DEAKIN UNIVERSITY AUSTRALIA

Financial Econometrics Series

SWP 2014/12

A Factor Analytical Approach to the Efficient Futures Market Hypothesis

J. Westerlund, M. Norkute, and P.K. Narayan

A FACTOR ANALYTICAL APPROACH TO THE EFFICIENT FUTURES MARKET HYPOTHESIS*

Joakim Westerlund† — Deakin University, Australia

Milda Norkute — Lund University, Sweden

Paresh Kumar Narayan — Deakin University, Australia

May 21, 2014

Abstract

Most empirical evidence suggests that the efficient futures market hypothesis, henceforth referred to as EFMH, stating that spot and futures prices should cointegrate with a unit slope on futures prices, does not hold, a finding at odds with many theoretical models. This paper argues that these results can be attributed in part to the low power of univariate tests, and that the use of panel data can generate more powerful tests. The current paper can be seen as a step in this direction. In particular, a newly developed factor analytical approach is employed, which is very general and, in addition, free of the otherwise so common incidental parameters bias in the presence of fixed effects. The approach is applied to a large panel covering 17 commodities between March 1991 and August 2012. The evidence suggests that the EFMH cannot be rejected once the panel evidence has been taken into account.

JEL Classification: C12; C13; C33; C36.

Keywords: Dynamic panel data models; Unit root; Factor analytical method; Efficient market hypothesis; Futures markets.

1 Introduction

The efficient market hypothesis (EMH) is based on the principle that asset prices reflect all publicly available information. Under the joint assumptions of risk neutrality and rational-

*A version of the paper was presented at a symposium on the Performance of Financial Markets and Credit Derivatives hosted by Deakin University. The authors would like to thank symposium participants, and in particular Mårten Nielsen and Robert Webb (Editor) for many valuable comments and suggestions.

†Corresponding author: Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia. Telephone: +61 3 924 46973. Fax: +61 3 924 46283. E-mail address: `j.westerlund@deakin.edu.au`.

ity, the expected returns to speculative activity in an efficient market should be zero. Thus, in a futures market, the current price of an asset for delivery at a specified date should be an unbiased predictor of the future spot rate. This is the efficient futures market hypothesis (EFMH). Despite the wide acceptance of this hypothesis in theory, however, the postulated long-run one-for-one relationship between spot and futures prices has proven very difficult to verify empirically. In fact, most studies tend to reject the EFMH (see Westerlund and Narayan, 2013, and the references provided therein).

A typical test of the EFMH involves taking the difference between the current spot price and past futures price (in logs), which is then subjected to a unit root test. The EFMH requires that the current spot price and past futures price are cointegrated with cointegrating vector $(1, -1)'$, which is tantamount to requiring that the difference between the two variables is stationary. There are basically two ways in which the weak evidence in favor of the EFMH (and also of the EMH) can be explained. One way is to take the empirical results for granted and modify the theoretical arguments. Fama (1970) argues that tests of the EMH are actually joint tests of market efficiency and the model of market equilibrium. There are two commonly accepted theories of commodity futures prices. These are the theory of storage, and the theory that futures prices consist of the expected future spot prices and an expected risk premium. Fama and French (1987) clarify that while the theory of storage is not controversial, there is little agreement on whether futures prices contain expected premiums. Persistent deviations between spot and future prices could therefore be due to the presence of such premiums (see Barkoulas et al., 2003, for a detailed discussion). The second explanation involves taking the theoretical arguments as given and instead focus on the econometric approach. In particular, it has been argued that the lack evidence in favor of the EFMH (and also of the EMH) can be attributed (at least partly) to the low power of the unit root test methodology employed (see, for example, Fama and French, 1987; Summers, 1986). This issue is discussed to some extent in Peroni and McNown (1998), who propose using more "informative" tests that are robust to possible endogeneity. However, as pointed out by Westerlund and Narayan (2013) such robustness corrections, while leading to tests with better size properties, are known to be quite costly in terms of power. Hence, accounting for endogeneity does not bring any more power into the analysis. As an alternative approach, Westerlund and Narayan (2013) suggest accounting for the information contained in the conditional heteroskedasticity of the data, which is shown, both analytically and via Monte

Carlo simulation, to lead to tests with higher power than that achievable using conventional tests that ignore this information (see also Westerlund and Narayan, 2012, 2014).

One drawback with the approach developed by Westerlund and Narayan (2013) is that it assumes that the time-variation is in the conditional variance only, which need not be the case in practice. Indeed, Pagan and Schwert (1990), Loretan and Phillips (1994), Watson (1999), and Busetti and Taylor (2003), to mention a few, all provide strong evidence against the conditional homoskedasticity assumption for most financial variables, including exchange rates, interest rates and stock returns. Hence, there is a need for procedures that are general enough to accommodate not only conditional but also unconditional heteroskedasticity.

Another source of information that is almost always there but is always ignored is that contained in the multiplicity of contracts/commodities usually considered (see, for example, Peroni and McNown, 1998; Westerlund and Narayan, 2013; Shawky et al., 2003). Put differently, while the data usually have a panel structure with multiple cross-section units, the econometric approach employed is a time series one, in which the testing is carried out in a unit-by-unity fashion. This means that when one unit is tested, the information contained in the other units is ignored. This is wasteful. Indeed, as is well-known, spot and futures prices tend to be highly correlated across contracts/commodities (see, for example, Barkoulas et al., 2003). Hence, ideally one would like to have a testing approach that takes full account not only of the unconditional heteroskedasticity, but also of the panel structure of the data. Such a strategy is likely to be more powerful, and thus more successful in evidencing the EFMH.

In this paper, we make an attempt in this direction by employing a dynamic panel approach, drawing upon a large data set covering 17 commodities over the March 1991–August 2012 period. In so doing, we pay special attention to the many features that characterize this type of financial data. For example, in addition to unconditional heteroskedasticity, given the high degree of dependence/heterogeneity that exists across both time and commodities, time- and commodity-specific fixed effects would seem to be necessary. Unfortunately, when taken together these considerations invalidate the use of most existing dynamic panel data approaches. The only exception known to us is the factor analytical (FA) approach of Bai (2013). As far as we are aware, this is the first application of FA and also one of the very first studies in the futures literature to consider a dynamic panel data approach to the EFMH. In fact, the only other dynamic panel data study known to us is that of Bernoth and von Hagen (2004), who consider the impact of the European central bank's policy announcements on the EFMH within the Euribor futures market.

The rest of the paper is organized as follows. Section 2 discusses the dynamic panel data model considered and puts it into context. The estimation of this model is discussed in Section 3. Section 4 reports the results from a small Monte Carlo simulation study. The empirical results are contained in Section 5. Section 6 concludes.

2 Model discussion

Consider the panel data variable $y_{i,t}$, observable for $t = 1, ..., T$ time series and $i = 1, ..., N$ cross-sectional units. In the empirical part, we consider four specifications of $y_{i,t}$, which are all designed to infer the EFMH by testing the joint hypothesis that spot and futures prices are cointegrated with cointegrating vector $(1, -1)'$. For example, in our first specification, denoted S1, $y_{i,t}$ is the log difference between the current spot and past futures prices. The following baseline dynamic panel data model can be seen as the kernel of all four specifications:

$$y_{i,t} = \mu_i + \delta_t + \rho y_{i,t-1} + \beta' x_{i,t-1} + \varepsilon_{i,t}, \tag{1}$$

where μ_i and δ_t are unit- and time-specific fixed effects, respectively, $x_{i,t-1}$ is a vector of predetermined regressors, and $\varepsilon_{i,t}$ is an error term that is assumed to be time and cross-section independent with $E(\varepsilon_{i,t}) = 0$, $E(\varepsilon_{i,t}^2) = \sigma_t^2 > 0$ and $E(\varepsilon_{i,t}^4) < \infty$.

Under the additional assumptions that $|\rho| < 1$ and $\sigma_t^2 = \sigma^2$ (1) represents what can only be described as the "classical" dynamic panel data model with unit-specific fixed effects, which has attracted considerable attention in the econometric literature. A major reason for this is the existence of the well-known incidental parameters bias, which arises because of an increasing number of fixed effects. In the micro panel setting with T small and N large this bias is rather devastating, as in this case ordinary least squares (OLS) is even inconsistent, which has in turn led to the development of alternative estimators. Prominent among these alternative estimators is generalized method of moments (GMM) (see, for example, Arellano and Bond, 1991; Arellano and Bover, 1995), which is now the most common approach in practical empirical work with dynamic panels. However, as noted by Blundell and Bond (1998), these estimators suffer from a weak instrument problem when ρ approaches unity, and when $\rho = 1$ the moment conditions are completely irrelevant. Allowing T to be large lessens the problem of bias; however, the asymptotic distributions of most known estimators are still miscentered, causing deceptive inference (see Moon et al., 2013a, Section 3.1.3).

The above discussion supposes that $\sigma_t^2 = \sigma^2$. If in addition σ_t^2 is unrestricted, then the estimation problem becomes even more complicated. In particular, unlike in standard regression analysis where heteroskedasticity is often a matter of efficiency rather than consistency, in dynamic panel data models omitted heteroskedasticity leads to inconsistency (see Bai, 2013). Also, while here we assume that $\varepsilon_{i,t}$ is cross-section homoskedastic, this is not necessary. Indeed, as Bai (2013) points out, if $\varepsilon_{i,t}$ is both time and cross-section heteroskedastic, then σ_t^2 simply represents the time-t cross-section average variance.

The above issue of bias recently motivated Bai (2013) to propose FA as an approach to the estimation of (1). The name stems from the fact that the estimator, which is based on quasi-maximum likelihood (quasi-ML), coincides with the one used in factor analysis (see, for example, Anderson and Amemiya, 1988). A key feature of FA is that it does not require estimation of the fixed effects themselves, but only estimation of the variance of μ_i, $S_\mu = \sum_{i=1}^N (\mu_i - \overline{\mu})^2/(N-1)$, where $\overline{\mu} = \sum_{i=1}^N \mu_i/N$. Since this is just a scalar the incidental parameter problem caused by the fixed effects is effectively removed, leading to an estimator that is completely bias-free. It is also instrumentation-free, which means that the difficulties associated with weak instruments and instrument proliferation in case of GMM do not arise.

Bai (2013) focuses on the classical setup with $|\rho| < 1$. This is enough in many applications; however, in the present case it is important to be able to allow also unit root behavior ($\rho = 1$). For example, if $y_{i,t}$ is the difference between the current spot price and past futures price, then $\rho = 1$ corresponds to a failure of the EFMH, and we do not want to assume a priori that the EFMH holds ($|\rho| < 1$). Fortunately, as Westerlund and Norkute (2014) show, FA can be extended to cover also the unit root case. Hence, in this study, the only requirement is that $\rho \in (-1, 1]$.

3 FA

Suppose for simplicity that $\beta = 0$, such that the predetermined regressors are absent from (1). It is useful to write the resulting model in matrix form, and for this purpose we need the following $T \times 1$ vectors: $y_i = (y_{i,1}, ..., y_{i,T})'$, $\varepsilon_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$, $\delta = (\delta_1, ..., \delta_T)'$ and $1_T =$

$(1, ..., 1)'$. The following matrices are also needed:

$$L = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho & 1 & 0 & \dots & 0 \\ \rho^2 & \rho & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-1} & \dots & \rho^2 & \rho & 1 \end{bmatrix},$$

which are both $T \times T$. Note how L and Γ can be seen as "lag" and "accumulation" matrices, respectively; the t-th row of Ly_i and Γy_i are given by $y_{i,t-1}$ and $\sum_{n=1}^{t} \rho^{n-1} y_{i,n}$, respectively. In this notation,

$$y_i = 1_T \mu_i + \delta + \rho L y_i + \varepsilon_i, \tag{2}$$

which can be solved for y_i, giving

$$y_i = \Gamma 1_T \mu_i + \Gamma \delta + \Gamma \varepsilon_i, \tag{3}$$

where Γ is related to ρ and L via $\Gamma = (I_T - \rho L)^{-1}$. This is the model considered in Bai (2013). However, he assumes that $|\rho| < 1$, which we have already argued need not be the case in practice, especially not the case of the EFMH. A complicating factor of allowing $\rho = 1$ is that the meaning of μ_i and δ changes. This is easily appreciated by considering the t-th row of y_i;

$$y_{i,t} = \mu_i t + \sum_{n=1}^{t} \delta_n + \sum_{n=1}^{t} \varepsilon_{i,n}.$$

Hence, what under $|\rho| < 1$ represent unit-specific fixed effects $(\mu_1, ..., \mu_N)$ are under $\rho = 1$ unit specific trend slopes. Therefore, in order to prevent such changes in the meaning of the model parameters, in what follows we consider the following modified version of (2):

$$y_i = \Gamma^{-1} 1_T \mu_i + \Gamma^{-1} \delta + \rho L y_i + \varepsilon_i, \tag{4}$$

such that

$$y_i = 1_T \mu_i + \delta + \Gamma \varepsilon_i. \tag{5}$$

The premultiplication of Γ^{-1} is under $\rho = 1$ tantamount to first-differencing, which removes the accumulation that occurs under $\rho = 1$. The model in (4) can be seen as emanating from the following components model, which is very common in the unit root literature (see Schmidt and Phillips, 1992; Westerlund and Breitung, 2013, for discussions): $y_i = 1_T \mu_i + \delta + s_i$, where $s_i = \rho L s_i + \varepsilon_i$.

We now describe the FA estimator of (4). The parameter vector of interest is given by $\theta = (S_\mu, \rho, \sigma_1^2, ..., \sigma_T^2)'$. Note how δ is not included. The reason for this is that since we are not really interested in making inference regarding this parameter vector anyways, we can remove it by taking differences from the cross-section average; $y_i - \overline{y} = 1_T(\mu_i - \overline{\mu}) + \Gamma(\varepsilon_i - \overline{\varepsilon})$, where $\overline{y} = \sum_{i=1}^N y_i/N$ with similar definitions of $\overline{\mu}$ and $\overline{\varepsilon}$. Let $S_y = \sum_{i=1}^N (y_i - \overline{y})(y_i - \overline{y})'/(N-1)$, which is $T \times T$. Under the assumptions placed on $\varepsilon_{i,t}$, it is not difficult to show that

$$E(S_y) = \Sigma(\theta) = 1_T 1_T' S_\mu + \Gamma\Phi\Gamma'^{-1} 1_T 1_T'^{-1'} S_\mu + \Phi)\Gamma'. \tag{6}$$

where $\Phi = \mathrm{diag}(\sigma_1^2, ..., \sigma_T^2)$ is $T \times T$ and S_μ is as in Section 2. The model in (5) can be seen as a common factor model with factor 1_T and loading μ_i, suggesting that the estimation can be carried out using methods designed for such models. This motivated Bai (2013) to consider the following "discrepancy function":

$$Q(\theta) = \log(|\Sigma(\theta)|) + \mathrm{tr}\,(S_y \Sigma(\theta)^{-1}), \tag{7}$$

which is often used in classical factor analysis (see, for example, Anderson and Amemiya, 1988). The FA estimator $\hat{\theta}$ of θ is the minimizer of $Q(\theta)$. Hence, viewing $Q(\theta)$ as a distance measure between S_y and $\Sigma(\theta)$, $\hat{\theta}$ is essentially a moment matching estimator based on the variance of $y_i - \overline{y}$. This is important because while the mean of $y_i - \overline{y}$ obviously depends on $\mu_1, ..., \mu_N$, the variance only depends on S_μ, which has the same dimension regardless of the value taken by N. Hence, unlike estimators based on the mean, such as OLS, in FA the unit-specific fixed effects are not incidental parameters. Of course, the same cannot be said about the error variances, $\sigma_1^2, ..., \sigma_T^2$, which obviously increase in number as T increases. Fortunately, as Bai (2013) shows, the estimation of these variances do not lead to an incidental parameters bias.

While Bai (2013) focuses on the model in (2), under $|\rho| < 1$ the modification in (4) does not affect the results. Let us therefore denote by $\hat{\rho}_{FA}$ the FA estimator of ρ. As Bai (2013) shows, as $N, T \to \infty$ with $N/T^3 \to 0$, provided that $|\rho| < 1$,

$$\sqrt{NT}(\hat{\rho}_{FA} - \rho) \to_d N(0, 1/\gamma), \tag{8}$$

where $\to_d$ signifies convergence in distribution and

$$\gamma = \lim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \sum_{n=1}^{t-1} \rho^{2(t-n-1)} \sigma_t^{-2} \sigma_n^2$$

Hence, in contrast to most existing estimators, with FA there is no asymptotic bias. The estimator is even efficient. Note also that under homoskedasticity, $\gamma = 1/(1-\rho^2)$, which is the same as for the OLS fixed effects estimator (see Hahn and Kuersteiner, 2002). In fact, the same is true even for the time series OLS estimator. The difference is the rate of consistency, which in the times series case is only $\sqrt{T} \leq \sqrt{NT}$ for all $N \geq 1$. Hence, since in this paper N is assumed to be large, the panel FA and OLS estimators are infinitely more efficient than times series OLS. In Section 4 we use Monte Carlo simulation to elaborate on this issue.

The result in (8) requires that $|\rho| < 1$. If $\rho = 1$, the above results change. Westerlund and Norkute (2014) only consider the case with homoskedasticity; however, their results can be easily extended to cover also the present more general setting. The relevant asymptotic distribution is given in Theorem 1.

Theorem 1. *Under $\rho = 1$ and the conditions laid out in the above, as $N, T \to \infty$ with $N/T^2 \to 0$,*

$$\sqrt{N}T(\hat{\rho}_{FA} - \rho) \to_d N(0, T/\gamma).$$

Proof: The proof of Theorem 1 follows by simple manipulations of the proof of Theorem 3 in Westerlund and Norkute (2014). It is therefore omitted. ■

The main difference between (8) and the result given in Theorem 1 is the rate of convergence, which is higher in Theorem 1. This is due to the fact that in this case $y_{i,t}$ has a unit root and therefore the estimator of ρ is "superconsistent", as is to be expected given the large literature on non-stationary panels (see Breitung and Pesaran, 2008; Baltagi, 2008, Chapter 12, for surveys). The value of ρ is therefore easier to discern if $\rho = 1$ than if $|\rho| < 1$. Another difference is the condition placed on the relative expansion rate of N and T, which is relatively stronger in Theorem 1. The reason for this is again the unit root property of the data, which leads to increased variance. The assumption that $N/T^2 \to 0$ prevents this variance from having a dominating effect. In practice this means that $T^2 > N$, which is not very restrictive. In our data set, $N = 17$ and $T = 258$, suggesting that $N/T^2 = 17/66\,564 \approx 0.0003$, which is a very small number indeed.

The second thing to note about Theorem 1 is that the variance has exactly the same form as in (8), except for the scaling by T, which is due to the increased rate of consistency under $\rho = 1$. In particular, note how under homoskedasticity, γ/T reduces to

$$\frac{1}{T}\gamma = \lim_{T\to\infty} \frac{1}{T^2} \sum_{t=2}^{T} (t-1) = \frac{1}{2},$$

which implies that the variance of $\sqrt{N}T(\hat{\rho}_{FA} - \rho)$ is given by $T/\gamma = 2$, which is again the same as for OLS without fixed effects (see, for example, Levin and Lin, 1992, Theorem 3.2). In the present case with fixed effects the variance of the OLS estimator of ρ is $51/5 \approx 10 > 2$ (see Hahn and Kuersteiner, 2002). Hence, the use of FA leads to a quite substantial reduction in variance when compared to OLS. Another difference is that, in contrast to FA, OLS is biased. Indeed, as Hahn and Kuersteiner (2002) show,

$$\sqrt{N}T(\hat{\rho}_{OLS} - \rho) + 3\sqrt{N} \to_d N(0, 51/5),$$

where $\hat{\rho}_{OLS}$ is the fixed effects OLS estimator of ρ. This result suggests that without bias correction, the OLS estimator is actually only T-consistent, which is the same rate as for the times series estimator.

The fact that the FA estimator is not only unbiased but also supports asymptotically normal inference for all values of ρ, including unity, makes it unique. In fact, the only other estimator with this property is the GMM estimator of Phillips and Han (2010). However, in contrast to FA, which is $\sqrt{N}T$-consistent, in the unit root case the estimator of Phillips and Han (2010) is only $\sqrt{NT}$-consistent. FA is therefore superior in this regard.

4 Monte Carlo study

A small-scale Monte Carlo simulation exercise was carried out to demonstrate the advantage of using FA when compared to more conventional pooled and time series (unit-by-unit) OLS. The data generating process is given by a simplified version of (4) that sets $\mu_i \sim U(-1, 1)$. Time series OLS can accommodate fixed effects that are unit-specific, but not fixed effects that are time-specific, which is just another way of saying that it is unable to explore the common variation in the data. In applications to spot and futures price data such variation is likely to be the rule rather than the exception, and in this section we therefore consider $\delta_t \sim U(-1, 1)$.[1] The allowance of heteroskedasticity is extremely time consuming and in fact not possible unless one is considering a single sample. In this section, we therefore assume that $\varepsilon_{i,t} \sim N(0, 1)$. In Section 5 we show how to implement the FA estimator in the presence of heteroskedasticity.

A large number of results were produced; however, in interest of space we focus on

[1]We also considered the case when $\delta_1 = ... = \delta_T = 0$. However, since the conclusions were not affected by this, in the paper we focus on the more realistic case when time effects are present.

the bias, absolute bias and root mean squared error (RMSE) of $\hat{\rho}_{FA}$, and the size and size-adjusted of a nominal 5% level t-test when the hypothesis is formulated as $H_0 : \rho = 1$. The number of replications is set to 1,000. The data set that we consider in the empirical part has $N = 17$ and $T = 258$, and we also consider subsamples with $T = 67$ and $T = 191$. In view of this, in the simulations we set $N \in \{10, 20\}$ and $T \in \{50, 100, 200\}$. All computational work was done in GAUSS 11. In implementing FA we used the BFGS algorithm for constrained optimization with non-negativity constraints on estimated variances. The standard errors are obtained from the inverse of the Hessian matrix evaluated at the estimated parameters. The results of the time series OLS estimator are averaged over the cross-section.

The results are reported in Table 1. The first thing to note is that the bias is generally much smaller for FA than for OLS. In fact, the pooled (time series) OLS bias can be as large as 200 (300) times the corresponding FA bias, and it is never smaller than eight (14) times the FA bias. The difference in RMSE is less dramatic; however, it is still sizable, especially when $\rho = 1$, in which case the RMSE of pooled (time series) OLS is never smaller than six (11) times that of FA. The RMSE of all three estimators should be decreasing in T; however, it is only for FA and pooled OLS that RMSE should decreasing also in N. The results reported in Table 1 are quite suggestive of this. We also see that the RMSE is generally lower when $\rho = 1$ than when $|\rho| < 1$, which is a reflection of the relatively high rate of consistency in this case. There is a big difference in size accuracy with the OLS-based t-tests exhibiting substantial distortions and FA having the best size accuracy by far. There is also a huge difference in terms of (size-adjusted) power with FA being uniformly more powerful then the OLS-based tests.

All-in-all, we find that the FA estimator leads to a substantial improvement when compared to both pooled and time series OLS. Since such differences are likely to be very important in practice in the empirical part we will only consider FA.[2]

[2]Some of the specifications considered in the empirical part include predetermined regressors. We therefore generated simulation results also for such models. The results were, however, almost identical to those reported in Table 1. We therefore omit them.

5 Empirical results

5.1 Data

Our empirical analysis is based on the following 17 commodities: soybeans, wheat, corn, cocoa, silver, soybean meal, copper, coffee, sugar, soybean oil, platinum, palladium, gold, cotton, canola, (west Texas intermediate) crude oil, and natural gas. The data are monthly and span the period March 1991 to August 2012. We also consider sub-samples, motivated by the global financial crisis: the pre-crisis sample covers the period from March 1991 to January 2007, while the post-crisis sample covers February 2007–August 2012 period. The full sample period has a total of $T = 258$ observations, while for the pre- and post-crisis samples we have 191 and 67 time series observations, respectively. All data are downloaded from the Commodity Research Bureau database. Both spot and futures prices are transformed by taking logs.

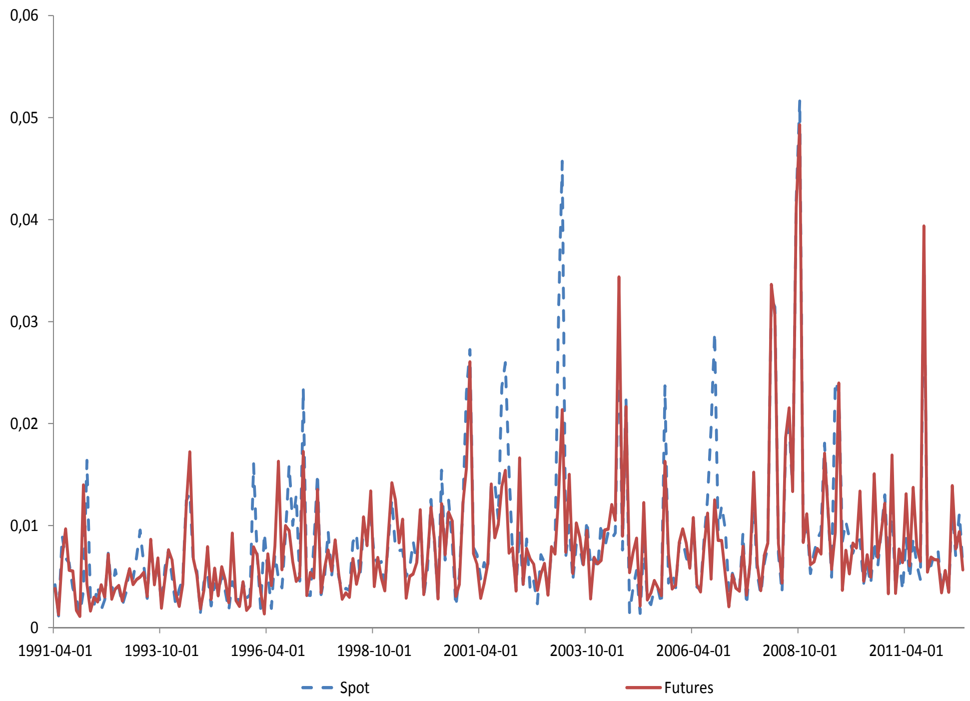
5.2 Preliminary results

Before we report the results from applying the FA approach we examine the extent of cross-section dependence and heteroskedasticity in the data.

In order to infer the significance of the cross-section correlation problem, we compute the pair-wise correlation coefficients of the spot and futures data. In particular, to gauge against possible unit roots the correlations are based on the first-differenced rather than the level data. The simple average of these correlations across all pairs of commodities, together with the associated CD test discussed in Pesaran et al. (2008), are given in Table 2. The average correlation coefficient ranges between 0.21 and 0.23, and the CD statistic is highly significant for both variables, which we take as strong evidence of cross-section dependence. In view of this, in the implementation of FA we will focus on the specification with common time effects included. To illustrate the impact of allowing for common time effects, Table 2 also reports some results based on the cross-section demeaned first-differenced data. As we can see, while the CD test is still significant, the average correlation is substantially reduced by the demeaning, from about 0.2 to about -0.05.

As a second preliminary we examine the extent of time series heteroskedasticity in the data. In Figure 1 we plot the cross-section variance of the first-differenced spot and futures prices for each point in time, where the first-differencing is again performed to gauge against possible unit roots. The first thing to note is that the variance is clearly not constant and that there are clusters of high/low variance observations. We also see that the variance profiles of the two variables are very similar. In fact, most of the time the two lines are almost on top of each other. In order to formally test the hypothesis of variance constancy we apply the partial sum of squares Q-statistic considered by, for example, Perron (2006). The test values for the spot and futures variables are 2.13 and 2.74, respectively. The appropriate right tail critical value at the conservative 1% level is 0.74, suggesting that the null hypothesis of constant variance can be easily rejected for both variables. Hence, as expected, the presence of heteroskedasticity cannot be ignored.

Figure 1: Cross-section variances for each point in time.



Note: The cross-section variances are for the first-differenced spot and futures prices.

The above results are relevant not only for the implementation of FA, but also for what they imply for the results reported in existing studies. Bernoth and von Hagen (2004) employ data on spot and multiple futures prices to test the EFMH in the Euribor market. They consider a dynamic panel data model that is similar to (1) but without common time effects, which is estimated by OLS with panel-corrected standard errors. However, while robust against cross-section dependence, the standard errors cannot handle heteroskedasticity over time and they are unsuitable in general for use in dynamic models (see Kristensen and Wawro, 2003). One reason for this is the incidental parameters bias caused by the unit-specific effects, a problem that is made worse by taking ρ closer to unity (see Moon et al., 2013). The FA approach considered here accounts for heteroskedasticity and is unbiased for all $\rho \in (-1, 1]$. It is therefore expected to lead to more reliable results.

5.3 Main results

Having considered briefly the heteroskedasticity and cross-section correlation properties of the data, we now turn to the test for predictability. Since both features seem important in this section we focus on the model considered in Sections 2 and 3 that allows both common time effects and heteroskedasticity, although results for the time series homoskedastic and unit-specific effects-only model are also reported for comparison. Denote by $s_{i,t}$ and $f_{i,t}$ the time-t log spot and futures prices, respectively. The particular model specifications that we will consider are as follows:

S1. $y_{i,t} = s_{i,t} - f_{i,t-1}$ and $\beta = 0$;

S2. $y_{i,t} = s_{i,t} - f_{i,t}$ and $\beta = 0$;

S3. $y_{i,t} = \Delta s_{i,t}$, $\rho = 0$ and $x_{i,t-1} = s_{i,t-1} - f_{i,t-1}$;

S4. $y_{i,t} = \Delta s_{i,t}$, $\rho = 0$ and $x_{i,t-1} = [(s_{i,t-1} - f_{i,t-2}), \Delta f_{i,t-1}]'$.

As mentioned in Section 1, the EFMH implies that $s_{i,t}$ and $f_{i,t-1}$ should be cointegrated with a unit slope on the latter variable. S1–S2 are designed to test the null hypothesis of no cointegration versus the alternative of cointegration. In S1, which can be seen as a unit root test regression for $s_{i,t} - f_{i,t-1}$, this is done by testing $\rho = 1$ versus $|\rho| < 1$. As Zivot (2000) points out, if $s_{i,t}$ and $f_{i,t-1}$ are cointegrated, then $s_{i,t}$ and $f_{i,t}$ should be cointegrated too. The purpose of S2 is to test this. S3 and S4 are error-correction models, so the appropriate restriction to test here is that the coefficient on $s_{i,t-1} - f_{i,t-1}$ and/or $s_{i,t-1} - f_{i,t-2}$ is zero. As with S1 and S2, S3 and S4 are very similar and are in fact equivalent representations. Hence, if there is error-correction in S3, then there should be error-correction also in S4, and vice versa.

Specifications S1–S4 are estimated while adjusting the discrepancy function in (7) accordingly (see Bai, 2013, for details). As in Bai (2013) the discrepancy function is made conditional on the initial observation, $y_{i,0}$. By doing so, we obtain an estimator that is robust to arbitrary initializations, which is a great advantage, as we have little or no priors regarding the initial values of spot and futures prices. The fixed-effects and the regressors in $x_{i,t-1}$ are correlated. To control for this we apply the Mundlak–Chamberlain projection. In our case, T is relatively large, which makes it possible to use a relatively simple version of the projection where the fixed-effects are projected on the regressors averaged over time (rather than on each time period; see Bai 2013). The numerical optimization of the objective function is carried out exactly as described previously in Section 4.

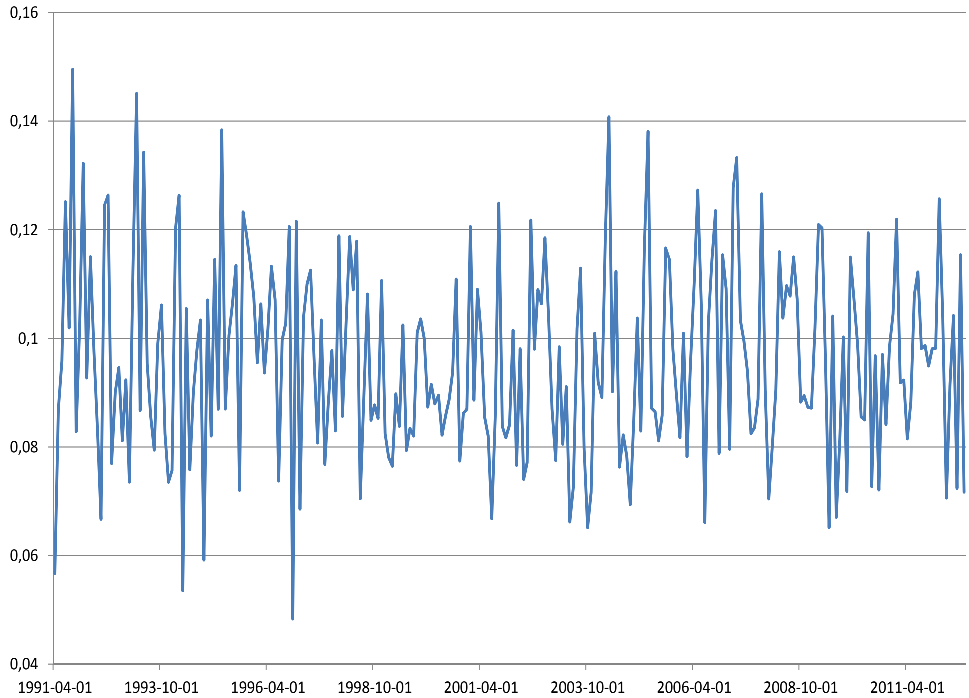
The results are reported in Table 3. We begin by considering the full-sample results for S1 and S2. We see that while the estimates of ρ are quite high, especially for S2, they are still far from one. This observation is supported by the reported t-statistics for the unit root restriction ($\rho = 1$), which are all highly significant, suggesting that in these cases $y_{i,t}$ is persistent but still mean reverting. This means we cannot reject that $s_{i,t}$ and $f_{i,t}$ or $f_{i,t-1}$ are cointegrated with cointegrating vector $(1, -1)'$, which we take as evidence in favor of the EFMH.

Consider next the (full-sample) results reported for S3 and S4. The estimated coefficients for $s_{i,t-1} - f_{i,t-1}$ and $s_{i,t-1} - f_{i,t-2}$ are well above zero and very similar in magnitude, as to be expected. The t-statistics for these coefficients are significant, but only marginally, at the 5% level or better for S4 and at the 10% for S3. Hence, while not as overwhelming as for S1 and S2, the evidence for S3 and S4 is still in favor of error-correction and hence of the EFMH.

According to theory, under the EFMH the fixed effects should all be zero. As a by-product of the FA procedure, we obtain estimates of the variance of the fixed effects, S_{μ}. These are very close to zero for all specifications considered and in fact never go above 0.001. In order to infer also the mean we computed the averages of the OLS fixed effects estimates. As expected, the averages were very close to zero, which we interpret as evidence in favor of the EFMH. This finding is in agreement with the results of Bernoth and von Hagen (2004) for the Euribor market.

The subsample results are generally in agreement with the overall conclusion for the full sample, that is, the evidence tends to favor of the EFMH. One important difference is the estimated coefficients which differ quite markedly across the two subsamples. In particular, while for S1 and S2 the estimated coefficients are relatively larger in the post-crisis subsample, for S3 and S4 it is the other way around, that is, the estimates are larger in the pre-crisis

Figure 2: Estimated variances for S1 at each point in time.



subsample. While this could be taken as a sign of model instability, we argue that it should not. In particular, as is well-known, the serial correlation properties of a stationary variable subject to structural change are akin to those of a random walk, and therefore tests for a unit root tend to have poor power against alternatives that are breaking but otherwise stationary. Hence, if the relationship had indeed undergone a structural change as a consequence of the global financial crisis, then the full sample tests should have rejected cointegration/error-correction, which is not what we observe. The observed differences are therefore more likely to be due to sample variation and/or the smallness of T, especially in the post-break subsample.

In order to infer the effect of common time effects and heteroskedasticity over time, Table 4 reports results obtained by assuming that these features are absent. The main result here is that while the coefficient estimates are very similar to those obtained in the presence of heteroskedasticity and common time effects, the standard errors are now much larger. The effect of this is to overturn the previous conclusions. That is, while the evidence under heteroskedasticity and common time effects tend to favor cointegration/error-correction, if these features are ignored, then the evidence goes in the other direction. This illustrates the importance of accounting for heteroskedasticity and time effects, which we have seen are present in the data. To get a feeling for the extent of the heteroskedasticity in the regression errors, Figure 2 plots as an example the estimated variances for S1. Consistent with the evidence reported in Figure 1 we see that the variance is not constant.

The above results are based on the model in (4). As explained in Section 3, the difference between this model and the one in (2) is that under $\rho = 1$ the fixed effects are restricted in (4) but not in (2). But this is the only difference. Indeed, under $|\rho| < 1$, the two models are indistinguishable. This fact can be used as a robustness check of the above findings; the EFMH holds, then the results obtained by fitting (2) should be very similar to those obtained by fitting (4). Our unreported results confirm this. In fact, the results are almost identical. The largest difference (in percentage terms) is obtained for S1 where the estimate of ρ is reduced from 0.244 to 0.210, which in absolute terms is really quite marginal.

6 Conclusion

This paper is about testing whether commodity markets satisfy the EFMH. The thrust of the paper is the econometric approach used to test this hypothesis, which is different from the otherwise so common time series unit root and cointegration approaches. In particular, a dynamic panel data approach is proposed. The advantage of such a panel approach, which seems to have gone largely unnoticed in the futures markets literature, is the increased number of observations that can be brought to bear on the EFMH. This is expected to lead to tests with higher precision when compared to more conventional time series tests, a result that is verified using Monte Carlo simulation. However, the particular panel approach considered here, which is based on the recently proposed FA estimator of Bai (2013), is superior not only when compared to time series approaches, but also when compared against existing panel approaches. Indeed, while existing panel data estimators are known to suffer from bias due to heteroskedasticity and heterogeneity, FA is completely bias-free. It is also unique in that it supports asymptotically normal inference for all values of ρ, including unity. This last feature of FA is particularly important when testing the EFMH, because if the hypothesis fails then the difference between the spot and futures prices is expected to follow a random walk (provided that spot and futures prices are themselves random walk processes). In view of these advantages, it is not surprising that when FA is applied to a panel of 17 commodities, we find strong support for the EFMH.

References

Anderson, T. W., and Y. Amemiya (1988). The Asymptotic Normal Distribution of Estimators in Factor Analysis Under General Conditions. *Annals of Statistics* **16**, 759–771.

Arellano, M., and S. Bond (1991). Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58**, 277–297.

Arellano, M., and O. Bover (1995). Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics* **68**, 29–51.

Bai, J. (2013). Fixed-Effects Dynamic Panel Models, a Factor Analytical Method. *Econometrica* **81**, 285–314.

Baltagi, B. (2008). *Econometric Analysis of Panel Data, Fourth Edition*. John Wiley and Sons, New York.

Barkoulas, J., C. F. Baum and A. Chakraborty (2003). Forward Premiums and Market Efficiency: Panel Unit-Root Evidence from the Term Structure of Forward Premiums. *Journal of Macroeconomics* **25**, 109–122.

Bernoth, K., and J. von Hagen (2004). The Euribor Futures Market: Efficiency and the Impact of ECB Policy Announcements. *International Finance* **7**, 1–24.

Blundell, R., and S. Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* **87**, 115–143.

Breitung J., and M. H. Pesaran (2008). Unit Roots and Cointegration in Panels. In Matyas, L., and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Klüwer Academic Publishers, 279–322.

Bun, M., and V. Sarafidis (2013). Dynamic Panel Data Models. Forthcoming in Baltagi, B. (Ed.), *Oxford Handbook on Panel Data*, Oxford University Press.

Busetti, F., and A. M. R. Taylor (2003). Variance Shifts, Structural Breaks and Stationarity Tests. *Journal of Business & Economic Statistics* **21**, 510–531.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* **25**, 383–417.

Fama, E. F., and K. R. French (1987). Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *Journal of Business* **60**, 55–73.

Hahn, J., and G. Kuersteiner (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T are Large. *Econometrica* **70**, 1639–1657.

Han, C., and P. C. B. Phillips (2010). GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity. *Econometric Theory* **26**, 119–151.

Kristensen, I., and G. Wawro (2003). Lagging the Dog?: The Robustness of Panel Corrected Standard Errors in the Presence of Serial Correlation and Observation Specific Effects. Unpublished manuscript.

Levin, A., and C. Lin (1992). Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. University of California working paper 92–23.

Loretan, M., and P. C. B. Phillips (1994). Testing Covariance Stationarity under Moment Condition Failure with an Application to Common Stock Returns. *Journal of Empirical Finance* **1**, 211–248.

Moon, H. R., B. Perron and P. C. B. Phillips (2013). Incidental Parameters and Dynamic Panel Modeling. Forthcoming in Baltagi, B. (Ed.), *Oxford Handbook on Panel Data*, Oxford University Press.

Pagan, A. R., and G. W. Schwert (1990). Testing for Covariance Stationarity in Stock Market Data. *Economics Letters* **33**, 165–170.

Peroni, E., and R. F. McNown (1998). Non Informative and Informative Tests of Efficiency in three Energy Futures Markets. *Journal of Futures Markets* **18**, 936–964

Perron, P. (2006). Dealing with Structural Breaks. In Patterson, K., and T. C. Mills (Eds.), *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, Palgrave Macmillan, 278–352.

Pesaran, H. M., A. Ullah and T. Yamagata (2008). A Bias-Adjusted LM Test Of Error Cross Section Independence. *Econometrics Journal* **11**, 105–127.

Schmidt, P. and P. C. B. Phillips (1992). LM Tests for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics* **54**, 257–287.

Shawky, H. A., A. Marathe, and C. L. Barrett (2003). A First Look at the Empirical Relation Between Spot and Futures Electricity Prices in the United States. *Journal of Futures Markets* **23**, 931–955.

Summers, L.H. (1986). Does the Stock Market Rationally Reflect Fundamental Values? *Journal of Finance* **41**, 591-601.

Watson, M. W. (1999). Explaining the Increased Variability in Long-Term Interest Rates. *Federal Reserve Bank of Richmond, Economic Quarterly* **85**, 71–96.

Westerlund, J., and P. Narayan (2012). Does the Choice of Estimator Matter when Forecasting Returns? *Journal of Banking and Finance* **36**, 2632–2640.

Westerlund, J., and P. Narayan (2013). Testing the Efficient Market Hypothesis in Conditionally Heteroskedastic Futures Markets. *Journal of Futures Markets* **33**, 1024–1045.

Westerlund, J., and P. Narayan (2014). Testing for Predictability in Conditionally Heteroskedastic Stock Returns. Forthcoming in *Journal of Financial Econometrics*.

Westerlund, J., and M. Norkute (2014). A Factor Analytical Approach to Interactive Effects Dynamic Panel Models with or without Unit Root. Unpublished manuscript.

Westerlund, J., and J. Breitung (2013). Lessons from a Decade of IPS and LLC. *Econometric Reviews* **32**, 547–591.

Zivot, E. (2000). Cointegration and Forward and Spot Exchange Rate Efficiency. *Journal of International Money and Finance* **19**, 785–812.

Table 1: Monte Carlo results.

		FA				Pooled OLS				Time series OLS			
N	T	Bias	Abs bias	RMSE	Rej	Bias	Abs bias	RMSE	Rej	Bias	Abs bias	RMSE	Rej
						$\rho = 1$							
10	50	−0.0034	0.0034	0.0119	5.8	−0.0639	0.0639	0.0674	98.6	−0.1031	0.1031	0.1321	30.2
10	100	−0.0018	0.0018	0.0059	6.6	−0.0321	0.0321	0.0340	98.8	−0.0526	0.0526	0.0684	29.8
10	200	−0.0008	0.0008	0.0029	6.9	−0.0162	0.0162	0.0171	97.7	−0.0266	0.0266	0.0345	30.3
20	50	−0.0018	0.0018	0.0072	5.2	−0.0609	0.0609	0.0624	100.0	−0.1030	0.1030	0.1326	29.9
20	100	−0.0008	0.0008	0.0038	6.7	−0.0309	0.0309	0.0318	100.0	−0.0524	0.0524	0.0678	29.9
20	200	−0.0004	0.0004	0.0018	6.2	−0.0157	0.0157	0.0162	100.0	−0.0265	0.0265	0.0342	30.5
						$\rho = 0.99$							
10	50	−0.0038	0.0038	0.0134	6.0	−0.0666	0.0666	0.0704	5.2	−0.1048	0.1048	0.1343	2.7
10	100	−0.0020	0.0020	0.0073	24.2	−0.0335	0.0335	0.0357	8.0	−0.0532	0.0532	0.0700	3.1
10	200	−0.0010	0.0010	0.0043	77.6	−0.0161	0.0161	0.0174	16.5	−0.0264	0.0264	0.0358	3.4
20	50	−0.0019	0.0019	0.0082	19.2	−0.0634	0.0634	0.0651	6.7	−0.1042	0.1042	0.1344	2.6
20	100	−0.0009	0.0009	0.0048	49.7	−0.0323	0.0323	0.0334	14.7	−0.0531	0.0531	0.0696	3.0
20	200	−0.0005	0.0005	0.0028	99.2	−0.0157	0.0157	0.0163	34.1	−0.0262	0.0262	0.0353	3.4
						$\rho = 0.95$							
10	50	−0.0034	0.0034	0.0184	85.2	−0.0615	0.0615	0.0669	22.4	−0.0987	0.0987	0.1344	4.0
10	100	−0.0003	0.0003	0.0120	100.0	−0.0271	0.0271	0.0312	68.5	−0.0458	0.0458	0.0687	6.1
10	200	0.0007	0.0007	0.0078	100.0	−0.0120	0.0120	0.0149	100.0	−0.0214	0.0214	0.0368	17.6
20	50	−0.0003	0.0003	0.0116	100.0	−0.0576	0.0576	0.0602	44.1	−0.0963	0.0963	0.1323	3.6
20	100	0.0011	0.0011	0.0079	100.0	−0.0258	0.0258	0.0277	97.0	−0.0458	0.0458	0.0687	6.1
20	200	0.0015	0.0015	0.0057	100.0	−0.0115	0.0115	0.0131	100.0	−0.0214	0.0214	0.0367	18.4

Notes: "OLS" and "RMSE" refer to the times series OLS estimator and root mean squared error, respectively. "Rej" refers to the rejection frequency for a nominal 5% level test when the null hypothesis is given by $H_0 : \rho = 1$. Therefore, when $\rho = 1$, "Rej" represents size, whereas when $\rho \neq 1$, then "Rej" represents size-adjusted power.

Table 2: Cross-correlation test results.

Variable	Time effects?	Correlation	CD	p-value
Spot prices	No	0.214	40.138	0.000
Futures prices	No	0.230	43.118	0.000
Spot prices	Yes	-0.049	-9.163	0.000
Futures prices	Yes	-0.053	-10.004	0.000

Notes: "CD" refers to the Pesaran et al. (2008) test of the null hypothesis of no cross-section correlation. The reported correlations are the average of the pair-wise correlation coefficients. The spot and futures prices have transformed by taking first differences.

Table 3: Estimation results.

Specification	Coefficient	Estimate	SE	$t(0)$	p-value	$t(1)$	p-value
			Full sample				
S1	ρ	0.210	0.044	4.737	0.000	-17.803	0.000
S2	ρ	0.701	0.099	7.044	0.000	-3.008	0.003
S3	β	0.199	0.111	1.802	0.072	-7.246	0.000
S4	β_1	0.197	0.078	2.507	0.012	-10.239	0.000
	β_2	0.223	0.091	2.439	0.015	-8.518	0.000
			Pre-crisis subsample				
S1	ρ	0.193	0.031	6.234	0.000	-25.995	0.000
S2	ρ	0.702	0.075	9.322	0.000	-3.963	0.000
S3	β	0.232	0.059	3.905	0.000	-12.933	0.000
S4	β_1	0.205	0.060	3.414	0.001	-13.240	0.000
	β_2	0.196	0.068	2.859	0.004	-11.762	0.000
			Post-crisis subsample				
S1	ρ	0.364	0.030	11.942	0.000	-20.872	0.000
S2	ρ	0.878	0.017	52.294	0.000	-7.297	0.000
S3	β	0.118	0.043	2.750	0.006	-20.606	0.000
S4	β_1	0.115	0.043	2.700	0.007	-20.705	0.000
	β_2	0.193	0.050	3.851	0.000	-16.127	0.000

Notes: "SE" refers to the estimated standard error. "$t(0)$" and "$t(1)$" refer to the t-statistics for testing the null hypothesis that the true coefficient is equal to zero and one, respectively. β_1 and β_1 are the coefficients of $s_{i,t-1} - f_{i,t-2}$ and $\Delta f_{i,t-1}$, respectively.

Table 4: Estimation results under the assumption of time-homoskedastic errors and no time effects.

Specification	Coefficient	Estimate	SE	$t(0)$	p-value	$t(1)$	p-value
S1	ρ	0.199	0.144	1.382	0.167	-5.566	0.000
S2	ρ	0.822	0.216	3.806	0.000	-0.823	0.410
S3	β	0.213	0.272	0.784	0.433	-2.896	0.004
S4	β_1	0.209	0.258	0.811	0.417	-3.067	0.002
	β_2	0.215	0.283	0.760	0.447	-2.779	0.005

Notes: See Table 3 for an explanation.