "differencing and recumulating" can accommodate both I(1) and I(0) errors. Thus the test results based on estimates of Bai and Ng (2004) should be more reliable than that based on estimates under the premise that the idiosyncratic errors are stationary. Thus there is evidence suggesting that the factor model assumption of Bai (2004) with stationary idiosyncratic errors is not appropriate for this Stock and Watson (2005) data set and the factor estimates using method from Bai (2004) may be misleading.

After verifying that there are unit roots in some of the idiosyncratic error terms of the large panel, we use the trace and maximal eigenvalue tests of Johansen (1988) to analyze the possible cointegration among the nonstationary factors. As discussed in Bai and Ng (2004), because $\hat{F}_t$ consistently estimates the space spanned by F_t, Johansen tests that assume F_t is observed remain valid when F_t is estimated using the "differencing and recumulating" method. The Johansen test results are summarized in Table 2. For the method of Bai (2004), given the number of factor is 4, the rank of cointegration among these factors is 1 by both tests, and hence the number of independent stochastic trends in the factors is 3. For the method of Bai and Ng (2004), given the number of factor is 5, the rank of cointegration among these factors is 4 by both tests, and hence the number of independent stochastic trends in the factors is 1. The cointegration test results indicate that there exists cointegration among estimated factors, no matter what the estimation method is. Hence, the factor model and the estimation method adopted in Banerjee, Marcellino, and Masten (2014a) that assume all of the factors are independent is not appropriate. Again, the estimation method of Bai and Ng (2004) which could accommodate cointegrated nonstationary factors shows its advantage over that in Bai (2004).

Next, we use the FADI forecasting method proposed in this paper to simulate real-time forecasts of four US real macroeconomic variables, i.e., Personal income less transfers (PI), Real manufacturing trade and sales (ManTr), Industrial Production (IP), and Employees on non-agriculture payrolls (Empl) over the sample 1970-1998, with estimation starting in 1960. As in Banerjee, Marcellino, and Masten (2014a), we use the iterated h-step-ahead forecasts (dynamic forecasts) instead of the direct h-step-ahead forecasts as in Stock and Watson (1998, 2002a, 2002b). The iterated h-step-ahead forecasts at time T of the FADI method is given by

$$\hat{y}^h_{T+h|T} = \hat{y}_T + \sum_{i=1}^{h} \Delta\hat{y}^1_{T+i|T+i-1}, \tag{17}$$

with

$$\Delta\hat{y}^1_{T+1|T} = \hat{c}_1 + \hat{\gamma}_1(y_T - \hat{\delta}^{'}_{FM}\hat{F}_T) + \sum_{j=1}^{k_1} \hat{\alpha}^{'}_{1j}\Delta y_{T-j+1} + \sum_{j=1}^{k_2} \hat{\beta}^{'}_{1j}\Delta\hat{F}_{T-j+1}. \tag{18}$$

The number of factors are kept fixed through all of the forecasting horizons with 4 independent I(1) factors in the nonstationary panel and 5 I(0) factors in the first-differenced panel. The factor estimates are updated recursively and model selection is conducted for each forecasting recursion. The above forecasting equation is estimated in two steps, with the cointegration relation estimated in the first step and the forecasting equation in the second step. In the first step, the cointegration test is conducted through the residual-based cointegration test at each forecasting recursion. In the second step, the lag lengths are selected based on the BIC at each forecasting recursion.

Before we conduct the recursive iterated forecasting, we use the whole sample period 1960:1-1998:12 to analyze the possible cointegration relation among these four real variables (PI, ManTr, IP, and Empl) and the factor estimates. These four variables are from the large panel X. From the unit root test results in Table 1, variables PI, ManTr, and IP exhibit unit roots while Empl may be considered as a stationary series. Also, the unit root testing of the idiosyncratic terms associated with there four real variables indicate that variables PI, ManTr, and Empl show cointegration relations with the non-stationary factors estimated by the method of Bai (2004). When the factors are estimated from Bai and Ng (2004), all of these four variables show cointegration relations with the factors.

In the first step of the FADI forecasting, to take advantage of all of the possible cointegration relations among these four real variables and the factors, we run the following least squares regressions and test the unit roots in the residuals:

$$y_{it} = \hat{\delta}' \hat{F}_t + \sum_{j=i+1}^{N} \gamma_j y_{jt} + \varepsilon_{it},$$

for $i = 1, 2, ..., N$ with N being the number of variables of interest, and

$$\hat{F}_{it} = \sum_{j=i+1}^{K} \kappa_j \hat{F}_{jt} + \tilde{\varepsilon}_{it},$$

for $i = 1, 2, ..., K - 1$ with K being the number of factors. In this empirical example, $y_t = (PI_t, ManTr_t, IP_t, Empl_t)'$, and thus the number of variables of interest is $N = 4$. Hence, in the first step, we run $N + K - 1$ least squares regressions and test the unit roots in the residuals. If the unit root testing suggests the residual is stationary, then we have found a cointegration relation among the variables involved in the regression and record the coefficients in the cointegration vector. The number of stationary residuals from the above $N + K - 1$ regressions is the cointegration rank r we use for the FADI method and we include all of the r error corrections terms in each forecasting equation of these four real variables.

We calculate the out-of-sample prediction mean squared errors (MSEs) of the FADI relative to the MSE of the AR at each horizon h for the real four variables under study, and list the relative MSEs of all of the models in Banerjee, Marcellino, and Masten (2014a) for comparison. Table 3 reports the forecasting results. The relative MSEs of the FADI method using factors estimated by the method in Bai (2004) are displayed in Column FADI, while the relative MSEs using factors estimated by the differencing and recummulating method in Bai and Ng (2004) are displayed in Column FADI2. By the unit root testing results in Table 1 and Table 2, our preferred method is FADI2 since the factors are estimated more consistently than that in FADI.

Comparing the relative MSEs of FADI to those of FECM, we find that the method FADI using factors from Bai (2004) rarely outperforms the method FECM. The relative MSEs for FADI is only smaller than those of FECM for 3 cases out of the 24 cases. The performances of FADI2 increase when we use the factors estimated by the method in Bai and Ng (2004). Even though the RMSEs are smaller only in 7 cases for FADI2, the forecasting are persistently better for variables IP and Empl at horizon h=1, 3, 6 and also better for Empl at h=12. As the forecasting horizon increases (for h=18 and 24), the FADI2 loses its forecasting advantage to FECM for all of the four variables. We also reestimate the FECM using factors estimates of Bai and Ng (2004), with Column FECM2 of Table 3 denoting the results. Generally speaking, FECM2 generates worse forecasting results than FECM, as found in Banerjee, Marcellino, and Masten (2014a). However, FADI2 generates better forecasting results than FADI in most of the cases. So consistent estimates of factors can improve the forecasting accuracy of the FADI method significantly.

A possible reason why FADI2 cannot improve over the FECM method at longer forecasting horizons for all of these four variables lies in the difference in cointegration estimation. The FECM method in Banerjee, Marcellino, and Masten (2014a) uses Johansen

(1988) Maximum Likelihood estimators of the cointegration ranks and the cointegration vectors, while our FADI method relies on least squares regressions and unit roots resting of the regression residuals. The last panel of Table 3 gives the information about the average cointegration ranks used by each method. The FADI method tends to overstate the number of cointegrations. The inclusion of extra error correction terms may lead to the under-performance of the method when forecasting horizon increases.

7 Conclusion

In this paper, we use FM-OLS method to directly estimate the cointegration relation between an integrated series of interest and a vector of possibly cointegrated nonstationary latent factors. Under some restrictions on the relative sample sizes, the kernel function, and the bandwidth expansion rates, we show that the estimation errors in the latent nonstationary factors do not affect the rate of convergence and the nuisance parameter-free limiting distribution of the FM-OLS estimators. Moreover, cointegration tests between the variable of interest and the vector of possibly cointegrated nonstationary latent factors have the usual limiting distributions when factors are consistently estimated. Given the existence of cointegration relation, the estimated cointegration relation can be used to form an error correction term, which could be added to the traditional diffusion index forecast model to improve forecasting accuracy.

Our empirical example on the Stock and Watson (2005) data set verifies that there are idiosyncratic nonstationarities in the nonstationary panel, and there are cointegration relations among these nonstationary factors themselves. Hence, the factor model in Bai and Ng (2005) is more appropriate for the Stock and Watson (2005) data set. We also show that the proposed Factor-Augmented Diffusion Index (FADI) forecasting method improves over the FECM method of Banerjee and Marcellino (2009) for variables Industrial Production (IP), and Employees on non-agriculture payrolls (Empl) at short horizons for forecasting period 1970-1998. Consistently estimated nonstationary factors improve the performance of the FADI method significantly. However, the overstated cointegration ranks by the FADI method may lead to inferior forecasting performance at longer forecasting horizons.

Acknowledgements

Chapter 3 is based on the working paper Kao, Lee, and Shen (2017).

Table 1: Unit root testing of Stock and Watson (2005): 1960:1-1998:12

Model	Bai (2004)			Bai and Ng (2004)		
	X_{it}	$\hat{F}_t$	$\hat{e}_{it}$	X_{it}	$\hat{F}_t$	$\hat{e}_{it}$
Number of series	104	4	104	104	5	104
Number of I(1) series	61	3	5	61	0	44

NOTE: X_{it} stands for the 104 I(1) variables from Stock and Watson (2005) for the time period of 1960:1 to 1998:12. Column Bai (2004) applies the principal component analysis to the level of X to get estimates $\hat{F}_t$ and $\hat{e}_{it}$, while Column Bai and Ng (2004) applies the principal component analysis to ΔX to get estimates $\hat{F}_t$ and $\hat{e}_{it}$. The number of factors are selected by the Bai and Ng (2002) PC2 criterion. The unit root tests of X_{it} and $\hat{F}_t$ are through the ADF regressions with a constant, while the unit root tests of $\hat{e}_{it}$ are through the ADF regressions without a constant. The numbers of lagged differences in the ADF regressions are selected by BIC criteria.

Table 2: Johansen tests of factors in Stock and Watson (2005): 1960:1-1998:12

Model	Tests stat.	r=0	r=1	r=2	r=3	r=4	Common Trends
Bai (2004)	trace	79.20***	23.12	8.1	1.49	-	3
	maximal	56.08***	15.01	6.61	1.49	-	3
Bai and Ng (2004)	trace	212.06***	123.84***	68.45***	30.61***	8.04	1
	maximal	88.22***	55.39***	37.85***	22.56***	8.04	1

NOTE: These above statistics are for testing the rank of cointegration in the estimated factors, with "trace" standing for the trace statistics and "maximal" for the maximal eigenvalue statistics of Johansen (1988). Test statistics in Column "r=0" are for the null hypothesis of the rank of cointegration among the factors is zero, etc.. In Row Bai (2004), factors are estimated by applying the principal component analysis to the level of X, and Row Bai and Ng (2004) applies the principal component analysis to ΔX to get estimates $\hat{F}_t$, while X stands for 104 I(1) series of Stock and Watson (2005) for the period of 1960:1 to 1998:12. The number of factors are selected by the Bai and Ng (2002) PC2 criterion. For the method of Bai (2004), given the number of factor is 4, the rank of cointegration among these factors is 1 by both tests, and hence the number of common trends in the factors is 3. For the method of Bai and Ng (2004), given the number of factor is 5, the rank of cointegration among these factors is 4 by both tests, and hence the number of common trends in the factors is 1. *** stands for 1 % significance level, with critical values from Table A2 of Johansen and Juselius (1990).

Table 3: Forecasting US real variables, forecasting period 1970-1998

h	log of	RMSE of AR	MSE relative to MSE of AR model								
			FAR	VAR	FAVAR	ECM	FECM	FADI	FECM2	FADI2	FECMc
1	PI	0.007	1.02	0.94	0.92	0.93	0.90	0.98	0.96	0.93	0.93
	ManTr	0.011	1.04	0.98	0.95	1.10	1.03	1.02*	1*	1.05	1.00
	IP	0.007	0.99	1.08	0.95	1.11	1.24	1.16*	1.19*	1.01*	1.15
	Empl	0.002	1.09	1.33	1.20	1.40	1.34	1.43	1.78	1.31*	1.40
3	PI	0.011	1.01	0.91	0.87	0.94	0.85	1.07	0.96	0.88	0.91
	ManTr	0.018	1.01	1.01	0.96	1.21	0.97	1.05	1.00	1.07	0.93
	IP	0.017	0.96	1.04	0.94	1.10	1.17	1.32	1.15*	1.07*	1.09
	Empl	0.005	1.12	1.51	1.40	1.64	1.52	1.81	1.98	1.39*	1.57
6	PI	0.016	1.00	0.94	0.92	1.02	0.86	1.24	0.99	0.93	0.95
	ManTr	0.029	1.01	1.01	0.98	1.17	0.89	1.09	1.00	1.03	0.87
	IP	0.029	0.97	1.00	0.96	1.08	1.08	1.33	1.06*	1.06*	1.02
	Empl	0.010	1.10	1.34	1.32	1.49	1.36	1.59	1.61	1.29*	1.37
12	PI	0.026	1.00	0.96	0.96	1.04	0.87	1.23	1.02	0.95	0.93
	ManTr	0.045	1.01	0.99	0.98	1.07	0.74	1.05	1.00	0.94	0.75
	IP	0.049	0.99	1.00	0.99	1.03	0.96	1.26	1.03	1.01	0.94
	Empl	0.020	1.02	1.11	1.12	1.25	1.10	1.2	1.23	1.09*	1.11
18	PI	0.036	1.01	0.98	0.98	1.09	0.89	1.15	1.01	0.99	0.96
	ManTr	0.058	1.00	1.00	0.99	1.06	0.71	1.01	1.00	0.94	0.73
	IP	0.065	1.00	1.00	1.00	1.08	0.93	1.25	1.01	1.04	0.96
	Empl	0.029	0.96	0.99	1.00	1.15	0.97	0.98	1.06	1.01	0.99
24	PI	0.042	1.01	0.99	0.99	1.07	0.90	1.09	1.01	0.99	0.96
	ManTr	0.069	1.01	1.00	1.01	0.99	0.64	0.9	1.00	0.89	0.66
	IP	0.076	1.01	0.99	1.00	1.07	0.90	1.2	0.99	1.01	0.95
	Empl	0.037	0.91	0.91	0.92	1.04	0.88	0.77*	0.96	0.89	0.91
Cointegration rank:				mean	min	max			mean	min	max
			FECM	3.75	1	4		FADI	5.97	4	7
			FECM2	4	4	4		FADI2	5.97	5	8

NOTE: h is the forecasting horizon. Results for model FAR, VAR, FAVAR, ECM, FECM, and FECMc are from Banerjee, Marcellino, and Masten (2014a). The FECM contains four I(1) factors, and FECMc contains five I(0) factors. FECM2 uses factor estimates of Bai and Ng (2004). FADI stands for the Factor-Augmented Diffusion Index method proposed in this paper using factor estimates of Bai (2004), while FADI2 is the FADI method using factor estimates of Bai and Ng (2004). Data: 1960:1-1998:12, forecasting: 1970:1-1998:12. Variables: Personal income less transfers (PI), Real manufacturing trade and sales (ManTr), Industrial Production (IP), and Employees on non-agriculture payrolls (Empl). Lag selection are based on the BIC. * stands for smaller RMSEs compared to Column FECM.

Appendix

Preliminaries for Lemma 1

As in Bai and Ng (2004), for notational simplicity, we assume there are T+1 observations (t=0, 1, ..., T) for this lemma. The differenced data have T observations so that x is $T \times n$. Let V_{nT} be the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(nT)^{-1}xx'$ in descending order. By the definition of eigenvectors and eigenvalues, we have $(nT)^{-1}xx'\hat{f} = \hat{f}V_{nT}$ or $(nT)^{-1}xx'\hat{f}V_{nT}^{-1} = \hat{f}$. We make use of an $r \times r$ matrix H defined as follows: $H = V_{nT}^{-1}(\hat{f}'\hat{f}/T)(\Lambda'\Lambda/N)$. Since the following proofs rely on results from Bai and Ng (2002), Bai (2003), and Bai and Ng (2004), we state some results of these papers explicitly as lemmas.

Lemma A.1

(Corresponds to Lemma 1 of Bai and Ng (2004)). Under Assumptions 1-5, considering estimation of $\hat{f}_t$ by the method of principal components, we have an H with rank r such that as $(n, T) \to \infty$,

(a) $min[n, T]T^{-1}\sum_{t=2}^{T} \|\hat{f}_t - Hf_t\|^2 = O_p(1)$,

(b) $min[\sqrt{n}, T](\hat{f}_t - Hf_t) = O_p(1)$, for each given t,

(c) $min[\sqrt{T}, n](\hat{\lambda}_i - H'^{-1}\lambda_i) = O_p(1)$, for each given i.

As is well known in factor analysis, λ_i and f_t are not directly identifiable. Therefore, when assessing the properties of the estimates, we can only consider the difference in the space spanned by $\hat{f}_t$ and f_t, and likewise between $\hat{\lambda}_i$ and λ_i. The matrix H is defined such that Hf_t is the projection of $\hat{f}_t$ on the space spanned by the factors, f_t. Result (a) is proved in Bai and Ng (2002), while (b) and (c) are proved in Bai (2003).

Lemma A.2

(Corresponds to Lemma 2 of Bai and Ng (2004)). Consider estimation of (7). Suppose Assumptions 1-5 hold. Then there exists an H with rank r such that as $(n, T) \to$

∞,

$$\max_{1\le t\le T}\|\hat{F}_t - HF_t + HF_1\| = O_p(T^{1/2}n^{-1/2}) + O_p(T^{-1/4}).$$

As stated in this lemma, $\hat{F}_t$ is uniformly consistent for HF_t (up to a shift factor HF_1) provided $T/n \to 0$ as $n, T \to \infty$. Without loss of generality, we assume that at t=1, $F_1 = 0$. Then we have $\max_{1\le t\le T}\|\hat{F}_t - HF_t\| = O_p(T^{1/2}N^{-1/2}) + O_p(T^{-1/4})$.

Lemma A.3

In order to prove Lemma 1 of this paper, we need one more lemma regarding the estimation errors of the factors, which we state here as Lemma A.3.

Consider estimation of (7). Recall the definition of the rotation matrix $A = [A_1, A_2]'$ (A_1 is $r \times r_1$ satisfying $A_1'A_1 = I_{r_1}$, and $A_1'A_2 = 0$), such that $F_{1t} = A_1'HF_t$ to be the r_1 common stochastic trends and $F_{2t} = A_2'HF_t$ to be the r_0 stationary elements resulting from such a rotation. Denote $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)'$, $\hat{F}_1 = \hat{F}A_1$ and $\hat{F}_2 = \hat{F}A_2$. Suppose Assumptions 1-5 hold. Then there exists an H with rank r such that as $(n, T) \to \infty$,

(a) $T^{-1}\hat{F}_1'(FH' - \hat{F}) = O_p(\frac{T}{\sqrt{n}})$;

(b) $T^{-1}\hat{F}_2'(FH' - \hat{F}) = O_p(\sqrt{\frac{T}{n}})$.

Proof. Let ϕ_t denote the estimation error of factors, i.e. $\phi_t = \hat{F}_t - HF_t$. Then

$$\frac{\hat{F}_1'(FH' - \hat{F})}{T} = \frac{\sum_{t=1}^{T}\hat{F}_{1t}(F_t'H' - \hat{F}_t')}{T} = -\frac{\sum_{t=1}^{T}\hat{F}_{1t}\phi_t'}{T},$$

and

$$\frac{\hat{F}_2'(FH' - \hat{F})}{T} = \frac{\sum_{t=1}^{T}\hat{F}_{2t}(F_t'H' - \hat{F}_t')}{T} = -\frac{\sum_{t=1}^{T}\hat{F}_{2t}\phi_t'}{T}.$$

From Lemma A.2 we have

$$\begin{aligned}\max_{1\le t\le T}\frac{1}{\sqrt{T}}\|\phi_t\| &= O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{T^{3/4}})\\ &= O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{\sqrt{T}}) = O_p(max(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}))\\ &= O_p(C_{nT}^{-1})\end{aligned}$$

where $C_{nT} = min[\sqrt{n}, \sqrt{T}]$. Then $\|\phi_t\|^2 = T \cdot O_p(max(\frac{1}{n}, \frac{1}{T})) = O_p(T/n)$ uniformly in t. We also have

$$\frac{1}{T}\sum_{t=1}^{T}\|\phi_t\|^2 = O_p(\frac{T}{n}).$$

Thus,

$$\begin{aligned}\|\frac{\sum_{t=1}^{T}\hat{F}_{1t}\phi_t^{'}}{T}\| &\leq \sqrt{T}(\frac{\sum_{t=1}^{T}\|\hat{F}_{1t}\|^2}{T^2})^{1/2}(\frac{\sum_{t=1}^{T}\|\phi_t\|^2}{T})^{1/2} \\ &= \sqrt{T}O_p(1)O_p(\sqrt{\frac{T}{n}}) = O_p(\frac{T}{\sqrt{n}}).\end{aligned}$$

Similarly, we have

$$\begin{aligned}\|\frac{\sum_{t=1}^{T}\hat{F}_{2t}\phi_t^{'}}{T}\| &\leq (\frac{\sum_{t=1}^{T}\|\hat{F}_{2t}\|^2}{T})^{1/2}(\frac{\sum_{t=1}^{T}\|\phi_t\|^2}{T})^{1/2} \\ &= O_p(1)O_p(\sqrt{\frac{T}{n}}) = O_p(\sqrt{\frac{T}{n}}).\end{aligned}$$

■

Proof of Lemma 1

Suppose Assumptions 1-5 and 6 (EC) hold. As $(n, T) \rightarrow \infty$, if $T/\sqrt{n} \rightarrow 0$,

(a) $TA_1^{'}(\hat{\delta} - \delta) \xrightarrow{d} (\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 dB_0 + \Delta_{10})$,

(b) $\sqrt{T}A_2^{'}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1})$.

Proof. Rewrite the cointegration equation as follows

$$\begin{aligned}y_t &= \alpha^{'}F_t + \varepsilon_t \\ &= \alpha^{'}H^{-1}\hat{F}_t + \varepsilon_t + \alpha^{'}H^{-1}(HF_t - \hat{F}_t).\end{aligned}$$

In matrix notation, $Y = \hat{F}\delta + \varepsilon + (FH^{'} - \hat{F})\delta$. It follows that

$$\hat{\delta} - \delta = (\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}\varepsilon + (\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}(FH^{'} - \hat{F})\delta.$$

Partitioning the coefficients into the nonstationary and stationary part, we have

$$A_1^{'}(\hat{\delta} - \delta) = A_1^{'}\left(\hat{F}^{'}\hat{F}\right)^{-1}\hat{F}^{'}\varepsilon + A_1^{'}\left(\hat{F}^{'}\hat{F}\right)^{-1}\hat{F}^{'}(FH^{'} - \hat{F})H^{-1'}\alpha,$$

and

$$A_2'(\hat{\delta}-\delta) = A_2'\left(\hat{F}'\hat{F}\right)^{-1}T^{-1}\hat{F}'\varepsilon + A_2'\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'(FH'-\hat{F})H^{-1'}\alpha.$$

Note that by partitioned inversion

$$\begin{aligned}
A_1'(\hat{F}'\hat{F})^{-1}\hat{F}'\varepsilon &= A_1'A'(A\hat{F}'\hat{F}A')^{-1}A\hat{F}'\varepsilon \\
&= \begin{bmatrix} I_{r_1} & 0 \end{bmatrix} \begin{bmatrix} \hat{F}_1'\hat{F}_1 & \hat{F}_1'\hat{F}_2 \\ \hat{F}_2'\hat{F}_1 & \hat{F}_2'\hat{F}_2 \end{bmatrix}^{-1} A\hat{F}'\varepsilon \\
&= \begin{bmatrix} I_{r_1} & 0 \end{bmatrix} \begin{bmatrix} (\hat{F}_1'Q_2\hat{F}_1)^{-1} & -(\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\hat{F}_2(\hat{F}_2'Q_1\hat{F}_2)^{-1} \\ -(\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\hat{F}_1(\hat{F}_1'\hat{F}_1)^{-1} & (\hat{F}_2'Q_1\hat{F}_2)^{-1} \end{bmatrix} A\hat{F}'\varepsilon \\
&= \begin{bmatrix} (\hat{F}_1'Q_2\hat{F}_1)^{-1} & -(\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\hat{F}_2(\hat{F}_2'Q_1\hat{F}_2)^{-1} \end{bmatrix} \begin{bmatrix} \hat{F}_1' \\ \hat{F}_2' \end{bmatrix} \varepsilon \\
&= (\hat{F}_1'Q_2\hat{F}_1)^{-1}\hat{F}_1'\varepsilon - (\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\hat{F}_2(\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\varepsilon
\end{aligned}$$

and

$$\begin{aligned}
A_2'(\hat{F}'\hat{F})^{-1}\hat{F}'\varepsilon &= A_2'A'(AF'FA')^{-1}A\hat{F}'\varepsilon \\
&= \begin{bmatrix} 0 & I_{r_0} \end{bmatrix} \begin{bmatrix} \hat{F}_1'\hat{F}_1 & \hat{F}_1'\hat{F}_2 \\ \hat{F}_2'\hat{F}_1 & \hat{F}_2'\hat{F}_2 \end{bmatrix}^{-1} A\hat{F}'\varepsilon \\
&= \begin{bmatrix} 0 & I_{r_0} \end{bmatrix} \begin{bmatrix} (\hat{F}_1'Q_2\hat{F}_1)^{-1} & -(\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\hat{F}_2(\hat{F}_2'Q_1\hat{F}_2)^{-1} \\ -(\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\hat{F}_1(\hat{F}_1'\hat{F}_1)^{-1} & (\hat{F}_2'Q_1\hat{F}_2)^{-1} \end{bmatrix} A\hat{F}'\varepsilon \\
&= \begin{bmatrix} -(\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\hat{F}_1(\hat{F}_1'\hat{F}_1)^{-1} & (\hat{F}_2'Q_1\hat{F}_2)^{-1} \end{bmatrix} \begin{bmatrix} \hat{F}_1' \\ \hat{F}_2' \end{bmatrix} \varepsilon \\
&= (\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\varepsilon - (\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\hat{F}_1(\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\varepsilon
\end{aligned}$$

where $Q_i = I - \hat{F}_i(\hat{F}_i^{'}\hat{F}_i)^{-1}\hat{F}_i^{'}$, i=1,2. Thus

$$\begin{aligned}
TA_1^{'}(\hat{\delta}-\delta) &= A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}T^{-1}\hat{F}^{'}\varepsilon + A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}T^{-1}\hat{F}^{'}(FH^{'}-\hat{F})H^{-1'}\alpha \\
&= (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T} - (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T} \\
&+ (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T} - (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T}.
\end{aligned}$$

Under Assumption 6 (EC), we have

$$(\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T} = (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T} + O_p(\frac{1}{T}) \xrightarrow{d} (\int B_1B_1^{'})^{-1}(\int_0^1 B_1 dB_0 + \Delta_{10}),$$

and

$$(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T} = (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{\sqrt{T}}\frac{1}{\sqrt{T}} = O_p(\frac{1}{\sqrt{T}}),$$

and

$$(\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T} = O_p(1)O_p(\frac{T}{\sqrt{n}}) = O_p(\frac{T}{\sqrt{n}}),$$

and

$$(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T} = O_p(1)O_p(1)O_p(1)O_p(\sqrt{\frac{T}{n}}) = O_p(\sqrt{\frac{T}{n}}).$$

The last two lines of proof also use results from Lemma A.3. Thus if we assume that $\frac{T}{\sqrt{n}} \to 0$ as $(n,T)\to\infty$, the last two expressions will be $o_p(1)$. Hence, we have

$$TA_1^{'}(\hat{\delta}-\delta) \xrightarrow{d} (\int B_1B_1^{'})^{-1}(\int_0^1 B_1 dB_0 + \Delta_{10}).$$

Similarly, we have

$$\begin{aligned}
\sqrt{T}A_2^{'}(\hat{\delta}-\delta) &= A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}T^{-1/2}\hat{F}^{'}\varepsilon + A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}T^{-1/2}\hat{F}^{'}(FH^{'}-\hat{F})H^{-1'}\alpha \\
&= (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{\sqrt{T}} - (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T}\frac{1}{\sqrt{T}} \\
&+ (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{\sqrt{T}} - (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T}\frac{1}{\sqrt{T}}.
\end{aligned}$$

Under Assumption 6 (EC), we have

$$(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{\sqrt{T}} = (\frac{\hat{F}_2^{'}\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{\sqrt{T}} + O_p(\frac{1}{T}) \xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1}),$$

and

$$(\frac{\hat{F}_2'Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2'\hat{F}_1}{T}(\frac{\hat{F}_1'\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1'\varepsilon}{T}\frac{1}{\sqrt{T}} = O_p(\frac{1}{\sqrt{T}}),$$

and

$$\begin{aligned}(\frac{\hat{F}_2'Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2'(FH' - \hat{F})H^{-1'}\alpha}{\sqrt{T}} &= (\frac{\hat{F}_2'Q_2\hat{F}_2}{T})^{-1}\frac{\hat{F}_2'(FH' - \hat{F})H^{-1'}\alpha}{T}\sqrt{T} \\ &= \sqrt{T}O_p(\sqrt{\frac{T}{n}}) = O_p(\frac{T}{\sqrt{n}}),\end{aligned}$$

and

$$(\frac{\hat{F}_2'Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2'\hat{F}_1}{T}(\frac{\hat{F}_1'\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1'(FH' - \hat{F})H^{-1'}\alpha}{T}\frac{1}{\sqrt{T}} = \frac{1}{\sqrt{T}}O_p(\frac{T}{\sqrt{n}}) = O_p(\sqrt{\frac{T}{n}}).$$

The last two lines of proof also use results from Lemma A.3. If $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$, the last two expressions will be $o_p(1)$. Hence, we have

$$\sqrt{T}A_2'(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1}).$$

■

Preliminaries for Theorem 1

As in Phillips (1995), to simplify the presentation of our arguments it will be convenient to assume in our proofs that we are working with long-run covariance matrix estimates that satisfy Assumption KL (a) and (b). This leads to estimates of the form

$$\hat{\Omega} = \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}(j), \text{ and } \hat{\Delta} = \sum_{j=0}^{K-1} \omega(j/K)\hat{\Gamma}(j),$$

which corresponds to (10) when the lag kernel is truncated as in KL (b), i.e. $\omega(x) = 0$, for $|x| > 1$. The proofs given below for Lemma A.4, A.5, and A.6 apply as they stand under KL (a) and (b) and therefore hold for the Parzen and Tukey-Hanning kernels, for example, which satisfy these conditions.

Lemma A.4

We have the following lemma adapted from Lemma 8.1 in the Appendix of Phillips (1995). Under Assumptions 1-5, Assumptions 6 (EC), 7 (KL), and 8 (BW(iv)), the following hold:

(a) $\hat{\Omega}_{\Delta u_2 \Delta u_2} = -K^{-2}\omega''(0)\Omega_{22} + o_p(K^{-2})$;

(b) $\hat{\Omega}_{\varepsilon \Delta u_2} = K^{-2}\omega''(0)\Phi_{02} + O_p(1/\sqrt{KT}) + o_p(K^{-2})$, where $\Phi_{02} = \sum_{j=-\infty}^{\infty}(j-1/2)\Gamma_{\varepsilon u_2(j)}$, and

$\hat{\Omega}_{u_1 \Delta u_2} = K^{-2}\omega''(0)\Phi_{12} + O_p(1/\sqrt{KT}) + o_p(K^{-2})$, where $\Phi_{12} = \sum_{j=-\infty}^{\infty}(j - 1/2)\Gamma_{u_1 u_2(j)}$;

(c) $\hat{\Omega}_{0\Delta \hat{u}_2} := \hat{\Omega}_{\hat{\varepsilon}\Delta u_2} = \hat{\Omega}_{\varepsilon \Delta u_2} + O_p(1/T)$;

(d*) $\hat{\Omega}_{\varepsilon a}\hat{\Omega}_{aa}^{-1} = \left[\begin{array}{cc} \Omega_{01}\Omega_{11}^{-1} + o_p(1), & -[\Phi_{02} - \Omega_{01}\Omega_{11}^{-1}\Phi_{12}]\Omega_{22}^{-1} + O_p(K^{3/2}/\sqrt{T}) + o_p(K^{3/2}/\sqrt{T}) \end{array}\right]$;

(e) $K^2[T^{-1}\Delta U_2' U_2 - \hat{\Delta}_{\Delta u_2 \Delta u_2}] \xrightarrow{p} \omega''(0)\{\Delta_{22} - (1/2)\Sigma_{22}\}$;

(f) $T^{-1}U_1' U_2 - \hat{\Delta}_{u_1 \Delta u_2} = K^{-2}\omega''(0)\Psi_{12} + O_p(1/\sqrt{KT}) + o_p(K^{-2})$, where

$\Psi_{12} = \sum_{j=1}^{\infty}(j-1/2)\Gamma_{u_1 u_2}(j)$;

(g) $T^{-1}\Delta U_2' F_1 - \hat{\Delta}_{\Delta u_2 u_1} = T^{-1}u_{2T}F_{1T}' + K^{-2}\omega''(0)\Psi_{21} + O_p(1/\sqrt{KT}) + o_p(K^{-2})$, where

$\Psi_{21} = \sum_{j=1}^{\infty}(j-1/2)\Gamma_{u_2 u_1}(j)$;

(h*) Under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have $\hat{\Delta}_{0\Delta \hat{u}_2} := \hat{\Delta}_{\hat{\varepsilon}\Delta \hat{u}_2} = O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}})$;

(i*) Under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have $\hat{\Delta}_{0\hat{u}_1} := \hat{\Delta}_{\hat{\varepsilon}\hat{u}_1} = \Delta_{01} + O_p((K/T)^{1/2}) + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right)$;

(j) $T^{-1}U_1' F_1 - \hat{\Delta}_{u_1 u_1} \xrightarrow{d} \int_0^1 dB_1 B_1'$;

(k) $T^{-1}F_1' \varepsilon - \hat{\Delta}_{u_1 \varepsilon} \xrightarrow{d} \int_0^1 B_1 dB_0$;

(l) $T^{-2}F_1' F_1 \xrightarrow{d} \int_0^1 B_1 B_1'$.

Proof.

Proof of (a)-(c), (e)-(g), and (j)-(l) are from Lemma 8.1 in the Appendix of Phillips (1995).

(d*).

$$\hat{\Omega}_{\varepsilon a}\hat{\Omega}_{aa}^{-1} = \begin{bmatrix} \hat{\Omega}_{\varepsilon u_1} & \hat{\Omega}_{\varepsilon \Delta u_2} \end{bmatrix} \begin{bmatrix} \hat{\Omega}_{u_1 u_1} & \hat{\Omega}_{u_1 u_1} \\ \hat{\Omega}_{\Delta u_2 u_1} & \hat{\Omega}_{\Delta u_2 \Delta u_2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \hat{\Omega}_{\varepsilon u_1} & \hat{\Omega}_{\varepsilon \Delta u_2} \end{bmatrix} \begin{bmatrix} \hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1} & -\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}\hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} \\ -\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1} & \hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} + \hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}\hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} X_1 & X_2 \end{bmatrix}$$

with $\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2} = \hat{\Omega}_{u_1 u_1} - \hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}$, $X_1 = \hat{\Omega}_{\varepsilon u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1} - \hat{\Omega}_{\varepsilon \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}$, and

$$X_2 = \hat{\Omega}_{\varepsilon \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} + \hat{\Omega}_{\varepsilon \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}\hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} - \hat{\Omega}_{\varepsilon u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}\hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}.$$

Using parts (a)-(c) of the lemma we find that

$$\begin{aligned}
\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2} &= \hat{\Omega}_{u_1 u_1} - \hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1} \\
&= \hat{\Omega}_{u_1 u_1} + O_p(K^{-2}) + O_p(K/T) + O_p(K^{-1/2}T^{-1/2}) \\
&= \hat{\Omega}_{u_1 u_1} + o_p(1) \xrightarrow{p} \Omega_{11} > 0,
\end{aligned}$$

and

$$\begin{aligned}
X_1 &= \hat{\Omega}_{\varepsilon u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1} - \hat{\Omega}_{\varepsilon \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1} \\
&= \Omega_{01}\Omega_{11}^{-1} + o_p(1) - [\omega''(0)\Phi_{02} + O_p(K^2/\sqrt{KT})][-\omega''(0)\Omega_{22} + o_p(K^{-2})]^{-1}[O_p(K^{-2}) + O_p(1/\sqrt{KT})] \\
&= \Omega_{01}\Omega_{11}^{-1} + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
X_2 &= \hat{\Omega}_{\varepsilon \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} + \hat{\Omega}_{\varepsilon \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1}\hat{\Omega}_{\Delta u_2 u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}\hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} - \hat{\Omega}_{\varepsilon u_1}\hat{\Omega}_{u_1 u_1 \cdot \Delta u_2}^{-1}\hat{\Omega}_{u_1 \Delta u_2}\hat{\Omega}_{\Delta u_2 \Delta u_2}^{-1} \\
&= [-\Phi_{02} + O_p(K^{3/2}/\sqrt{T})]\Omega_{22}^{-1} + [-\Phi_{02} + O_p(K^{3/2}/\sqrt{T})]\Omega_{22}^{-1}[O_p(K^{-2}) + O_p(K/T)][\Omega_{11} + o_p(1)]^{-1} \\
&\times [-\Phi_{12} + O_p(K^{3/2}/\sqrt{T})]\Omega_{22}^{-1} - [\Omega_{01} + o_p(1)][\Omega_{11} + o_p(1)]^{-1}[-\Phi_{12} + O_p(K^{3/2}/\sqrt{T})]\Omega_{22}^{-1} \\
&= -[\Phi_{02} - \Omega_{01}\Omega_{11}^{-1}\Phi_{12}]\Omega_{22}^{-1} + O_p(K^{3/2}/\sqrt{T}) + o_p(K^{3/2}/\sqrt{T}).
\end{aligned}$$

(h*). To prove part (h*) we write

$$\begin{aligned}
\hat{\Delta}_{\hat{\varepsilon}\Delta \hat{u}_2} &= \sum_{j=0}^{K-1} \omega(j/K)\hat{\Gamma}_{\hat{\varepsilon}\Delta \hat{u}_2}(j) = \sum_{j=0}^{K-1} \omega(j/K)[\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) - \hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j+1)] \\
&= \sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) + \hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(0) - \omega((K-1)/K)\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(K)
\end{aligned}$$

$$= \sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) + O_p(K^{-2}T^{-1/2}),$$

since $\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(0) = T^{-1}\hat{\varepsilon}^{'}\hat{u}_2 = T^{-1}\hat{\varepsilon}^{'}\hat{F}_2 = 0$ by least squares orthogonality, $\omega((K-1)/K) = O(K^{-2})$ and $\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(K) = O_p(T^{-1/2})$.

Since $\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) = \hat{\Gamma}_{\varepsilon u_2}(j) - \delta^{'}\hat{\Gamma}_{\phi u_2}(j) - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}u_2}(j) + \hat{\Gamma}_{\varepsilon\phi}(j)A_2 - \delta^{'}\hat{\Gamma}_{\phi\phi}(j)A_2 - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\phi}(j)A_2$, we have

$$\begin{aligned}
\hat{\Delta}_{\hat{\varepsilon}\Delta\hat{u}_2} &= \sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) + O_p(K^{-2}T^{-1/2}) \\
&= \sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\varepsilon u_2}(j) - \delta^{'}\sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\phi u_2}(j) \\
&- (\hat{\delta}-\delta)^{'}\sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\hat{F}u_2}(j) + \sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j)A_2 \\
&- \delta^{'}\sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\phi\phi}(j)A_2 - (\hat{\delta}-\delta)^{'}\sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\hat{F}\phi}(j)A_2 \quad \text{(A-1)} \\
&+ O_p(K^{-2}T^{-1/2}).
\end{aligned}$$

The first term in (A-1) has mean zero because

$$\Gamma_{\varepsilon u_2}(j) = 0 \text{ for all } j \geq 0$$

in view of Assumption 6 (EC). Next we consider the variance matrix of the first term, i.e.,

$$\sum_{j=1}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\varepsilon u_2}(j) = K^{-1}\sum_{j=1}^{K-1} \omega^{'}((j-1)/K)\hat{\Gamma}_{\varepsilon u_2}(j)[1 + O(K^{-1})],$$

and

$$\begin{aligned}
&\lim_{T\to\infty} KTvar[vec\{K^{-1}\sum_{j=1}^{K-1} \omega^{'}((j-1)/K)\hat{\Gamma}_{\varepsilon u_2}(j)\}] \\
&= \lim_{T\to\infty} \frac{T}{K} var[vec\{\sum_{j=1}^{K-1} \omega^{'}((j-1)/K)\hat{\Gamma}_{\varepsilon u_2}(j)\}] = constant.
\end{aligned}$$

The last line of proof is following Theorem 9 of Hannan (1970, p. 280) on the asymptotic covariance matrix of spectral estimates because of the first term of (A-1) has the same form as a spectral estimate at the origin and $\omega^{'}(x)$ is continuous and uniformly bound under KL. Hence,

$$\sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon u_2}(j)=O_p(1/\sqrt{KT}).$$

For the second term in (A-1), under Assumption 7 (KL) we can use the following Taylor expansion for $\omega(j/K)$ as $K\to\infty$:

$$\omega(j/K)-\omega((j-1)/K)=K^{-1}\omega^{'}((j-1)/K)(1+O(1/K)).$$

Thus

$$\sum_{j=1}^{K-1}(\omega(j/K)-\omega((j-1)/K))\hat{\Gamma}_{\phi u_2}(j)=K^{-1}\sum_{j=-K+2}^{K-1}\omega^{'}((j-1)/K)\hat{\Gamma}_{\phi u_2}(j)(1+O(1/K)).$$

The modulus of $K^{-1}\sum_{j=1}^{K-1}\omega^{'}((j-1)/K)\hat{\Gamma}_{\phi u_2}(j)$ is dominated above by

$$\begin{aligned}&(sup_{|j|\leq K}|\omega^{'}(\theta_j)|)K^{-1}\sum_{j=1}^{K-1}||\hat{\Gamma}_{\phi u_2}(j)||\\&\leq \text{constant } K^{-1}\sum_{j=1}^{K-1}||\hat{\Gamma}_{\phi u_2}(j)||\\&=O_p(\sqrt{\frac{T}{n}}).\end{aligned}$$

So the second term $\sum_{j=1}^{K-1}(\omega(j/K)-\omega((j-1)/K))\hat{\Gamma}_{\phi u_2}(j)=O_p(\sqrt{\frac{T}{n}})$.

By a similarly reasoning, for the fourth term we have $\sum_{j=1}^{K-1}(\omega(j/K)-\omega((j-1)/K))\hat{\Gamma}_{\varepsilon\phi}(j)=O_p(\sqrt{\frac{T}{n}})$, and for the fifth term we have $\sum_{j=1}^{K-1}(\omega(j/K)-\omega((j-1)/K))\hat{\Gamma}_{\phi\phi}(j)=O_p(\frac{T}{n})$.

For the third term,

$$\begin{aligned}&(\hat{\delta}-\delta)^{'}\sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\hat{F}u_2}(j)\\&=\sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)](\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}u_2}(j)\\&=\sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)](\hat{\delta}_1-\delta_1)^{'}\hat{\Gamma}_{\hat{F}_1u_2}(j)\\&+\sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)](\hat{\delta}_2-\delta_2)^{'}\hat{\Gamma}_{\hat{F}_2u_2}(j).\end{aligned}\tag{A-2}$$

Using the fact that $\hat{\delta}_2-\delta_2=A_2^{'}(\hat{\delta}-\delta)=O_p(T^{-1/2})$ under the assumption that $T/\sqrt{n}\to 0$ as

$(n,T) \to \infty$, we find that the second term on the last line is $O_p(1/\sqrt{KT})O_p(1/\sqrt{T}) = o_p(T^{-1})$. For the first term on the last line we note that $\hat{\delta}_1 - \delta_1 = A_1^{'}(\hat{\delta} - \delta) = O_p(T^{-1})$ under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$ and

$$\sum_{j=1}^{K-1}[\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\hat{F}_2 u_2}(j) = K^{-1}\sum_{j=1}^{K-1}\omega^{'}(\theta_j)\hat{\Gamma}_{\hat{F}_2 u_2}(j) = O_p(1).$$

So the third term in (A-1) is at most $O_p(T^{-1})$. Similarly, for the last term in (A-1) is also at most $O_p(T^{-1})$. To conclude, under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$

$$\begin{aligned}\hat{\Delta}_{\hat{\varepsilon}\Delta\hat{u}_2} &= O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}}) + O_p(T^{-1}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{T}{n}) + O_p(T^{-1}) + O_p(K^{-2}T^{-1/2}). \\ &= O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}}).\end{aligned}$$

(i*). By definition

$$\hat{\Delta}_{0\hat{u}_1} := \hat{\Delta}_{\hat{\varepsilon}\hat{u}_1} = \sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_1}(j), \text{ where } \hat{\Gamma}_{\hat{\varepsilon}\hat{u}_1}(j) = T^{-1}\sum_{1\le t,t+j\le T}\hat{\varepsilon}_{t+j}\hat{u}_{1t}^{'},$$

and $\hat{u}_{1t} = A_1^{'}\hat{f}_t = A_1^{'}(\hat{f}_t - Hf_t + Hf_t) = A_1^{'}\varphi_t + u_{1t}$, in which $\varphi_t = \hat{f}_t - Hf_t$ and $u_{1t} = A_1^{'}Hf_t$ by definition, and $\hat{\varepsilon}_t$ is the residual from a preliminary least squares regression of y_t on $\hat{F}_t$. Noticing that $y_t = \hat{F}_t^{'}\delta + \varepsilon_t + \alpha^{'}H^{-1}(HF_t - \hat{F}_t)$, we have

$$\begin{aligned}\hat{\varepsilon}_t &= y_t - \hat{F}_t^{'}\hat{\delta} \\ &= \hat{F}_t^{'}\delta + \varepsilon_t + \alpha^{'}H^{-1}(HF_t - \hat{F}_t) - \hat{F}_t^{'}\hat{\delta} \\ &= \varepsilon_t - \delta^{'}\phi_t - \hat{F}_t^{'}(\hat{\delta} - \delta),\end{aligned}$$

in which $\delta^{'} = \alpha^{'}H^{-1}$, $\hat{\delta} = (\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}Y$ and $\phi_t = \hat{F}_t - HF_t$. This leads to

$$\begin{aligned}\hat{\varepsilon}_{t+j}\hat{u}_{1t}^{'} &= \{\varepsilon_{t+j} - \delta^{'}\phi_{t+j} - \hat{F}_{t+j}^{'}(\hat{\delta} - \delta)\}\{A_1^{'}(\hat{f} - Hf_t + Hf_t)\}^{'} \\ &= \{\varepsilon_{t+j} - \delta^{'}\phi_{t+j} - \hat{F}_{t+j}^{'}(\hat{\delta} - \delta)\}\{u_{1t} + A_1^{'}\varphi_t\}^{'} \\ &= \varepsilon_{t+j}u_{1t}^{'} - \delta^{'}\phi_{t+j}u_{1t}^{'} - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}u_{1t}^{'} \\ &+ \varepsilon_{t+j}\varphi_t^{'}A_1 - \delta^{'}\phi_{t+j}\varphi_t^{'}A_1 - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}\varphi_t^{'}A_1.\end{aligned}$$

By definition

$$
\begin{aligned}
\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_1}(j) &= T^{-1} \sum_{1 \le t, t+j \le T} \hat{\varepsilon}_{t+j} \hat{u}_{1t}^{'} \\
&= T^{-1} \sum_{1 \le t, t+j \le T} \varepsilon_{t+j} u_{1t}^{'} - \delta^{'} T^{-1} \sum_{1 \le t, t+j \le T} \phi_{t+j} u_{1t}^{'} \\
&- (\hat{\delta} - \delta)^{'} T^{-1} \sum_{1 \le t, t+j \le T} \hat{F}_{t+j} u_{1t}^{'} \\
&+ T^{-1} \sum_{1 \le t, t+j \le T} \varepsilon_{t+j} \varphi_t^{'} A_1 - \delta^{'} T^{-1} \sum_{1 \le t, t+j \le T} \phi_{t+j} \varphi_t^{'} A_1 \\
&- (\hat{\delta} - \delta)^{'} T^{-1} \sum_{1 \le t, t+j \le T} \hat{F}_{t+j} \varphi_t^{'} A_1 \\
&= \hat{\Gamma}_{\varepsilon u_1}(j) - \delta^{'} \hat{\Gamma}_{\phi u_1}(j) - (\hat{\delta} - \delta)^{'} \hat{\Gamma}_{\hat{F} u_1}(j) \\
&+ \hat{\Gamma}_{\varepsilon \varphi}(j) A_1 - \delta^{'} \hat{\Gamma}_{\phi \varphi}(j) A_1 - (\hat{\delta} - \delta)^{'} \hat{\Gamma}_{\hat{F} \varphi}(j) A_1,
\end{aligned}
$$

where $\hat{\Gamma}_{\phi u_1}(j)$, $\hat{\Gamma}_{\hat{F} u_1}(j)$, $\hat{\Gamma}_{\varepsilon \varphi}(j)$, $\hat{\Gamma}_{\phi \varphi}(j)$ and $\hat{\Gamma}_{\hat{F} \varphi}(j)$ are defined similarly to $\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_1}(j)$.

Then we have

$$
\begin{aligned}
\hat{\Delta}_{\hat{\varepsilon}\hat{u}_1} &= \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\hat{\varepsilon}\hat{u}_1}(j) \\
&= \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\varepsilon u_1}(j) - \delta^{'} \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\phi u_1}(j) \\
&- (\hat{\delta} - \delta)^{'} \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\hat{F} u_1}(j) \\
&+ \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\varepsilon \varphi}(j) A_1 - \delta^{'} \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\phi \varphi}(j) A_1 \\
&- (\hat{\delta} - \delta)^{'} \sum_{j=0}^{K-1} \omega(j/K) \hat{\Gamma}_{\hat{F} \varphi}(j) A_1.
\end{aligned}
$$

Notice the first term in the above equation

$$
\sum_{j=-K+1}^{K-1} \omega(j/K) \hat{\Gamma}_{\varepsilon u_1}(j) = \hat{\Delta}_{\varepsilon u_1}.
$$

For any given j, $\hat{\Gamma}_{\phi u_1}(j) \;=\; T^{-1} \sum_{1 \le t, t+j \le T} \phi_{t+j} u_{1t}^{'}$, whose modulus satisfies

$$||\hat{\Gamma}_{\phi u_1}(j)||^2 \leq (1/T \sum_{1\leq t,t+j\leq T} ||u_{1t}||^2)(1/T \sum_{1\leq t,t+j\leq T} ||\phi_{t+j}||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\leq t,t+j\leq T}||u_{1t}||^2 = O_p(1)$. According to Lemma A.3, under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^T ||\phi_t||^2) = O_p(T/n)$. So for any given j,

$$||\hat{\Gamma}_{\phi u_1}(j)||^2 \leq O_p(1)O_p(\frac{T}{n}) = O_p(\frac{T}{n}), \text{ and } ||\hat{\Gamma}_{\phi u_1}(j)|| = O_p\left(\sqrt{\frac{T}{n}}\right).$$

Similarly, we have for any given j,

$$||\hat{\Gamma}_{\varepsilon\varphi}(j)||^2 \leq (1/T \sum_{1\leq t,t+j\leq T} ||\varepsilon_{t+j}||^2)(1/T \sum_{1\leq t,t+j\leq T} ||\varphi_t||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\leq t,t+j\leq T}||\varepsilon_{t+j}||^2 = O_p(1)$. According to Lemma 1 of Bai and Ng (2004), under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^T ||\varphi_t||^2) = (\frac{1}{T}\sum_{t=1}^T ||\hat{f}_t - Hf_t||^2) = O_p(D_{nT}^{-1})$ with $D_{nT} = min\{n, T\}$. So for any given j,

$$||\hat{\Gamma}_{\varepsilon\varphi}(j)||^2 \leq O_p(1)O_p(\frac{1}{D_{nT}}) = O_p(\frac{1}{D_{nT}}), \text{ and } ||\hat{\Gamma}_{\varepsilon\varphi}(j)|| = O_p\left(\frac{1}{\sqrt{D_{nT}}}\right).$$

Similarly,

$$||\hat{\Gamma}_{\phi\varphi}(j)||^2 \leq (1/T \sum_{1\leq t,t+j\leq T} ||\phi_{t+j}||^2)(1/T \sum_{1\leq t,t+j\leq T} ||\varphi_t||^2) = O_p(\frac{T}{nD_{nT}}),$$

and

$$||\hat{\Gamma}_{\phi\varphi}(j)|| \leq (1/T \sum_{1\leq t,t+j\leq T} ||\phi_{t+j}||^2)^{1/2}(1/T \sum_{1\leq t,t+j\leq T} ||\varphi_t||^2)^{1/2} = O_p(\frac{\sqrt{T}}{\sqrt{nD_{nT}}}).$$

To summarize,

$$||\hat{\Gamma}_{\phi u_1}(j)|| = O_p\left(\sqrt{\frac{T}{n}}\right), \ ||\hat{\Gamma}_{\varepsilon\varphi}(j)|| = O_p\left(\frac{1}{\sqrt{D_{nT}}}\right), \text{ and } ||\hat{\Gamma}_{\phi\varphi}(j)|| = O_p\left(\frac{\sqrt{T}}{\sqrt{nD_{nT}}}\right).$$

So for the term $\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi u_1}(j)$, its modulus is dominated by

$$\begin{aligned}
&(sup_{|j|\leq K}|\omega(\theta_j)|)\sum_{j=0}^{K-1}||\hat{\Gamma}_{\phi u_1}(j)|| \\
&\leq \text{constant} \sum_{j=0}^{K-1}||\hat{\Gamma}_{\phi u_1}(j)|| \\
&= O_p\left(K\sqrt{\frac{T}{n}}\right).
\end{aligned}$$

Thus the second term $\delta' \sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi u_1}(j) = O_p\left(K\sqrt{\frac{T}{n}}\right)$.

For the term $\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)$, since $\varphi_t = \phi_t - \phi_{t+1}$, we have $\hat{\Gamma}_{\varepsilon\varphi}(j) = \hat{\Gamma}_{\varepsilon\Delta\phi}(j)$ and

$$\begin{aligned}\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j) &= \sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\Delta\phi}(j)\\ &= \sum_{j=0}^{K-1}\omega(j/K)[\hat{\Gamma}_{\varepsilon\phi}(j)-\hat{\Gamma}_{\varepsilon\phi}(j+1)]\\ &= \sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j)+\\ &\quad \omega(0/K)\hat{\Gamma}_{\varepsilon\phi}(0)-\omega((K-1)/K)\hat{\Gamma}_{\varepsilon\phi}(K).\end{aligned}$$

Assumption KL implies that

$$\omega(0/K)=1, \omega((K-1)/K)=O(K^{-2}),$$

so that the second and third terms are $O_p(\sqrt{\frac{T}{n}})$ and $o_p(K^{-2})$. This leaves us with the first term which we write as

$$\sum_{j=1}^{K-1}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j) = \Big(\sum_{\mathfrak{B}_*}+\sum_{\mathfrak{B}^*}\Big)[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j) \tag{A-3}$$

where $\mathfrak{B}_* = \{j : |j| \le K^*\}$ and $\mathfrak{B}^* = \{j : |j| > K^*, 1 \le j \le K-1\}$ for some $K^* = K^b$ with $0 < b < 1$. Under KL we can use the following Taylor development for $\omega(j/K)$ when $|j| \le K^*$ and $K \to \infty$ to get

$$\begin{aligned}\omega(j/K)-\omega((j-1)/K) &= \omega^{'}((j-1)/K)(1/K)+(1/2)\omega^{''}(0)(1/K^2)(1+o(1))\\ &= \omega^{''}(0)((j-1)/K^2)(1+o(1))+(1/2)\omega^{''}(0)(1/K^2)(1+o(1)).\end{aligned}$$

The first sum in (A-4) is then

$$\sum_{\mathfrak{B}_*}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j) = K^{-2}\omega^{''}(0)\{\sum_{|j|\le K^*}(j-1)\hat{\Gamma}_{\varepsilon\phi}(j)+(1/2)\sum_{|j|\le K^*}\hat{\Gamma}_{\varepsilon\phi}(j)\}(1+o(1)).$$

The mean of the term in braces of above expression is

$$\begin{aligned}&\sum_{|j|\le K^*}(j-1)(1-|j|/T)\Gamma_{\varepsilon\phi}(j)+(1/2)\sum_{|j|\le K^*}(1-|j|/T)\Gamma_{\varepsilon\phi}(j)\\ &\to \sum_{j=-\infty}^{\infty}(j-1/2)\Gamma_{\varepsilon\phi}(j).\end{aligned}$$

Thus the first sum in (A-4) is $O_p(K^{-2})$. The second sum in (A-4) is

$$\sum_{\mathfrak{B}^*}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j)=K^{-1}\sum_{\mathfrak{B}^*}\omega^{'}(\theta_j)\hat{\Gamma}_{\varepsilon\phi}(j)$$

where $(j-1)/K<\theta_j<j/K$. This expression has mean given by

$$K^{-1}\sum_{\mathfrak{B}^*}\omega^{'}(\theta_j)(1-|j|/T)\Gamma_{\varepsilon\phi}(j)$$

whose modulus is dominated by

$$\begin{aligned}
&(sup_{|j|\leq K}|\omega^{'}(\theta_j)|)K^{-1}\sum_{|j|>K^*}||\Gamma_{\varepsilon\phi}(j)||\\
&\leq \text{constant}K^{-1}\sum_{j=-K+2}^{K-1}||\Gamma_{\varepsilon\phi}(j)||\\
&=O_p\left(\sqrt{\frac{T}{n}}\right).
\end{aligned}$$

Thus the fourth term $\delta^{'}\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)=O_p(K^{-2})+O_p\left(\sqrt{\frac{T}{n}}\right)$. By the same reasoning, for the fifth term we have $\delta^{'}\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi\varphi}(j)=O_p(K^{-2})+O_p\left(\frac{T}{n}\right)$.

For the third term $(\hat{\delta}-\delta)^{'}\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}u_1}(j)$, we have

$$(\hat{\delta}-\delta)^{'}\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}u_1}(j)=(\hat{\delta}-\delta)^{'}A_1\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1u_1}(j)+(\hat{\delta}-\delta)^{'}A_2\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2u_1}(j).$$

By definition

$$\begin{aligned}
\hat{\Gamma}_{\hat{F}_1u_1}(j)=T^{-1}\sum_{1\leq t,t+j\leq T}\hat{F}_{1,t+j}u_{1t}^{'}&=T^{-1}\sum_{1\leq t,t+j\leq T}A_1^{'}(\hat{F}_{t+j}-HF_{t+j}+HF_{t+j})u_{1t}^{'}\\
&=T^{-1}\sum_{1\leq t,t+j\leq T}A_1^{'}\phi_{t+j}u_{1t}^{'}+T^{-1}\sum_{1\leq t,t+j\leq T}A_1^{'}HF_{t+j}u_{1t}^{'}\\
&=A_1^{'}\tilde{\Gamma}_{\phi u_1}(j)+H\hat{\Gamma}_{F_1u_1}(j)=O_p\left(\sqrt{\frac{T}{n}}\right)+O_p(1),
\end{aligned}$$

and

$$\begin{aligned}
\hat{\Gamma}_{\hat{F}_2u_1}(j)=T^{-1}\sum_{1\leq t,t+j\leq T}\hat{F}_{2,t+j}u_{1t}^{'}&=T^{-1}\sum_{1\leq t,t+j\leq T}A_2^{'}(\hat{F}_{t+j}-HF_{t+j}+HF_{t+j})u_{1t}^{'}\\
&=T^{-1}\sum_{1\leq t,t+j\leq T}A_2^{'}\phi_{t+j}u_{1t}^{'}+T^{-1}\sum_{1\leq t,t+j\leq T}A_2^{'}HF_{t+j}u_{1t}^{'}
\end{aligned}$$

$$= A_2'\tilde{\Gamma}_{\phi u_1}(j) + H\hat{\Gamma}_{F_2 u_1}(j) = O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(\frac{1}{\sqrt{T}}).$$

So for the term $\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1 u_1}(j)$, its modulus is dominated by

$$(sup_{|j|\leq K}|\omega(\theta_j)|)\sum_{j=0}^{K-1}||\hat{\Gamma}_{\hat{F}_1 u_1}(j)|| \leq \text{constant}\sum_{j=0}^{K-1}||\hat{\Gamma}_{\hat{F}_1 u_1}(j)|| = O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(K),$$

and for the term $\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2 u_1}(j)$, its modulus is dominated by

$$(sup_{|j|\leq K}|\omega(\theta_j)|)\sum_{j=0}^{K-1}||\hat{\Gamma}_{\hat{F}_2 u_1}(j)|| \leq \text{constant}\sum_{j=0}^{K-1}||\hat{\Gamma}_{\hat{F}_1 u_1}(j)|| = O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(\frac{K}{\sqrt{T}}).$$

Thus under the assumption that $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$, we have

$$\begin{aligned}(\hat{\delta}-\delta)'\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}u_1}(j) &= (\hat{\delta}-\delta)'A_1\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1 u_1}(j) + (\hat{\delta}-\delta)'A_2\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2 u_1}(j)\\ &= O_p(\frac{1}{T})(O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(K)) + O_p(\frac{1}{\sqrt{T}})(O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(\frac{K}{\sqrt{T}}))\\ &= O_p\left(\frac{K}{\sqrt{Tn}}\right) + O_p(\frac{K}{T}) + O_p\left(\frac{K}{\sqrt{n}}\right).\end{aligned}$$

Similarly, for the sixth term, we have

$$\begin{aligned}&(\hat{\delta}-\delta)'\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}\varphi}(j) = (\hat{\delta}-\delta)'A_1\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1\varphi}(j) + (\hat{\delta}-\delta)'A_2\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2\varphi}(j)\\ &= O_p(\frac{1}{T})(O_p\left(\frac{K}{\sqrt{D_{nT}}}\sqrt{\frac{T}{n}}\right) + O_p(K)) + O_p(\frac{1}{\sqrt{T}})(O_p\left(\frac{K}{\sqrt{D_{nT}}}\sqrt{\frac{T}{n}}\right) + O_p(\frac{K}{\sqrt{T}}))\\ &= O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{Tn}}\right) + O_p(\frac{K}{T}) + O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{n}}\right).\end{aligned}$$

To conclude, under the assumption that $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$,

$$\begin{aligned}\hat{\Delta}_{\hat{\varepsilon}\hat{u}_1} &= \sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon u_1}(j) - \delta'\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi u_1}(j) - (\hat{\delta}-\delta)'\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}u_1}(j)\\ &+ \sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)A - \delta'\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi\varphi}(j)A_1 - (\hat{\delta}-\delta)'\sum_{j=0}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}\varphi}(j)A_1\\ &= \hat{\Delta}_{\varepsilon u_1} + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{\sqrt{Tn}}\right) + O_p(\frac{K}{T}) + O_p\left(\frac{K}{\sqrt{n}}\right) + O_p(K^{-2}) + O_p\left(\sqrt{\frac{T}{n}}\right)\\ &+ O_p(K^{-2}) + O_p\left(\frac{T}{n}\right) + O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{Tn}}\right) + O_p(\frac{K}{T}) + O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{n}}\right)\\ &= \hat{\Delta}_{\varepsilon u_1} + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right)\\ &= \Delta_{01} + O_p((K/T)^{1/2}) + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right).\end{aligned}$$

The last line of above proof is because $\hat{\Delta}_{\varepsilon u_1} = \Delta_{01} + O_p((K/T)^{1/2})$.

■

Lemma A.5

Under Assumptions 1-5, 6 (EC), 7 (KL), and 8 (BW), we have:

(a) Under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$

$$\begin{aligned}\hat{\Gamma}_{0\hat{u}_2}(j) := \hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) &= \hat{\Gamma}_{\varepsilon u_2}(j) - \delta^{'}\hat{\Gamma}_{\phi u_2}(j) - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}u_2}(j) \\ &+ \hat{\Gamma}_{\varepsilon\phi}(j)A_2 - \delta^{'}\hat{\Gamma}_{\phi\phi}(j)A_2 - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\phi}(j)A_2 \\ &= \hat{\Gamma}_{\varepsilon u_2}(j) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(T^{-1/2}) + O_p(\frac{T}{n});\end{aligned}$$

(b) Under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$

$$\begin{aligned}\hat{\Gamma}_{0\hat{a}}(j) := \hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) &= \hat{\Gamma}_{\varepsilon a}(j) - \delta^{'}\hat{\Gamma}_{\phi a}(j) - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}a}(j) \\ &+ \hat{\Gamma}_{\varepsilon\varphi}(j)A^{'} - \delta^{'}\hat{\Gamma}_{\phi\varphi}(j)A^{'} - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'} \\ &= \hat{\Gamma}_{\varepsilon a}(j) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(T^{-1/2}) + O_p\left(\frac{1}{\sqrt{D_{nT}}}\right) + O_p(\frac{\sqrt{T}}{\sqrt{nD_{nT}}});\end{aligned}$$

(c) Under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$,

$$\hat{\Omega}_{0\hat{a}} := \hat{\Omega}_{\hat{\varepsilon}\hat{a}} = \hat{\Omega}_{\varepsilon a} + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right);$$

(d) $\hat{\Omega}_{\hat{a}\hat{a}} = \hat{\Omega}_{aa} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}})$;

(e)

$$\begin{aligned}\hat{\Delta}_{\hat{a}\hat{a}} := \begin{bmatrix} \hat{\Delta}_{\hat{u}_1\hat{u}_1} & \hat{\Delta}_{\hat{u}_1\Delta\hat{u}_2} \\ \hat{\Delta}_{\Delta\hat{u}_2\hat{u}_1} & \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} \end{bmatrix} &= \begin{bmatrix} \hat{\Delta}_{u_1u_1} & \hat{\Delta}_{u_1\Delta u_2} \\ \hat{\Delta}_{\Delta u_2u_1} & \hat{\Delta}_{\Delta u_2\Delta u_2} \end{bmatrix} + O_p(K^{-2}) + O_p\left(\sqrt{\frac{T}{n}}\right) \\ &= \hat{\Delta}_{aa} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}});\end{aligned}$$

(f)

$$T^{-1}\Delta\hat{U}_2^{'}\hat{U}_2 - \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} = O_p(K^{-2}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}});$$

(g)

$$T^{-1}\hat{U}_1'\hat{U}_2 - \hat{\Delta}_{\hat{u}_1\Delta\hat{u}_2} = O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}});$$

(h)

$$\begin{aligned} T^{-1}\Delta\hat{U}_2'\hat{F}_1 - \hat{\Delta}_{\Delta\hat{u}_2\hat{u}_1} &= T^{-1}u_{2T}F_{1T}' + T^{-1}\phi_T F_{1T}' - T^{-1}\phi_0 F_{1,1}' \\ &+ O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\sqrt{\frac{T}{nD_{nT}}}) + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}) \\ &= O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(T^{-1/2}) + O_p(\sqrt{\frac{T}{n}}); \end{aligned}$$

(i) When $T/n \to 0$ as $(n,T) \to \infty$, we have

$T^{-1}\hat{U}_1'\hat{F}_1 - \hat{\Delta}_{\hat{u}_1\hat{u}_1} := N_{11T} \xrightarrow{d} \int_0^1 dB_1 B_1'$;

(j) When $K\sqrt{T}/\sqrt{n} \to 0$, $K/T \to 0$ and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$,

$T^{-1}\hat{F}_1'\varepsilon - \hat{\Delta}_{\hat{u}_1 0} \xrightarrow{d} \int_0^1 B_1 dB_0$;

with $D_{nT} = min\{n,T\}$.

Proof. (a).

By definition

$$\hat{\Omega}_{\varepsilon u_2} = \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon u_2}(j), \text{ where } \hat{\Gamma}_{\varepsilon u_2}(j) = T^{-1}\sum_{1\le t,t+j\le T} \varepsilon_{t+j}u_{2t}',$$

and

$$\hat{\Omega}_{\hat{\varepsilon}\hat{u}_2} = \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j), \text{ where } \hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) = T^{-1}\sum_{1\le t,t+j\le T} \hat{\varepsilon}_{t+j}\hat{u}_{2t}',$$

where $\hat{u}_{2t} = \hat{F}_{2t} = A_2'\hat{F}_t$ and $\hat{\varepsilon}_t$ is the residual from a preliminary least squares regression of y_t on $\hat{F}_t$. Noticing that $y_t = \hat{F}_t'\delta + \varepsilon_t + \alpha' H^{-1}(HF_t - \hat{F}_t)$, we have

$$\begin{aligned} \hat{\varepsilon}_t &= y_t - \hat{F}_t'\hat{\delta} \\ &= \hat{F}_t'\delta + \varepsilon_t + \alpha' H^{-1}(HF_t - \hat{F}_t) - \hat{F}_t'\hat{\delta} \\ &= \varepsilon_t - \delta'\phi_t - \hat{F}_t'(\hat{\delta} - \delta), \end{aligned}$$

in which $\delta' = \alpha' H^{-1}$, $\hat{\delta} = (\hat{F}'\hat{F})^{-1}\hat{F}'Y$ and $\phi_t = \hat{F}_t - HF_t$. This leads to

$$\hat{\varepsilon}_{t+j}\hat{u}_{2t}' = \{\varepsilon_{t+j} - \delta'\phi_{t+j} - \hat{F}_{t+j}'(\hat{\delta} - \delta)\}\{A_2'(\hat{F} - HF_t + HF_t)\}'$$

$$= \{\varepsilon_{t+j} - \delta^{'}\phi_{t+j} - \hat{F}_{t+j}^{'}(\hat{\delta} - \delta)\}\{u_{2t} + A_2^{'}\phi_t\}^{'}$$

$$= \varepsilon_{t+j}u_{2t}^{'} - \delta^{'}\phi_{t+j}u_{2t}^{'} - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}u_{2t}^{'}$$

$$+ \varepsilon_{t+j}\phi_t^{'}A_2 - \delta^{'}\phi_{t+j}\phi_t^{'}A_2 - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}\phi_t^{'}A_2.$$

By definition

$$\begin{aligned}
\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) &= T^{-1}\sum_{1\leq t,t+j\leq T}\hat{\varepsilon}_{t+j}\hat{u}_t^{'} \\
&= T^{-1}\sum_{1\leq t,t+j\leq T}\varepsilon_{t+j}u_{2t}^{'} - \delta^{'}T^{-1}\sum_{1\leq t,t+j\leq T}\phi_{t+j}u_{2t}^{'} \\
&- (\hat{\delta} - \delta)^{'}T^{-1}\sum_{1\leq t,t+j\leq T}\hat{F}_{t+j}u_{2t}^{'} \\
&+ T^{-1}\sum_{1\leq t,t+j\leq T}\varepsilon_{t+j}\phi_t^{'}A_2 - \delta^{'}T^{-1}\sum_{1\leq t,t+j\leq T}\phi_{t+j}\phi_t^{'}A_2 \\
&- (\hat{\delta} - \delta)^{'}T^{-1}\sum_{1\leq t,t+j\leq T}\hat{F}_{t+j}\phi_t^{'}A_2 \\
&= \hat{\Gamma}_{\varepsilon u_2}(j) - \delta^{'}\hat{\Gamma}_{\phi u_2}(j) - (\hat{\delta} - \delta)^{'}\hat{\Gamma}_{\hat{F}u_2}(j) \\
&+ \hat{\Gamma}_{\varepsilon\phi}(j)A_2 - \delta^{'}\hat{\Gamma}_{\phi\phi}(j)A_2 - (\hat{\delta} - \delta)^{'}\hat{\Gamma}_{\hat{F}\phi}(j)A_2,
\end{aligned}$$

where $\hat{\Gamma}_{\phi u_2}(j)$, $\hat{\Gamma}_{\hat{F}u_2}(j)$, $\hat{\Gamma}_{\varepsilon\phi}(j)$, $\hat{\Gamma}_{\phi\phi}(j)$ and $\hat{\Gamma}_{\hat{F}\phi}(j)$ are defined similarly to $\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j)$.

For any given j, $\hat{\Gamma}_{\phi u_2}(j) = T^{-1}\sum_{1\leq t,t+j\leq T}\phi_{t+j}u_{2t}^{'}$, whose modulus satisfies

$$||\hat{\Gamma}_{\phi u_2}(j)||^2 \leq (1/T\sum_{1\leq t,t+j\leq T}||u_{2t}||^2)(1/T\sum_{1\leq t,t+j\leq T}||\phi_{t+j}||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\leq t,t+j\leq T}||u_{2t}||^2 = O_p(1)$. According to Lemma 2, under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^{T}||\phi_t||^2) = O_p(T/n)$. So for any given j,

$$||\hat{\Gamma}_{\phi u_2}(j)||^2 \leq O_p(1)O_p(\frac{T}{n}) = O_p(\frac{T}{n}), \text{ and } ||\hat{\Gamma}_{\phi u_2}(j)|| = O_p\left(\sqrt{\frac{T}{n}}\right).$$

Similarly, we have for any given j,

$$||\hat{\Gamma}_{\varepsilon\phi}(j)||^2 \leq (1/T\sum_{1\leq t,t+j\leq T}||\varepsilon_{t+j}||^2)(1/T\sum_{1\leq t,t+j\leq T}||\phi_t||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\leq t,t+j\leq T}||\varepsilon_{t+j}||^2 = O_p(1)$. According to Lemma 2, under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^{T}||\phi_t||^2) = O_p(T/n)$. So for any given j,

$$||\hat{\Gamma}_{\varepsilon\phi}(j)||^2 \leq O_p(1)O_p(\frac{T}{n}) = O_p(\frac{T}{n}), \text{ and } ||\hat{\Gamma}_{\varepsilon\phi}(j)|| = O_p\left(\sqrt{\frac{T}{n}}\right).$$

Similarly,

$$||\hat{\Gamma}_{\phi\phi}(j)||^2 \leq (1/T\sum_{1\leq t,t+j\leq T}||\phi_{t+j}||^2)(1/T\sum_{1\leq t,t+j\leq T}||\phi_t||^2) = O_p(\frac{T^2}{n^2}),$$

and

$$||\hat{\Gamma}_{\phi\phi}(j)|| \leq (1/T\sum_{1\leq t,t+j\leq T}||\phi_{t+j}||^2)^{1/2}(1/T\sum_{1\leq t,t+j\leq T}||\phi_t||^2)^{1/2} = O_p(\frac{T}{n}).$$

For the third and sixth terms, under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$(\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}u_2}(j) = O_p(T^{-1/2})O_p(1) = O_p(T^{-1/2})$$

$$(\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\phi}(j)A_2 = O_p(T^{-1/2})O_p(1) = O_p(T^{-1/2}).$$

To conclude, under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$

$$\begin{aligned}\hat{\Gamma}_{\hat{\varepsilon}\hat{u}_2}(j) &= \hat{\Gamma}_{\varepsilon u_2}(j) - \delta^{'}\hat{\Gamma}_{\phi u_2}(j) - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}u_2}(j) \\ &+ \hat{\Gamma}_{\varepsilon\phi}(j)A_2 - \delta^{'}\hat{\Gamma}_{\phi\phi}(j)A_2 - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\phi}(j)A_2 \\ &= \hat{\Gamma}_{\varepsilon u_2}(j) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(T^{-1/2}) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(\frac{T}{n}) + O_p(T^{-1/2}) \\ &= \hat{\Gamma}_{\varepsilon u_2}(j) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(T^{-1/2}) + O_p(\frac{T}{n}).\end{aligned}$$

(b).

By definition

$$\hat{\Omega}_{0\hat{a}} := \hat{\Omega}_{\hat{\varepsilon}\hat{a}} = \sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j), \text{ where } \hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) = T^{-1}\sum_{1\leq t,t+j\leq T}\hat{\varepsilon}_{t+j}\hat{u}_{at}^{'},$$

and $\hat{u}_{at} = A\hat{f}_t = A(\hat{f}_t - Hf_t + Hf_t) = A\varphi_t + u_{at}$, in which $\varphi_t = \hat{f}_t - Hf_t$ and $u_{at} = AHf_t$ by definition. From the proof of part (a) we have $\hat{\varepsilon}_t = \varepsilon_t - \delta^{'}\phi_t - \hat{F}_t^{'}(\hat{\delta}-\delta)$. Thus

$$\begin{aligned}\hat{\varepsilon}_{t+j}\hat{u}_{at}^{'} &= \{\varepsilon_{t+j} - \delta^{'}\phi_{t+j} - \hat{F}_{t+j}^{'}(\hat{\delta}-\delta)\}\{A\varphi_t + u_{at}\}^{'} \\ &= \varepsilon_{t+j}u_{at}^{'} - \delta^{'}\phi_{t+j}u_{at}^{'} - (\hat{\delta}-\delta)^{'}\hat{F}_{t+j}u_{at}^{'}\end{aligned}$$

$$+ \varepsilon_{t+j}\varphi_t^{'}A^{'} - \delta^{'}\phi_{t+j}\varphi_t^{'}A^{'} - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}\varphi_t^{'}A^{'}.$$

Then

$$\begin{aligned}
\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) &= T^{-1}\sum_{1\le t,t+j\le T}\hat{\varepsilon}_{t+j}\hat{u}_{at}^{'} \\
&= T^{-1}\sum_{1\le t,t+j\le T}\varepsilon_{t+j}u_{at}^{'} - \delta^{'}T^{-1}\sum_{1\le t,t+j\le T}\phi_{t+j}u_{at}^{'} \\
&- (\hat{\delta}-\delta)^{'}T^{-1}\sum_{1\le t,t+j\le T}\hat{F}_{t+j}u_{at}^{'} \\
&+ T^{-1}\sum_{1\le t,t+j\le T}\varepsilon_{t+j}\varphi_t^{'}A^{'} - \delta^{'}T^{-1}\sum_{1\le t,t+j\le T}\phi_{t+j}\varphi_t^{'}A^{'} \\
&- (\hat{\delta}-\delta)^{'}T^{-1}\sum_{1\le t,t+j\le T}\hat{F}_{t+j}\varphi_t^{'}A^{'} \\
&= \hat{\Gamma}_{\varepsilon a}(j) - \delta^{'}\hat{\Gamma}_{\phi a}(j) - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}a}(j) \\
&+ \hat{\Gamma}_{\varepsilon\varphi}(j)A^{'} - \delta^{'}\hat{\Gamma}_{\phi\varphi}(j)A^{'} - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'},
\end{aligned}$$

where $\hat{\Gamma}_{\phi a}(j)$, $\hat{\Gamma}_{\hat{F}a}(j)$, $\hat{\Gamma}_{\varepsilon\varphi}(j)$, $\hat{\Gamma}_{\phi\varphi}(j)$ and $\hat{\Gamma}_{\hat{F}\varphi}(j)$ are defined similarly to $\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j)$.

For any given j, $\hat{\Gamma}_{\phi a}(j) = T^{-1}\sum_{1\le t,t+j\le T}\phi_{t+j}u_{at}^{'}$, whose modulus satisfies

$$||\hat{\Gamma}_{\phi a}(j)||^2 \le (1/T\sum_{1\le t,t+j\le T}||u_{at}||^2)(1/T\sum_{1\le t,t+j\le T}||\phi_{t+j}||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\le t,t+j\le T}||u_{at}||^2 = O_p(1)$. According to Lemma 2, under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^{T}||\phi_t||^2) = O_p(T/n)$. So for any given j,

$$||\hat{\Gamma}_{\phi a}(j)||^2 \le O_p(1)O_p(\frac{T}{n}) = O_p(\frac{T}{n}), \text{ and } ||\hat{\Gamma}_{\phi a}(j)|| = O_p\left(\sqrt{\frac{T}{n}}\right).$$

Similarly, we have for any given j,

$$||\hat{\Gamma}_{\varepsilon\varphi}(j)||^2 \le (1/T\sum_{1\le t,t+j\le T}||\varepsilon_{t+j}||^2)(1/T\sum_{1\le t,t+j\le T}||\varphi_t||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\le t,t+j\le T}||\varepsilon_{t+j}||^2 = O_p(1)$. According to Lemma 1 of Bai and Ng (2004), under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^{T}||\varphi_t||^2) = (\frac{1}{T}\sum_{t=1}^{T}||\hat{f}_t - Hf_t||^2) = O_p(D_{nT}^{-1})$

with $D_{nT} = min\{n, T\}$. So for any given j,

$$||\hat{\Gamma}_{\varepsilon\varphi}(j)||^2 \leq O_p(1)O_p(\frac{1}{D_{nT}}) = O_p(\frac{1}{D_{nT}}), \text{ and } ||\hat{\Gamma}_{\varepsilon\varphi}(j)|| = O_p\left(\frac{1}{\sqrt{D_{nT}}}\right).$$

Similarly,

$$||\hat{\Gamma}_{\phi\varphi}(j)||^2 \leq (1/T \sum_{1\leq t,t+j\leq T} ||\phi_{t+j}||^2)(1/T \sum_{1\leq t,t+j\leq T} ||\varphi_t||^2) = O_p(\frac{T}{nD_{nT}}),$$

and

$$||\hat{\Gamma}_{\phi\varphi}(j)|| \leq (1/T \sum_{1\leq t,t+j\leq T} ||\phi_{t+j}||^2)^{1/2}(1/T \sum_{1\leq t,t+j\leq T} ||\varphi_t||^2)^{1/2} = O_p(\frac{\sqrt{T}}{\sqrt{nD_{nT}}}).$$

For the third and sixth terms, under the assumption that $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$, we have

$$(\hat{\delta} - \delta)^{'}\hat{\Gamma}_{\hat{F}a}(j) = O_p(T^{-1/2})O_p(1) = O_p(T^{-1/2})$$

$$(\hat{\delta} - \delta)^{'}\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'} = O_p(T^{-1/2})O_p(1) = O_p(T^{-1/2}).$$

To conclude, under the assumption that $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$,

$$\begin{aligned}\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) &= \hat{\Gamma}_{\varepsilon a}(j) - \delta^{'}\hat{\Gamma}_{\phi a}(j) - (\hat{\delta} - \delta)^{'}\hat{\Gamma}_{\hat{F}a}(j)\\ &+ \hat{\Gamma}_{\varepsilon\varphi}(j)A^{'} - \delta^{'}\hat{\Gamma}_{\phi\varphi}(j)A^{'} - (\hat{\delta} - \delta)^{'}\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'}\\ &= \hat{\Gamma}_{\varepsilon a}(j) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(T^{-1/2}) + O_p\left(\frac{1}{\sqrt{D_{nT}}}\right) + O_p(\frac{\sqrt{T}}{\sqrt{nD_{nT}}}) + O_p(T^{-1/2})\\ &= \hat{\Gamma}_{\varepsilon a}(j) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(T^{-1/2}) + O_p\left(\frac{1}{\sqrt{D_{nT}}}\right) + O_p(\frac{\sqrt{T}}{\sqrt{nD_{nT}}}).\end{aligned}$$

(c). By definition

$$\hat{\Omega}_{0\hat{a}} := \hat{\Omega}_{\hat{\varepsilon}\hat{a}} = \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j), \text{ where } \hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) = T^{-1} \sum_{1\leq t,t+j\leq T} \hat{\varepsilon}_{t+j}\hat{u}_{at}^{'},$$

and $\hat{u}_{at} = A\hat{f}_t = A(\hat{f}_t - Hf_t + Hf_t) = A\varphi_t + u_{at}$, in which $\varphi_t = \hat{f}_t - Hf_t$ and $u_{at} = AHf_t$ by definition. From the proof of part (a) we have $\hat{\varepsilon}_t = \varepsilon_t - \delta^{'}\phi_t - \hat{F}_t^{'}(\hat{\delta} - \delta)$. Thus

$$\begin{aligned}\hat{\varepsilon}_{t+j}\hat{u}_{at}^{'} &= \{\varepsilon_{t+j} - \delta^{'}\phi_{t+j} - \hat{F}_{t+j}^{'}(\hat{\delta} - \delta)\}\{A\varphi_t + u_{at}\}^{'}\\ &= \varepsilon_{t+j}u_{at}^{'} - \delta^{'}\phi_{t+j}u_{at}^{'} - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}u_{at}^{'}\\ &+ \varepsilon_{t+j}\varphi_t^{'}A^{'} - \delta^{'}\phi_{t+j}\varphi_t^{'}A^{'} - (\hat{\delta} - \delta)^{'}\hat{F}_{t+j}\varphi_t^{'}A^{'}.\end{aligned}$$

Then

$$\begin{aligned}
\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) &= T^{-1} \sum_{1\le t,t+j\le T} \hat{\varepsilon}_{t+j}\hat{u}_{at}^{'} \\
&= T^{-1} \sum_{1\le t,t+j\le T} \varepsilon_{t+j}u_{at}^{'} - \delta^{'} T^{-1} \sum_{1\le t,t+j\le T} \phi_{t+j}u_{at}^{'} \\
&- (\hat{\delta}-\delta)^{'} T^{-1} \sum_{1\le t,t+j\le T} \hat{F}_{t+j}u_{at}^{'} \\
&+ T^{-1} \sum_{1\le t,t+j\le T} \varepsilon_{t+j}\varphi_t^{'}A^{'} - \delta^{'} T^{-1} \sum_{1\le t,t+j\le T} \phi_{t+j}\varphi_t^{'}A^{'} \\
&- (\hat{\delta}-\delta)^{'} T^{-1} \sum_{1\le t,t+j\le T} \hat{F}_{t+j}\varphi_t^{'}A^{'} \\
&= \hat{\Gamma}_{\varepsilon a}(j) - \delta^{'}\hat{\Gamma}_{\phi a}(j) - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}a}(j) \\
&+ \hat{\Gamma}_{\varepsilon\varphi}(j)A^{'} - \delta^{'}\hat{\Gamma}_{\phi\varphi}(j)A^{'} - (\hat{\delta}-\delta)^{'}\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'},
\end{aligned}$$

where $\hat{\Gamma}_{\phi a}(j)$, $\hat{\Gamma}_{\hat{F}a}(j)$, $\hat{\Gamma}_{\varepsilon\varphi}(j)$, $\hat{\Gamma}_{\phi\varphi}(j)$ and $\hat{\Gamma}_{\hat{F}\varphi}(j)$ are defined similarly to $\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j)$.

Then we have

$$\begin{aligned}
\hat{\Omega}_{\hat{\varepsilon}\hat{a}} &= \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\hat{\varepsilon}\hat{a}}(j) \\
&= \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon a}(j) - \delta^{'} \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\phi a}(j) \\
&- (\hat{\delta}-\delta)^{'} \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\hat{F}a}(j) \\
&+ \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)A - \delta^{'} \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\phi\varphi}(j)A \\
&- (\hat{\delta}-\delta)^{'} \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'}.
\end{aligned}$$

Notice the first term in the above equation

$$\sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon a}(j) = \hat{\Omega}_{\varepsilon a}.$$

From the proof of part (b), we have

$$||\hat{\Gamma}_{\phi a}(j)|| = O_p\left(\sqrt{\frac{T}{n}}\right),\ ||\hat{\Gamma}_{\varepsilon\varphi}(j)|| = O_p\left(\frac{1}{\sqrt{D_{nT}}}\right),\ \text{and } ||\hat{\Gamma}_{\phi\varphi}(j)|| = O_p\left(\frac{\sqrt{T}}{\sqrt{nD_{nT}}}\right).$$

So for the term $\sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\phi a}(j)$, its modulus is dominated by

$$\begin{aligned}
&(sup_{|j|\leq K}|\omega(\theta_j)|) \sum_{j=-K+1}^{K-1} ||\hat{\Gamma}_{\phi a}(j)|| \\
&\leq \text{constant} \sum_{j=-K+1}^{K-1} ||\hat{\Gamma}_{\phi a}(j)|| \\
&= O_p\left(K\sqrt{\frac{T}{n}}\right).
\end{aligned}$$

Thus the second term $\delta^{'} \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\phi a}(j) = O_p\left(K\sqrt{\frac{T}{n}}\right)$. For the term $\sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)$, since $\varphi_t = \phi_t - \phi_{t+1}$, we have $\hat{\Gamma}_{\varepsilon\varphi}(j) = \hat{\Gamma}_{\varepsilon\Delta\phi}(j)$ and

$$\begin{aligned}
\sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j) &= \sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{\varepsilon\Delta\phi}(j) \\
&= \sum_{j=-K+1}^{K-1} \omega(j/K)[\hat{\Gamma}_{\varepsilon\phi}(j) - \hat{\Gamma}_{\varepsilon\phi}(j+1)] \\
&= \sum_{j=-K+2}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j) + \\
&\quad \omega((-K+1)/K)\hat{\Gamma}_{\varepsilon\phi}(-K+1) - \omega((K-1)/K)\hat{\Gamma}_{\varepsilon\phi}(K).
\end{aligned}$$

Assumption KL implies that

$$\omega((-K+1)/K), \omega((K-1)/K) = O(K^{-2}),$$

so that the second and third terms are $o_p(K^{-2})$. This leaves us with the first term which we write as

$$\sum_{j=-K+2}^{K-1} [\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j) = \big(\sum_{\mathfrak{B}_*} + \sum_{\mathfrak{B}^*}\big)[\omega(j/K) - \omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j) \qquad \text{(A-4)}$$

where $\mathfrak{B}_* = \{j : |j| \leq K^*\}$ and $\mathfrak{B}^* = \{j : |j| > K^*, -K+1 \leq j \leq K-2\}$ for some $K^* = K^b$ with $0 < b < 1$. Under KL we can use the following Taylor development for $\omega(j/K)$ when $|j| \leq K^*$ and $K \to \infty$ to get

$$\begin{aligned}
\omega(j/K) - \omega((j-1)/K) &= \omega^{'}((j-1)/K)(1/K) + (1/2)\omega^{''}(0)(1/K^2)(1 + o(1)) \\
&= \omega^{''}(0)((j-1)/K^2)(1 + o(1)) + (1/2)\omega^{''}(0)(1/K^2)(1 + o(1)).
\end{aligned}$$

The first sum in (A-4) is then

$$\sum_{\mathfrak{B}_*}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j)=K^{-2}\omega^{''}(0)\{\sum_{|j|\le K^*}(j-1)\hat{\Gamma}_{\varepsilon\phi}(j)+(1/2)\sum_{|j|\le K^*}\hat{\Gamma}_{\varepsilon\phi}(j)\}(1+o(1)).$$

The mean of the term in braces of above expression is

$$\begin{aligned}&\sum_{|j|\le K^*}(j-1)(1-|j|/T)\Gamma_{\varepsilon\phi}(j)+(1/2)\sum_{|j|\le K^*}(1-|j|/T)\Gamma_{\varepsilon\phi}(j)\\&\to\sum_{j=-\infty}^{\infty}(j-1/2)\Gamma_{\varepsilon\phi}(j).\end{aligned}$$

Thus the first sum in (A-4) is $O_p(K^{-2})$. The second sum in (A-4) is

$$\sum_{\mathfrak{B}^*}[\omega(j/K)-\omega((j-1)/K)]\hat{\Gamma}_{\varepsilon\phi}(j)=K^{-1}\sum_{\mathfrak{B}^*}\omega^{'}(\theta_j)\hat{\Gamma}_{\varepsilon\phi}(j)$$

where $(j-1)/K<\theta_j<j/K$. This expression has mean given by

$$K^{-1}\sum_{\mathfrak{B}^*}\omega^{'}(\theta_j)(1-|j|/T)\Gamma_{\varepsilon\phi}(j)$$

whose modulus is dominated by

$$\begin{aligned}&(sup_{|j|\le K}|\omega^{'}(\theta_j)|)K^{-1}\sum_{|j|>K^*}||\Gamma_{\varepsilon\phi}(j)||\\&\le \text{constant}K^{-1}\sum_{j=-K+2}^{K-1}||\Gamma_{\varepsilon\phi}(j)||\\&=O_p\left(\sqrt{\frac{T}{n}}\right).\end{aligned}$$

Thus the fourth term $\delta^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)=O_p(K^{-2})+O_p\left(\sqrt{\frac{T}{n}}\right)$. By the same reasoning, for the fifth term we have $\delta^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi\varphi}(j)=O_p(K^{-2})+O_p\left(\frac{T}{n}\right)$.

For the third term $(\hat{\delta}-\delta)^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}a}(j)$, we have

$$(\hat{\delta}-\delta)^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}a}(j)=(\hat{\delta}-\delta)^{'}A_1\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1a}(j)+(\hat{\delta}-\delta)^{'}A_2\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2a}(j).$$

By definition

$$\begin{aligned}\hat{\Gamma}_{\hat{F}_1a}(j)=T^{-1}\sum_{1\le t,t+j\le T}\hat{F}_{1,t+j}u_{at}^{'}&=T^{-1}\sum_{1\le t,t+j\le T}A_1^{'}(\hat{F}_{t+j}-HF_{t+j}+HF_{t+j})u_{at}^{'}\\&=T^{-1}\sum_{1\le t,t+j\le T}A_1^{'}\phi_{t+j}u_{at}^{'}+T^{-1}\sum_{1\le t,t+j\le T}A_1^{'}HF_{t+j}u_{at}^{'}\end{aligned}$$

$$= A_1^{'}\tilde{\Gamma}_{\phi u_a}(j) + H\hat{\Gamma}_{F_1 u_a}(j) = O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(1),$$

and

$$\hat{\Gamma}_{\hat{F}_2 a}(j) = T^{-1}\sum_{1\le t,t+j\le T}\hat{F}_{2,t+j}u_{at}^{'} = T^{-1}\sum_{1\le t,t+j\le T}A_2^{'}(\hat{F}_{t+j} - HF_{t+j} + HF_{t+j})u_{at}^{'}$$

$$= T^{-1}\sum_{1\le t,t+j\le T}A_2^{'}\phi_{t+j}u_{at}^{'} + T^{-1}\sum_{1\le t,t+j\le T}A_2^{'}HF_{t+j}u_{at}^{'}$$

$$= A_2^{'}\tilde{\Gamma}_{\phi u_a}(j) + H\hat{\Gamma}_{F_2 u_a}(j) = O_p\left(\sqrt{\frac{T}{n}}\right) + O_p(\frac{1}{\sqrt{T}}).$$

So for the term $\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1 a}(j)$, its modulus is dominated by

$$(sup_{|j|\le K}|\omega(\theta_j)|)\sum_{j=-K+1}^{K-1}||\hat{\Gamma}_{\hat{F}_1 a}(j)|| \le \text{constant}\sum_{j=-K+1}^{K-1}||\hat{\Gamma}_{\hat{F}_1 a}(j)|| = O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(K),$$

and for the term $\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2 a}(j)$, its modulus is dominated by

$$(sup_{|j|\le K}|\omega(\theta_j)|)\sum_{j=-K+1}^{K-1}||\hat{\Gamma}_{\hat{F}_2 a}(j)|| \le \text{constant}\sum_{j=-K+1}^{K-1}||\hat{\Gamma}_{\hat{F}_1 a}(j)|| = O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(\frac{K}{\sqrt{T}}).$$

Thus under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$(\hat{\delta}-\delta)^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}a}(j) = (\hat{\delta}-\delta)^{'}A_1\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1 a}(j) + (\hat{\delta}-\delta)^{'}A_2\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2 a}(j)$$

$$= O_p(\frac{1}{T})(O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(K)) + O_p(\frac{1}{\sqrt{T}})(O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p(\frac{K}{\sqrt{T}}))$$

$$= O_p\left(\frac{K}{\sqrt{Tn}}\right) + O_p(\frac{K}{T}) + O_p\left(\frac{K}{\sqrt{n}}\right).$$

Similarly, for the sixth term, we have

$$(\hat{\delta}-\delta)^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}\varphi}(j) = (\hat{\delta}-\delta)^{'}A_1\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_1\varphi}(j) + (\hat{\delta}-\delta)^{'}A_2\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}_2\varphi}(j)$$

$$= O_p(\frac{1}{T})(O_p\left(\frac{K}{\sqrt{D_{nT}}}\sqrt{\frac{T}{n}}\right) + O_p(K)) + O_p(\frac{1}{\sqrt{T}})(O_p\left(\frac{K}{\sqrt{D_{nT}}}\sqrt{\frac{T}{n}}\right) + O_p(\frac{K}{\sqrt{T}}))$$

$$= O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{Tn}}\right) + O_p(\frac{K}{T}) + O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{n}}\right).$$

To conclude, under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$,

$$\hat{\Omega}_{\hat{\varepsilon}\hat{a}} = \sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon a}(j) - \delta^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi a}(j) - (\hat{\delta}-\delta)^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}a}(j)$$

$$+\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)A-\delta^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\phi\varphi}(j)A-(\hat{\delta}-\delta)^{'}\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{F}\varphi}(j)A^{'}$$

$$=\hat{\Omega}_{\varepsilon a}+O_p\left(K\sqrt{\frac{T}{n}}\right)+O_p\left(\frac{K}{\sqrt{Tn}}\right)+O_p(\frac{K}{T})+O_p\left(\frac{K}{\sqrt{n}}\right)+O_p(K^{-2})+O_p\left(\sqrt{\frac{T}{n}}\right)$$

$$+O_p(K^{-2})+O_p\left(\frac{T}{n}\right)+O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{Tn}}\right)+O_p(\frac{K}{T})+O_p\left(\frac{K}{\sqrt{D_{nT}}\sqrt{n}}\right)$$

$$=\hat{\Omega}_{\varepsilon a}+O_p\left(K\sqrt{\frac{T}{n}}\right)+O_p\left(\frac{K}{T}\right)+O_p(K^{-2})+O_p\left(\frac{T}{n}\right).$$

(d). By definition

$$\hat{\Omega}_{\hat{a}\hat{a}}:=\hat{\Omega}_{\hat{a}\hat{a}}=\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{a}\hat{a}}(j),\ \text{where } \hat{\Gamma}_{\hat{a}\hat{a}}(j)=T^{-1}\sum_{1\le t,t+j\le T}\hat{u}_{a,t+j}\hat{u}_{at}^{'},$$

and $\hat{u}_{at}=A\hat{f}_t=A(\hat{f}_t-Hf_t+Hf_t)=A\varphi_t+u_{at}$, in which $\varphi_t=\hat{f}_t-Hf_t$ and $u_{at}=AHf_t$ by definition. Thus

$$\begin{aligned}\hat{u}_{a,t+j}\hat{u}_{at}^{'}&=\{A\varphi_{t+j}+u_{a,t+j}\}\{A\varphi_t+u_{at}\}^{'}\\&=u_{a,t+j}u_{at}^{'}+u_{a,t+j}\varphi_t^{'}A^{'}+A\varphi_{t+j}u_{at}^{'}+A\varphi_{t+j}\varphi_t^{'}A^{'}.\end{aligned}$$

Then

$$\begin{aligned}\hat{\Gamma}_{\hat{a}\hat{a}}(j)&=T^{-1}\sum_{1\le t,t+j\le T}\hat{u}_{a,t+j}\hat{u}_{at}^{'}\\&=T^{-1}\sum_{1\le t,t+j\le T}u_{a,t+j}u_{at}^{'}+T^{-1}\sum_{1\le t,t+j\le T}u_{a,t+j}\varphi_t^{'}A^{'}\\&+AT^{-1}\sum_{1\le t,t+j\le T}\varphi_{t+j}u_{at}^{'}+AT^{-1}\sum_{1\le t,t+j\le T}\varphi_{t+j}\varphi_t^{'}A^{'}\\&=\hat{\Gamma}_{aa}(j)+\hat{\Gamma}_{a\varphi}(j)A^{'}+A\hat{\Gamma}_{\varphi a}(j)+A\hat{\Gamma}_{\varphi\varphi}(j)A^{'},\end{aligned}$$

where $\hat{\Gamma}_{aa}(j)$, $\hat{\Gamma}_{a\varphi}(j)$, $\hat{\Gamma}_{\varphi a}(j)$, and $\hat{\Gamma}_{\varphi\varphi}(j)$ are defined similarly to $\hat{\Gamma}_{\hat{a}\hat{a}}(j)$.

Then we have

$$\begin{aligned}\hat{\Omega}_{\hat{a}\hat{a}}&=\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{a}\hat{a}}(j)\\&=\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{aa}(j)+\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{a\varphi}(j)A^{'}\\&+A\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varphi a}(j)+A\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varphi\varphi}(j)A^{'}.\end{aligned}$$

Notice the first term in the above equation

$$\sum_{j=-K+1}^{K-1} \omega(j/K)\hat{\Gamma}_{aa}(j) = \hat{\Omega}_{aa}.$$

For any given j, $\hat{\Gamma}_{a\varphi}(j) = T^{-1}\sum_{1\le t,t+j\le T} u_{a,t+j}\varphi_t^{'}$, whose modulus satisfies

$$||\hat{\Gamma}_{a\varphi}(j)||^2 \le (1/T\sum_{1\le t,t+j\le T}||u_{a,t+j}||^2)(1/T\sum_{1\le t,t+j\le T}||\varphi_t||^2).$$

Assumption 6 (EC) insures that $1/T\sum_{1\le t,t+j\le T}||u_{at}||^2 = O_p(1)$. According to Lemma 1 of Bai and Ng (2004), under Assumptions 1-5, $(\frac{1}{T}\sum_{t=1}^{T}||\varphi_t||^2) = (\frac{1}{T}\sum_{t=1}^{T}||\hat{f}_t - Hf_t||^2) = O_p(D_{nT}^{-1})$ with $D_{nT} = min\{n,T\}$. So for any given j,

$$||\hat{\Gamma}_{a\varphi}(j)||^2 \le O_p(1)O_p(\frac{1}{D_{nT}}) = O_p(\frac{1}{D_{nT}}), \text{ and } ||\hat{\Gamma}_{a\varphi}(j)|| = O_p\left(\frac{1}{\sqrt{D_{nT}}}\right).$$

Similarly,

$$||\hat{\Gamma}_{\varphi\varphi}(j)||^2 \le (1/T\sum_{1\le t,t+j\le T}||\varphi_{t+j}||^2)(1/T\sum_{1\le t,t+j\le T}||\varphi_t||^2) = O_p(\frac{1}{D_{nT}^2}),$$

and

$$||\hat{\Gamma}_{\varphi\varphi}(j)|| \le (1/T\sum_{1\le t,t+j\le T}||\varphi_{t+j}||^2)^{1/2}(1/T\sum_{1\le t,t+j\le T}||\varphi_t||^2)^{1/2} = O_p(\frac{1}{D_{nT}}).$$

So the modulus of the second term $\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{a\varphi}(j)$ can be shown to be $O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}})$ as in the term $\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varepsilon\varphi}(j)$ in the proof of part (c). And similarly, $\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varphi\varphi}(j) = O_p(K^{-2}) + O_p(\sqrt{\frac{T}{nD_{nT}}})$.

To conclude,

$$\begin{aligned}
\hat{\Omega}_{\hat{a}\hat{a}} &= \sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\hat{a}\hat{a}}(j) \\
&= \sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{aa}(j) + \sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{a\varphi}(j)A^{'} + A\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varphi a}(j) + A\sum_{j=-K+1}^{K-1}\omega(j/K)\hat{\Gamma}_{\varphi\varphi}(j)A^{'} \\
&= \hat{\Omega}_{aa} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}) + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}) + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{nD_{nT}}}) \\
&= \hat{\Omega}_{aa} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}).
\end{aligned}$$

(e). Similar to the proof of part (d).

(f).

Lemma 8.1 (e) in the Appendix of Phillips (1995) states that

$$K^2[T^{-1}\Delta U_2' U_2 - \hat{\Delta}_{\Delta u_2 \Delta u_2}] \xrightarrow{p} \omega''(0)\{\Delta_{22} - (1/2)\Sigma_{22}\}.$$

Thus $T^{-1}\Delta U_2' U_2 - \hat{\Delta}_{\Delta u_2 \Delta u_2} = O_p(K^{-2})$. From Lemma A.5 (e), we have $\hat{\Delta}_{\Delta \hat{u}_2 \Delta \hat{u}_2} = \hat{\Delta}_{\Delta u_2 \Delta u_2} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}})$, and from the proof of Lemma A.5 (a), we know that $\hat{u}_{2t} = \hat{F}_{2t} = A_2'\hat{F}_t = A_2'(\hat{F} - HF_t + HF_t) = u_{2t} + A_2'\phi_t$ with $\phi_t = \hat{F}_t - HF_t$, and $\Delta\hat{u}_{2t} = \hat{f}_{2t} = A_2'\hat{f}_t = A_2'(\hat{f} - Hf_t + Hf_t) = \Delta u_{2t} + A_2'\varphi_t$ with $\varphi_t = \hat{f}_t - Hf_t$. This leads to

$$\begin{aligned}\Delta\hat{u}_{2t}\hat{u}_{2t}' &= (\Delta u_{2t} + A_2'\varphi_t)(u_{2t} + A_2'\phi_t)' \\ &= \Delta u_{2t}u_{2t}' + A_2'\varphi_t u_{2t}' + \Delta u_{2t}\phi_t A_2 + A_2'\varphi_t\phi_t' A_2,\end{aligned}$$

and

$$\begin{aligned}T^{-1}\Delta\hat{U}_2'\hat{U}_2 &= \frac{\sum_{t=1}^T \Delta\hat{u}_{2t}\hat{u}_{2t}'}{T} \\ &= \frac{\sum_{t=1}^T \Delta u_{2t}u_{2t}'}{T} + A_2'\frac{\sum_{t=1}^T \varphi_t u_{2t}'}{T} + \frac{\sum_{t=1}^T \Delta u_{2t}\phi_t'}{T}A_2 + A_2'\frac{\sum_{t=1}^T \varphi_t\phi_t'}{T}A_2.\end{aligned}$$

It can be easily shown that $\frac{\sum_{t=1}^T \Delta u_{2t}\phi_t'}{T} = O_p(\sqrt{\frac{T}{n}})$, and $\frac{\sum_{t=1}^T \varphi_t\phi_t'}{T} = O_p(\sqrt{\frac{T}{nD_{nT}}})$.

We also have

$$\begin{aligned}\frac{\sum_{t=1}^T \varphi_t u_{2t}'}{T} &= \frac{\sum_{t=1}^T (\phi_t - \phi_{t-1})u_{2t}'}{T} \\ &= T^{-1}\phi_T u_{2T}' - T^{-1}\phi_0 u_{2,1}' - \frac{\sum_{s=1}^{T-1}\phi_s(u_{2s}' - u_{2,s+1}')}{T} \\ &= T^{-1}\phi_T u_{2T}' - T^{-1}\phi_0 u_{2,1}' + \frac{\sum_{s=1}^{T-1}\phi_s\Delta u_{2,s+1}'}{T} \\ &= T^{-1}\phi_T u_{2T}' - T^{-1}\phi_0 u_{2,1}' + O_p(\sqrt{\frac{T}{n}}) \\ &= O_p(\frac{1}{T}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) = O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}),\end{aligned}$$

with the last line from the fact that $T^{-1}\phi_T u_{2T}' = O_p(1/T)$, $T^{-1}\phi_0 u_{2,1}' = O_p(1/T)$, and $\frac{\sum_{s=1}^{T-1}\phi_s\Delta u_{2,s+1}'}{T} = O_p(\sqrt{\frac{T}{n}})$. Thus

$$\begin{aligned}T^{-1}\Delta\hat{U}_2'\hat{U}_2 - \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} &= \frac{\sum_{t=1}^T \Delta\hat{u}_{2t}\hat{u}_{2t}'}{T} - \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} \\ &= T^{-1}\Delta U_2' U_2 - \hat{\Delta}_{\Delta u_2\Delta u_2} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\sqrt{\frac{T}{nD_{nT}}})\end{aligned}$$

$$= O_p(K^{-2}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}).$$

Hence, if $T/n \to 0$ as $(n,T) \to \infty$, we have $T^{-1}\Delta\hat{U}_2'\hat{U}_2 - \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} = T^{-1}\Delta U_2' U_2 - \hat{\Delta}_{\Delta u_2 \Delta u_2} + o_p(1)$.

(g).

Lemma 8.1 (f) in the Appendix of Phillips (1995) states that

$$T^{-1}U_1^{'}U_2 - \hat{\Delta}_{u_1\Delta u_2} = K^{-2}\omega^{''}(0)\Psi_{12} + O_p(1/\sqrt{KT}) + o_p(K^{-2})$$

where $\Psi_{12} = \sum_{j=1}^{\infty}(j-1/2)\Gamma_{u_1u_2}(j)$. Thus $T^{-1}U_1^{'}U_2 - \hat{\Delta}_{u_1\Delta u_2} = O_p(K^{-2}) + O_p(1/\sqrt{KT}) + o_p(K^{-2})$. From Lemma A.5 (e), we have $\hat{\Delta}_{\hat{u}_1\Delta\hat{u}_2} = \hat{\Delta}_{u_1\Delta u_2} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}})$. From the proof of Lemma A.5 (a), we know that $\hat{u}_{2t} = \hat{F}_{2t} = A_2^{'}\hat{F}_t = A_2^{'}(\hat{F} - HF_t + HF_t) = u_{2t} + A_2^{'}\phi_t$ with $\phi_t = \hat{F}_t - HF_t$, and $\Delta\hat{u}_{2t} = \hat{f}_{2t} = A_2^{'}\hat{f}_t = A_2^{'}(\hat{f} - Hf_t + Hf_t) = \Delta u_{2t} + A_2^{'}\varphi_t$ with $\varphi_t = \hat{f}_t - Hf_t$. And $\hat{u}_{1t} = \hat{f}_{1t} = A_1^{'}\hat{f}_t = A_1^{'}(\hat{f} - Hf_t + Hf_t) = u_{1t} + A_1^{'}\varphi_t$ with $\varphi_t = \hat{f}_t - Hf_t$. This leads to

$$\begin{aligned}\hat{u}_{1t}\hat{u}_{2t}^{'} &= (u_{1t} + A_1^{'}\varphi_t)(u_{2t} + A_2^{'}\phi_t)^{'} \\ &= u_{1t}u_{2t}^{'} + A_1^{'}\varphi_t u_{2t}^{'} + u_{1t}\phi_t^{'}A_2 + A_1^{'}\varphi_t\phi_t^{'}A_2,\end{aligned}$$

and

$$\begin{aligned}T^{-1}\hat{U}_1^{'}\hat{U}_2 &= \frac{\sum_{t=1}^{T}\hat{u}_{1t}\hat{u}_{2t}^{'}}{T} \\ &= \frac{\sum_{t=1}^{T}u_{1t}u_{2t}^{'}}{T} + A_1^{'}\frac{\sum_{t=1}^{T}\varphi_t u_{2t}^{'}}{T} + \frac{\sum_{t=1}^{T}u_{1t}\phi_t^{'}}{T}A_2 + A_1^{'}\frac{\sum_{t=1}^{T}\varphi_t\phi_t^{'}}{T}A_2.\end{aligned}$$

It can be shown that $\frac{\sum_{t=1}^{T}\varphi_t u_{2t}^{'}}{T} = O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}})$, $\frac{\sum_{t=1}^{T}u_{1t}\phi_t^{'}}{T} = O_p(\sqrt{\frac{T}{n}})$, and $\frac{\sum_{t=1}^{T}\varphi_t\phi_t^{'}}{T} = O_p(\sqrt{\frac{T}{nD_{nT}}})$. Thus

$$\begin{aligned}T^{-1}\hat{U}_1^{'}\hat{U}_2 - \hat{\Delta}_{\hat{u}_1\Delta\hat{u}_2} &= \frac{\sum_{t=1}^{T}\hat{u}_{1t}\hat{u}_{2t}^{'}}{T} - \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} \\ &= T^{-1}U_1^{'}U_2 - \hat{\Delta}_{\Delta u_2\Delta u_2} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\sqrt{\frac{T}{nD_{nT}}}) \\ &= O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}).\end{aligned}$$

Hence, if $T/n \to 0$ as $(n,T) \to \infty$, we have $T^{-1}\hat{U}_1^{'}\hat{U}_2 - \hat{\Delta}_{\hat{u}_1\Delta\hat{u}_2} = T^{-1}U_1^{'}U_2 - \hat{\Delta}_{u_1\Delta u_2} + o_p(1)$.

(h). Lemma 8.1 (g) in the Appendix of Phillips (1995) states that

$$T^{-1}\Delta U_2^{'}F_1 - \hat{\Delta}_{\Delta u_2 u_1} = T^{-1}u_{2T}F_{1T}^{'} + K^{-2}\omega^{''}(0)\Psi_{21} + O_p(1/\sqrt{KT}) + o_p(K^{-2}),$$

where $\Psi_{21} = \sum_{j=1}^{\infty}(j-1/2)\Gamma_{u_2u_1}(j)$. Thus $T^{-1}\Delta U_2^{'}F_1 - \hat{\Delta}_{\Delta u_2 u_1} = O_p(K^{-2}) + O_p(1/\sqrt{KT}) + o_p(K^{-2})$. From Lemma A.5 (e), we have $\hat{\Delta}_{\Delta \hat{u}_2 \hat{u}_1} = \hat{\Delta}_{\Delta u_2 u_1} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}})$. We also have

$$\begin{aligned}
\Delta\hat{u}_{2t}\hat{F}_{1t}^{'} &= (\Delta u_{2t} + A_2^{'}\varphi_t)(A_1^{'}(\hat{F}_t - HF_t + F_t))^{'} \\
&= (\Delta u_{2t} + A_2^{'}\varphi_t)(F_{1t}^{'} + \phi_t^{'}A_1) \\
&= \Delta u_{2t}F_{1t}^{'} + A_2^{'}\varphi_t F_{1t}^{'} + \Delta u_{2t}\phi_t^{'}A_1 + A_2^{'}\varphi_t\phi_t^{'}A_1,
\end{aligned}$$

and

$$\begin{aligned}
T^{-1}\Delta\hat{U}_2^{'}\hat{F}_1 &= \frac{\sum_{t=1}^{T}\Delta\hat{u}_{2t}\hat{F}_{1t}^{'}}{T} \\
&= \frac{\sum_{t=1}^{T}\Delta u_{2t}F_{1t}^{'}}{T} + A_2^{'}\frac{\sum_{t=1}^{T}\varphi_t F_{1t}^{'}}{T} + \frac{\sum_{t=1}^{T}\Delta u_{2t}\phi_t^{'}}{T}A_1 + A_2^{'}\frac{\sum_{t=1}^{T}\varphi_t\phi_t^{'}}{T}A_1.
\end{aligned}$$

It can be easily shown that $\frac{\sum_{t=1}^{T}\Delta u_{2t}\phi_t^{'}}{T} = O_p(\sqrt{\frac{T}{n}})$, and $\frac{\sum_{t=1}^{T}\varphi_t\phi_t^{'}}{T} = O_p(\sqrt{\frac{T}{nD_{nT}}})$. We also have

$$\begin{aligned}
\frac{\sum_{t=1}^{T}\varphi_t F_{1t}^{'}}{T} &= \frac{\sum_{t=1}^{T}(\phi_t - \phi_{t-1})F_{1t}^{'}}{T} \\
&= T^{-1}\phi_T F_{1T}^{'} - T^{-1}\phi_0 F_{1,1}^{'} - \frac{\sum_{s=1}^{T-1}\phi_s(F_{1s}^{'} - F_{1,s+1}^{'})}{T} \\
&= T^{-1}\phi_T F_{1T}^{'} - T^{-1}\phi_0 F_{1,1}^{'} + \frac{\sum_{s=1}^{T-1}\phi_s u_{1,s+1}^{'}}{T} \\
&= T^{-1}\phi_T F_{1T}^{'} - T^{-1}\phi_0 F_{1,1}^{'} + O_p(\sqrt{\frac{T}{n}}) \\
&= O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) = O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}),
\end{aligned}$$

with the last line from the fact that $T^{-1}\phi_T F_{1T}^{'} = O_p(1/\sqrt{T})$, $T^{-1}\phi_0 F_{1,1}^{'} = O_p(1/T)$, and $\frac{\sum_{s=1}^{T-1}\phi_s u_{1,s+1}^{'}}{T} = O_p(\sqrt{\frac{T}{n}})$. Thus

$$\begin{aligned}
T^{-1}\Delta\hat{U}_2^{'}\hat{F}_1 - \hat{\Delta}_{\Delta\hat{u}_2\hat{u}_1} &= \frac{\sum_{t=1}^{T}\Delta\hat{u}_{2t}\hat{F}_{1t}^{'}}{T} - \hat{\Delta}_{\Delta\hat{u}_2\hat{u}_1} \\
&= T^{-1}\Delta U_2^{'}F_1 + T^{-1}\phi_T F_{1T}^{'} - T^{-1}\phi_0 F_{1,1}^{'} + O_p(\sqrt{\frac{T}{n}}) + O_p(\sqrt{\frac{T}{nD_{nT}}})
\end{aligned}$$

$$- \hat{\Delta}_{\Delta u_2 u_1} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}})$$

$$= T^{-1} u_{2T} F_{1T}^{'} + T^{-1} \phi_T F_{1T}^{'} - T^{-1} \phi_0 F_{1,1}^{'}$$

$$+ O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}})$$

$$= O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(T^{-1/2}) + O_p(\sqrt{\frac{T}{n}}).$$

(i). We have

$$\begin{aligned} T^{-1}\hat{U}_1^{'}\hat{F}_1 &= \frac{\sum_{t=1}^{T} \hat{u}_{1t}\hat{F}_{1t}^{'}}{T} \\ &= \frac{\sum_{t=1}^{T} (u_{1t} + A_1^{'}\varphi_t)(F_{1t}^{'} + \phi_t^{'}A_1)}{T} \\ &= \frac{\sum_{t=1}^{T} u_{1t}F_{1t}^{'}}{T} + A_1^{'}\frac{\sum_{t=1}^{T} \varphi_t F_{1t}^{'}}{T} + \frac{\sum_{t=1}^{T} u_{1t}\phi_t^{'}}{T}A_1 + A_1^{'}\frac{\sum_{t=1}^{T} \varphi_t\phi_t^{'}}{T}A_1 \\ &= T^{-1}U_1^{'}F_1 + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\sqrt{\frac{T}{nD_{nT}}}). \end{aligned}$$

with the last line from the fact that $\frac{\sum_{t=1}^{T} u_{1t}\phi_t^{'}}{T} = O_p(\sqrt{\frac{T}{n}})$, $\frac{\sum_{t=1}^{T} \varphi_t\phi_t^{'}}{T} = O_p(\sqrt{\frac{T}{nD_{nT}}})$, and $\frac{\sum_{t=1}^{T} \varphi_t F_{1t}^{'}}{T} = O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}})$ from part (h). Hence when $T/n \to 0$ as $(n,T) \to \infty$, we have

$$\begin{aligned} T^{-1}\hat{U}_1^{'}\hat{F}_1 - \hat{\Delta}_{\hat{u}_1\hat{u}_1} &= \frac{\sum_{t=1}^{T} \hat{u}_{1t}\hat{F}_{1t}^{'}}{T} - \hat{\Delta}_{\hat{u}_1\hat{u}_1} \\ &= T^{-1}U_1^{'}F_1 - \Delta_{u_1u_1} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\sqrt{\frac{T}{nD_{nT}}}) \\ &\xrightarrow{d} \int_0^1 dB_1 B_1^{'}. \end{aligned}$$

(j).

$$\begin{aligned} T^{-1}\hat{F}_1^{'}\varepsilon - \hat{\Delta}_{\hat{u}_1 0} &= \frac{\sum_{t=1}^{T} \hat{F}_{1t}\varepsilon_t}{T} - \hat{\Delta}_{\hat{u}_1 0} \\ &= \frac{\sum_{t=1}^{T} (F_{1t} + A_1^{'}\phi_t)\varepsilon_t}{T} - \hat{\Delta}_{\hat{u}_1 0} \\ &= \frac{\sum_{t=1}^{T} F_{1t}\varepsilon_t}{T} + A_1^{'}\frac{\sum_{t=1}^{T} \phi_t\varepsilon_t}{T} - \hat{\Delta}_{\hat{u}_1 0} \\ &= T^{-1}F_1^{'}\varepsilon - \Delta_{u_1 0} \\ &+ O_p(\sqrt{\frac{T}{n}}) + O_p((K/T)^{1/2}) + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right) \\ &= T^{-1}F_1^{'}\varepsilon - \Delta_{u_1 0} + O_p(K^{-2}) + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right). \end{aligned}$$

because $\frac{\sum_{t=1}^{T}\phi_t\varepsilon_t}{T} = O_p(\sqrt{\frac{T}{n}})$, and according to Lemma A.4 (i*), we have $\hat{\Delta}_{0\hat{u}_1} = \Delta_{\varepsilon u_1} +$ $+O_p((K/T)^{1/2}) + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right)$ when $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$. So when $K\sqrt{T}/\sqrt{n} \to 0$, $K/T \to 0$ and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$T^{-1}\hat{F}_1^{'}\varepsilon - \hat{\Delta}_{\hat{u}_1 0} = T^{-1}F_1^{'}\varepsilon - \Delta_{u_1\varepsilon} + o_p(1)$$
$$\xrightarrow{d} \int_0^1 B_1 dB_0.$$

■

Lemma A.6

Under Assumptions 1-5, 6 (EC), 7 (KL), 8(BW), $K\sqrt{T}/\sqrt{n} \to 0$, $K/T \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have:

(a)

$$\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_1 - \hat{\Delta}_{\hat{a}\hat{u}_1}] = \Omega_{01}\Omega_{11}^{-1}N_{11T} + O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}}{\sqrt{n}}),$$

where $N_{11T} \xrightarrow{d} \int_0^1 dB_1 B_1^{'}$; Under the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $K\sqrt{T}/\sqrt{n} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_1 - \hat{\Delta}_{\hat{a}\hat{u}_1}] = \Omega_{01}\Omega_{11}^{-1}N_{11T} + o_p(1) \xrightarrow{d} \Omega_{01}\Omega_{11}^{-1}\int_0^1 dB_1 B_1^{'};$$

(b)

$$\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_2 - \hat{\Delta}_{\hat{a}\Delta\hat{u}_2}] = O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{K^{3/2}}{T^{3/2}}) + O_p(\frac{K^{3/2}}{\sqrt{n}});$$

In the above expression, under the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $K\sqrt{T}/\sqrt{n} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have $\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_2 - \hat{\Delta}_{\hat{a}\Delta\hat{u}_2}] = o_p(1)$.

Furthermore, under the assumption that $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$, $K^{3/2}\sqrt{\frac{T}{n}} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$T^{1/2} \cdot \hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_2 - \hat{\Delta}_{\hat{a}\Delta\hat{u}_2}] = O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}})$$

$$= o_p(1);$$

(c) When $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$, we have

$$T^{1/2}[T^{-1}\varepsilon^{'}\hat{F}_2 - \hat{\Delta}_{0\Delta\hat{u}_2}] = T^{-1/2}\varepsilon^{'}F_2 + O_p(K^{-1/2}) \xrightarrow{d} N(0, \Omega_{\psi\psi});$$

Proof. From Lemma A.5 (c) and (d), we have

$$\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1} = (\hat{\Omega}_{\varepsilon a} + O_p\left(K\sqrt{\frac{T}{n}}\right) + O_p\left(\frac{K}{T}\right) + O_p(K^{-2}) + O_p\left(\frac{T}{n}\right))(\hat{\Omega}_{aa} + O_p(K^{-2}) + O_p(\sqrt{\frac{T}{n}}))^{-1}$$

$$= (\hat{\Omega}_{\varepsilon a} + o_p(1))(\hat{\Omega}_{aa} + o_p(1))^{-1} = \hat{\Omega}_{\varepsilon a}\hat{\Omega}_{aa}^{-1} + o_p(1),$$

when $K\sqrt{T}/\sqrt{n} \to 0$, $K/T \to 0$ and $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$. By Lemma A.4 part (d*) we have

$$\hat{\Omega}_{\varepsilon a}\hat{\Omega}_{aa}^{-1} = \begin{bmatrix} \Omega_{01}\Omega_{11}^{-1} + o_p(1), & -[\Phi_{02} - \Omega_{01}\Omega_{11}^{-1}\Phi_{12}]\Omega_{22}^{-1} + O_p(K^{3/2}/\sqrt{T}) + o_p(K^{3/2}/\sqrt{T}) \end{bmatrix}.$$

Thus,

$$\begin{aligned}
&\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F} - \hat{\Delta}_{\hat{a}\hat{a}}] = \hat{\Omega}_{\varepsilon a}\hat{\Omega}_{aa}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F} - \hat{\Delta}_{\hat{a}\hat{a}}] + o_p(1) \\
&= \begin{bmatrix} \Omega_{01}\Omega_{11}^{-1} + o_p(1), & -[\Phi_{02} - \Omega_{01}\Omega_{11}^{-1}\Phi_{12}]\Omega_{22}^{-1} + O_p(K^{3/2}/\sqrt{T}) + o_p(K^{3/2}/\sqrt{T}) \end{bmatrix} \\
&\times \begin{bmatrix} T^{-1}\hat{U}_1^{'}\hat{F}_1 - \hat{\Delta}_{\hat{u}_1\hat{u}_1} & T^{-1}\hat{U}_1^{'}\hat{U}_2 - \hat{\Delta}_{\hat{u}_1\Delta\hat{u}_2} \\ T^{-1}\Delta\hat{U}_2^{'}\hat{F}_1 - \hat{\Delta}_{\Delta\hat{u}_2\hat{u}_1} & T^{-1}\Delta\hat{U}_2^{'}\hat{U}_2 - \hat{\Delta}_{\Delta\hat{u}_2\Delta\hat{u}_2} \end{bmatrix} + o_p(1) \\
&= \begin{bmatrix} \Omega_{01}\Omega_{11}^{-1} + o_p(1), & -[\Phi_{02} - \Omega_{01}\Omega_{11}^{-1}\Phi_{12}]\Omega_{22}^{-1} + O_p(K^{3/2}/\sqrt{T}) + o_p(K^{3/2}/\sqrt{T}) \end{bmatrix} \\
&\times \begin{bmatrix} N_{11T} & O_p(K^{-2}) + O_p(1/\sqrt{KT}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) \\ O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) & O_p(K^{-2}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) \end{bmatrix} + o_p(1) \\
&= \begin{bmatrix} I \mid II \end{bmatrix}
\end{aligned}$$

with

$$\begin{aligned}
I &= \hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_1 - \hat{\Delta}_{\hat{a}\hat{u}_1}] \\
&= \Omega_{01}\Omega_{11}^{-1}N_{11T} + O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) \\
&+ O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{K}{T}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}}{\sqrt{n}}) \\
&= \Omega_{01}\Omega_{11}^{-1}N_{11T} + O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}}{\sqrt{n}}),
\end{aligned}$$

and

$$II = \hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_2 - \hat{\Delta}_{\hat{a}\Delta\hat{u}_2}]$$

$$= O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{K^{3/2}}{T^{3/2}}) + O_p(\frac{K^{3/2}}{\sqrt{n}})$$
$$= O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{T}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{K^{3/2}}{T^{3/2}}) + O_p(\frac{K^{3/2}}{\sqrt{n}}).$$

Under the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$I = \hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a'\hat{F}_1 - \hat{\Delta}_{\hat{a}\hat{u}_1}] = \Omega_{01}\Omega_{11}^{-1}N_{11T} + o_p(1) \xrightarrow{d} \Omega_{01}\Omega_{11}^{-1}\int_0^1 dB_1B_1'.$$

For term II, we have

$$\begin{aligned} T^{1/2}\cdot II &= T^{1/2}\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a'\hat{F}_2 - \hat{\Delta}_{\hat{a}\Delta\hat{u}_2}] \\ &= O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(K^{3/2}\sqrt{\frac{T}{n}}) \\ &= O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(K^{3/2}\sqrt{\frac{T}{n}}). \end{aligned}$$

Notice that when $K = O_e(T^k)$ for some $k > 1/4$, the term $O_p(\frac{T^{1/2}}{K^2}) = o_p(1)$, and when $k < 2/3$, the term $O_p(\frac{K^{3/2}}{T}) = o_p(1)$. We also require that $K^{3/2}\sqrt{\frac{T}{n}} \to 0$ to make the last error term negligible in the limit.

To conclude, under the assumption that $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$, and $K^{3/2}\sqrt{\frac{T}{n}} \to 0$ as $(n,T) \to \infty$, we have

$$T^{1/2}\cdot II = O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}}) = o_p(1).$$

For part (c), from Lemma A.4 (h*), under the assumption that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have $\hat{\Delta}_{0\Delta\hat{u}_2} := \hat{\Delta}_{\hat{\varepsilon}\Delta\hat{u}_2} = O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}})$. From the proof of Lemma A.5 (a), we know that $\hat{u}_{2t} = \hat{F}_{2t} = A_2'\hat{F}_t = A_2'(\hat{F} - HF_t + HF_t) = u_{2t} + A_2'\phi_t$ with $\phi_t = \hat{F}_t - HF_t$. This leads to

$$\begin{aligned} \varepsilon_t\hat{F}_{2t}' &= \varepsilon_t(u_{2t} + A_2'\phi_t)' \\ &= \varepsilon_t u_{2t}' + \varepsilon_t\phi_t' A_2, \end{aligned}$$

and

$$T^{-1/2}\varepsilon'\hat{F}_2 = \frac{\sum_{t=1}^T \varepsilon_t\hat{F}_{2t}'}{\sqrt{T}}$$

$$= \frac{\sum_{t=1}^{T} \varepsilon_t u_{2t}^{'}}{\sqrt{T}} + \frac{\sum_{t=1}^{T} \varepsilon_t \phi_t^{'}}{\sqrt{T}} A_2.$$

It can be shown that $\frac{\sum_{t=1}^{T} \varepsilon_t \phi_t^{'}}{\sqrt{T}} = O_p(\frac{T}{\sqrt{n}})$, and noticing Assumption 6 (EC), we have

$$T^{1/2}[T^{-1}\varepsilon \hat{F}_2 - \hat{\Delta}_{0\Delta\hat{u}_2}] = T^{-1/2}\varepsilon F_2 + O_p(\frac{T}{\sqrt{n}}) + O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}})$$

$$\xrightarrow{d} N(0, \Omega_{\psi\psi}),$$

under the assumption that $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$.

Discussion: (i) part (a) and part (b) comprise of the matrix $\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F} - \hat{\Delta}_{\hat{a}\hat{a}}]$, which corresponds to the separation of the FM correction terms into those that relate to the stationary and nonstationary coefficients, respectively. Part (b) gives the stationary coefficient correction more explicitly (and when it is scaled by $T^{1/2}$), as it is in the analysis of the limit distribution of the FM estimates of the stationary coefficients). The correction term in this case has magnitude $O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}})$, which is $o_p(1)$ when the bandwidth expansion rate $K = O_e(T^k)$ satisfies $1/4 < k < 2/3$ and $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$.

(ii) The condition $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ is more strict than the requirement that $K\sqrt{\frac{T}{n}} \to 0$, in which the latter condition is needed in the consistency of the long-run covariance estimates $\hat{\Omega}_{0\hat{a}}$. This condition $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ could be written as $\sqrt{\frac{T^3}{n}} \to 0$ since $O_p(\frac{K^{3/2}}{T}) = o_p(1)$ under the assumption that $K = O_e(T^k)$ satisfies $1/4 < k < 2/3$. This bandwidth expansion rate along with the extra requirement that $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ is different than that in Phillips (1995) because of the extra error terms $O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}})$. These terms are the results of the estimation error in the factors. In order to guarantee that the estimation error in the factors does not contaminate the limiting behavior of the FM estimates, we need more strict requirement on the relative rate of the bandwidth expansion rate, the cross sectional and time series sample sizes, i.e., $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ as $(n, T) \to \infty$.

(iii) Part (c) shows that the FM correction term for serial correlation (in the case of the stationary coefficients) also has no effect asymptotically and is $O_p(K^{-1/2})$. The submatrix that appears in part (a) relates to the FM endogeneity correction for the nonstationary coefficients. For the endogeneity correction to work we want this matrix to be $O_p(1)$ and to be as close to its dominating term, viz. $\Omega_{01}\Omega_{11}^{-1}N_{11T}$, as possible. Note that the error in this case involves a term of order $O_p(\frac{1}{K^2}) + O_p(\frac{1}{\sqrt{KT}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}}{\sqrt{n}})$. Thus this correction term operates satisfactorily provided $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$. Notice that we have used the fact that $O_p(\frac{K^{3/2}}{\sqrt{n}}) = O_p(\frac{K^{3/2}}{T}\frac{T}{\sqrt{n}}) = o_p(1)$ under the condition that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $T/\sqrt{n} \to 0$.

(iv) Combining the effects of the error terms for the stationary and the nonstationary coefficients we see that the correction terms work satisfactorily provided the bandwidth expansion rate $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$. ■

Proof of Theorem 1

Under Assumptions 1-5, 6 (EC), 7 (KL), and 8 (BW),

(a) under the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $K\sqrt{T}/\sqrt{n} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$TA_1'(\hat{\delta}_{FM} - \delta) \xrightarrow{d} (\int B_1 B_1')^{-1} \int_0^1 B_1 dB_{0\cdot 1};$$

(b) under the assumption that $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$, $K^{3/2}\sqrt{\frac{T}{n}} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$,

$$\sqrt{T}A_2'(\hat{\delta}_{FM} - \delta) \xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1}),$$

where $B_{0\cdot 1} = B_0 - \Omega_{01}\Omega_{11}^{-1}B_1 \equiv BM(\sigma_{00\cdot 1}^2)$ in which $\sigma_{00\cdot 1}^2 = \Omega_{00} - \Omega_{01}\Omega_{11}^{-1}\Omega_{10}$.

Proof. (a) From the endogeneity correction

$$y_t^+ = y_t - \hat{\Omega}_{0\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\Delta\hat{F}_t = y_t - \Delta\hat{F}_t'\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0}.$$

In matrix form, we have $Y^+ = Y - \Delta\hat{F}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0}$, in which $\Delta\hat{F} = (\hat{f}_1, \hat{f}_2, ..., \hat{f}_T)'$. Let $\phi_t = \hat{F}_t - HF_t$ denotes the estimation error of the factors and $\phi = (\phi_1, \phi_2, ...\phi_T)'$.

We can rewrite the cointegration regression as

$$\begin{aligned} y_t^+ &= \alpha' F_t + \varepsilon_t^+ \\ &= \alpha' H^{-1}\hat{F}_t + \varepsilon_t^+ + \alpha' H^{-1}(HF_t - \hat{F}_t) \\ &= \hat{F}_t'\delta + \varepsilon_t^+ + \alpha' H^{-1}(HF_t - \hat{F}_t), \end{aligned}$$

where $\varepsilon_t^+ = \varepsilon_t - \hat{\Omega}_{0\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\Delta\hat{F}_t$. In matrix notation, $Y^+ = \hat{F}\delta + \varepsilon^+ + (FH' - \hat{F})H^{-1'}\alpha$, where $Y^+ = (y_1^+, ..., y_T^+)'$, $\varepsilon^+ = (\varepsilon_1^+, ..., \varepsilon_T^+)'$, and $\hat{F} = (\hat{F}_1, ..., \hat{F}_{T-h})'$.

By definition

$$\begin{aligned} \hat{\delta}_{FM} &= (\hat{F}'\hat{F})^{-1}(\hat{F}'Y^+ - T\hat{\Delta}_{\hat{f}0}^+) \\ &= (\hat{F}'\hat{F})^{-1}(\hat{F}'(\hat{F}\delta + \varepsilon^+ + (FH' - \hat{F})H^{-1'}\alpha) - T\hat{\Delta}_{\hat{f}0}^+) \\ &= \delta + (\hat{F}'\hat{F})^{-1}\hat{F}'\varepsilon^+ + (\hat{F}'\hat{F})^{-1}\hat{F}'(FH' - \hat{F})H^{-1'}\alpha - (\hat{F}'\hat{F})^{-1}T\hat{\Delta}_{\hat{f}0}^+. \end{aligned}$$

So we have

$$A_1'(\hat{\delta}_{FM} - \delta) = A_1'\left(\hat{F}'\hat{F}\right)^{-1}\left(\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+\right) + A_1'\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'\left(FH' - \hat{F}\right)H^{-1'}\alpha,$$

and

$$A_2'(\hat{\delta}_{FM} - \delta) = A_2'\left(\hat{F}'\hat{F}\right)^{-1}\left(\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+\right) + A_2'\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'\left(FH' - \hat{F}\right)H^{-1'}\alpha.$$

By the partitioned inversion, we have

$$\begin{aligned} &A_1'(\hat{F}'\hat{F})^{-1}(\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+) = A_1'A'(A\hat{F}'\hat{F}A')^{-1}A(\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+) \\ &= \begin{bmatrix} I_{r_1} & 0 \end{bmatrix} \begin{bmatrix} \hat{F}_1'\hat{F}_1 & \hat{F}_1'\hat{F}_2 \\ \hat{F}_2'\hat{F}_1 & \hat{F}_2'\hat{F}_2 \end{bmatrix}^{-1} A(\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+) \\ &= \begin{bmatrix} I_{r_1} & 0 \end{bmatrix} \begin{bmatrix} (\hat{F}_1'Q_2\hat{F}_1)^{-1} & -(\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\hat{F}_2(\hat{F}_2'Q_1\hat{F}_2)^{-1} \\ -(\hat{F}_2'Q_1\hat{F}_2)^{-1}\hat{F}_2'\hat{F}_1(\hat{F}_1'\hat{F}_1)^{-1} & (\hat{F}_2'Q_1\hat{F}_2)^{-1} \end{bmatrix} A(\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+) \\ &= \begin{bmatrix} (\hat{F}_1'Q_2\hat{F}_1)^{-1} & -(\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1'\hat{F}_2(\hat{F}_2'Q_1\hat{F}_2)^{-1} \end{bmatrix} \begin{bmatrix} A_1' \\ A_2' \end{bmatrix} (\hat{F}'\varepsilon^+ - T\hat{\Delta}_{\hat{f}0}^+) \end{aligned}$$

$$= (\hat{F}_1^{'} Q_2 \hat{F}_1)^{-1} A_1^{'} (\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta \hat{F} \hat{\Omega}_{\hat{f}\hat{f}}^{-1} \hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+})$$
$$- (\hat{F}_1^{'}\hat{F}_1)^{-1} \hat{F}_1^{'} \hat{F}_2 (\hat{F}_2^{'} Q_1 \hat{F}_2)^{-1} A_2^{'} (\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta \hat{F} \hat{\Omega}_{\hat{f}\hat{f}}^{-1} \hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+}).$$

For the second last term in the last line, we have

$$A_1^{'}(\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta\hat{F}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+}) = \hat{F}_1^{'}\varepsilon - \hat{F}_1^{'}\Delta\hat{F}A^{'}(A\hat{\Omega}_{\hat{f}\hat{f}}A^{'})^{-1}A\hat{\Omega}_{\hat{f}0} - TA_1^{'}(\hat{\Delta}_{\hat{f}0} - \hat{\Delta}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0})$$
$$= \hat{F}_1^{'}\varepsilon - \hat{F}_1^{'}\Delta\hat{F}A^{'}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} - TA_1^{'}A^{'}A\hat{\Delta}_{\hat{f}0} + TA_1^{'}A^{'}(A\hat{\Delta}_{\hat{f}\hat{f}}A^{'})AA^{'}(A\hat{\Omega}_{\hat{f}\hat{f}}A^{'})^{-1}A\hat{\Omega}_{\hat{f}0}$$
$$= \hat{F}_1^{'}\varepsilon - \hat{F}_1^{'}\hat{U}_a\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} - T\begin{bmatrix} I_{r_1} & 0 \end{bmatrix}\hat{\Delta}_{\hat{a}0} + T\begin{bmatrix} I_{r_1} & 0 \end{bmatrix}\hat{\Delta}_{\hat{a}\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}$$
$$= \hat{F}_1^{'}\varepsilon - \hat{F}_1^{'}\hat{U}_a\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} - T\hat{\Delta}_{\hat{u}_1 0} + T\hat{\Delta}_{\hat{u}_1\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}$$
$$= (\hat{F}_1^{'}\varepsilon - T\hat{\Delta}_{\hat{u}_1 0}) - (\hat{F}_1^{'}\hat{U}_a - T\hat{\Delta}_{\hat{u}_1\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0},$$

where $\hat{U}_a = \Delta\hat{F}A^{'}$. Similarly, we have

$$A_2^{'}(\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta\hat{F}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+}) = \hat{F}_2^{'}\varepsilon - \hat{F}_2^{'}\Delta\hat{F}A^{'}(A\hat{\Omega}_{\hat{f}\hat{f}}A^{'})^{-1}A\hat{\Omega}_{\hat{f}0} - TA_2^{'}(\hat{\Delta}_{\hat{f}0} - \hat{\Delta}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0})$$
$$= \hat{F}_2^{'}\varepsilon - \hat{F}_2^{'}\Delta\hat{F}A^{'}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} - TA_2^{'}A^{'}A\hat{\Delta}_{\hat{f}0} + TA_2^{'}A^{'}(A\hat{\Delta}_{\hat{f}\hat{f}}A^{'})AA^{'}(A\hat{\Omega}_{\hat{f}\hat{f}}A^{'})^{-1}A\hat{\Omega}_{\hat{f}0}$$
$$= \hat{F}_2^{'}\varepsilon - \hat{F}_2^{'}\hat{U}_a\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} - T\begin{bmatrix} 0 & I_{r_2} \end{bmatrix}\hat{\Delta}_{\hat{a}0} + T\begin{bmatrix} 0 & I_{r_2} \end{bmatrix}\hat{\Delta}_{\hat{a}\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{a0}$$
$$= \hat{F}_2^{'}\varepsilon - \hat{F}_2^{'}\hat{U}_a\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} - T\hat{\Delta}_{\Delta\hat{u}_2 0} + T\hat{\Delta}_{\Delta\hat{u}_2\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}$$
$$= (\hat{F}_2^{'}\varepsilon - T\hat{\Delta}_{\Delta\hat{u}_2 0}) - (\hat{F}_2^{'}\hat{U}_a - T\hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}.$$

Then for FM-OLS estimates corresponding to the nonstationary factors, we have

$$TA_1^{'}(\hat{\delta}_{FM} - \delta) = A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}\left(\frac{\hat{F}^{'}\varepsilon^{+}}{T} - \hat{\Delta}_{\hat{f}0}^{+}\right) + A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}T^{-1}\hat{F}^{'}\left(FH^{'} - \hat{F}\right)H^{-1'}\alpha$$
$$= I + II,$$

in which

$$I = A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}\left(\frac{\hat{F}^{'}\varepsilon^{+}}{T} - \hat{\Delta}_{\hat{f}0}^{+}\right)$$
$$= (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}T^{-1}A_1^{'}(\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta\hat{F}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+})$$
$$- (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}T^{-1}A_2^{'}(\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta\hat{F}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+})$$

$$= (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}[(\frac{\hat{F}_1^{'}\varepsilon}{T} - \hat{\Delta}_{\hat{u}_1 0}) - (\frac{\hat{F}_1^{'}\hat{U}_a}{T} - \hat{\Delta}_{\hat{u}_1\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}]$$

$$- (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}[(\frac{\hat{F}_2^{'}\varepsilon}{T} - \hat{\Delta}_{\Delta\hat{u}_2 0}) - (\frac{\hat{F}_2^{'}\hat{U}_a}{T} - \hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}]$$

$$= I_1 - I_2,$$

with

$$I_1 = (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}[(\frac{\hat{F}_1^{'}\varepsilon}{T} - \hat{\Delta}_{\hat{u}_1 0}) - (\frac{\hat{F}_1^{'}\hat{U}_a}{T} - \hat{\Delta}_{\hat{u}_1\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}]$$

$$\xrightarrow{d} (\int B_1B_1^{'})^{-1}(\int_0^1 B_1 dB_0 - \int_0^1 B_1 dB_1^{'}\Omega_{11}^{-1}\Omega_{10})$$

$$= (\int B_1B_1^{'})^{-1}(\int_0^1 B_1 dB_{0\cdot 1}),$$

where $B_{0\cdot 1} = B_0 - \Omega_{01}\Omega_{11}^{-1}B_1$. By Lemma A.6 (a), we have $(\frac{\hat{F}_1^{'}\hat{U}_a}{T} - \hat{\Delta}_{\hat{u}_1\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} = N_{11T}^{'}\Omega_{11}^{-1}\Omega_{01} + o_p(1)$ with $N_{11T} \xrightarrow{d} \int_0^1 dB_1B_1^{'}$. Thus we have $(\frac{\hat{F}_1^{'}\hat{U}_a}{T} - \hat{\Delta}_{\hat{u}_1\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} \xrightarrow{d} \int_0^1 B_1 dB_1^{'}\Omega_{11}^{-1}\Omega_{01}$. By Lemma A.5 (j), we have $\frac{\hat{F}_1^{'}\varepsilon}{T} - \hat{\Delta}_{\hat{u}_1 0} \xrightarrow{d} \int_0^1 B_1 dB_0$. And

$$I_2 = (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}[(\frac{\hat{F}_2^{'}\varepsilon}{T} - \hat{\Delta}_{\Delta\hat{u}_2 0}) - (\frac{\hat{F}_2^{'}\hat{U}_a}{T} - \hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}]$$

$$= O_p(1)o_p(1).$$

By Lemma A.6 (b), we have $(\frac{\hat{F}_2^{'}\hat{U}_a}{T} - \hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0} = o_p(1)$ under the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$. By Lemma A.4 (h*), we have $\hat{\Delta}_{\Delta\hat{u}_2 0} = O_p(1/\sqrt{KT}) + O_p(\sqrt{\frac{T}{n}})$. And noticing $\frac{\hat{F}_2^{'}\varepsilon}{T} = \frac{1}{\sqrt{T}}\frac{\hat{F}_2^{'}\varepsilon}{\sqrt{T}} = O_p(1/\sqrt{T})$, we have the above result that $I_2 = o_p(1)$. From the proof of Lemma 1, we have $II = A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1} T^{-1}\hat{F}^{'}\left(FH^{'} - \hat{F}\right)H^{-1'}\alpha = O_p(\frac{T}{\sqrt{n}}) + O_p(\sqrt{\frac{T}{n}})$.

To conclude, under the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $K\sqrt{T}/\sqrt{n} \to 0$, and $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$, we have

$$TA_1^{'}(\hat{\delta}_{FM} - \delta) = A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}\left(\frac{\hat{F}^{'}\varepsilon^{+}}{T} - \hat{\Delta}_{\hat{f}0}^{+}\right) + A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1} T^{-1}\hat{F}^{'}\left(FH^{'} - \hat{F}\right)H^{-1'}\alpha$$

$$= I_1 - I_2 + II$$

$$\xrightarrow{d} (\int B_1B_1^{'})^{-1}(\int_0^1 B_1 dB_{0\cdot 1}).$$

The assumption $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $K\sqrt{T}/\sqrt{n} \to 0$, and $T/\sqrt{n} \to 0$ as $(n, T) \to \infty$, comes from combining the assumption $K\sqrt{T}/\sqrt{n} \to 0$, $K/T \to 0$ and $T/\sqrt{n} \to 0$

as $(n,T) \to \infty$ in Lemma A.5 (j), the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$ in Lemma A.6 (a), and the assumption that $K = O_e(T^k)$ for some $k \in (0, 2/3)$, $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$ as in Lemma A.6 (b).

Notice that the assumption $K = O_e(T^k)$ for some $k \in (0, 2/3)$ as $(n,T) \to \infty$ is the same as $K = O_e(T^k)$ for some $k \in (0, 2/3)$ as $T \to \infty$ in Phillips (1995) for the nonstationary estimates. However, we require the extra condition that $K\sqrt{T}/\sqrt{n} \to 0$ in addition to the condition that $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$. In Lemma 8.1 of Phillips (1995, p.1058), which shows the consistency of the kernel estimates with observable regressors, the only requirement on the bandwidth expansion rate is the one stated in Assumption 8 (BW). But with estimation errors in the factors (converge at rate $O_p(\sqrt{T/n})$), the induced errors in the kernel estimates will accumulate at rate $O_p(K\sqrt{T/n})$. Thus in order to guarantee the consistency of the kernel estimates, the extra restriction $K\sqrt{T/n} \to 0$ should be imposed. In another words, using estimated factors does not affect the consistency of the kernel estimates as long as the estimation errors of the factors converge to zero fast enough relative to the bandwidth expansion rate.

(b) For FM-OLS estimates corresponding to the stationary factors (cointegrated factors), we have

$$\begin{aligned}\sqrt{T}A_2'(\hat{\delta}_{FM} - \delta) &= A_2'\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\left(\hat{F}'\varepsilon^+ - T\hat{\Delta}^+_{\hat{f}0}\right) + A_2'\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\hat{F}'\left(FH' - \hat{F}\right)H^{-1'}\alpha \\ &= III + IV,\end{aligned}$$

with

$$\begin{aligned}III &= A_2'\left(\frac{\hat{F}'\hat{F}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\left(\hat{F}'\varepsilon^+ - T\hat{\Delta}^+_{\hat{f}0}\right), \\ &= \left(\frac{\hat{F}_2'Q_1\hat{F}_2}{T}\right)^{-1}\sqrt{T}A_2'\left(\frac{\hat{F}'\varepsilon}{T} - \frac{\hat{F}'\Delta\hat{F}}{T}\hat{\Omega}^{-1}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}0} - \hat{\Delta}^+_{\hat{f}0}\right) \\ &\quad - \frac{1}{T}\left(\frac{\hat{F}_2'Q_1\hat{F}_2}{T}\right)^{-1}\frac{\hat{F}_2'\hat{F}_1}{T}\left(\frac{\hat{F}_1'\hat{F}_1}{T^2}\right)^{-1}\sqrt{T}A_1'\left(\frac{\hat{F}'\varepsilon}{T} - \frac{\hat{F}'\Delta\hat{F}}{T}\hat{\Omega}^{-1}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}0} - \hat{\Delta}^+_{\hat{f}0}\right) \\ &= \left[\left(\frac{\hat{F}_2'Q_1\hat{F}_2}{T}\right)^{-1}A_2' - O_p(T^{-1})A_1'\right]\sqrt{T}\left(\frac{\hat{F}'\varepsilon}{T} - \frac{\hat{F}'\Delta\hat{F}}{T}\hat{\Omega}^{-1}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}0} - \hat{\Delta}^+_{\hat{f}0}\right) \\ &= \left(\frac{\hat{F}_2'Q_1\hat{F}_2}{T}\right)^{-1}\sqrt{T}A_2'\left(\frac{\hat{F}'\varepsilon}{T} - \frac{\hat{F}'\Delta\hat{F}}{T}\hat{\Omega}^{-1}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}0} - \hat{\Delta}^+_{\hat{f}0}\right) + O_p(T^{-1/2}),\end{aligned}$$

and

$$IV = A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\hat{F}^{'}\left(FH^{'}-\hat{F}\right)H^{-1'}\alpha$$
$$= (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{\sqrt{T}} - (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T}\frac{1}{\sqrt{T}}$$
$$= O_p(\frac{T}{\sqrt{n}}) + O_p(\sqrt{\frac{T}{n}}),$$

with the last line from the proof of Lemma 1.

Recall that

$$A_2^{'}(\hat{F}^{'}\varepsilon - \hat{F}^{'}\Delta\hat{F}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0} - T\hat{\Delta}_{\hat{f}0}^{+}) = \hat{F}_2^{'}\varepsilon - \hat{F}_2^{'}\Delta\hat{F}A^{'}(A\hat{\Omega}_{\hat{f}\hat{f}}A^{'})^{-1}A\hat{\Omega}_{\hat{f}0} - TA_2^{'}(\hat{\Delta}_{\hat{f}0} - \hat{\Delta}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0})$$
$$= (\hat{F}_2^{'}\varepsilon - T\hat{\Delta}_{\Delta\hat{u}_2 0}) - (\hat{F}_2^{'}\hat{U}_a - T\hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}.$$

Thus

$$\sqrt{T}A_2^{'}(\frac{\hat{F}^{'}\varepsilon}{T} - \frac{\hat{F}^{'}\Delta\hat{F}}{T}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0} - \hat{\Delta}_{\hat{f}0}^{+}) = \frac{\hat{F}_2^{'}\varepsilon}{\sqrt{T}} - \frac{\hat{F}_2^{'}\Delta\hat{F}A^{'}}{\sqrt{T}}(A\hat{\Omega}_{\hat{f}\hat{f}}A^{'})^{-1}A\hat{\Omega}_{\hat{f}0} - \sqrt{T}A_2^{'}(\hat{\Delta}_{\hat{f}0} - \hat{\Delta}_{\hat{f}\hat{f}}\hat{\Omega}_{\hat{f}\hat{f}}^{-1}\hat{\Omega}_{\hat{f}0})$$
$$= \sqrt{T}(\frac{\hat{F}_2^{'}\varepsilon}{T} - \hat{\Delta}_{\Delta\hat{u}_2 0}) - \sqrt{T}(\frac{\hat{F}_2^{'}\hat{U}_a}{T} - \hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}.$$

From Lemma A.6 (c), we have when $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$T^{1/2}[T^{-1}\varepsilon^{'}\hat{F}_2 - \hat{\Delta}_{0\Delta\hat{u}_2}] = T^{-1/2}\varepsilon^{'}F_2 + O_p(K^{-1/2}) \xrightarrow{d} N(0,\Omega_{\psi\psi}),$$

and Lemme A.6 (b) states that under the assumption that $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$, $K^{3/2}\sqrt{\frac{T}{n}} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$T^{1/2}\cdot\hat{\Omega}_{0\hat{a}}\hat{\Omega}_{\hat{a}\hat{a}}^{-1}[T^{-1}\hat{U}_a^{'}\hat{F}_2 - \hat{\Delta}_{\hat{a}\Delta\hat{u}_2}] = O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}})$$
$$= o_p(1).$$

Hence, under the assumption that $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$, $K^{3/2}\sqrt{\frac{T}{n}} \to 0$, and $T/\sqrt{n} \to 0$ as $(n,T) \to \infty$, we have

$$\sqrt{T}A_2^{'}(\delta_{FM} - \delta) = A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\left(\hat{F}^{'}\varepsilon^{+} - T\hat{\Delta}_{\hat{f}0}^{+}\right) + A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\hat{F}^{'}\left(FH^{'}-\hat{F}\right)H^{-1'}\alpha$$
$$= (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\left(\sqrt{T}(\frac{\hat{F}_2^{'}\varepsilon}{T} - \hat{\Delta}_{\Delta\hat{u}_2 0}) - \sqrt{T}(\frac{\hat{F}_2^{'}\hat{U}_a}{T} - \hat{\Delta}_{\Delta\hat{u}_2\hat{a}})\hat{\Omega}_{\hat{a}\hat{a}}^{-1}\hat{\Omega}_{\hat{a}0}\right) + O_p(T^{-1/2})$$
$$+ O_p(\frac{T}{\sqrt{n}}) + O_p(\sqrt{\frac{T}{n}})$$

$$\xrightarrow{d} N(0, \Sigma_{22}^{-1}\Omega_{\psi\psi}\Sigma_{22}^{-1}).$$

Notice that the assumption $K = O_e(T^k)$ for some $k \in (1/4, 2/3)$ as $(n, T) \to \infty$ is tighter than that $K = O_e(T^k)$ for some $k \in (1/4, 1)$ as $T \to \infty$ in Phillips (1995) for the stationary estimates. This tighter bandwidth expansion rate comes from Lemma A.6 (b), which gives the stationary coefficient correction more explicitly (and when it is scaled by $T^{1/2}$). The correction term in this case has magnitude $O_p(\frac{T^{1/2}}{K^2}) + O_p(\frac{1}{\sqrt{K}}) + O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}})$, which is $o_p(1)$ when the bandwidth expansion rate $K = O_e(T^k)$ satisfies $1/4 < k < 2/3$ and $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$.

This tighter bandwidth expansion rate comes from the accumulation of estimation errors in the factors across the summation of K sample covariances. Thus, to guarantee the estimation error in the factors does not contaminate the limiting behavior of the long-run covariance estimates, we do not allow the Bandwidth expansion rate to be too large. We also impose the more strict relative expansion rate $K^{3/2}\sqrt{\frac{T}{n}} \to 0$ for the stationary FM estimates than for the nonstationary FM estimates (which only requires $K\sqrt{\frac{T}{n}} \to 0$, which is needed in the consistency of the long-run covariance estimates $\hat{\Omega}_{0\hat{a}}$). This condition $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ could be written as $\sqrt{\frac{T^3}{n}} \to 0$ since $O_p(\frac{K^{3/2}}{T}) = o_p(1)$ under the assumption that $K = O_e(T^k)$ satisfies $1/4 < k < 2/3$. This bandwidth expansion rate along with the extra requirement that $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ is different than that in Phillips (1995) because of the extra error terms $O_p(\frac{T}{\sqrt{n}}) + O_p(\frac{K^{3/2}}{T}) + O_p(\frac{K^{3/2}T^{1/2}}{\sqrt{n}})$. These terms are the results of the estimation error in the factors. In order to guarantee that the estimation error in the factors does not contaminate the limiting behavior of the FM estimates, we need more strict requirement on the relative rate of the bandwidth expansion rate, the cross sectional and time series sample sizes, i.e., $\frac{K^{3/2}T^{1/2}}{\sqrt{n}} \to 0$ as $(n, T) \to \infty$.

■

Proof of Lemma 2

Suppose Assumptions 1-5 and Assumption 9 (EC') hold. As $(n,T) \to \infty$, if $T/n \to 0$,

(a) $A_1^{'}(\hat{\delta}-\delta) \xrightarrow{d} (\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$,

(b) $A_2^{'}(\hat{\delta}-\delta) \xrightarrow{d} \Sigma_{22}^{-1}(\int_0^1 dB_2 B_0 + \Delta_{20}) - \Sigma_{22}^{-1}(\int_0^1 dB_2 B_1 + \Delta_{21})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$. **Proof.**

Rewrite the original regression equation as follows

$$\begin{aligned} y_t &= \alpha^{'} F_t + \varepsilon_t \\ &= \alpha^{'} H^{-1}\hat{F}_t + \varepsilon_t + \alpha^{'} H^{-1}(HF_t - \hat{F}_t). \end{aligned}$$

In matrix notation, $Y = \hat{F}\delta + \varepsilon + (FH^{'} - \hat{F})\delta$. It follows that

$$\hat{\delta} - \delta = (\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}\varepsilon + (\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}(FH^{'} - \hat{F})\delta.$$

Partitioning the coefficients into the nonstationary and stationary part, we have

$$A_1^{'}(\hat{\delta}-\delta) = A_1^{'}\left(\hat{F}^{'}\hat{F}\right)^{-1}\hat{F}^{'}\varepsilon + A_1^{'}\left(\hat{F}^{'}\hat{F}\right)^{-1}\hat{F}^{'}(FH^{'} - \hat{F})H^{-1'}\alpha,$$

and

$$A_2^{'}(\hat{\delta}-\delta) = A_2^{'}\left(\hat{F}^{'}\hat{F}\right)^{-1}T^{-1}\hat{F}^{'}\varepsilon + A_2^{'}\left(\hat{F}^{'}\hat{F}\right)^{-1}\hat{F}^{'}(FH^{'} - \hat{F})H^{-1'}\alpha.$$

By the proof of Lemma 1, we have

$$A_1^{'}(\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}\varepsilon = (\hat{F}_1^{'}Q_2\hat{F}_1)^{-1}\hat{F}_1^{'}\varepsilon - (\hat{F}_1^{'}\hat{F}_1)^{-1}\hat{F}_1^{'}\hat{F}_2(\hat{F}_2^{'}Q_1\hat{F}_2)^{-1}\hat{F}_2^{'}\varepsilon$$

and

$$A_2^{'}(\hat{F}^{'}\hat{F})^{-1}\hat{F}^{'}\varepsilon = (\hat{F}_2^{'}Q_1\hat{F}_2)^{-1}\hat{F}_2^{'}\varepsilon - (\hat{F}_2^{'}Q_1\hat{F}_2)^{-1}\hat{F}_2^{'}\hat{F}_1(\hat{F}_1^{'}\hat{F}_1)^{-1}\hat{F}_1^{'}\varepsilon$$

where $Q_i = I - \hat{F}_i(\hat{F}_i^{'}\hat{F}_i)^{-1}\hat{F}_i^{'}$, i=1,2. Thus

$$\begin{aligned} A_1^{'}(\hat{\delta}-\delta) &= A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}T^{-2}\hat{F}^{'}\varepsilon + A_1^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T^2}\right)^{-1}T^{-2}\hat{F}^{'}(FH^{'} - \hat{F})H^{-1'}\alpha \\ &= (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T^2} - (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T^2} \\ &+ (\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'} - \hat{F})H^{-1'}\alpha}{T^2} - (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'} - \hat{F})H^{-1'}\alpha}{T^2}. \end{aligned}$$

in which

$$(\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T^2} = (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T^2} + O_p(\frac{1}{T^2}) \xrightarrow{d} (\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0),$$

and

$$(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T^2} = (\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T}\frac{1}{T} = O_p(\frac{1}{T}),$$

and

$$(\frac{\hat{F}_1^{'}Q_2\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T^2} = O_p(1)O_p(\frac{T}{\sqrt{n}}\times\frac{1}{T}) = O_p(\frac{1}{\sqrt{n}}),$$

and

$$(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\hat{F}_2}{T}(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T^2} = O_p(1)O_p(1)O_p(1)O_p(\sqrt{\frac{T}{n}}\times\frac{1}{T}) = O_p(\frac{1}{\sqrt{nT}}).$$

The last two results come from Lemma A.2. Hence, we have

$$A_1^{'}(\hat{\delta}-\delta) \xrightarrow{d} (\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0).$$

Similarly, we have

$$\begin{aligned}
A_2^{'}(\hat{\delta}-\delta) &= A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}T^{-1}\hat{F}^{'}\varepsilon + A_2^{'}\left(\frac{\hat{F}^{'}\hat{F}}{T}\right)^{-1}T^{-1}\hat{F}^{'}(FH^{'}-\hat{F})H^{-1'}\alpha \\
&= (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T} - (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T^2} \\
&+ (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T} - (\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T}\frac{1}{T},
\end{aligned}$$

in which

$$(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T} = (\frac{\hat{F}_2^{'}\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\varepsilon}{T} + O_p(\frac{1}{T}) \xrightarrow{d} \Sigma_{22}^{-1}(\int_0^1 dB_2B_0 + \Delta_{20}),$$

and

$$(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}\varepsilon}{T^2} \xrightarrow{d} \Sigma_{22}^{-1}(\int_0^1 dB_2B_1^{'} + \Delta_{21})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0),$$

and

$$(\frac{\hat{F}_2^{'}Q_2\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T} = O_p(\sqrt{\frac{T}{n}}),$$

and

$$(\frac{\hat{F}_2^{'}Q_1\hat{F}_2}{T})^{-1}\frac{\hat{F}_2^{'}\hat{F}_1}{T}(\frac{\hat{F}_1^{'}\hat{F}_1}{T^2})^{-1}\frac{\hat{F}_1^{'}(FH^{'}-\hat{F})H^{-1'}\alpha}{T}\frac{1}{T} = \frac{1}{T}O_p(\frac{T}{\sqrt{n}}) = O_p(\frac{1}{\sqrt{n}}).$$

The last two expressions come from Lemma A.2. If $T/n \to 0$ as $(n,T) \to \infty$, the last two expressions will be $o_p(1)$.

Hence, we have

$$A_2'(\hat{\delta} - \delta) \xrightarrow{d} \Sigma_{22}^{-1}(\int_0^1 dB_2 B_0 + \Delta_{20}) - \Sigma_{22}^{-1}(\int_0^1 dB_2 B_1' + \Delta_{21})(\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0).$$

■

Proof of Theorem 2

Suppose Assumptions 1-5 and Assumption (9) hold. As $(n,T) \to \infty$, if $T/n \to 0$,

(a)

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0 + \Lambda_{\tilde{0}\tilde{0}}}{\int_0^1 \tilde{B}_0 \tilde{B}_0},$$

where $\tilde{B}_0 = B_0 - (\int_0^1 B_0 B_1')(\int B_1 B_1')^{-1} B_1$. The Brownian motion $\tilde{B}_0$ has long-run covariance matrix $\Omega_{\tilde{0}\tilde{0}} = \Omega_{00} - (\int_0^1 B_0 B_1')(\int B_1 B_1')^{-1}\Omega_{10} - \Omega_{01}(\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0) + (\int_0^1 B_0 B_1')(\int B_1 B_1')^{-1}\Omega_{11}(\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0)$ and one-sided long-run covariance $\Lambda_{\tilde{0}\tilde{0}} = \Lambda_{00} - (\int_0^1 B_0 B_1')(\int B_1 B_1')^{-1}\Lambda_{10} - \Lambda_{01}(\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0) + (\int_0^1 B_0 B_1')(\int B_1 B_1')^{-1}\Lambda_{11}(\int B_1 B_1')^{-1}(\int_0^1 B_1 B_0)$.

(b) If $q \to \infty$ as $T \to \infty$ but $q/T \to 0$, then the statistic $Z_{\rho,T}$ satisfies

$$Z_{\rho,T} \xrightarrow{d} Z_n,$$

where

$$Z_n = \frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0}{\int_0^1 \tilde{B}_0 \tilde{B}_0} = \frac{\int_0^1 W(r) dW(r)}{\int_0^1 W(r) W(r) dr},$$

in which $W(r)$ is a one dimensional standard Brownian motion.

Proof. (a) Under the null hypothesis that there is no cointegration relation between y_t and F_t, i.e., there is a unit root in ε_t,

$$T(\hat{\rho}_T - 1) = \frac{T^{-1}\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})\hat{\varepsilon}_{t-1}}{T^{-2}\sum_{t=2}^T \hat{\varepsilon}_{t-1}^2}.$$

Notice that

$$
\begin{aligned}
\hat{\varepsilon}_t &= y_t - \hat{\delta}^{'}\hat{F}_t \\
&= (\alpha^{'}H^{-1}HF_t + \varepsilon_t) - \hat{\delta}^{'}\hat{F}_t \\
&= \delta^{'}HF_t + \varepsilon_t - \hat{\delta}^{'}HF_t + \hat{\delta}^{'}HF_t - \hat{\delta}^{'}\hat{F}_t \\
&= \varepsilon_t - (\hat{\delta}-\delta)^{'}HF_t - \hat{\delta}^{'}(\hat{F}_t - HF_t) \\
&= \varepsilon_t - (\hat{\delta}-\delta)^{'}A_1 \cdot F_{1t} - (\hat{\delta}-\delta)^{'}A_2 \cdot F_{2t} - \hat{\delta}^{'}\phi_t,
\end{aligned}
$$

in which $\phi_t = \hat{F}_t - HF_t$. We have

$$
T^{-2}\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2 = T^{-2}\sum_{t=2}^{T}\{\varepsilon_{t-1} - (\hat{\delta}-\delta)^{'}A_1 \cdot F_{1,t-1} - (\hat{\delta}-\delta)^{'}A_2 \cdot F_{2,t-1} - \hat{\delta}^{'}\phi_{t-1}\}^2
$$

$$
= \frac{1}{T^2}\sum_{t=2}^{T}\varepsilon_{t-1}^2 + (\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}F_{1,t-1}^{'}}{T^2}A_1^{'}(\hat{\delta}-\delta) + (\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{F_{2,t-1}F_{2,t-1}^{'}}{T^2}A_2^{'}(\hat{\delta}-\delta) + \hat{\delta}^{'}\sum_{t=2}^{T}\frac{\phi_{t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta}
$$

$$
- 2\sum_{t=2}^{T}\frac{\varepsilon_{t-1}F_{1,t-1}^{'}}{T^2}A_1^{'}(\hat{\delta}-\delta) - 2\sum_{t=2}^{T}\frac{\varepsilon_{t-1}F_{2,t-1}^{'}}{T^2}A_2^{'}(\hat{\delta}-\delta) - 2\sum_{t=2}^{T}\frac{\varepsilon_{t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta}
$$

$$
+ 2(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}F_{2,t-1}^{'}}{T^2}A_2^{'}(\hat{\delta}-\delta) + 2(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta}
$$

$$
+ 2(\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{F_{2,t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta},
$$

For the first four squared terms, we have

$$
\frac{1}{T^2}\sum_{t=2}^{T}\varepsilon_{t-1}^2 \xrightarrow{d} \int_0^1 B_0B_0,
$$

and

$$
\begin{aligned}
(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}F_{1,t-1}^{'}}{T^2}A_1^{'}(\hat{\delta}-\delta) &\xrightarrow{d} (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) \\
&= (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0),
\end{aligned}
$$

and

$$
(\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{F_{2,t-1}F_{2,t-1}^{'}}{T^2}A_2^{'}(\hat{\delta}-\delta) = O_p(1)O_p(\frac{1}{T})O_p(1) = O_p(\frac{1}{T}),
$$

and

$$
\hat{\delta}^{'}\sum_{t=2}^{T}\frac{\phi_{t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta} = O_p(1)O_p(\frac{T}{n}\times\frac{1}{T})O_p(1) = O_p(\frac{1}{n}).
$$

For the next six cross-product terms, we have

$$2\sum_{t=2}^{T}\frac{\varepsilon_{t-1}F_{1,t-1}^{'}}{T^2}A_1^{'}(\hat{\delta}-\delta)\xrightarrow{d}2(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0),$$

and

$$2\sum_{t=2}^{T}\frac{\varepsilon_{t-1}F_{2,t-1}^{'}}{T^2}A_2^{'}(\hat{\delta}-\delta)=2\frac{1}{T}\sum_{t=2}^{T}\frac{\varepsilon_{t-1}F_{2,t-1}^{'}}{T}A_2^{'}(\hat{\delta}-\delta)=O_p(\frac{1}{T}),$$

and

$$\|\sum_{t=2}^{T}\frac{\varepsilon_{t-1}\phi_{t-1}^{'}}{T^2}\|\leq\frac{1}{\sqrt{T}}(\sum_{t=2}^{T}\frac{\varepsilon_{t-1}^2}{T^2})^{1/2}(\sum_{t=2}^{T}\frac{\|\phi_{t-1}\|^2}{T})^{1/2}=\frac{1}{\sqrt{T}}O_p(1)O_p(\sqrt{\frac{T}{n}})=O_p(\frac{1}{\sqrt{n}}),$$

and

$$2(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}F_{2,t-1}^{'}}{T^2}A_2^{'}(\hat{\delta}-\delta)=O_p(\frac{1}{T}),$$

and

$$2(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta}\leq O_p(\frac{1}{\sqrt{n}}),$$

and

$$2(\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{F_{2,t-1}\phi_{t-1}^{'}}{T^2}\hat{\delta}\leq O_p(\frac{1}{\sqrt{Tn}}).$$

Hence,

$$\begin{aligned}
&T^{-2}\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2=T^{-2}\sum_{t=2}^{T}\{\varepsilon_{t-1}-(\hat{\delta}-\delta)^{'}A_1\cdot F_{1,t-1}-(\hat{\delta}-\delta)^{'}A_2\cdot F_{2,t-1}-\hat{\delta}^{'}\phi_{t-1}\}^2\\
&=\frac{1}{T^2}\sum_{t=2}^{T}\varepsilon_{t-1}^2+(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{F_{1,t-1}F_{1,t-1}^{'}}{T^2}A_1^{'}(\hat{\delta}-\delta)+O_p(\frac{1}{T})+O_p(\frac{1}{n})\\
&-2\sum_{t=2}^{T}\frac{\varepsilon_{t-1}F_{1,t-1}^{'}}{T^2}A_1^{'}(\hat{\delta}-\delta)-O_p(\frac{1}{T})-O_p(\frac{1}{\sqrt{n}})\\
&+O_p(\frac{1}{T})+O_p(\frac{1}{\sqrt{n}})\\
&+O_p(\frac{1}{\sqrt{Tn}})\\
&\xrightarrow{d}\int_0^1 B_0B_0-(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0)\\
&=\int_0^1\tilde{B}_0\tilde{B}_0.
\end{aligned}$$

in which $\tilde{B}_0 = B_0-(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}B_1$. The Brownian motion $\tilde{B}_0$

has long-run covariance matrix $\Omega_{\tilde{0}\tilde{0}} = \Omega_{00} - (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1}\Omega_{10} - \Omega_{01}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0) + (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1}\Omega_{11}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$ and one-sided long-run covariance $\Lambda_{\tilde{0}\tilde{0}} = \Lambda_{00} - (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1}\Lambda_{10} - \Lambda_{01}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0) + (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1}\Lambda_{11}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$.

For the numerator in $T(\hat{\gamma}-1)$, we have

$$
\begin{aligned}
&T^{-1}\sum_{t=2}^{T}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})\hat{\varepsilon}_{t-1} = T^{-1}\sum_{t=2}^{T}(\Delta\varepsilon_t - (\hat{\delta}-\delta)^{'}A_1\cdot\Delta F_{1t} - (\hat{\delta}-\delta)^{'}A_2\cdot\Delta F_{2t} - \hat{\delta}^{'}\Delta\phi_t) \\
&\cdot(\varepsilon_{t-1} - (\hat{\delta}-\delta)^{'}A_1\cdot F_{1,t-1} - (\hat{\delta}-\delta)^{'}A_2\cdot F_{2,t-1} - \hat{\delta}^{'}\phi_{t-1}) \\
&= \sum_{t=2}^{T}\frac{\Delta\varepsilon_t\varepsilon_{t-1}}{T} - (\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}\varepsilon_{t-1}}{T} - (\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}\varepsilon_{t-1}}{T} - \hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t\varepsilon_{t-1}}{T} \\
&- \sum_{t=2}^{T}\frac{\Delta\varepsilon_t F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) + (\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) \\
&+ (\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) - \hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) \\
&- \sum_{t=2}^{T}\frac{\Delta\varepsilon_t F_{2,t-1}^{'}}{T}A_2^{'}(\hat{\delta}-\delta) + (\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}F_{2,t-1}^{'}}{T}A_2^{'}(\hat{\delta}-\delta) \\
&+ (\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}F_{2,t-1}^{'}}{T}A_2^{'}(\hat{\delta}-\delta) - \hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t F_{2,t-1}^{'}}{T}A_2^{'}(\hat{\delta}-\delta) \\
&- \sum_{t=2}^{T}\frac{\Delta\varepsilon_t\phi_{t-1}^{'}}{T}\hat{\delta} - (\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}\phi_{t-1}^{'}}{T}\hat{\delta} - (\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}\phi_{t-1}^{'}}{T}\hat{\delta} - \hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t\phi_{t-1}^{'}}{T}\hat{\delta}.
\end{aligned}
$$

For the first terms, we have

$$
\sum_{t=2}^{T}\frac{\Delta\varepsilon_t\varepsilon_{t-1}}{T} \xrightarrow{d} \int_0^1 B_0 dB_0 + \Lambda_{00}.
$$

Consider the second term together:

$$
(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}\varepsilon_{t-1}}{T} \xrightarrow{d} (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1}(\int_0^1 dB_1 B_0 + \Lambda_{10}),
$$

and the fifth term:

$$
\sum_{t=2}^{T}\frac{\Delta\varepsilon_t F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) \xrightarrow{d} (\int_0^1 dB_0 B_1^{'} + \Lambda_{01})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0).
$$

Consider the second and the fifth term together:

$$
(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}\varepsilon_{t-1}}{T} + \sum_{t=2}^{T}\frac{\Delta\varepsilon_t F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta)
$$

$$= \frac{\varepsilon_{T-1}F_{1T}^{'} - \varepsilon_1 F_{11}^{'} - \sum_{t=2}^{T-1}\Delta\varepsilon_t F_{1t}^{'}}{T} A_1^{'}(\hat{\delta}-\delta) + \sum_{t=2}^{T}\frac{\Delta\varepsilon_t F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta)$$
$$= \{\frac{\varepsilon_{T-1}F_{1T}^{'} - \varepsilon_1 F_{11}^{'} - \sum_{t=2}^{T-1}\Delta\varepsilon_t F_{1t}^{'}}{T} + \frac{\sum_{t=2}^{T}\Delta\varepsilon_t F_{1,t-1}^{'}}{T}\}A_1^{'}(\hat{\delta}-\delta)$$
$$= \{\frac{\varepsilon_{T-1}F_{1T}^{'} - \varepsilon_1 F_{11}^{'} + \Delta\varepsilon_T F_{1,T-1}^{'}}{T} - \frac{\sum_{t=2}^{T-1}\Delta\varepsilon_t \Delta F_{1,t}^{'}}{T}\}A_1^{'}(\hat{\delta}-\delta)$$
$$\xrightarrow{d} (\Omega_{01} - O_p(\frac{1}{T}) - O_p(\frac{1}{\sqrt{T}}) - \Sigma_{01})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$
$$= \Lambda_{01}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0),$$

in which we have used the fact that $\frac{\varepsilon_{T-1}F_{1T}^{'}}{T} \xrightarrow{d} \Omega_{01}$, and $\frac{\sum_{t=2}^{T-1}\Delta\varepsilon_t\Delta F_{1,t}^{'}}{T} \xrightarrow{d} \Sigma_{01}$, and $\frac{\varepsilon_1 F_{11}^{'}}{T} = O_p(\frac{1}{T})$, and $\frac{\Delta\varepsilon_T F_{1,T-1}^{'}}{T} = O_p(\frac{1}{\sqrt{T}})$. For the third term, we have

$$(\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}\varepsilon_{t-1}}{T} = (\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{(F_{2t}-F_{2,t-1})\varepsilon_{t-1}}{T}$$
$$= (\hat{\delta}-\delta)^{'}A_2(\frac{F_{2,T}\varepsilon_{T-1}}{T} - \frac{F_{2,1}\varepsilon_1}{T} - \frac{\sum_{t=2}^{T-1}F_{2,t}\Delta\varepsilon_t}{T})$$
$$\xrightarrow{d} O_p(1)(O_p(\frac{1}{\sqrt{T}}) - O_p(\frac{1}{T}) - 0) = O_p(\frac{1}{\sqrt{T}}),$$

where the last is from the fact that $\frac{F_{2,T}\varepsilon_{T-1}}{T} = O_p(\frac{1}{\sqrt{T}})$ and $\frac{\sum_{t=2}^{T-1}F_{2,t}\Delta\varepsilon_t}{T} \xrightarrow{d} E(u_{2t}\otimes\Delta\varepsilon_t) = 0$ by Assumption 9 (EC').

For the forth term, we have

$$\hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t\varepsilon_{t-1}}{T} = \hat{\delta}^{'}\sum_{t=2}^{T}\frac{(\phi_t-\phi_{t-1})\varepsilon_{t-1}}{T}$$
$$= \hat{\delta}^{'}(\frac{\phi_T\varepsilon_{T-1}}{T} - \frac{\phi_1\varepsilon_1}{T} - \frac{\sum_{t=2}^{T-1}\phi_t\Delta\varepsilon_t}{T})$$
$$\xrightarrow{d} O_p(1)(O_p(\frac{1}{\sqrt{T}}) - O_p(\frac{1}{T}) - O_p(\sqrt{\frac{T}{n}}) = O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{T}{n}}),$$

where the last line is from the fact that $\frac{\phi_T\varepsilon_{T-1}}{T} = O_p(\frac{1}{\sqrt{T}})$ and $\|\frac{\sum_{t=2}^{T-1}\phi_t\Delta\varepsilon_t}{T}\| \leq (\frac{\sum_{t=2}^{T-1}\|\phi_t\|^2}{T})^{1/2}(\frac{\sum_{t=2}^{T-1}\Delta\varepsilon_t^2}{T})^{1/2} = O_p(\sqrt{\frac{T}{n}})$.

For the sixth term, we have

$$(\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) \xrightarrow{d} (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1}(\int dB_1 B_1^{'} + \Lambda_{11})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$

Consider the seventh and tenth terms together:

$$(\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}F_{1,t-1}^{'}}{T}A_1^{'}(\hat{\delta}-\delta) + (\hat{\delta}-\delta)^{'}A_1\sum_{t=2}^{T}\frac{\Delta F_{1t}F_{2,t-1}^{'}}{T}A_2^{'}(\hat{\delta}-\delta)$$

$$= (\hat{\delta}-\delta)^{'}A_2\{\frac{F_{2T}F^{'}_{1,T-1}-F_{21}F^{'}_{11}-\sum_{t=2}^{T-1}F_{2t}\Delta F^{'}_{1t}}{T}+\sum_{t=2}^{T}\frac{F_{2,t-1}\Delta F^{'}_{1t}}{T}\}A^{'}_1(\hat{\delta}-\delta)$$

$$= (\hat{\delta}-\delta)^{'}A_2\{\frac{F_{2T}F^{'}_{1,T-1}-F_{21}F^{'}_{11}+F_{2,T-1}\Delta F^{'}_{Tt}}{T}-\frac{\sum_{t=2}^{T-1}\Delta F_{2t}\Delta F^{'}_{1,t}}{T}\}A^{'}_1(\hat{\delta}-\delta)$$

$$\xrightarrow{d} O_p(1)(O_p(\frac{1}{\sqrt{T}})-O_p(\frac{1}{T})+O_p(\frac{1}{T})-O_p(\frac{1}{\sqrt{T}}))O_p(1)$$

$$= O_p(\frac{1}{\sqrt{T}}).$$

For the eighth term, we have

$$\hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t F_{1,t-1}}{T}A^{'}_1(\hat{\delta}-\delta) = \hat{\delta}^{'}\sum_{t=2}^{T}\frac{(\phi_t-\phi_{t-1})F_{1,t-1}}{T}A^{'}_1(\hat{\delta}-\delta)$$
$$= \hat{\delta}^{'}(\frac{\phi_T F_{1,T-1}}{T}-\frac{\phi_1 F_{11}}{T}-\frac{\sum_{t=2}^{T-1}\phi_t\Delta F_{1t}}{T})A^{'}_1(\hat{\delta}-\delta)$$
$$\xrightarrow{d} O_p(1)(O_p(\frac{1}{\sqrt{T}})-O_p(\frac{1}{T})-O_p(\sqrt{\frac{T}{n}}) = O_p(\frac{1}{\sqrt{T}})+O_p(\sqrt{\frac{T}{n}}),$$

where the last line is from the fact that $\frac{\phi_T F_{1,T-1}}{T} = O_p(\frac{1}{\sqrt{T}})$ and $\|\frac{\sum_{t=2}^{T-1}\phi_t\Delta F_{1t}}{T}\| \leq (\frac{\sum_{t=2}^{T-1}\|\phi_t\|^2}{T})^{1/2}(\frac{\sum_{t=2}^{T-1}\Delta F^2_{1t}}{T})^{1/2} = O_p(\sqrt{\frac{T}{n}})$.

For the ninth term, we have

$$\sum_{t=2}^{T}\frac{\Delta\varepsilon_t F^{'}_{2,t-1}}{T}A^{'}_2(\hat{\delta}-\delta) \xrightarrow{d} O_p(\frac{1}{\sqrt{T}})O_p(1) = O_p(\frac{1}{\sqrt{T}}),$$

and for the eleventh term, we have

$$(\hat{\delta}-\delta)^{'}A_2\sum_{t=2}^{T}\frac{\Delta F_{2t}F^{'}_{2,t-1}}{T}A^{'}_2(\hat{\delta}-\delta) \xrightarrow{d} O_p(1)O_p(\frac{1}{\sqrt{T}})O_p(1) = O_p(\frac{1}{\sqrt{T}}).$$

For the 12th term, we have

$$\hat{\delta}^{'}\sum_{t=2}^{T}\frac{\Delta\phi_t F^{'}_{2,t-1}}{T}A^{'}_2(\hat{\delta}-\delta) \xrightarrow{d} O_p(1)O_p(\frac{1}{min[\sqrt{n},\sqrt{T}]})O_p(1) = O_p(\frac{1}{min[\sqrt{n},\sqrt{T}]}),$$

since $\|\frac{\sum_{t=2}^{T-1}\Delta\phi_t F^{'}_{2,t-1}}{T}\| \leq (\frac{\sum_{t=2}^{T-1}\|\Delta\phi_t\|^2}{T})^{1/2}(\frac{\sum_{t=2}^{T-1}\|F_{2,t-1}\|^2}{T})^{1/2} = O_p(\frac{1}{min[\sqrt{n},\sqrt{T}]})$.

For the 13th term, we have

$$\|\sum_{t=2}^{T}\frac{\Delta\varepsilon_t\phi^{'}_{t-1}}{T}\hat{\delta}\| \leq (\frac{\sum_{t=2}^{T-1}\|\Delta\varepsilon_t\|^2}{T})^{1/2}(\frac{\sum_{t=2}^{T-1}\|\phi_{t-1}\|^2}{T})^{1/2} = O_p(\sqrt{\frac{T}{n}}).$$

Likewise, we can show that the 14th term is also at the speed of $O_p(\sqrt{\frac{T}{n}})$.

Hence,

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T}(\hat{\varepsilon}_t-\hat{\varepsilon}_{t-1})\hat{\varepsilon}_{t-1} \xrightarrow{d} & \int_0^1 B_0 dB_0 + \Lambda_{00} \\
& -(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 dB_1B_0+\Lambda_{10}) \\
& -(\int_0^1 dB_0B_1^{'}+\Lambda_{01})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) \\
& +(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int dB_1B_1^{'}+\Lambda_{11})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) \\
& = \int_0^1 B_0dB_0 - (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 dB_1B_0+\int_0^1 B_1dB_0) \\
& +(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int dB_1B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) \\
& +\Lambda_{00}-2(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{10}+(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{11}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) \\
& = \int_0^1 \tilde{B}_0 d\tilde{B}_0 + \Lambda_{\tilde{0}\tilde{0}},
\end{aligned}
$$

where we have used the fact that

$$
\begin{aligned}
\int_0^1 \tilde{B}_0 d\tilde{B}_0 = & = \int_0^1 B_0dB_0 - (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 dB_1B_0+\int_0^1 B_1dB_0) \\
& +(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}(\int dB_1B_1^{'})(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0),
\end{aligned}
$$

and

$$
\Lambda_{\tilde{0}\tilde{0}} = \Lambda_{00}-2(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{10}+(\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{11}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0).
$$

In all, we have proved that

$$
\begin{aligned}
T(\hat{\rho}_T-1) &= \frac{T^{-1}\sum_{t=2}^{T}(\hat{\varepsilon}_t-\hat{\varepsilon}_{t-1})\hat{\varepsilon}_{t-1}}{T^{-2}\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2} \\
&\xrightarrow{d} \frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0+\Lambda_{\tilde{0}\tilde{0}}}{\int_0^1 \tilde{B}_0\tilde{B}_0},
\end{aligned}
$$

in which $\tilde{B}_0 = B_0 - (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}B_1$. The Brownian motion $\tilde{B}_0$ has long-run covariance matrix $\Omega_{\tilde{0}\tilde{0}} = \Omega_{00} - (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Omega_{10} - \Omega_{01}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) + (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Omega_{11}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0)$ and one-sided long-run covariance $\Lambda_{\tilde{0}\tilde{0}} = \Lambda_{00} - (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{10} - \Lambda_{01}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0) + (\int_0^1 B_0B_1^{'})(\int B_1B_1^{'})^{-1}\Lambda_{11}(\int B_1B_1^{'})^{-1}(\int_0^1 B_1B_0)$.

(b) Firstly, notice that

$$(T-1)^2\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2 = \frac{1}{(T-1)^{-2}\sum_{t=2}^T \hat{\varepsilon}_{t-1}^2}$$
$$\xrightarrow{d} \frac{1}{\int_0^1 \tilde{B}_0\tilde{B}_0}.$$

Secondly,

$$\hat{c}_{j,T} = (T-1)^{-1}\sum_{t=j+2}^T \hat{\varkappa}_t\hat{\varkappa}_{t-j}$$
$$= (T-1)^{-1}\sum_{t=j+2}^T (\hat{\varepsilon}_t - \hat{\rho}_T\hat{\varepsilon}_{t-1})(\hat{\varepsilon}_{t-j} - \hat{\rho}_T\hat{\varepsilon}_{t-j-1})$$
$$= (T-1)^{-1}\sum_{t=j+2}^T (\Delta\hat{\varepsilon}_t - (\hat{\rho}_T-1)\hat{\varepsilon}_{t-1})(\Delta\hat{\varepsilon}_{t-j} - (\hat{\rho}_T-1)\hat{\varepsilon}_{t-j-1})$$
$$= \frac{1}{T-1}\sum_{t=j+2}^T \Delta\hat{\varepsilon}_t\Delta\hat{\varepsilon}_{t-j} - \frac{1}{T-1}\sum_{t=j+2}^T(\hat{\rho}_T-1)\hat{\varepsilon}_{t-1}\Delta\hat{\varepsilon}_{t-j} - \frac{1}{T-1}\sum_{t=j+2}^T(\hat{\rho}_T-1)\Delta\hat{\varepsilon}_t\hat{\varepsilon}_{t-j-1}$$
$$+ \frac{1}{T-1}\sum_{t=j+2}^T(\hat{\rho}_T-1)^2\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-j-1}.$$

Result (a) implies that $T(\hat{\rho}_T - 1) = O_p(1)$, and it is easy to show that $\frac{1}{T-1}\sum_{t=j+2}^T(\hat{\rho}_T - 1)\hat{\varepsilon}_{t-1}\Delta\hat{\varepsilon}_{t-j} = O_p(\frac{1}{T})$, and $\frac{1}{T-1}\sum_{t=j+2}^T(\hat{\rho}_T - 1)\Delta\hat{\varepsilon}_t\hat{\varepsilon}_{t-j-1} = O_p(\frac{1}{T})$, and $\frac{1}{T-1}\sum_{t=j+2}^T(\hat{\rho}_T - 1)^2\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-j-1} = O_p(\frac{1}{T^2})$. Hence, we have

$$\hat{c}_{j,T} \xrightarrow{p} \frac{1}{T-1}\sum_{t=j+2}^T \Delta\hat{\varepsilon}_t\Delta\hat{\varepsilon}_{t-j}$$
$$= \frac{1}{T-1}\sum_{t=j+2}^T (\Delta\varepsilon_t - (\hat{\delta}-\delta)^{'}A_1\cdot\Delta F_{1t} - (\hat{\delta}-\delta)^{'}A_2\cdot\Delta F_{2t} - \hat{\delta}^{'}\Delta\phi_t)$$
$$\cdot(\Delta\varepsilon_{t-j} - (\hat{\delta}-\delta)^{'}A_1\cdot\Delta F_{1,t-j} - (\hat{\delta}-\delta)^{'}A_2\cdot\Delta F_{2,t-j} - \hat{\delta}^{'}\Delta\phi_{t-j})$$
$$= \frac{1}{T-1}\sum_{t=j+2}^T \Delta\varepsilon_t\Delta\varepsilon_{t-j} - (\hat{\delta}-\delta)^{'}A_1\frac{1}{T-1}\sum_{t=j+2}^T\Delta F_{1t}\Delta\varepsilon_{t-j}$$
$$- (\hat{\delta}-\delta)^{'}A_2\frac{1}{T-1}\sum_{t=j+2}^T\Delta F_{2t}\Delta\varepsilon_{t-j} - \hat{\delta}^{'}\frac{1}{T-1}\sum_{t=j+2}^T\Delta\phi_t\Delta\varepsilon_{t-j}$$
$$- \frac{1}{T-1}\sum_{t=j+2}^T\Delta\varepsilon_t\Delta F_{1,t-j}^{'}A_1^{'}(\hat{\delta}-\delta) + (\hat{\delta}-\delta)^{'}A_1\frac{1}{T-1}\sum_{t=j+2}^T\Delta F_{1t}\Delta F_{1,t-j}^{'}A_1^{'}(\hat{\delta}-\delta)$$
$$+ (\hat{\delta}-\delta)^{'}A_2\frac{1}{T-1}\sum_{t=j+2}^T\Delta F_{2t}\Delta F_{1,t-j}^{'}A_1^{'}(\hat{\delta}-\delta) + \hat{\delta}^{'}\frac{1}{T-1}\sum_{t=j+2}^T\Delta\phi_t\Delta F_{1,t-j}^{'}A_1^{'}(\hat{\delta}-\delta)$$
$$- \frac{1}{T-1}\sum_{t=j+2}^T\Delta\varepsilon_t\Delta F_{2,t-j}^{'}A_2^{'}(\hat{\delta}-\delta) + (\hat{\delta}-\delta)^{'}A_1\frac{1}{T-1}\sum_{t=j+2}^T\Delta F_{1t}\Delta F_{2,t-j}^{'}A_2^{'}(\hat{\delta}-\delta)$$

$$+ (\hat{\delta} - \delta)^{'} A_2 \frac{1}{T-1} \sum_{t=j+2}^{T} \Delta F_{2t} \Delta F_{2,t-j}^{'} A_2^{'} (\hat{\delta} - \delta) + \hat{\delta}^{'} \frac{1}{T-1} \sum_{t=j+2}^{T} \Delta \phi_t \Delta F_{2,t-j}^{'} A_2^{'} (\hat{\delta} - \delta)$$
$$- \frac{1}{T-1} \sum_{t=j+2}^{T} \Delta \varepsilon_t \Delta \phi_{t-j}^{'} \hat{\delta} + (\hat{\delta} - \delta)^{'} A_1 \frac{1}{T-1} \sum_{t=j+2}^{T} \Delta F_{1t} \Delta \phi_{t-j}^{'} \hat{\delta}$$
$$+ (\hat{\delta} - \delta)^{'} A_2 \frac{1}{T-1} \sum_{t=j+2}^{T} \Delta F_{2t} \Delta \phi_{t-j}^{'} \hat{\delta} + \hat{\delta}^{'} \frac{1}{T-1} \sum_{t=j+2}^{T} \Delta \phi_t \Delta \phi_{t-j}^{'} \hat{\delta}$$
$$\xrightarrow{p} E(\Delta \varepsilon_t \Delta \varepsilon_{t-j}) - plim\{(\hat{\delta} - \delta)^{'} A_1\} E(\Delta F_{1t} \Delta \varepsilon_{t-j}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{min[\sqrt{n}, \sqrt{T}]})$$
$$- E(\Delta \varepsilon_t \Delta F_{1,t-j}) plim\{A_1^{'}(\hat{\delta} - \delta)\} + plim\{(\hat{\delta} - \delta)^{'} A_1\} E(\Delta F_{1t} \Delta F_{1,t-j}^{'}) plim\{A_1^{'}(\hat{\delta} - \delta)\} + O_p(\frac{1}{\sqrt{T}})$$
$$+ O_p(\frac{1}{min[\sqrt{n}, \sqrt{T}]}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{min[\sqrt{n}, \sqrt{T}]})$$
$$+ O_p(\frac{1}{min[\sqrt{n}, \sqrt{T}]}) + O_p(\frac{1}{min[\sqrt{n}, \sqrt{T}]}) + O_p(\frac{1}{min[\sqrt{n}, \sqrt{T}]}) + O_p(\frac{1}{min[n, T]})$$
$$= E(\Delta \varepsilon_t \Delta \varepsilon_{t-j}) - plim\{(\hat{\delta} - \delta)^{'} A_1\} E(\Delta F_{1t} \Delta \varepsilon_{t-j})$$
$$- E(\Delta \varepsilon_t \Delta F_{1,t-j}) plim\{A_1^{'}(\hat{\delta} - \delta)\} + plim\{(\hat{\delta} - \delta)^{'} A_1\} E(\Delta F_{1t} \Delta F_{1,t-j}^{'}) plim\{A_1^{'}(\hat{\delta} - \delta)\}.$$

Hence, we have proved that

$$\hat{c}_{j,T} \xrightarrow{p} E(\Delta \varepsilon_t \Delta \varepsilon_{t-j}) - plim\{(\hat{\delta} - \delta)^{'} A_1\} E(\Delta F_{1t} \Delta \varepsilon_{t-j})$$
$$- E(\Delta \varepsilon_t \Delta F_{1,t-j}) plim\{A_1^{'}(\hat{\delta} - \delta)\} + plim\{(\hat{\delta} - \delta)^{'} A_1\} E(\Delta F_{1t} \Delta F_{1,t-j}^{'}) plim\{A_1^{'}(\hat{\delta} - \delta)\}$$
$$= E(\Delta \varepsilon_t \Delta \varepsilon_{t-j}) - (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1} E(\Delta F_{1t} \Delta \varepsilon_{t-j}) - E(\Delta \varepsilon_t \Delta F_{1,t-j})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$
$$+ (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1} E(\Delta F_{1t} \Delta F_{1,t-j}^{'})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0).$$

Thus,

$$1/2\{\hat{\lambda}_T^2 - \hat{c}_{0,T}\} = \sum_{j=1}^{q} [1 - j/(q+1)] \hat{c}_{j,T}$$
$$\xrightarrow{d} \sum_{j=1}^{\infty} E(\Delta \varepsilon_t \Delta \varepsilon_{t-j}) - (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1} \sum_{j=1}^{\infty} E(\Delta F_{1t} \Delta \varepsilon_{t-j})$$
$$- \sum_{j=1}^{\infty} E(\Delta \varepsilon_t \Delta F_{1,t-j})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$
$$+ (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1} \sum_{j=1}^{\infty} E(\Delta F_{1t} \Delta F_{1,t-j}^{'})(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$
$$= \Lambda_{00} - (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1} \Lambda_{10} - \Lambda_{01}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$
$$- (\int_0^1 B_0 B_1^{'})(\int B_1 B_1^{'})^{-1} \Lambda_{11}(\int B_1 B_1^{'})^{-1}(\int_0^1 B_1 B_0)$$

$$= \Lambda_{\tilde{0}\tilde{0}},$$

as $q \to \infty$ and $q/T \to 0$.

To conclude, if $q \to \infty$ as $T \to \infty$ and $q/T \to 0$, then the Phillips's Z_ρ statistic (1987) satisfies:

$$\begin{aligned} Z_{\rho,T} &= T(\hat{\rho}_T - 1) - 1/2\{(T-1)^2\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2\}\{\hat{\lambda}^2_T - \hat{c}_{0,T}\} \\ &\xrightarrow{d} \frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0 + \Lambda_{\tilde{0}\tilde{0}}}{\int_0^1 \tilde{B}_0\tilde{B}_0} - \frac{\Lambda_{\tilde{0}\tilde{0}}}{\int_0^1 \tilde{B}_0\tilde{B}_0} \\ &= \frac{\int_0^1 \tilde{B}_0 d\tilde{B}_0}{\int_0^1 \tilde{B}_0\tilde{B}_0}. \end{aligned}$$

The last limit could also be written as $\frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)W(r)dr}$ for a scalar standard Brownian motion $W(r)$. ■

References

[1] Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

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

[3] —(2004), "Estimating Cross-section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics,* 122, 137-183.

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

[5] —(2004), "A PANIC Attack on Unit Roots and Cointegration," *Econometrica,* 72, 1127–1177.

[6] —(2006), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-augmented Regressions," *Econometrica,* 74, 1133-1150.

[7] Banerjee, A., and Marcellino, M. (2009), "Factor-Augmented Error Correction Models," in *The Methodology and Practice of Econometrics - A Festschrift for David Hendry*, eds. J.L. Castle and N. Shephard, Oxford: Oxford University Press, 227-254.

[8] Banerjee, A., Marcellino, M., and Masten, I. (2014a), "Forecasting with Factor-Augmented Error Correction Models," *International Journal of Forecasting*, 30, 589-612.

[9] —(2014b), "Structural FECM: Cointegration in Large-Scale Structural FAVAR Models," CEPR Discussion paper No. 9858.

[10] Bernanke, B. S., Boivin, J., and Eliasz, P. (2005), "Measuring the Effects of Monetary Policy: a Factor-augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387-422.

[11] Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

[12] Johansen, S., and Juselius, K. (1990), "Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

[13] Miller, J. I. (2010), "Cointegrating Regressions with Messy Regressors and An Application to Mixed-Frequency Series," *Journal of Time Series Analysis*, 31, 255-277.

[14] Pagan, A. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221-247.

[15] Park, J. Y. (1992), "Canonical Cointegrating Regressions," *Econometrica,* 60, 119-143.

[16] Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics,* 28, 329-348.

[17] Phillips, P. C. B., and Durlauf, S. N. (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies,* 53, 473-495.

[18] Phillips, P. C. B., and Hansen, B. E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies,* 57, 99-125.

[19] Phillips, P. C. B. (1995), "Fully Modified Least Squares and Vector Autoregression," *Econometrica,* 59, 1023-1078.

[20] Priestley, M. B. (1981), *Spectral Analysis and Time Series* (Vol. 1, Vol. 2), New York: Academic Press.

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

[22] —(2002a), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics,* 20, 147-162.

[23] —(2002b), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association,* 97, 1167-1179.

[24] Vogelsang, T. J., and Wagner, M. (2014), "Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions," *Journal of Econometrics,* 178, 741-760.

VITA

NAME OF AUTHOR: Shulin Shen

PLACE OF BIRTH: Ruichang, China

DATE OF BIRTH: October, 1988

EDUCATION:

2011 B.S. Xi’an Jiaotong University, Shaanxi, China

2013 M.S. Xi’an Jiaotong University, Shaanxi, China

RESEARCH EXPERIENCE AND EMPLOYMENT:

2013 – 2018 Teaching Assistant, Department of Economics, Syracuse University

2016 – 2018 Research Associate, Center for Policy Research, Syracuse University

AWARDS AND HONORS:

2013 – 2018 Syracuse University Graduate Assistantship, Syracuse University

2012 National Scholarship, Xi’an Jiaotong University

2008 National Scholarship, Xi’an Jiaotong University