A Factor Analytical Approach to Price Discovery

Joakim Westerlund
Simon Reese
Paresh Narayan

February 2015
A FACTOR ANALYTICAL APPROACH TO PRICE DISCOVERY∗

Joakim Westerlund†
Lund University
and
Financial Econometrics Group
Centre for Research in Economics and Financial Econometrics
Deakin University

Paresh Narayan
Financial Econometrics Group
Centre for Research in Economics and Financial Econometrics
Deakin University

Simon Reese
Lund University

December 30, 2014

Abstract

Existing econometric approaches for studying price discovery presume that the number of markets are small, and their properties become suspect when this restriction is not met. They also require making identifying restrictions and are in many cases not suitable for statistical inference. The current paper takes these shortcomings as a starting point to develop a factor analytical approach that makes use of the cross-sectional variation of the data, yet is very user-friendly in that it does not involve any identifying restrictions or obstacles to inference.

JEL Classification: C12; C13; C33.

Keywords: Price discovery; panel data; common factor models; cross-unit cointegration.

∗Earlier versions of this paper were presented at the 1st Conference on Recent Developments in Financial Econometrics and Applications in Geelong and at the 8th International Conference on Computational and Financial Econometrics in Pisa. The authors would like to thank seminar participants, and in particular Morten Ørregaard Nielsen for many useful comments and suggestions. Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1.

†Corresponding author: Department of Economics, Lund University, P. O. Box 7082, S–220 07 Lund, Sweden. E-mail address: joakim.westerlund@nek.lu.se.
1 Introduction

Financial markets incorporate new information into asset prices by matching buyers and sellers. They thereby facilitate the discovery of what the price of an asset should be. This price discovery role of financial markets can take place across separate exchanges and instruments, as many securities and derivatives based on the same underlying asset may trade on multiple venues. In the case of such a multiplicity, there may be variation in the share with which each market’s trades contribute to discovering the one true price of the underlying asset. The present paper is about the modeling and measuring of these so-called “information shares” (ISs), which are important for both investors concerned with price informativeness and adverse selection risk, as well as policy makers investigating the determinants of price efficiency.

The measurement of price discovery requires isolating informative price movements from noise. Observed price changes constitute the most obvious indicator of price discovery. However, they form an imperfect measure, as observed prices are susceptible to transitory mispricing, caused by noise trading or temporary order imbalances, for example. In contrast, when security prices absorb new information due to informed trading, these price changes last permanently. To formalize these ideas, let us denote by \( P_{i,t} \) the price of a particular asset on market \( i = 1, \ldots, N \) at time period \( t = 1, \ldots, T \), and let \( P^*_t \) be the fundamental value of the same asset at the same time period. The structural model we have in mind is taken from the market microstructure literature (see Madhavan, 2000, for a survey), as is standard in studies of price discovery (see De Jong, 2002; De Jong and Schotman, 2010; Harris et al., 2002; Hasbrouck, 1993; Lehman, 2002, to mention a few). It is given by

\[
P_{i,t} = P^*_t + E_{i,t},
\]

where \( E_{i,t} \) is an idiosyncratic (or market-specific) term capturing microstructure effects. While \( P^*_t \) is assumed to follow a random walk, \( E_{i,t} \) is stationary. Hence, while shocks to the fundamental value have a permanent effect on prices, the effect of idiosyncratic shocks is transitory. The transitory nature of \( E_{i,t} \) implies that \( P_{i,t} \) will adjust to the fundamental value over time. In fact, as Hasbrouck (1995) points out, under said assumptions, \( P_{i,t} \) and \( P_{j,t} \) are cointegrated with cointegrating vector \((1, -1)'\).

A market is relatively efficient in the price discovery process if it incorporates a larger amount of fundamental shocks than other markets. Hasbrouck (1995) defines the IS of a particular market as the variance share of that market that is attributable to the fundamental value.
Unfortunately, as (1) makes clear, the fundamental value is not directly observed, but is instead contaminated by additive market-specific pricing errors. Therefore, legitimate inference on the price discovery process cannot be made until the confounding effects of these errors have been appropriately accounted for. The most common approach by far is the one of Hasbrouck (1995), in which the effects of the shocks are retrieved from an estimated reduced form vector error correction model (VECM). This VECM route to the IS has in recent years become very popular, and is by now the workhorse of the industry with a huge number of applications (see, for example, Brenner and Kroner, 1995, for a survey).

But while popular, the VECM approach also has its fair share of drawbacks. First, since the VECM is just a reduced form model, the fundamental shocks cannot be retrieved without suitable identifying restrictions. Hasbrouck (1995) uses Cholesky factorization, which makes the IS dependent on the ordering of the series. With $N$ markets, there are no less than $N!$ such orderings, suggesting that without prior knowledge about the appropriate order, the IS is likely to be an uninformative measure. As a remedy, Hasbrouck (1995) suggests reporting upper and lower bounds, as obtained by considering all possible orderings. The resulting largest and smallest ISs for each market constitute the upper and lower bounds, respectively. These bounds can, however, be quite far apart (see, for example, Huang, 2002; Hasbrouck, 2002). This is true when $N$ is relatively small, and the bounds become wider when $N$ increases. Specifically, the width depends critically on the covariances of the shocks, whose number increases at the rate $N^2$. The number of degrees of freedom will therefore decrease very rapidly with $N$, leading to increased estimation uncertainty, and hence wider bounds. Booth et al. (2002) use trading intervals averaging about 30 minutes for only two markets, namely, the upstairs and downstairs markets on the Helsinki stock exchange. According to their results, the average IS interval for the downstairs market is $[13\%, 99.2\%]$, which is clearly too wide for any interesting conclusions (see, for example, Baillie et al., 2002; Martens, 1998; Tse, 1999, for similar findings). Second, the asymptotic standard errors of the ISs are difficult to obtain, complicating inference. Sapp (2002) proposes bootstrapping the standard errors. However, while certainly feasible, the bootstrap is computationally relatively costly, making it unattractive from an applied point of view. Third, well-specified VECMs tend to be heavily parameterized, making them infeasible in large-$N$ samples, which of course restricts their applicability. Fourth, while $P_1^*$ is likely to follow a random walk, the assumption that $E_{i,t}$ is stationary seems premature. That is, the restriction that $P_{i,t}$ and $P_{j,t}$ should be cointegrated with cointegrating vector $(1, -1)'$ need not
be true.

The above drawbacks recently motivated De Jong and Schotman (2010) to develop a generalized method of moments (GMM)-based approach that does not rely on reduced forms, but instead seeks to infer the structural model in (1) directly. In addition to the advantage of not having to identify the shocks, the main advantage of estimating (1) is that it is parsimonious, as opposed to reduced form VECMs. The GMM approach is therefore suitable even when $N$ is relatively large. The implementation is quite complicated, though, which is probably also the main reason for why the approach has not received much interest in the (applied) literature.

In the present paper we take the complicated implementation of the GMM approach as our starting point. The purpose is to develop a new approach to price discovery that is simple yet does not suffer from the drawbacks of the VECM approach. As a source of inspiration we consider the growing literature on large-dimension common factor models (see Bai and Ng, 2008, for a survey), within which (1) can be seen as a restricted common factor model with a single non-stationary common factor, or cross-section common stochastic trend, $P_t^*$, and unit factor loading. This suggests that (1) can be estimated using existing methods for such common factor models. The simplest method by far is the cross-section average (CA) approach first considered by Pesaran (2006) in the context of a factor-augmented panel regression model in stationary variables. As the name suggests, the basic idea is to use the cross-section average of the observed data as an estimator of the common factor. In this paper we extend the CA approach to the price discovery context. Specifically, a statistical toolbox is provided that enables not only estimation but also testing, and this under fairly general conditions. A GAUSS code containing this toolbox is available online at https://sites.google.com/site/perjoakimwesterlund/, making for easy implementation. Unlike VECMs, the new approach exploits the information contained in the cross-sectional dimension of the panel, making it particularly well-suited for large-$N$ panels, although $N$ can also be “small”.

The usefulness of the new toolbox is illustrated using two data sets. The first data set covers crude oil prices listed on three exchanges, namely, the US, the UK and Oman exchanges. The findings suggest that the price discovery process is dominated by the Oman market, followed by the UK market. The US market contributes less than 4% to price discovery. This result is rather interesting in that it represents the first piece of evidence on the oil price discovery

---

1See Lien and Shrestha (2009), and Yan and Zivot (2010) for other attempts to alleviate the identification problem of the VECM approach.
process across countries. Our second application is motivated by the equity prices of firms that are cross-listed. We pick a firm that has listings on a large number of exchanges. We select Arcelor Mittal, a major steel company, which has its primary listing on the Luxembourg stock exchange and has secondary listings on another eight exchanges. Our results reveal that while the primary exchange dominates, the contribution to price discovery is only around 30%, suggesting that the bulk of price discovery takes place on other exchanges.

The balance of the paper is organized as follows. In Section 2, we formalize the common factor model setup and explain how it relates to the IS. Section 3 presents the econometric results. This section is divided into two parts. The first part presents the asymptotic theory, while the second part is concerned with the small-sample accuracy of this theory. Section 4 contains the empirical results. Section 5 concludes. Proofs of important results are given in Appendix.

2 The common factor model

2.1 Assumptions

Consider the panel data variable \( X_{i,t} \), observable for \( i = 1, \ldots, N \) cross-section units and \( t = 1, \ldots, T \) time periods. The data generating process (DGP) of this variable is assumed to be given by the following dynamic common factor model:

\[
X_{i,t} = \lambda_i F_t + U_{i,t},
\]

where \( F_t \) is a common factor with \( \lambda_i \) being the associated factor loading, and \( U_{i,t} \) is an idiosyncratic error term. In terms of the model in (1), we have \( X_{i,t} = P_{i,t}, F_t = P_t^* \), \( \lambda_1 = \ldots = \lambda_N = 1 \) and \( U_{i,t} = E_{i,t} \). The dynamics of \( F_t \) and \( U_{i,t} \) are governed by the following equations:

\[
F_t = F_{t-1} + \eta_t, \quad (3)
\]

\[
U_{i,t} = \rho_i U_{i,t-1} + \epsilon_{i,t}, \quad (4)
\]

where \( \rho_i \in (-1, 1] \), and \( \eta_t \) and \( \epsilon_{i,t} \) are error terms that are supposed to satisfy Assumption 1.

Assumption 1.

(i) \( \eta_t \) is independently and identically distributed (iid) across \( t \) with \( E(\eta_t) = 0, \ E(\eta_t^2) = \sigma^2_\eta > 0 \) and \( E(|\eta_t|^4) < \infty \);
(ii) $\epsilon_{i,t}$ is iid across $i$ and $t$ with $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = \sigma_{\epsilon_i}^2 > 0$ and $E(|\epsilon_{i,t}|^8) < \infty$;

(iii) $E(|F_0|) < \infty$ and $E(|U_{1,0}|, \ldots, E(|U_{N,0}|) < \infty$;

(iv) $\lambda_i$ is either deterministic such that $|\lambda_i| < \infty$ or stochastic such that $E(|\lambda_i|^4) < \infty$ and $\lambda = N^{-1} \sum_{i=1}^{N} \lambda_i \neq 0$ for all $N$, including $N \to \infty$;

(v) $\epsilon_{i,t}$, $\eta_t$ and $\lambda_i$ are mutually independent.

As alluded to in the above, the model in (2)–(4) is almost identical to the one in (1), in which $X_{i,t}$ and $F_t$ represent the observed price and its fundamental value, respectively. The only differences are that; (i) $\lambda_i$ is not restricted to be equal to one, and (ii) $U_{i,t}$ need not be stationary. As mentioned in Section 1, the otherwise so common stationarity requirement on $U_{i,t}$ implies that prices are cointegrated across markets. Indeed, since under $|\rho_1|, \ldots, |\rho_N| < 1$, $X_{i,t} - \lambda_i \lambda_i^{-1} U_{i,t}$ is stationary, $X_{i,t}$ and $X_{j,t}$ are cointegrated with cointegrating vector $(1, -\lambda_i \lambda_i^{-1})'$. Hence, in the terminology of the time series literature, (2)–(4) constitute a “common trend” representation of $X_{1,t}, \ldots, X_{N,t}$ (see Stock and Watson, 1988). This also highlights the meaning of the unit loading assumption, as a restriction on the cointegrating vector. The fact that $\lambda \neq 0$ means that not all loadings can be zero. Zero loadings are not ruled out, however, provided that there is a non-negligible fraction of loadings for which the average is nonzero. This seems relevant in practice, as some markets may not contribute to price discovery. The assumption that $\epsilon_{i,t}$ and $\eta_t$ are independent is similar to the restriction employed by Watson (1986) (see Hasbrouck, 1993, for a discussion). In Section 2.2, we discuss the meaning of this assumption in terms of the structural model in (1).

The quantity of interest is the IS, which in the seminal work of Hasbrouck (1995) is defined as the fraction of the variance of the innovation to the fundamental price component that can be attributed to a particular market. As De Jong and Schotman (2010) show, in the current context, the IS for cross-section unit $i$ is given by

$$IS_i = \frac{\lambda_i^2 \sigma_{\epsilon_i}^2}{\sigma^2 + \sum_{n=1}^{N} \lambda_n^2 \sigma_{U,n}^2} = \frac{\lambda_i^2 \sigma_{\epsilon_i}^2 \sigma_{\eta}^{-2}}{1 + \sum_{n=1}^{N} \lambda_n^2 \sigma_{\eta}^2 \sigma_{U,n}^{-2}},$$

where $\sigma_{U,i}^2 = E(U_{i,t}^2)$ and $\sigma_{\eta}^2 = 1$ by Assumption 1. The derivation of the formula in (5) is quite involved, and is provided in Appendix A of this paper. However, we can see that it makes sense. For example, the less the noise in market $i$, as measured by $\sigma_{U,i}^2$, the higher the IS. Similarly, the stronger the covariance between the price of market $i$ and the efficient price, as
captured by $\lambda_i$, the higher the IS. For a given $N$, the measure does not sum up to one. However, asymptotically it does;

$$\sum_{i=1}^{N} IS_i = \frac{\sum_{i=1}^{N} \lambda_i^2 \sigma_{U.i}^{-2}}{\sum_{i=1}^{N} \lambda_i^2 \sigma_{U.i}^{-2} [1 + (\sum_{n=1}^{N} \lambda_n^2 \sigma_{U,n}^{-2})^{-1}]} = \frac{\sum_{i=1}^{N} \lambda_i^2 \sigma_{U.i}^{-2}}{\sum_{n=1}^{N} \lambda_n^2 \sigma_{U,n}^{-2} [1 + o(1)]} \to 1$$

as $N \to \infty$. This suggests that in the large-$N$ scenario considered here the following panel IS (PIS) may be used:

$$PIS_i = \frac{\lambda_i^2 \sigma_{U.i}^{-2}}{\sum_{n=1}^{N} \lambda_n^2 \sigma_{U,n}^{-2}}.$$  

(6)

As with $IS_i$, $PIS_i$ measures the relative variance of the efficient price innovations in market $i$ as a fraction of the total variation. The purpose of this paper is to infer $PIS_i$.

### 2.2 Comparison with existing models

As mentioned in Section 1, most (if not all) of the existing price discovery literature is based on the structural model in (1). This is therefore the model of interest. However, since the efficient price is unobservable, suitable identification and estimation strategies must be put in place before any meaningful measure of price discovery can be constructed. The question is: how to best go about this business? The most common approach is to employ the VECM approach of Hasbrouck (1995). Hence, from a popularity point of view, this is the relevant benchmark. However, we start by considering the approach of Gonzalo and Granger (1995), which is almost as popular and is in many regards very similar to the one of Hasbrouck (1995) (see Lehmann, 2002, for a detailed discussion).

The Gonzalo and Granger (1995) approach is based on the so-called “permanent–transitory” (PT) decomposition of the data (see Quah, 1992). It says that if $X_{i,t}$ is unit root non-stationary, then $X_t = (X_{1,t}, ..., X_{N,t})'$ may be decomposed as

$$X_t = PE_t + TR_t,$$

(7)

where $PE_t$ ($TR_t$) is unit root non-stationary (stationary). This decomposition is very similar to the common factor model in (2)–(4). In fact, setting $PE_t = \lambda F_t$ and $TR_t = U_t = (U_{1,t}, ..., U_{N,t})'$, where $\lambda = (\lambda_1, ..., \lambda_N)'$, we see that the two models coincide. Even the purpose of Gonzalo and Granger (1995), to estimate $F_t$, is very similar to the purpose of the present study. The authors recognize the potential of the common factor model approach. However, because of the stationarity restriction that was at the time required for its estimation, the approach was
not pursued. As we explain in detail in Section 3, the factor analytical approach considered in this paper is very general in this regard, and can be used even if there is uncertainty over the order of integration of $U_{i,t}$.

A crucial difference between the PT approach of Gonzalo and Granger (1995), and the one considered here is how the components of the data are identified. According to the structural model in (1), $P_t^*$ is unit root non-stationary and $E_{i,t}$ is stationary. In the PT approach, it is this difference in the order of integration that enables separation between $PE_t = P_t^*$ and $TR_t = E_t$. By contrast, in the approach considered here it is the common versus idiosyncratic variation that is key; while $F_t$ is assumed to be common, $U_{i,t}$ is purely idiosyncratic. It is important to note that the separation into common and idiosyncratic components can be achieved regardless of the order of integration of these components. The factor analytical approach considered here can therefore be thought of as a two-step procedure, in which the extraction of the common and idiosyncratic components of the data is just a first step. The second step is then to test the extracted components for unit roots. The fact that nothing is assumed regarding the order of integration of the components is a clear advantage when compared to the PT identification scheme. The common factor representation of the PT decomposition further requires that $F_t$ is a linear combination of $X_{1,t}, \ldots, X_{N,t}$ (Gonzalo and Granger, 1995, Proposition 2), which is clearly very restrictive when compared to our factor analytical approach, in which $F_t$ is treated as an unknown parameter to be estimated along with the other parameters of the model.

An advantage of the PT decomposition is that it does not require that $\Delta PE_t$ and $TR_t$ are independent. In particular, $TR_t$ may be correlated with the lags of $\Delta PE_t$. However, because of the way that $PE_t$ ($TR_t$) is identified as unit root non-stationary (stationary), $\Delta PE_t$ cannot be correlated with the lags of $TR_t$, since then $TR_t$ has a permanent effect on $X_t$. For example, if $\Delta P_t^* = \gamma' E_{t-1} + e_t$, where $\gamma$ is $N \times 1$ and $e_t$ is uncorrelated with $E_{t-1}$, then $P_t^* = P_0^* + \sum_{n=1}^{t} (\gamma' E_{n-1} + e_n)$. Thus, since shocks to $E_n$ have a permanent effect on $P_t^*$, and hence also on $X_t$, the PT decomposition will be unable to separate $P_t^*$ from $E_t$. Certain types of correlated innovations can be accommodated also within the present framework. For example, if $E_{i,t} = \lambda_i \Delta P_{t-1}^* + e_{i,t}$, where $e_{i,t}$ is idiosyncratic, then (2) still holds with $F_t = P_t^* + \Delta P_{t-1}^*$ and $U_{i,t} = e_{i,t}$. However, because of the common–idiosyncratic identification scheme, allowing for correlated innovations in the common factor framework is in general more difficult than in the PT framework. Both frameworks therefore have their pros and cons.

The VECM approach of Hasbrouck (1995) is based on the Beveridge–Nelson (BN) decom-
position (see Phillips and Solo, 1992, Lemma 2.1), which is in turn similar to the PT decomposition. It starts from the VECM representation of $\Delta X_t$, which can be written alternatively as the vector moving average $\Delta X_t = B(L)\nu_t$, where $B(L) = \sum_{n=0}^{\infty} B_n L^n$, $B_0 = I_N$ and $\nu_t$ is iid with mean zero and covariance matrix $\Sigma_\nu$. By the BN decomposition, $B(L) = B(1) + (1 - L)B^*(L)$, where $B^*(L)$ has all its roots outside the unit circle. By using this result, backwards substitution and then $X_0 = 0_{N \times 1}$, we obtain the following representation for $X_t$, which is identically the common trends representation of Stock and Watson (1988):

$$X_t = B(1) \sum_{n=1}^{\infty} \nu_n + B^*(L)\nu_t. \tag{8}$$

Now, since $X_{1,t}, ..., X_{N,t}$ are cointegrated, the rank of $B(1)$ is one, that is, there exists an $N \times (N - 1)$ matrix $\beta$ such that $\beta' B(1) = 0_{(N-1) \times N}$. Hence, $\beta' X_t = \beta' X_0 + \beta' B^*(L)\nu_t$ is stationary. It is important to note that the above representation has the same form as the PT decomposition with $P_t = B(1) \sum_{n=1}^{\infty} \nu_n$ and $T_t = B^*(L)\nu_t$. Many of the above mentioned drawbacks of the PT approach therefore applies also in case of the VECM approach. The main difference is that unlike in the former approach, in the latter the innovations driving $P_t$ are the same as those driving $T_t$, that is, in the VECM approach $\Delta P_t$ and $T_t$ are perfectly correlated. The drawback of this requirement, which was mentioned also in Section 1, is that the shocks are no longer separable, at least not without imposing further identifying assumptions (see, for example, De Jong and Schotman, 2010; Lehmann, 2002, for discussions).

In view of the above mentioned drawbacks of the PT and VECM approaches, and the results of studies such as Hasbrouck (2002), suggesting that neither approach seem to be very accurate, it seems natural to seek out other alternatives. The factor analytical approach considered here can be seen as one step in this direction. It should be pointed out, however, that the main advantage is not how the data are decomposed into common and idiosyncratic components, but rather how the new approach enables easy estimation and inference even when $N$ is relatively large. In Section 3 we elaborate in this.
3 Econometric results

3.1 Asymptotic theory

The purpose of this section is to infer $PIS_i$. The idea is the following. Consider the following first-differenced version of (2):

$$x_{i,t} = \lambda_i f_t + u_{i,t}, \quad (9)$$

where $x_{i,t} = \Delta X_{i,t}, f_t = \Delta F_t$ and $u_{i,t} = \Delta U_{i,t}$, which are defined for $t = 2, ..., T$. It is important to note that $f_t = \eta_t$, that is, the common factor in (9) is the innovation that drives $F_t$. Estimation of this model therefore leads naturally to an estimate $\hat{\lambda}_i \hat{f}_t$ of $\lambda_i f_t$ which we can use to infer $\hat{\lambda}_i \sigma^2 \eta_t$ as $\hat{\lambda}_i \sigma^2 = T^{-1} \sum_{t=2}^{T} \hat{f}_t^2$. The estimation of (9) does not lead to an estimator of $U_{i,t}$, however, it does lead to an estimator $\hat{u}_{i,t} = x_{i,t} - \hat{\lambda}_i \hat{f}_t$ of $u_{i,t}$, which can be accumulated up to levels. This leads to the following estimator of $U_{i,t}$: $\hat{U}_{i,t} = \sum_{n=2}^{t} \hat{u}_{i,n}$, which can in turn be used to obtain $\hat{\sigma}^2_{U_{i,t}} = T^{-1} \sum_{t=2}^{T} \hat{U}_{i,t}^2$. The proposed estimator of $PIS_i$ is given by

$$\hat{PIS}_i = \frac{\hat{\lambda}_i^2 \hat{\sigma}^2 \eta_t}{\sum_{n=1}^{N} \hat{\lambda}_i^2 \hat{\sigma}^2_{U_{i,n}}}. \quad (10)$$

Of course, since in the present paper $U_{i,t}$ is not necessarily stationary, the meaning of $\hat{PIS}_i$ is not obvious. Therefore, in practice some pre-testing will in general be necessary to ensure valid interpretation. This is discussed in Section 3.2. However, we start by discussing the construction of $\hat{PIS}_i$.

Since (9) is nothing but a static common factor model in stationary variables, the estimation can in principle be carried out using the principal components method, which has a long tradition in econometrics and statistics (see Bai, 2003, Section 1, for a brief review of this literature). However, preliminary Monte Carlo results suggest that this estimator suffers from poor small-sample performance, especially in the empirically relevant case when $T > N$. In the present paper we therefore consider an alternative estimator that not only has good small-sample properties, but that is also computationally very convenient. In fact, it is difficult to think of a simpler estimator. The idea is to follow Pesaran (2006) and to use $\hat{f}_t = \bar{x}_t = N^{-1} \sum_{i=1}^{N} x_{i,t}$ as an estimator of $f_t$. The estimator of $\lambda_i$ is given by $\hat{\lambda}_i = (\sum_{t=2}^{T} \hat{f}_t^2)^{-1} \sum_{t=2}^{T} x_{i,t} \hat{f}_t = \hat{\sigma}^{-2} T^{-1} \sum_{t=2}^{T} x_{i,t} \hat{f}_t$. Given $\hat{f}_t$ and $\hat{\lambda}_i$, we define $\hat{U}_{i,t} = \sum_{n=2}^{t} \hat{u}_{i,n}$, where

\[\text{Note that the quantity estimated here is } \hat{f}_t, \text{ not } f_t. \text{ As we explain later in this section, } \hat{f}_t = \sum_{n=2}^{t} \hat{f}_n \text{ may be used as an estimator of } f_t.\]
\( \hat{u}_{i,t} = x_{i,t} - \hat{\lambda}_i \hat{f}_t \) for \( t = 2, \ldots, T \). From \( \hat{f}_t \) and \( \hat{U}_{i,t} \) we compute \( \hat{\sigma}_{U,i}^2 \) and \( \hat{\sigma}_\eta^2 \), and then \( \hat{PIS}_i \) using the formula in (10). We are now ready for our first main result.

**Proposition 1.** Under Assumption 1, as \( N, T \to \infty \),

\[
N | \hat{PIS}_i - PIS_i | = o_p(1).
\]

Proposition 1 can be interpreted as follows. Note that \( \sum_{n=1}^N \lambda_i^2 \sigma_{U,n}^2 \sigma_{\eta,n}^{-2} = O_p(N) \), and therefore \( PIS_i = o_p(1) \). Hence, as expected, in the current large-\( N \) setting, the weight attached to each cross-section unit converges to zero. In fact, as we show in Appendix B (Proof of Proposition 1), \( N \cdot PIS_i \) converges to a constant. Proposition 1 says that \( N \cdot \hat{PIS}_i \) converges to the same constant and therefore \( N | \hat{PIS}_i - PIS_i | \) converges to zero. In this sense, \( \hat{PIS}_i \) is consistent for \( PIS_i \).

**Remark 2.** The large-\( N \) asymptotic framework used here may appear somewhat strange, in the sense that in applications \( N \) is always finite. The reason for working with framework is the same as why one in the time series literature tend to assume that \( T \) is large, that is, it facilitates an analysis of what happens as \( N \) increases. In Section 3.2, we show that while the performance improves as \( N \) increases, the new approach works well even when \( N \) is relatively small. Hence, as in the empirical VECM-based literature where a large-\( T \) method is applied to finite-\( T \) data, the approach developed here can and should of course be applied even if \( N \) and \( T \) are finite. An important difference here is that our approach explores the information contained in both the time series and cross-sectional dimensions of the panel, as compared to VECMs, which only explore the time series information and whose performance actually decreases with increased cross-section variation.

Proposition 1 is a pure consistency result that is silent regarding the asymptotic distribution of \( \hat{PIS}_i \). The next proposition can be used as a basis for testing the significance of the individual PISs. As is well known, \( f_t \) and \( \lambda_i \) are not separately identifiable. This is true in the present panel common factor model context, but is known also from the time series cointegration literature (see Escribano and Peña, 1994, for a detailed discussion). In Appendix B we show that \( \hat{\lambda}_i (\hat{f}_t) \) is consistent for \( \lambda_i \hat{X}^{-1} (\hat{X}_f) \). The fact that \( f_t \) and \( \lambda_i \) are not separately identifiable is unproblematic for two reasons. Firstly, the product \( \hat{\lambda}_i \hat{f}_t \) still estimates \( \lambda_i f_t \), which is what we need. Secondly, similar to, for example, Figuerola-Ferretti and Gonzalo (2010), our
interest lies only in testing $\lambda_i = 0$, which is of course unaffected by scaling. A more serious
problem, however, is that for $\sqrt{T}(\hat{\lambda}_i - \lambda_i \bar{\lambda}^{-1})$ to be correctly centered, it is necessary to assume
that $\sqrt{T}/N \to 0$, or $N > \sqrt{T}$, which of course need not be the case in practice. In fact, most
(if not all) existing works on price discovery involve samples where $T >> N$. Our approach
is therefore based on bias-adjustment. Define $\hat{\sigma}_{u,i}^2 = T^{-1} \sum_{t=2}^{T} \hat{u}_{i,t}^2$ and $\hat{\sigma}_u^2 = N^{-1} \sum_{i=1}^{N} \hat{\sigma}_{u,i}^2$. The bias-adjusted estimator of $\lambda_i \bar{\lambda}^{-1}$ is given by

$$\hat{\lambda}_{BA,i} = \hat{\lambda}_i - N^{-1} \hat{\sigma}_u^{-2}(\hat{\sigma}_{u,i}^2 - \hat{\lambda}_i \hat{\sigma}_u^2).$$

The asymptotic distribution of $\sqrt{T}(\hat{\lambda}_{BA,i} - \lambda_i \bar{\lambda}^{-1})$ is given in Proposition 2. Before we come to
the proposition, however, it is useful to introduce some notation. For a product $a_{i,t}b_{j,t}$ of any
two variables $a_{i,t}$ and $b_{j,t}$, we define

$$\omega_{ab,ij}^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(a_{i,t}b_{j,t}a_{i,s}b_{j,s})$$

as the “long-run variance" of $a_{i,t}b_{j,t}$. If $i = j$, then we write $\omega_{ab,ii}^2 = \omega_{ab,i}^2$. Similarly, if $b_{j,t} = 1$, then we write $\omega_{ab,ij}^2 = \omega_{ab,i}^2$. It is also convenient to introduce $v_{i,t} = (u_{i,t}^2 - \sigma_{u,i}^2)$, $\phi_{uu,i}^2 = \lim_{N \to \infty} N^{-1} \sum_{j=1}^{N} \omega_{u,i,j}^2$, $\sigma_{u,i}^2 = E(u_{i,t}^2)$ and $\sigma_u^2 = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sigma_{u,i}^2$. 

**Proposition 2.** Under Assumption 1, as $N$, $T \to \infty$ with $\sqrt{T}/N^{3/2} \to 0$,

$$\bar{\lambda}^2 \omega_{\eta}^2 [\lambda^2 + N^{-1}(\lambda^2 \omega_{\eta}^2 \sigma_{u,i}^2 + \phi_{uu,i}^2) + N^{-2} \omega_{u,i}^2]^{-1/2} \sqrt{T}(\hat{\lambda}_{BA,i} - \lambda_i \bar{\lambda}^{-1}) \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ denotes convergence in distribution.

**Remark 3.** As the formula for $\hat{\lambda}_{BA,i}$ suggests, the bias is decreasing in $N$, and according to
Proposition 2, so is the variance. The fact that both the bias and variance of $\hat{\lambda}_{BA,i}$ are decreasing in $N$ is in sharp contrast to the VECM approach, in which the estimation uncertainty is increasing in $N$. In other words, in contrast to the VECM approach, as mentioned in Remark 2, the factor analytical approach considered here exploits the cross-sectional variation of the panel.

Proposition 2 is useful for constructing test statistics of the null hypothesis of $PIS_i = 0$. Indeed, since $\sigma_{\eta}^2$ and $\sigma_{UU,i}^2$ are both positive, for a given $N$, testing $PIS_i = 0$ is the same as testing $H_0 : \lambda_i = 0$. Proposition 2 implies that under this null,

$$\bar{\lambda}^2 \omega_{\eta}^2 [\lambda^2 + N^{-1}\phi_{uu,i}^2 + N^{-2} \omega_{u,i}^2]^{-1/2} \sqrt{T}(\hat{\lambda}_{BA,i} \to d N(0,1) \tag{11}$$

12
as \( N, T \to \infty \) with \( \sqrt{T}/N^{3/2} \to 0 \). The requirement that \( \sqrt{T}/N^{3/2} \to 0 \), or \( N^3 > T \), represents an improvement over the \( \hat{\lambda}_i \) assumption that \( \sqrt{T}/N \to 0 \). Interestingly, there is one instance when also the \( \sqrt{T}/N^{3/2} \to 0 \) requirement can be relaxed, and that is when testing \( H_0 \), as opposed to any other value of \( \lambda_i \). In particular, let us define the following restricted version of \( \hat{\lambda}_{BA,i} \):

\[
\hat{\lambda}_{RBA,i} = \hat{\lambda}_i - N^{-1} \hat{\sigma}^{-2}_{\hat{v},i} \hat{v}_{ui,i}.
\]

The asymptotic distribution of this estimator under \( H_0 \) is given in the following corollary to Proposition 2.

**Corollary 1.** Under Assumption 1 and \( H_0 \), as \( N, T \to \infty \),

\[
\sqrt{N} \hat{\sigma}^2_{\hat{v},i} \left( \frac{\hat{\lambda}_i^2}{\hat{\sigma}^2_{\hat{v},i}} + \frac{N^{-1} \hat{\phi}^2_{uu,i} + N^{-2} \hat{\omega}^2_{ij}}{\hat{\sigma}^2_{\hat{v},i}} \right)^{-1/2} \sqrt{T} \hat{\lambda}_{RBA,i} \to_d N(0,1).
\]

A remarkable feature about Corollary 1 is that it holds regardless of the relative expansion rate of \( N \) and \( T \). Hence, when using \( \hat{\lambda}_{RBA,i} \) to infer \( H_0 \) there are no restrictions on \( N \) and \( T \).

Inference based on Proposition 2 and/or Corollary 1 requires consistent estimators of \( \hat{\phi}^2_{uu,i} \) and \( \hat{\omega}^2_{ij} \). Natural candidates are given by heteroskedasticity and autocorrelation consistent (HAC) estimators similar to those considered by Bai (2003, Section 5). Let us therefore denote by \( \hat{a}_{il} (\hat{b}_{ij}) \) any consistent estimator of \( a_{il} (b_{ij}) \). The estimator of \( \hat{\omega}^2_{ij} \) is given by

\[
\hat{\omega}^2_{ij} = \frac{1}{T} \sum_{t=1}^{T} \hat{a}^2_{il} \hat{b}^2_{ij} + 2 \sum_{k=2}^{K} \left( 1 - \frac{k}{K+1} \right) \frac{1}{T} \sum_{t=k+1}^{T} \hat{a}_{il} \hat{b}_{ij} \hat{a}_{il-k} \hat{b}_{ij-k},
\]

where \( K \) is a bandwidth parameter satisfying \( K/T^{1/4} \to 0 \) as \( K, T \to \infty \). Hence, \( \hat{\omega}^2_{ij} \) is \( \hat{\omega}^2_{ij} \) with \( \hat{a}_{il} \hat{b}_{ij} = \hat{a}_{il} \hat{a}_{il} \hat{b}_{ij} \). Let \( \hat{\phi}^2_{uu,i} = N^{-1} \sum_{\ell \neq i} \hat{\omega}^2_{uu,ij} \). Note that \( \hat{f}_i \) estimates \( \hat{f}_i \), not \( f_i \). Therefore, \( \hat{\omega}^2_{ij} \) estimates \( \Lambda^2 v_i^2 \). The appropriate statistic to consider for the test of \( H_0 : \lambda_i = 0 \) (\( PIS_i = 0 \)) is consequently given by

\[
t_{RBA,i} = \hat{\omega}^2_{ij} \left( \hat{\sigma}^2_{\hat{v},i} + N^{-1} \hat{\phi}^2_{uu,i} + N^{-2} \hat{\omega}^2_{ij} \right)^{-1/2} \sqrt{T} \hat{\lambda}_{RBA,i},
\]

which has a limiting \( N(0,1) \) distribution under \( H_0 \) as \( N, T, K \to \infty \) with \( K/T^{1/4} \to 0 \). The asymptotic distribution of \( t_{BA,i} \), the \( t \)-statistic based on \( \hat{\lambda}_{BA,i} \), is the same but only under the additional requirement that \( \sqrt{T}/N^{3/2} \to 0 \).

The above \( t \)-statistic is suitable for testing \( \lambda_i = 0 \) for a single cross-section unit \( i \). Testing multiple cross-section units is, however, just as easy. Suppose therefore that we are interested
in testing $H_0 : \lambda_i = 0$ for $i \in S \subset \{1, \ldots, N\}$. Let $n < N$ denote the cardinality of $S$. This restriction can be tested using the following Wald-type test statistic:

$$W_{RBA,S} = \sum_{i \in S} t_{RBA,i}^2,$$

which has a limiting $\chi^2(n)$ distribution under $H_0$.

**Remark 4.** As a measure the fraction of units for which $\lambda_i \neq 0$ one may consider the empirical rejection frequency of $t_{RBA,1}, \ldots, t_{RBA,N}$, as defined by

$$\hat{\alpha}_{RBA} = \frac{1}{N} \sum_{i=1}^{N} 1(|t_{RBA,i}| > c_\alpha),$$

where $1(A)$ is the indicator function for the event $A$ and $c_\alpha$ is the appropriate right-tail $\alpha$-level critical value from $N(0,1)$.

**Remark 5.** The simplicity by which the significance of the PISs can be tested stands in sharp contrast to the standard approach of Hasbrouck (1995), which is not well-suited for testing (see, for example, Figuerola-Ferretti and Gonzalo, 2010, for a discussion).

### 3.2 Testing for unit roots

The above asymptotic results hold regardless of the value taken by $\rho_i \in (-1, 1]$. For purpose of interpretation, however, it is important that $|\rho_i| < 1$, because only if $U_{i,t}$ is stationary is it meaningful to refer to $F_t$ as a “fundamental price”. But $T^{-1/2} |\hat{U}_{i,t} - U_{i,t}| = o_p(1)$ (under $\rho_i = 1$), which means that testing for a unit root in $\hat{U}_{i,t}$ is asymptotically the same as testing for a unit root in $U_{i,t}$. The testing of $\hat{U}_{i,t}$ therefore does not require any specialized techniques, but can be carried out using any existing unit root test, such as the augmented Dickey–Fuller (ADF) test. The fact that $U_{1,t}, \ldots, U_{N,t}$ are independent means that $\hat{U}_{1,t}, \ldots, \hat{U}_{N,t}$ are independent too (although strictly speaking the latter independence only holds asymptotically). Hence, when testing $\hat{U}_{1,t}, \ldots, \hat{U}_{N,t}$ one can also consider so-called “first-generation” panel unit root tests for cross-section independent data, such as the Im–Pesaran–Shin (IPS) test. In Section 4 we elaborate on this.

That $F_t$ is unit root non-stationary is necessary for the asymptotic results reported in Section 3.1 to hold, and is in this sense similar to the assumptions of the previous literature. However, as alluded to in Section 2.2, in contrast to the PT approach, here the non-stationarity of $F_t$ is not an identifying assumption, which means that it is testable. Let $\hat{F}_t = \sum_{n=2}^{t} \hat{f}_n$. By using the
results provided in Appendix B of this paper, it can be shown that \( T^{-1/2}|\hat{F}_t - \bar{\lambda}F_t| = o_p(1) \). Hence, as expected given that \( \lambda_i \) and \( f_t \) are not separately identifiable, \( F_t \) can only be estimated up to a multiplicative scalar. Of course, if \( F_t \) has a unit root, then so does \( \bar{\lambda}F_t \). In other words, for the purpose of testing the unit root restriction, the fact that \( F_t \) is not identified is not a problem. As with \( \hat{U}_{i,t} \), the testing can be carried out using any existing unit root test, such as the ADF test.

### 3.3 Monte Carlo simulations

A small-scale Monte Carlo study was undertaken to investigate the small-sample properties of the proposed PIS; however, we start by considering the bias-adjusted (restricted and unrestricted) estimators of \( \lambda_i \), which are of course key in inferring the PIS. The DGP considered for this purpose is given by a simplified version of (2)–(4), in which both \( \eta_t \) and \( \epsilon_{t,i} \) are drawn from \( N(0, 1) \). Furthermore, we specify \( \rho_1 = ... = \rho_N = 0 \), draw \( \lambda_2, ..., \lambda_N \) from \( N(1, 1) \), and set \( \lambda_1 \) to one of 0, 0.25, 0.5, 0.75 or 1, depending on the purpose of the experiment. In interest of space we focus on the bias and root mean squared error (RMSE) of the estimators, and the size and power of a nominal 5% level test of the null hypothesis of \( H_0 : \lambda_1 = 0 \).

Table 1 reports size, bias and RMSE for \( \hat{\lambda}_{BA,1} \) and \( \hat{\lambda}_{RBA,1} \) when \( \lambda_1 = 0 \), which is the relevant scenario when wanting to infer if \( PIS_1 = 0 \). The first thing to notice is that the test statistics of both estimators have a tendency to be oversized; however, this is mainly among the smaller values of \( N \) and \( T \). Specifically, while the accuracy of the restricted estimator improves very quickly with increases in both \( N \) and \( T \), for the unrestricted estimator the performance is quite flat in \( T \). The reason for the distortions can be found by looking at the bias results; unlike the restricted estimator, for the unrestricted estimator \( N = 10 \) is apparently too small, and therefore the bias remains even if \( T \) increases. Of course, given the \( \sqrt{T}/N^{3/2} \rightarrow 0 \) condition, the fact that the restricted estimator requires \( N > 10 \) does not come as a surprise.

The power of the tests for different values of \( \lambda_1 \neq 0 \) are reported in Table 2. Consider the \( t \)-statistic for testing \( H_0 \) based on \( \hat{\lambda}_{BA,1} \). The numerator of this test statistic is given by

\[
\sqrt{T}\hat{\lambda}_{BA,1} = \sqrt{T}(\hat{\lambda}_{BA,1} - \lambda_1\bar{\lambda}^{-1}) + \sqrt{T}\lambda_1\bar{\lambda}^{-1},
\]

where the first term is normally distributed by Proposition 2, while the second term is \( O_p(\sqrt{T}) \) whenever \( \lambda_1 \neq 0 \). The same is true for the \( t \)-statistic based on \( \hat{\lambda}_{RBA,1} \). It follows that under \( H_1 : \lambda_1 \neq 0 \) both \( t \)-statistics diverge at the rate \( \sqrt{T} \). Power should therefore be increasing in \( T \) but not in \( N \), which is also what we see in the table. We also see that the power of \( \hat{\lambda}_{BA,1} \) is generally higher than that of \( \hat{\lambda}_{RBA,1} \), which is
partly expected given its size distortions under $H_0$ and the fact that the reported powers are not corrected for size.

Figure 1: Mean absolute bias of $N \cdot \hat{PIS}_1$ as a function of $N$ and $T$.

Note: The $x$-, $y$- and $z$-axes display the value of $N$, $T$ and the absolute bias, respectively.

The results for the PIS are summarized in Figure 1, which reports the absolute bias of $N \cdot \hat{PIS}_1$ as a function of $N$ and $T$ for the above DGP when $\lambda_1 = 1$. In the proof of Proposition 1, we show that $N|\hat{PIS}_1 - PIS_1| = O_p(T^{-1/2}) + O_p(N^{-1/2})$. In accordance with this, we see that the bias is decreasing in both $N$ and $T$, and also that the rate at which this happens is roughly the same for the two indices.

In sum, we find that the proposed estimators of $\lambda_i$ have good small-sample properties. The restricted estimator performs particularly well, and in fact works quite well even when $N$ and $T$ are as small as 5, which of course makes it widely applicable. This good performance is reflected in the PIS, which is very accurate. In our empirical illustrations, $N < 10$ and $T > 1,000$, suggesting that inference should be based on the restricted estimator. This is confirmed by Table 3, which contains 5% size and power results for some constellations of $N$ and $T$ with $N^3 < T$ (such that the $\sqrt{T}/N^{3/2} \rightarrow 0$ requirement is not met).
4 Empirical illustrations

We demonstrate the usefulness of our new toolbox for price discovery using two different data sets. Our first data set, denoted “OIL”, contains crude oil prices. These data emanate from different countries. Specifically, we consider three crude oil price series listed on the US, the UK and Oman exchanges. The data are 5-day daily covering the period 1/06/2007 to 12/10/2012, culminating into 1,401 time series observations per exchange. Studies of oil are relatively rare in the price discovery literature, and as far as we know none has yet studied oil price discovery from a cross-country perspective. This first data set is therefore interesting from an empirical perspective, and also because it has the potential to stimulate more research on price discovery in markets other than cross-listed equities.

Our second data set, denoted “EQUITY”, is more conventional in the sense that we have equity prices of a particular firm that is listed in more than one exchange. We choose Arcelor Mittal, a major steel company, which has its primary listing on the Luxembourg stock exchange and has secondary listings on another eight exchanges. The stock price data are 5-day daily covering the period 30/11/2010 to 1/05/2014 for a total of 1,249 observations.

In Table 4 we report some descriptive statistics on the two data sets. We begin by considering the results for OIL. The mean and median crude oil price (measured in US dollars) is highest for Oman and lowest for the UK. Likewise, the standard deviation of oil price is highest for the UK, followed by Oman, and the US. The lowest (highest) oil price is recorded for the US (UK) at US$33.98 (US$146.08). These differences in the descriptive statistics are suggestive of different price setting mechanisms. Consider next the results for EQUITY. The first thing to note is that the cross-sectional variation of EQUITY is somewhat smaller than that of OIL. For example, while in case of OIL the mean price ranges between 81.249 and 87.919, in case of EQUITY the range is between 14.595 and 14.782. We also see that the highest price is recorded at the TOM MTF (the Netherlands) exchange and that the lowest price of US$5.094 is recorded at the BATS Europe exchange.

Table 5 reports some ADF and IPS unit root test results for the estimated common and idiosyncratic components respectively. Both tests allow for a constant and a linear trend. As expected, we see that while for the common component the null hypothesis of a unit root is not rejected, for the idiosyncratic component the evidence against the unit root null is much stronger. The latter evidence is, however, not that overwhelming, and the conclusion in case
of OIL depends on the chosen significance level. Indeed, only at the 10% level is the evidence against the unit root null significant. This illustrates the importance of being able to test rather than just assume that the idiosyncratic component is in fact stationary. However, at the 10% level, we may conclude that both oil and equity prices are cross-unit cointegrated with a common stochastic trend, as predicted by theory.

Consider next the evidence on price discovery. The results from the (best performing) restricted estimator are reported in Table 6. We begin by looking at the results for OIL. The Oman exchange has the highest PIS of 51%, followed by the UK exchange with a PIS of about 45%. The US exchange ends up last with a PIS of 3.6% only. The main implication here is that the Oman and UK exchanges dominate the price discovery in the market for crude oil, and that the contribution of the US exchange is only very marginal. These results are supported by a formal test for no price discovery ($\lambda_i = 0$). Indeed, while for Oman and the UK the hypothesis is comfortably rejected, for the US the evidence is less strong, being significant at the 10% level only. The fact that the PIS is highest in Oman is not unexpected. Indeed, the rapid decline in Dubai crude oil output has increased the (potential) importance of Oman in pricing crude oil. Oman has some of the characteristics to enable it to play a dominating role in the price discovery process. For example, the production is not subject to Organization of Petroleum Exporting Countries (OPEC) quotas as Oman is not a member of OPEC, there are no destination restrictions, Oman’s government offers incentives to international oil companies for extraction and development activities, and Oman has a strong market presence in Asia, one of the fastest growing regions in the world.\(^3\)

The results for EQUITY show how the price discovery process for Arcelor Mittal is dominated by the primary exchange, the Luxembourg stock exchange, which has a PIS of 30.4%. That price discovery is dominated by the primary exchange is consistent with the previous literature (see, for example, Bacidore and Sofianos, 2002; Solnik, 1996). However, while the primary exchange dominates, in agreement with the results of, for example, Chan et al. (2013), and Eun and Sabherwal (2003), the contribution of the other exchanges is still important. Indeed, the contribution of the XETRA exchange (Germany-based) is almost as high as for the primary exchange with a PIS of 22.4%, and four of five UK exchanges contribute around 10% each. The evidence of price discovery process is therefore less concentrated for EQUITY than for OIL. We also notice from the results for EQUITY that the null hypothesis of no price dis-

\(^3\)In 2012 over 95% of Oman’s oil was exported to Asia with about half of this exported to China.
covery is comfortably rejected at the 1% level for all nine exchanges, which is again in contrast to the results for OIL.

5 Conclusion

Increasing globalization and financial integration have, together with recent major events such as the global financial crisis, sparked interest in the issue of price discovery. The question is: to what extent do markets contribute to the (fundamental) price of cross-listed assets? The standard econometric approach by which researchers have been trying to answer this question is based on a fitted VECM, from which the ISs of different markets can be obtained. But while very popular, this approach suffers from a number of important shortcomings that are likely to become even more important in the future as more data become available. Specifically, being multivariate time series, the VECM approach is not equipped to handle data where \( N \) is “large”, which obviously puts a restriction on applicability. Also, the resulting ISs depend critically on the appropriateness of the required identifying restrictions. The current paper can be seen as a reaction to these shortcomings. The purpose is to develop a new reduced-form approach that makes use of the cross-sectional variation of the data. It should also be user-friendly, and enable both estimation and statistical inference of the ISs. The solution is a factor analytical approach that has its roots in the large and growing literature on large-dimensional common factor models. The asymptotic properties of the approach are derived and evaluated in small samples using Monte Carlo simulation. In the empirical part of the paper we consider two data sets; (i) crude oil prices listed at three national exchanges, and (ii) Arcelor Mittal, a major steel company, whose equity is listed at nine exchanges.
References


Appendix A: Derivation of (5)

The IS is obtained as the explanatory power of observed price changes in a regression. Let us use \( v_{it} = \lambda_i \eta_t + U_{it} \) to denote the sum of the noise coming from the efficient price and market microstructure components of the model. Hence, \( v_{it} \) can be seen as the shocks to the observed price, \( X_{it} \). In vector notation, we have

\[
v_t = \Lambda \eta_t + U_t, \tag{A1}\]

where \( v_t = (v_{1,t}, ..., v_{N,t})' \), \( \Lambda = (\lambda_1, ..., \lambda_N)' \) and \( U_t = (U_{1,t}, ..., U_{N,t})' \) are all \( N \times 1 \). The \( N \times N \) covariance matrix of \( v_t \) is given by

\[
\Sigma_v = \sigma_\eta^2 \Lambda \Lambda' + \Sigma_U, \tag{A2}\]

where, under Assumption 1, \( \Sigma_U \) is diagonal; \( \Sigma_U = \text{diag}(\sigma_{U_1}^2, ..., \sigma_{U_N}^2) \). As in Hasbrouck (1995), De Jong and Schotman (2010) consider the following (reverse) population regression relationship between \( \eta_t \) and \( v_t \):

\[
\eta_t = \gamma' v_t + e_t, \tag{A3}\]

where \( e_t \) is the part of the innovation in the efficient price that is not due to shocks in the observed price, and \( \gamma = (\gamma_1, ..., \gamma_N)' = \Sigma_v^{-1} \Lambda \sigma_\eta^2 \). The \( R^2 \) measure from the above regression is

\[
R^2 = \frac{\gamma' \Sigma_v \gamma}{\sigma_\eta^2} = \gamma' \Lambda = \sum_{i=1}^N \gamma_i \lambda_i. \tag{A4}\]

The IS of market \( i \) is given simply by \( IS_i = \gamma_i \lambda_i \). In what remains, we show that this formula is in fact equal to the one in (5). We begin by using the fact that for a nonsingular \( N \times N \) matrix \( A \) and an \( N \times 1 \) vector \( a \), we have \((A + aa')^{-1} = A^{-1} - (1 + a'A^{-1}a)^{-1}A^{-1}a'aA^{-1}\) (see Abadir and Magnus, 2005, Exercise 4.28). Applying this to \( \Sigma_v^{-1} \) in \( \gamma \), we obtain

\[
\Sigma_v^{-1} = \Sigma_U^{-1} - \frac{1}{1 + \sigma^2_\eta^2 \Lambda' \Sigma^{-1}_u \Lambda} \sigma^2_\eta \Sigma^{-1}_{u} \Lambda \Lambda' \Sigma^{-1}_u, \tag{A5}\]

which in turn implies

\[
\gamma = \Sigma_v^{-1} \Lambda \sigma_\eta^2 = \sigma_\eta^2 \Sigma_u^{-1} \Lambda - \frac{1}{1 + \sigma^2_\eta^2 \Lambda' \Sigma^{-1}_u \Lambda} \sigma^2_\eta \Sigma^{-1}_{u} \Lambda \Lambda' \Sigma^{-1}_u \Lambda \sigma^2_\eta = \sigma^2_\eta \Sigma_u^{-1} \Lambda \left( 1 - \frac{1}{1 + \sigma^2_\eta^2 \Lambda' \Sigma^{-1}_u \Lambda} \sigma^2_\eta \Lambda' \Sigma^{-1}_u \Lambda \right) = \frac{\sigma^2_\eta \Sigma_u^{-1} \Lambda}{1 + \sigma^2_\eta \Lambda' \Sigma^{-1}_u \Lambda}. \tag{A6}\]
But $\Sigma_U$ is diagonal and therefore $\Sigma_U^{-1} = \text{diag}(\sigma_{U,1}^{-2}, ..., \sigma_{U,N}^{-2})$. Hence, denoting by $[a]_n$ the $n$-th element of the vector $a$, we can show that

$$IS_i = \gamma_i \lambda_i = \frac{\sigma^2_{\eta} [\Sigma_U^{-1} \Lambda]_i \lambda_i}{1 + \sigma^2_{\eta} \Lambda' \Sigma_U^{-1} \Lambda} = \frac{\sigma^2_{\eta} \sigma_{U,i}^{-2} \lambda_i^2}{1 + \sum_{n=1}^{N} \sigma^2_{\eta} \sigma_{U,n}^{-2} \lambda_n^2},$$

(A7)

which is the formula given in (5).

### Appendix B: Proofs

We start with some notation. The model for $x_{i,t}$ can be written in matrix notation as

$$x_i = f \lambda_i + u_i,$$

(A8)

where $x_i = (x_{i,2}, ..., x_{i,T})'$, $f = (f_2, ..., f_T)'$ and $u_i = (u_{i,2}, ..., u_{i,T})'$ are $(T - 1) \times 1$, while $\lambda_i$ is $1 \times 1$. Alternatively, the model can be written as the following $N$-dimensional system:

$$x_t = \lambda f_t + u_t,$$

(A9)

where $x_t = (x_{1,t}, ..., x_{N,t})'$ and $u_t = (u_{1,t}, ..., u_{N,t})'$ are $N \times 1$, and $\lambda = (\lambda_1, ..., \lambda_N)'$ is $N \times 1$. The matrix notation

$$x = f \lambda' + u$$

(A10)

will also be used, where $x = (x_1, ..., x_N)$ and $u = (u_1, ..., u_N)$ are $(T - 1) \times N$. In what follows the representations in (A8)-(A10) will be used interchangeably.

Denote by $\bar{a} = \frac{1}{N} \sum_{i=1}^{N} a_i$ the cross-section average of the generic variable $a_i$. Many of the results can be expressed in terms of $\hat{f} - f \bar{\lambda} = \bar{u}$, whose $t$-th element is given by $\hat{f}_t - \bar{\lambda} f_t = \bar{u}_t$.

Before we come to the proofs of Propositions 1 and 2, we state a preliminary lemma.

**Lemma A.1.** Under Assumption 1, as $N, T \to \infty$,

$$\sqrt{N T^{-1/2} f' \pi} \to_d N(0, \sigma^2_{\eta} \sigma^2_{\mu}),$$

$$NT^{-1} u'_i \pi \sim \sigma^2_{u,i} + (T^{-1} \omega^2_{u,i} + NT^{-1} \phi_{aa,i})^{1/2} N(0, 1),$$

$$NT^{-1} \bar{\pi} \pi = \sigma^2_{\pi} + O_p(N^{-1/2}) + O_p(T^{-1/2}),$$

24
where \( v_{i,t} = (u_{i,t}^2 - \sigma^2_{u,i}) \) and

\[
\begin{align*}
\sigma^2_u &= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{u,i}, \\
\omega^2_{i,j} &= \lim_{T \to \infty} \frac{1}{T} \sum_{t=2}^{T} \sum_{s=2}^{T} E(v_{i,s}v_{i,t}), \\
\omega^2_{uu,ij} &= \lim_{T \to \infty} \frac{1}{T} \sum_{t=2}^{T} \sum_{s=2}^{T} E(u_{i,s}u_{i,t}u_{j,t}), \\
\phi^2_{uu,i} &= \lim_{N \to \infty} \frac{1}{N} \sum_{j \neq i} \omega^2_{uu,ij}.
\end{align*}
\]

**Proof of Lemma A.1.**

Consider the first result. Note that

\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} f' u_i \right)^2 \right] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} E(f'u_i u'_j f) = \frac{1}{NT} \sum_{i=1}^{N} E(f'u_i u'_i f)
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=2}^{T} E(f_s u_{i,s} f_t u_{i,t})
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=2}^{T} E(f'_i E(u_{i,t}^2)) + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=2}^{T} E(f_s) E(u_{i,s} u_{i,t}) E(f_t)
\]

\[
= \sigma^2_\eta \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{u,i} \to \sigma^2_\eta \sigma^2_u
\]

as \( N \to \infty \). Application of the Lindeberg–Feller central limit theorem now yields

\[
\sqrt{NT^{-1/2}} f' \tilde{\mu} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} f' u_i \to_d N(0, \sigma^2_\eta \sigma^2_u)
\]

(A11)

as \( N, T \to \infty \).

For the second result, from

\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{t=2}^{T} v_{i,t} \right)^2 \right] = \frac{1}{T} \sum_{t=2}^{T} \sum_{s=2}^{T} E(v_{i,s} v_{i,t}),
\]

\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{j \neq i} u_{i,t} u_{j,t} \right)^2 \right] = \frac{1}{NT} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{j \neq i} \sum_{i \neq j} E(u_{i,s} u_{i,t} u_{j,t} u_{j,s})
\]

\[
= \frac{1}{NT} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{j \neq i} \sum_{i \neq j} E(u_{i,t} u_{j,t}) E(u_{j,t} u_{j,s}),
\]

25
we obtain
\[
NT^{-1}u_i^T \Pi = \frac{1}{T} \sum_{t=2}^{T} u_i t N \Pi t = \frac{1}{T} \sum_{t=2}^{T} N \sum_{j=1}^{N} u_{ij} u_{ij, t} = \frac{1}{T} \sum_{t=2}^{T} u_i^T + \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=2}^{T} \sum_{j \neq i} u_{ij} u_{ij, t} \\
= \sigma^2_{u,i} + \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \sigma_{\hat{\lambda}, t} + \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=2}^{T} \sum_{j \neq i} u_{ij} u_{ij, t} \\
\sim \sigma^2_{u,i} + (T^{-1} \omega_{\hat{\lambda}, i}^2 + NT^{-1} \phi_{u,i,t}^2)^{1/2} N(0, 1), \tag{A12}
\]
which requires $N, T \to \infty$.

The third and final result follows from noting that
\[
NT^{-1} \hat{\Pi} \hat{\Pi} = \frac{N}{T} \sum_{t=2}^{T} \hat{\Pi} \hat{\Pi}_t = \frac{1}{NT} \sum_{t=2}^{T} N \sum_{j=1}^{N} u_{ij} u_{ij, t} \\
= \frac{1}{NT} \sum_{t=2}^{T} N \sum_{j=1}^{N} u_{ij}^2 + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{t=2}^{T} \sum_{j \neq i} u_{ij} u_{ij, t} \\
= \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{u,i} + \frac{1}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=2}^{T} u_{i,t}^2 - \sigma^2_{u,i} \right) + O_p(T^{-1/2}) \\
= \sigma^2_u + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A13}
\]

This completes the proof of the lemma.

Proof of Proposition 1.

We begin by showing the consistency of $\hat{\lambda}_i^2 \hat{\sigma}_q^2$. The estimator of $f_t$ is given by
\[
f_t = \bar{x}_t = \bar{x} f_t + \Pi_t = \bar{x} f_t + O_p(N^{-1/2}), \tag{A14}
\]
which holds point-wise in $t$. However, it is easy to show that the result also holds uniformly in $t$. Given $\hat{f} = (\hat{f}_{2, \ldots, \hat{f}_T})'$, we compute $\hat{\lambda}_i = \hat{\sigma}_q^{-2} T^{-1} \hat{f}' x_i$. By Lemma A.1,
\[
\hat{\sigma}_q^2 = T^{-1} \hat{f}' \hat{f} \\
= T^{-1} (\bar{x}^2 f' f + 2 \bar{x} f' \Pi + \Pi' \Pi) \\
= \bar{x}^2 T^{-1} f' f + O_p((NT)^{-1/2}) + O_p(N^{-1}) \\
= \bar{x}^2 \sigma_q^2 + T^{-1/2} \bar{x} \sqrt{T} T^{-1} f' f - \sigma_q^2 + O_p((NT)^{-1/2}) + O_p(N^{-1}) \\
= \bar{x}^2 \sigma_q^2 + O_p(T^{-1/2}) + O_p(N^{-1}). \tag{A15}
\]
Hence, by Taylor expansion,
\[
\hat{\sigma}_q^{-2} = \bar{x}^{-2} \sigma_q^{-2} + O_p(T^{-1/2}) + O_p(N^{-1}). \tag{A16}
\]
It is easy to show that $T^{-1}f'u_i = O_p(T^{-1/2})$ and $T^{-1}u'_i \pi = O_p(N^{-1}) + O_p((NT)^{-1/2})$, giving

$$
\hat{\lambda}_i = \hat{\sigma}^{-2} T^{-1} x'_i f
$$

$$
= \hat{\sigma}^{-2} T^{-1} (\lambda_i f + u_i)'(\bar{x}f + \bar{\pi})
$$

$$
= \hat{\sigma}^{-2} T^{-1} (\bar{x} \lambda_i f'f + \lambda_i f' \pi + \bar{x} u'_i f + u'_i \pi)
$$

$$
= \lambda_i \bar{x}^{-1} \sigma^{-2} T^{-1} f'f + O_p(T^{-1/2}) + O_p(N^{-1})
$$

$$
= \lambda_i \bar{x}^{-1} + O_p(T^{-1/2}) + O_p(N^{-1}).
$$

(A17)

This result, together with the fact that $\hat{f}_t = \bar{x} f_t + O_p(N^{-1/2})$, implies

$$
\hat{\lambda}_i \hat{f}_t = [\lambda_i \bar{x}^{-1} + O_p(T^{-1/2}) + O_p(N^{-1})][\bar{x} f_t + O_p(N^{-1/2})]
$$

$$
= \lambda_i f_t + O_p(N^{-1/2}) + O_p(T^{-1/2}).
$$

(A18)

By taking squares and then averaging over time, we obtain

$$
\hat{\lambda}_i^2 \hat{\sigma}^2 = \lambda_i T^{-1} f'f + O_p(N^{-1/2}) + O_p(T^{-1/2})
$$

$$
= \lambda_i \sigma_u^2 + T^{-1/2} \lambda_i \sqrt{T} (T^{-1} f'f - \sigma^2) + O_p(N^{-1/2}) + O_p(T^{-1/2})
$$

$$
= \lambda_i \sigma_u^2 + O_p(N^{-1/2}) + O_p(T^{-1/2}),
$$

(A19)

which establishes the consistency of $\hat{\lambda}_i^2 \hat{\sigma}^2$.

Let $\hat{r}_i = \hat{\lambda}_i^2 \hat{\sigma}^2 \hat{\sigma}^{-2} u_{i,i}$ and $\hat{R} = N^{-1} \sum_{i=1}^{N} \hat{r}_i$ with analogous definitions of $r_i$ and $R$, but without hats. It can be shown that under $|\rho_i| < 1$, $\sigma_{u,i}^2 = \sigma_{\bar{u},i}^2 + O_p(T^{-1/2})$. This result, together with the one for $\hat{\lambda}_i^2 \hat{\sigma}^2$, implies

$$
\hat{r}_i - r_i = O_p(N^{-1/2}) + O_p(T^{-1/2}),
$$

with $\hat{R} - R$ being of the same order. Also, by Taylor expansion,

$$
\hat{R}^{-1} = R^{-1} + O_p(N^{-1/2}) + O_p(T^{-1/2}),
$$

(A20)

which in turn implies

$$
N(\hat{PIS}_i - PIS_i) = \hat{r}_i \hat{R}^{-1} - r_i R^{-1} = N^{-1} (\hat{r}_i - r_i) R^{-1} + (\hat{R}^{-1} - R^{-1}) N^{-1} \hat{r}_i
$$

$$
= O_p(N^{-1/2}) + O_p(T^{-1/2}),
$$

(A21)

as was to be shown.

Proof of Proposition 2.
Consider \( \lambda_i = \theta^{-2}T^{-1}f'x_i \). The numerator is

\[
T^{-1}f'x_i = T^{-1}f'x_i + T^{-1}f'u_i = T^{-1}f'x_i(T^{-1}(\lambda f + \eta)\nu\lambda^{-1} \lambda_i + T^{-1}(\lambda f + \eta)u_i)
\]

which we can rewrite in the following way:

\[
T^{-1/2}f'x_i - \sqrt{T}x^{-1} \lambda_i \theta^{-2} \]

\[
= \sqrt{T}N^{-1}(\sigma^2_{\nu,i} - \lambda_i \nu^{-1} \sigma^2_{\nu}) - \lambda_i \nu^{-1} N^{-3/2} \sqrt{NT}(NT^{-1}\nu'\nu - \sigma^2_{\nu}) + \lambda_i NT^{-1/2}f'u_i
\]

By using calculations similar to those used in Proof of Lemma A.1, it is possible to show that \( \sqrt{NT}(NT^{-1}\nu'\nu - \sigma^2_{\nu}) = O_p(\sqrt{T}) + O_p(\sqrt{N}) \), implying

\[
T^{-1/2}f'x_i - \sqrt{T}x^{-1} \lambda_i \theta^{-2} \]

\[
= \sqrt{T}N^{-1}(\sigma^2_{\nu,i} - \lambda_i \nu^{-1} \sigma^2_{\nu}) - \lambda_i \nu^{-1} N^{-3/2} \sqrt{NT}(NT^{-1/2}\nu'f)
\]

\[
- \lambda_i N^{-1/2}(\sqrt{NT^{-1/2}\nu'u_i} - \sigma^2_{\nu,i}) + O_p(N^{-3/2} \sqrt{T}) + O_p(N^{-1}).
\]  

By assumption, \( T^{-1/2}f'u_i \to_d N(0, \sigma^2_{\nu,i}) \) as \( T \to \infty \), and by Lemma A.1, \( \sqrt{NT^{-1/2}\nu'f} \to_d N(0, \sigma^2_{\nu,i}) \) and \( NT^{-1}\nu'u_i - \sigma^2_{\nu,i} \sim (T^{-1}\sigma^2_{\nu,i} + NT^{-1}\phi^2_{\nu,i})^{1/2}N(0, 1) \). We now show that these three terms are asymptotically uncorrelated. We begin with the covariance between \( T^{-1/2}f'u_i \) and \( N^{-1/2}(\sqrt{NT^{-1/2}\nu'f}) \), which is given by

\[
E[(T^{-1/2}f'u_i)N^{-1/2}(\sqrt{NT^{-1/2}\nu'f})] = E\left(\frac{1}{T} \sum_{i=2}^{T} \sum_{s=2}^{T} f_iu_{i,s}\nu'sf_s\right)
\]

\[
= E\left(\frac{1}{NT} \sum_{i=2}^{T} \sum_{s=2}^{T} \sum_{j=1}^{N} f_iu_{i,s}u_{s,j}f_s\right)
\]

\[
= 1 \frac{1}{N} \sum_{i=2}^{T} E(f_i^2u_{i,i}^2) + 1 \sum_{i=2}^{T} \sum_{s \neq j \neq i} E(f_iu_{i,s}u_{s,j}f_s)
\]

\[
= N^{-1}\sigma^2_{\nu,i}\sigma^2_{\nu,i} = O(N^{-1}),
\]
which is obviously negligible. Also, $NT^{-1}\eta_i - \sigma^2_{u,i}$ is uncorrelated with $T^{-1/2}\eta f$ and $T^{-1/2}f' u_i$;

$$E[(T^{-1/2}\eta_i - \sigma^2_{u,i})] = E\left(\frac{N}{T^{3/2}} \sum_{i=2}^{T} \sum_{t=2}^{T} \sum_{j=1}^{N} E(u_{i,t} u_{i,t} E(f_i)) - T^{-1/2}\sigma^2_{u,i} E(f_i) f_i) = 0,$$

$$E[(T^{-1/2}f' u_i)(NT^{-1}\eta_i - \sigma^2_{u,i})] = E\left(\frac{N}{T^{3/2}} \sum_{i=2}^{T} \sum_{t=2}^{T} \sum_{j=1}^{N} E(u_{i,t} u_{i,t} E(f_i) f_i) = 0.$$

These results imply that

$$T^{-1/2}f' \eta_i - \sqrt{T} \lambda_i \lambda^{-1} \sigma^2_{u,i} - \sqrt{T} N^{-1}(\sigma^2_{u,i} - \lambda_i \lambda^{-1} \sigma^2_{u,i})$$

$$= \lambda T^{-1/2}f' \eta_i + \lambda_i N^{-1/2}(\sqrt{N} T^{-1/2}\eta f) + \sqrt{T} N^{-1}(\lambda_i - \lambda_i \lambda^{-1} \sigma^2_{u,i}) + O_p(N^{-3/2})$$

$$\sim \left[\lambda^2 \sigma^2_{u,i} + N^{-1}(\lambda_i^2 \sigma^2_{u} + \phi_{u,i}^2) + N^{-2} \omega^2_{u,i}\right]^{1/2} N(0, 1). \quad (A24)$$

By using the results obtained in Proof of Proposition 1,

$$\sqrt{T} N^{-1}(\sigma^2_{u,i} - \lambda_i \lambda^{-1} \sigma^2_{u}) = \sqrt{T} N^{-1}(\sigma^2_{u,i} - \lambda_i \lambda^{-1} \sigma^2_{u}) - \sqrt{T} (\sigma^2_{u,i} - \sigma^2_{u})$$

$$+ N^{-3/2} \lambda_i \lambda^{-1} \sqrt{N} (\sigma^2_{u,i} - \sigma^2_{u}) + N^{-1} \sqrt{T} (\lambda_i - \lambda_i \lambda^{-1}) \sigma^2_{u,i}$$

$$= \sqrt{T} N^{-1}(\sigma^2_{u,i} - \lambda_i \lambda^{-1} \sigma^2_{u}) + O_p(N^{-1}) + O_p(\sqrt{T} N^{-3/2}). \quad (A25)$$

Hence, provided that $\sqrt{T} N^{-3/2} = o(1),$

$$T^{-1/2}f' \eta_i - \sqrt{T} \lambda_i \lambda^{-1} \sigma^2_{u,i} - \sqrt{T} N^{-1}(\sigma^2_{u,i} - \lambda_i \lambda^{-1} \sigma^2_{u,i})$$

$$\sim \left[\lambda^2 \sigma^2_{u,i} + N^{-1}(\lambda_i^2 \sigma^2_{u} + \phi_{u,i}^2) + N^{-2} \omega^2_{u,i}\right]^{1/2} N(0, 1). \quad (A26)$$

Another application of the results provided as a part of Proof of Proposition 1 yields $\delta_{u,i}^2 = \lambda_i - \delta_{u,i}^2 N^{-1}( \lambda_i^2 \sigma^2_{u} + \phi_{u,i}^2) + N^{-2} \omega^2_{u,i},$ (A27)
which again requires $\sqrt{T}N^{-3/2} = o(1)$. The required result is implies by this.

Proof of Corollary 1.

This proof follows almost immediately from that of Proposition 2. The only difference is that the bias-correction in $\hat{\lambda}_{RBA,j}$ does not involve $\hat{\lambda}_i$, which in turn eliminates the need for the requirement that $\sqrt{T}N^{-3/2} = o(1)$. 

<table>
<thead>
<tr>
<th>Running</th>
<th>Size</th>
<th>Bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>T = 10, N runs</td>
<td>N = 10, T runs</td>
<td>N and T</td>
<td>T = 10, N runs</td>
</tr>
<tr>
<td>5</td>
<td>0.207</td>
<td>0.176</td>
<td>0.245</td>
</tr>
<tr>
<td>10</td>
<td>0.129</td>
<td>0.129</td>
<td>0.129</td>
</tr>
<tr>
<td>15</td>
<td>0.095</td>
<td>0.112</td>
<td>0.087</td>
</tr>
<tr>
<td>20</td>
<td>0.102</td>
<td>0.104</td>
<td>0.080</td>
</tr>
<tr>
<td>25</td>
<td>0.077</td>
<td>0.114</td>
<td>0.081</td>
</tr>
<tr>
<td>30</td>
<td>0.081</td>
<td>0.102</td>
<td>0.063</td>
</tr>
<tr>
<td>40</td>
<td>0.073</td>
<td>0.105</td>
<td>0.057</td>
</tr>
<tr>
<td>50</td>
<td>0.078</td>
<td>0.118</td>
<td>0.061</td>
</tr>
<tr>
<td>75</td>
<td>0.072</td>
<td>0.125</td>
<td>0.061</td>
</tr>
<tr>
<td>100</td>
<td>0.079</td>
<td>0.114</td>
<td>0.052</td>
</tr>
<tr>
<td>125</td>
<td>0.078</td>
<td>0.122</td>
<td>0.049</td>
</tr>
<tr>
<td>150</td>
<td>0.068</td>
<td>0.118</td>
<td>0.051</td>
</tr>
</tbody>
</table>

Table 1: Size, bias and RMSE for $\hat{\lambda}_{BA,1}$ and $\hat{\lambda}_{RBA,1}$ when $\lambda_1 = 0$. 

$\lambda_{BA,1}$

$\lambda_{RBA,1}$
<table>
<thead>
<tr>
<th>Running/\lambda_1</th>
<th>T = 10, N runs</th>
<th>N = 10, T runs</th>
<th>N and T runs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25 0.5 0.75 1</td>
<td>0.25 0.5 0.75 1</td>
<td>0.25 0.5 0.75 1</td>
</tr>
<tr>
<td>5</td>
<td>(\hat{\lambda}_{BA,1})</td>
<td>(\hat{\lambda}_{RBA,1})</td>
<td>(\hat{\lambda}_{BA,1})</td>
</tr>
<tr>
<td>0.25</td>
<td>0.263 0.368 0.497 0.619</td>
<td>0.201 0.253 0.350 0.446</td>
<td>0.271 0.342 0.423 0.513</td>
</tr>
<tr>
<td>0.5</td>
<td>0.169 0.276 0.436 0.590</td>
<td>0.169 0.276 0.436 0.590</td>
<td>0.169 0.276 0.436 0.590</td>
</tr>
<tr>
<td>0.75</td>
<td>0.144 0.258 0.417 0.586</td>
<td>0.179 0.343 0.542 0.729</td>
<td>0.154 0.339 0.555 0.745</td>
</tr>
<tr>
<td>1</td>
<td>0.131 0.247 0.422 0.590</td>
<td>0.205 0.423 0.659 0.837</td>
<td>0.147 0.390 0.638 0.839</td>
</tr>
<tr>
<td>25</td>
<td>0.117 0.225 0.405 0.579</td>
<td>0.220 0.510 0.740 0.900</td>
<td>0.161 0.414 0.724 0.903</td>
</tr>
<tr>
<td>30</td>
<td>0.115 0.225 0.390 0.573</td>
<td>0.243 0.545 0.796 0.928</td>
<td>0.168 0.482 0.787 0.944</td>
</tr>
<tr>
<td>40</td>
<td>0.109 0.212 0.377 0.555</td>
<td>0.297 0.639 0.891 0.970</td>
<td>0.206 0.595 0.894 0.986</td>
</tr>
<tr>
<td>50</td>
<td>0.125 0.228 0.401 0.573</td>
<td>0.336 0.696 0.936 0.989</td>
<td>0.242 0.680 0.943 0.993</td>
</tr>
<tr>
<td>75</td>
<td>0.111 0.213 0.382 0.566</td>
<td>0.460 0.868 0.988 0.997</td>
<td>0.321 0.854 0.994 1</td>
</tr>
<tr>
<td>100</td>
<td>0.111 0.211 0.388 0.581</td>
<td>0.538 0.925 0.998 1</td>
<td>0.422 0.931 1 1</td>
</tr>
<tr>
<td>125</td>
<td>0.109 0.200 0.382 0.573</td>
<td>0.606 0.963 1 1</td>
<td>0.503 0.975 1 1</td>
</tr>
<tr>
<td>150</td>
<td>0.110 0.228 0.401 0.583</td>
<td>0.664 0.989 0.999 0.999</td>
<td>0.562 0.989 1 1</td>
</tr>
</tbody>
</table>

**Table 2:** Power for \(\hat{\lambda}_{BA,1}\) and \(\hat{\lambda}_{RBA,1}\) when \(\lambda_1 = 0.25, 0.5, 0.75, 1\).
Table 3: Size and power when $T < N^3$.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$N$</th>
<th>$0$</th>
<th>$0.25$</th>
<th>$0.5$</th>
<th>$0.75$</th>
<th>$1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\lambda}_{BA,1}$</td>
<td>1,000</td>
<td>5</td>
<td>0.700</td>
<td>0.999</td>
<td>0.997</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>7</td>
<td>0.479</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>9</td>
<td>0.330</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1,250</td>
<td>5</td>
<td>0.725</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>1,250</td>
<td>7</td>
<td>0.523</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1,250</td>
<td>9</td>
<td>0.338</td>
<td>0.999</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>5</td>
<td>0.762</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>7</td>
<td>0.562</td>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>9</td>
<td>0.400</td>
<td>0.999</td>
<td>1</td>
<td>0.999</td>
</tr>
</tbody>
</table>

| $\hat{\lambda}_{BA,1}$ | 1,000 | 5 | 0.107 | 0.961 | 0.974 | 0.981 | 0.987 |
| | 1,000 | 7 | 0.061 | 0.961 | 0.993 | 0.994 | 0.996 |
| | 1,000 | 9 | 0.033 | 0.973 | 0.995 | 0.996 | 0.996 |
| | 1,250 | 5 | 0.121 | 0.974 | 0.984 | 0.987 | 0.988 |
| | 1,250 | 7 | 0.044 | 0.978 | 0.990 | 0.993 | 0.993 |
| | 1,250 | 9 | 0.034 | 0.986 | 0.997 | 0.999 | 0.999 |
| | 1,500 | 5 | 0.154 | 0.979 | 0.984 | 0.989 | 0.991 |
| | 1,500 | 7 | 0.056 | 0.982 | 0.992 | 0.993 | 0.996 |
| | 1,500 | 9 | 0.046 | 0.986 | 0.996 | 0.997 | 0.998 |

Table 4: Descriptive statistics.

<table>
<thead>
<tr>
<th>Country</th>
<th>Exchange</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>OIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oman</td>
<td></td>
<td>87.919</td>
<td>22.617</td>
<td>86.7</td>
<td>141.2</td>
<td>35</td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td>85.489</td>
<td>20.071</td>
<td>85.745</td>
<td>145.29</td>
<td>33.98</td>
</tr>
<tr>
<td>UK</td>
<td></td>
<td>81.249</td>
<td>23.11</td>
<td>76.99</td>
<td>146.08</td>
<td>36.61</td>
</tr>
<tr>
<td>EQUITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>XETRA ARRB GR</td>
<td>14.715</td>
<td>5.331</td>
<td>12.535</td>
<td>27.940</td>
<td>8.523</td>
</tr>
<tr>
<td>UK</td>
<td>TOM MTF MT</td>
<td>14.782</td>
<td>5.479</td>
<td>12.535</td>
<td>29.270</td>
<td>8.417</td>
</tr>
<tr>
<td>UK</td>
<td>BATS Europe MTS</td>
<td>14.595</td>
<td>5.094</td>
<td>12.595</td>
<td>28.130</td>
<td>8.487</td>
</tr>
</tbody>
</table>

Notes: “SD” refers to the estimated standard deviation.
### Table 5: Unit root test results.

<table>
<thead>
<tr>
<th>Country</th>
<th>Test</th>
<th>Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OIL</td>
<td>Common ADF</td>
<td>-1.617</td>
<td>&gt;0.10</td>
</tr>
<tr>
<td></td>
<td>Idiosyncratic IPS</td>
<td>-1.346</td>
<td>0.089</td>
</tr>
<tr>
<td>EQUITY</td>
<td>Common ADF</td>
<td>-1.351</td>
<td>&gt;0.10</td>
</tr>
<tr>
<td></td>
<td>Idiosyncratic IPS</td>
<td>-7.445</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: “ADF” and “IPS” refer to the augmented Dickey– Fuller and Im– Pesaran–Shin unit root tests, respectively.

### Table 6: PIS results.

<table>
<thead>
<tr>
<th>Country</th>
<th>Exchange</th>
<th>$\hat{PIS}_i$</th>
<th>$\hat{\lambda}_{RBA,i}$</th>
<th>SE</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oman</td>
<td>XETRA ARRB GR</td>
<td>0.5097</td>
<td>0.9698</td>
<td>0.0360</td>
<td>26.9578</td>
<td>0.0000</td>
</tr>
<tr>
<td>USA</td>
<td>Luxembourg SE MT</td>
<td>0.0360</td>
<td>0.1126</td>
<td>0.0641</td>
<td>1.7555</td>
<td>0.0792</td>
</tr>
<tr>
<td>UK</td>
<td>NYSE Euronext Amsterdam MT</td>
<td>0.4543</td>
<td>1.0703</td>
<td>0.0395</td>
<td>27.0631</td>
<td>0.0000</td>
</tr>
<tr>
<td>Germany</td>
<td>Luxembourg SE MT</td>
<td>0.2244</td>
<td>0.9571</td>
<td>0.0198</td>
<td>48.2879</td>
<td>0.0000</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>Luxembourg SE MT</td>
<td>0.3039</td>
<td>1.0039</td>
<td>0.0179</td>
<td>56.1553</td>
<td>0.0000</td>
</tr>
<tr>
<td>Netherlands</td>
<td>NYSE Euronext Amsterdam MT</td>
<td>0.0522</td>
<td>1.0645</td>
<td>0.0137</td>
<td>77.5905</td>
<td>0.0000</td>
</tr>
<tr>
<td>UK</td>
<td>TOM MTF MT</td>
<td>0.0073</td>
<td>1.1293</td>
<td>0.0951</td>
<td>11.8703</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>BATS Europe MTS</td>
<td>0.0005</td>
<td>0.4958</td>
<td>0.0369</td>
<td>13.4304</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>BATS Europe MT</td>
<td>0.1049</td>
<td>1.0449</td>
<td>0.0134</td>
<td>78.1444</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Chi–X Europe MTS</td>
<td>0.0960</td>
<td>0.9455</td>
<td>0.0174</td>
<td>54.4109</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Chi–X Europe MTAS</td>
<td>0.1015</td>
<td>1.0457</td>
<td>0.0133</td>
<td>78.7682</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Turquoise MT</td>
<td>0.1093</td>
<td>1.0440</td>
<td>0.0134</td>
<td>78.1465</td>
<td>0.0000</td>
</tr>
</tbody>
</table>