Financial Econometrics Series

SWP 2014/10

A Random Coefficient Approach to the Predictability of Stock Returns in Panels.

J. Westerlund and P.K. Narayan

The working papers are a series of manuscripts in their draft form. Please do not quote without obtaining the author’s consent as these works are in their draft form. The views expressed in this paper are those of the author and not necessarily endorsed by the School or IBISWorld Pty Ltd.
A RANDOM COEFFICIENT APPROACH TO THE PREDICTABILITY OF STOCK RETURNS IN PANELS∗

Joakim Westerlund† Paresh Narayan
Deakin University Deakin University
Australia Australia

December 24, 2013

Abstract

Most studies of the predictability of returns are based on time series data, and whenever panel data are used, the testing is almost always conducted in an unrestricted unit-by-unit fashion, which makes for a very heavy parametrization of the model. On the other hand, the few panel tests that exist are too restrictive in the sense that they are based on homogeneity assumptions that might not be true. As a response to this, the current paper proposes new predictability tests in the context of a random coefficient panel data model, in which the null of no predictability corresponds to the joint restriction that the predictive slope has zero mean and variance. The tests are applied to a large panel of stocks listed at the New York Stock Exchange. The results suggest that while the predictive slopes tend to average to zero, in case of book-to-market and cash flow-to-price the variance of the slopes is positive, which we take as evidence of predictability.

JEL Classification: C22; C23; G1; G12.

Keywords: Panel data; Predictive regression; Stock return predictability.

1 Introduction

Consider a panel of returns, \( y_{i,t} \), observable for \( t = 1, \ldots, T \) time series and \( i = 1, \ldots, N \) cross-sectional units. Recent years have witnessed an immense proliferation of research asking

∗The authors would like to thank Frederico Bandi (Co-Editor), an Associate Editor, and two anonymous referees for many valuable comments and suggestions.

†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.
whether $y_{i,t}$ can be predicted using past values of other financial variables such as the book-to-market ratio, the dividend–price ratio, the earnings–price ratio, and various interest rates. The conventional way in which earlier studies have been trying to test the predictability hypothesis is to first run a time series regression of $y_{i,t}$ onto a constant and one lag of the financial variable, $x_{i,t-1}$ say, and then to test whether the so-called predictive slope, $\beta_i$ say, is zero or not by using a conventional $t$-test (see, for example, Ang and Bekaert, 2007; Polk et al., 2006). This test is then repeated for each unit in the sample, each time using only the sample information for that particular unit.

In a recent paper Hjalmarsson (2010) questions this unit-by-unit approach and suggests combining the sample information obtained from the time series dimension with that obtained from the cross-sectional (see also Hjalmarsson, 2008; Kauppi, 2001). There are many advantages of doing this. First, in contrast to, for example, cross-country panels where the unit of observation is of some interest, the behavior of individual stocks is relatively uninteresting, which means that little is lost by taking the panel perspective. Second, the use of panel rather than time series data not only increases the total number of observations and their variation, but also reduces the noise coming from the individual time series regressions. This is reflected in the power of the resulting panel predictability test, which is increasing in both $N$ and $T$, as opposed to a time series/unit-by-unit approach where power is only increasing in $T$. Thus, from a power/precision point of view, a joint (panel) approach is always preferred. Third, since power is increasing in both $N$ and $T$, this means that in panels one can effectively compensate for a relatively small $T$ by having a relatively large $N$, and vice versa. Fourth, unlike the unit-by-unit approach, the joint panel approach accounts for the multiplicity of the testing problem. It is therefore correctly sized.

However, while appealing in many regards, the panel approach of Hjalmarsson (2010) also has its fair share of drawbacks. The main drawback is that the individual predictive slopes are restricted to be the same for all units (see also Hjalmarsson, 2008; Kauppi, 2001). Let us use $\beta$ to denote this common slope value. The homogeneity restriction makes sense under the null hypothesis that $\beta_1 = \ldots = \beta_N = 0$, but the alternative that $\beta_1 = \ldots = \beta_N = \beta \neq 0$ is too strong to be held in practice. In other words, while the predictability can certainly be similar across units, this cannot be a priori assumed.

In this paper we take the two opposing unit-by-unit and panel approaches as our starting point. Our mind set is the same as that of Hjalmarsson (2010), that is, a researcher that is
interested in making inference on the overall panel level, which seems like the most relevant consideration when using disaggregated firm-level data where the behavior of individual firms is not that interesting. In such cases the main drawback of the unit-to-unit approach is that the information contained in the unit-specific predictive t-statistics is not used in an efficient way. If the null is accepted, we conclude that the predictability is absent, whereas if it is rejected, we conclude that there is at least some units for which returns can be predicted, although the information we have to our disposal actually allows us to identify exactly the units that caused the rejection. Put in another way, the same conclusion could have been reached using less information. The fact that usually we are only interested in determining whether the null holds or not leads naturally to the consideration of a random specification for $\beta_i$, in which case the no predictability restriction corresponds to the joint null that the mean and variance of $\beta_i$ are zero, while the alternative is that the mean and/or the variance is different from zero. Hence, in contrast to the unit-by-unit approach, here the parameters considered are just enough to infer the no predictability null.

Taking the random coefficient model as our starting point, the goal of this paper is to design a procedure to test the joint null hypothesis that both the mean and variance of $\beta_i$ are zero, which has not been considered before. Our testing methodology is rooted in the Lagrange multiplier (LM) principle, which is very convenient because only estimation of the model parameters under the null hypothesis is required. This is in contrast to Wald tests, which are based on unrestricted estimates, and likelihood ratio tests which require both restricted and unrestricted estimates.

The form of the LM test statistic delivers significant insight regarding the predictability hypothesis. It has two parts; one tests the null hypothesis that the mean of $\beta_i$ is zero given that the variance is zero, while the other tests the null hypothesis of zero variance given a zero mean. The first part therefore tests the null hypothesis of predictability when the predictive slopes are assumed to be homogenous, which is one of the testing problems considered by Hjalmarsson (2010). By contrast, when the second part of the test statistic is used, the same null is tested against the alternative that there is predictability, but not on average, which seems like a very plausible scenario in practice. That is, while the individual predictive slopes are different from zero, positive and negative values tend to cancel out, making the predictability difficult to detect at the aggregate panel level. In fact, existing test statistics, including the first part of the LM test statistic, have no power against alternatives.
of this type. Thus, because these partial testing problems are interesting in their own right, in the paper we consider all three tests. The limiting distributions of the test statistics are derived and evaluated in small samples using Monte Carlo simulation.

In the empirical part of the paper we consider a large panel consisting of monthly observations from August 1996 to August 2010 on 1,559 firms. In contrast to, for example, cross-country panels where the unit of observation is of some interest, as already mentioned, the behavior of individual firms is relatively uninteresting, which means that little is lost by taking the panel perspective. On the other hand, the full panel is maybe too heterogeneous, and we therefore consider grouping the firms into 15 roughly homogenous sectors. For each sector we have six predictors; the book-to-market ratio, the cash flow-to-price ratio, the dividend–price ratio, dividend yield, the price–earnings ratio, and dividend–payout. The results suggest that while the first two are useful for forecasting returns, this is not the case for the other predictors. This is true for all sectors considered. Moreover, whenever predictability is found, the predictive slopes seem to average to zero, which means that existing panel tests for predictability based on estimates of the average predictive slope are likely to erroneously accept the no predictability null.

The rest of the paper is organized as follows. Sections 2–4 present the model, the test statistics, and their asymptotic distributions, respectively, which are evaluated using simulations in Section 5. Section 6 reports the results from the empirical application. Section 7 concludes. Proofs and derivations of important results are provided in Appendix.

2 The model

The data generating process of \( y_{i,t} \) is assumed to be given by

\[
\begin{align*}
    y_{i,t} &= \alpha_i + \beta_i x_{i,t-1} + u_{i,t}, \\
    x_{i,t} &= \delta_i (1 - \rho_i) + \rho_i x_{i,t-1} + \epsilon_{x_{i,t}}.
\end{align*}
\]  

This is a panel extension of the prototypical predictive regression model that has been widely used in the time series literature, in which \( x_{i,t} \) is a variable believed to be able to predict \( y_{i,t} \). In our case, \( x_{i,t} \) will be a financial ratio. As in previous studies, it is reasonable to assume that \( u_{i,t} \) is negatively correlated with \( \epsilon_{x_{i,t}} \). For example, if \( x_{i,t} \) is dividend yield, then an increase in the stock price will lower dividends and raise returns. Assumption 1 takes this into account.
Assumption 1.

\[ u_{i,t} = \gamma_i \epsilon_{x_{i,t}} + \epsilon_{y_{i,t}}, \]  

where \( \epsilon_{i,t} = (\epsilon_{x_{i,t}}, \epsilon_{y_{i,t}}) \) is independently and identically distributed (iid) with mean zero, covariance matrix \( \Sigma_\epsilon = \text{diag}(\sigma^2_{x_i}, \sigma^2_{y_i}) > 0 \) and finite fourth-order moments.

The assumption that \( \epsilon_{i,t} \) is iid (across both \( i \) and \( t \)) is for ease of exposure and is not necessary; in Section 3.3 we show how to relax this assumption. Assumption 2 summarizes the conditions placed on the coefficients of (1) and (2), which are all assumed to be random.

Assumption 2.

\[ \beta_i = \beta + N^{-p} T^{-q} \sigma_{yi} \sigma_{xi}^{-1} c_{\beta i}, \]  
\[ \alpha_i = \alpha + T^{-1/2} \sigma_{yi} c_{ai}, \]  
\[ \rho_i = 1 + N^{-1/2} T^{-1} \sigma_{yi} \sigma_{xi}^{-1} c_{\rho i}, \]  
\[ \delta_i = \delta + c_{\delta i}, \]  

where \( p \geq 0 \) and \( q \geq 0 \) are real numbers, \( c_i = (c_{\beta i}, c_{ai}, c_{\rho i}, c_{\delta i})' \) is iid with mean \( \mu_c = (\mu_{\beta}, \mu_{\alpha}, \mu_{\rho}, 0)' \) and covariance matrix \( \Sigma_c = \text{diag}(\sigma^2_{\beta}, \sigma^2_{\alpha}, \sigma^2_{\rho}, \sigma^2_{\delta}) > 0 \). \( c_i \) and \( \epsilon_{i,t} \) are mutually independent.

We start by discussing (4), which governs the main parameter of interest; then we also have some general remarks and also some remarks regarding (5)–(7). The null hypothesis of interest is that of no predictability, which can be formulated as \( H_0 : \beta_1 = \ldots = \beta_N = 0 \). A common way to formulate the alternative hypothesis is to assume that \( \beta_i \neq 0 \) is “non-local” in the sense that the degree of the predictability is not allowed to depend on \( N \) and \( T \) (see, for example, Lewellen, 2004). However, with such a specification we only learn if the test is consistent and, if so, at what rate. Therefore, to be able to evaluate the power analytically, in this paper we consider an alternative in which \( \beta_i \) is “local-to-constant” as \( N, T \to \infty \). This is captured by (4).

Now, since the main interest here lies in the testing of the hypothesis of no predictability, unless otherwise stated, we are going to assume that \( \beta = 0 \), and use \( c_{\beta i} \) (or, rather, \( \mu_{\beta} \) and \( \sigma^2_{\beta} \)) to measure the extent of the predictability. In this case, we will typically refer to \( \beta_i \) as being “local-to-zero”, rather than local-to-constant. The specification in (4) with \( \beta = 0 \) is extremely convenient because it means that the original \( N \)-dimensional problem of testing
whether $\beta_1 = \ldots = \beta_N = 0$ can be reformulated using only two parameters, $\mu_\beta$ and $\sigma_\beta^2$. The null hypothesis of no predictability can be stated as

$$H_0 : \mu_\beta = \sigma_\beta^2 = 0,$$

while the alternative can be stated as

$$H_1 : \mu_\beta \neq 0 \text{ and/or } \sigma_\beta^2 > 0.$$

The powers $p$ and $q$ determine the rate at which $\beta_i$ shrinks towards its hypothesized value under the null. On the one hand, if $p = q = 0$, then $\beta_i$ is independent of $N$ and $T$, and so we are back in the usual consideration of a non-local alternative. On the other hand, if $p > 0$ and/or $q > 0$, then $\beta_i$ is local-to-zero in the sense that $\beta_i \to 0$ as $N, T \to \infty$. For example, if $p = 0$ and $q = 1$, then (4) corresponds to the specification considered by Jansson and Moreira (2008) in the pure time series case. Of course, one of the main advantages of using panels rather than single time series is the greater information content, which should make it possible to detect even smaller deviations from the null. That is, in panels $p$ need not be zero. Because $N^{-p}T^{-q} < T^{-q}$ whenever $p > 0$, this means that we are now considering even smaller deviations from zero. This is important because whenever predictability is found, the evidence is usually weak, suggesting that the deviations from the no predictability null are not large. Local-to-constant specifications like the one in (4) are not only very flexible in the type of alternatives that can be accommodated, but have also been shown to provide very accurate approximations in small samples. In fact, local-to-constant modeling is in part motivated by the poor small-sample performance of non-local approximations.

Remarks.

1. One advantage of (4) is that under $H_1$ there is no need for any sign or homogeneity restrictions on $\beta_i$. Consider, as an example, the pooled $t$-tests discussed in Hjalmarsson (2008, 2010) and Kauppi (2001), in which the null is tested against the homogenous alternative that $\beta_1 = \ldots = \beta_N = \beta \neq 0$. The null makes sense, but it is unrealistic to assume a priori that all the units have the same degree of predictability in case of a rejection. While $\beta_i \to 0$ as $N, T \to \infty$, for a fixed sample size the above local-to-zero model accommodates a much wider range of values for $\beta_i$ as $c_{\beta_i}$ varies, including both predictive and non-predictive possibilities. Thus, for a fixed sample size one can
view the model in (4) as a conventional non-local alternative in which the deviation from the null is very small.

2. The assumption that $c_\beta_i$ is independent of the other random elements of the model can be relaxed at the expense of more complicated proofs. In particular, note that since under $H_0 \ c_\beta_1 = \ldots = c_\beta_N = 0$, independence is only an issue under $H_1$.

3. The assumption in (4) (together with the conditions places on $c_\beta_i$) is enough to infer $H_0/H_1$. However, as pointed out by Hjalmarsson (2010), it might also be interesting to test the hypothesis regarding $\alpha_i$. In particular, being a measure of expected return in the absence of predictability, the potential homogeneity of $\alpha_i$ can be a rather relevant restriction to test. For this reason, following the work of Orme and Yamagata (2006), we will make use of the local-to-constant specification in (5). The main motivation for this is the same as for $\beta_i$; that is, it enables us to evaluate the power analytically. This is done in Section 4.

4. Consider (2), which governs the behavior of the predictor. Since many of the predictors considered in the empirical part are known to be quite persistent, $\rho_i$ is modeled as being local-to-unity. Note in particular how (6) is nothing but the standard local-to-unity model in the panel unit root literature (see Moon et al., 2007), in which $c_{\rho_i}$ measures the deviation from a unit root. If $c_{\rho_i} < 0$, then $\rho_i$ approaches one from below and so $x_{it}$ might be said to be “locally stationary”, whereas if $c_{\rho_i} > 0$, then $\rho_i$ approaches one from above and so $x_{it}$ is “locally explosive”.

5. The intercept in (2) is not of any particular interest to us, and we are therefore simply going to assume that it is randomly distributed. This is captured by (7).

3 Tests of predictability

In this section, we first consider the true LM test statistics for the null hypothesis of no predictability (Section 3.1), which are based on the assumption that all parameters except $\beta_i$ are known. We then show how this analysis extends to the more realistic case when the parameters are unknown (Section 3.2). The section is concluded with a discussion of the case when the iid part of Assumption 1 fails (Section 3.3). Throughout we assume that $\beta = 0$, so that the testing problem can be expressed in terms of $\mu_\beta$ and $\sigma^2_\beta$ only.
3.1 The infeasible test statistics

It can be shown (a formal proof is available upon request) that the true LM test statistic for testing $H_0$ (under $\rho_1 = ... = \rho_N = 1$) is given by

$$
\frac{(A_0^0)^2}{B_0^0} + \frac{1}{2} \frac{(A_0^2)^2}{B_0^2},
$$

where

$$
A_0^\mu = \sum_{i=1}^N \sum_{t=2}^T \sigma_{xi}^{-1} \sigma_{yi}^{-1} r_{yi,t} x_{i,t-1},
$$

$$
B_0^\mu = \sum_{i=1}^N \sum_{t=2}^T \sigma_{xi}^{-2} x_{i,t-1}^2,
$$

$$
A_0^\sigma^2 = \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi}^{-2} \sigma_{xi}^{-2} (r_{yi,t}^2 - \sigma_{yi}^2) x_{i,t-1}^2,
$$

$$
B_0^\sigma^2 = \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi}^{-4} \sigma_{xi}^{-4} (2r_{yi,t}^2 - \sigma_{yi}^2) x_{i,t-1}^4,
$$

where $r_{yi,t} = y_{i,t} - \alpha_i - \gamma_i \Delta x_{i,t}$ is $\epsilon_{yi,t}$ with a unit root predictor imposed. In practice, it is more convenient to work with the following slightly modified version:

$$
LM^0 = LM_0^\mu + LM_0^\sigma^2,
$$

where

$$
LM_0^\mu = \frac{(A_0^0)^2}{B_0^0},
$$

$$
LM_0^\sigma^2 = \frac{12}{5(\kappa_y - 1)} \frac{(A_0^2)^2}{B_0^2},
$$

with $\kappa_y = \sigma_{yi}^{-4} E(\epsilon_{yi,t}^4)$.

The formula for $LM_0^\mu$ is a very simple and intuitive.\footnote{Note how $LM_0^\mu$ can be seen as the squared panel equivalent of the time series predictability test considered by Campbell and Yogo (2008).} In fact, a close inspection reveals that $LM_0^\mu$ is nothing but the LM test statistic for testing $H_0$ versus the alternative that $\mu_\beta \neq 0$ given $\sigma_\beta^2 = 0$. That is, $LM_0^\mu$ is the LM predictability statistic based on the assumption that $\beta_1 = ... = \beta_N = 0$. Similarly, $LM_0^\sigma^2$ is the LM statistic for testing $H_0$ versus the alternative that $\sigma_\beta^2 > 0$ given $\mu_\beta = 0$. In other words, $LM_0^\sigma^2$ tests $H_0$ versus the alternative that there is predictability at the level of the individual unit, but not on average. Thus, in contrast to, for
example, Hjalmarsson (2008, 2010) and Kauppi (2001), with our approach there is not just one way in which the no predictability null can be tested, but several.

Even if the error terms are normally distributed the exact distributions of the LM statistics are untractable. In this paper we therefore use asymptotic theory to obtain their limiting distributions. For simplicity, because of the additive structure of the joint test, we only present the results for \( LM_\mu^0 \) and \( LM_{\sigma^2}^0 \). The asymptotic distribution of the joint test statistic can then be obtained by simply adding the asymptotic distributions of \( LM_\mu^0 \) and \( LM_{\sigma^2}^0 \).

**Theorem 1.** Under Assumptions 1 and 2, with \( p = 1/2 \) and \( q = 1 \), as \( N, T \to \infty \) with \( N/T \to 0 \),

\[
LM_\mu^0 \to_d \frac{(\mu_\beta - \overline{\gamma} \mu_\rho)^2}{2} + \sqrt{2}(\mu_\beta - \overline{\gamma} \mu_\rho)Z_1 + Z_1^2,
\]

\[
LM_{\sigma^2}^0 \to_d Z_2^2,
\]

where the symbol \( \to_d \) signifies convergence in distribution, \( \overline{\gamma} = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \gamma_i \), and \( Z_1 \) and \( Z_2 \) are generic \( N(0,1) \) variables that are independent.

In order to appreciate fully the implications of these results it is instructive to consider some special cases depending on the values taken by \( \mu_\beta \) and \( \mu_\rho \).

1. If \( \mu_\beta = 0 \) (\( H_0 \) holds) and \( \mu_\rho = 0 \) (the predictor has a unit root on average), then \( LM_\mu^0 \to_d Z_1^2 \sim \chi^2(1) \), suggesting that the appropriate critical value for use with \( LM_\mu^0 \) can be obtained from the chi-squared distribution with one degree of freedom. The same applies to \( LM_{\sigma^2}^0 \). Hence, since the asymptotic null distribution of \( LM^0 \) is just the sum of the asymptotic null distributions of \( LM_\mu^0 \) and \( LM_{\sigma^2}^0 \) (which are independent), we have \( LM^0 \to_d Z_1^2 + Z_2^2 \sim \chi^2(2) \).

2. If \( \mu_\beta = 0 \) but \( \mu_\rho \neq 0 \), then \( LM_{\sigma^2}^0 \) again converges to its asymptotic distribution under \( H_0 \). Hence, \( LM_{\sigma^2}^0 \) is asymptotically invariant with respect to \( \mu_\rho \). However, this is not the case for \( LM_\mu^0 \), whose asymptotic distribution in this case is given by \( (\overline{\gamma} \mu_\rho)^2/2 + \sqrt{2}\overline{\gamma} \mu_\rho Z_1 + Z_1^2 \). Thus, unless \( \overline{\gamma} = 0 \), the asymptotic distribution of \( LM_\mu^0 \) will depend on both \( \overline{\gamma} \) and \( \mu_\rho \), which is in agreement with the time series literature (see, for example, Campbell and Yogo, 2006; Elliott and Stock, 1994). The presence of \( \mu_\rho \) and \( \overline{\gamma} \) has two effects. The first is to shift the mean of the limiting distribution. Specifically, since \( \mu_\rho^2 > 0 \), this means that the mean shifts to the left as we move away from \( H_0 \). The
second effect, which is captured by $\sqrt{2}\gamma \mu_{\rho} Z_1$, is to increase the variance of the limiting distribution. In the current setting with known parameters this is not a problem. However, in general with unknown parameters, this is a major complicating factor, as in this case a rejection need not be due to genuine predictability, but could also be due to the presence of nuisance parameters.

3. If $\mu_{\beta} \neq 0$ ($H_1$ holds) but $\mu_{\rho} = 0$, then the asymptotic distribution of $LM_{\sigma^2}$ is again the same as under $H_0$, suggesting that with $p = 1/2$ and $q = 1$ this test has no power under the particular local alternative given in (4). The asymptotic distribution of $LM_{\mu}^0$ in this case is given by $\mu_{\beta}/2 + \sqrt{2}\mu_{\rho} Z_1 + Z_1^2$, suggesting that, in contrast to $LM_{\sigma^2}^0$, $LM_{\mu}^0$ has non-negligible power.

4. If $\mu_{\beta} \neq 0$ and $\mu_{\rho} \neq 0$, while $LM_{\sigma^2}^0$ is unaffected, unless $\mu_{\beta} = \gamma \mu_{\rho}$, the asymptotic distribution of $LM_{\mu}^0$ will now depend on $\mu_{\beta}$, $\gamma$, and $\mu_{\rho}$. The fact that $LM_{\sigma^2}^0$ does not have any local power requires some discussion. The simple reason is that the rate of shrinking of the local alternative is too fast for $\mu_{\beta}$ and $\sigma_{\beta}^2$ to manifest themselves in the limiting distribution of $LM_{\sigma^2}^0$. Generally speaking, the rate of shrinking of the local alternative is determined by the probabilistic order of the numerator of the test statistic, here represented by $A_{\mu}^0$ and $A_{\sigma^2}^0$. Thus, with a composite test statistic like ours the appropriate rate of shrinking for the two parts need not be the same. Indeed, while the order of $A_{\mu}^0$ is given by $O_p(NT^2)$, that of $A_{\sigma^2}^0$ is given by $O_p(\sqrt{NT^3/2})$. Since $O_p(\sqrt{NT^3/2}) < O_p(NT^2)$, this means that $LM_{\mu}^0$ will dominate. Hence, in order for the deviations from $H_0$ to be detectable using $LM_{\sigma^2}^0$, they must be “larger” (as measured by $p$ and $q$) than before. This is shown in Proposition 1.

**Proposition 1.** Under the Assumptions 1 and 2, with $p = 1/4$ and $q = 3/4$, as $N, T \to \infty$ with $N/T \to 0$,

$$LM_{\sigma^2}^0 \to_d \frac{12(\mu_{\beta}^2 + \sigma_{\beta}^2)^2}{5(\kappa_y - 1)} + \frac{4\sqrt{3}(\mu_{\beta}^2 + \sigma_{\beta}^2)}{\sqrt{5(\kappa_y - 1)}} Z_1 + Z_1^2.$$

**Remarks.**

1. In contrast to $LM_{\mu}^0$, the asymptotic distribution of $LM_{\sigma^2}^0$ does not depend on $c_{\rho i}$ or $\gamma_i$. Note in particular that if $H_0$ holds so that $\mu_{\beta}^2 = \sigma_{\beta}^2 = 0$, then $LM_{\sigma^2}^0 \to_d Z_1^2 \sim$
\(\chi^2(1)\), which is completely free of nuisance parameters (including \(c_{\rho i}\) and \(\gamma_i\)). This is a great advantage, especially in view of the problematic dependence of the asymptotic distribution of \(LM^0_{\mu}\) on \(\mu_{\rho}\) and \(\bar{\gamma}\) (see remark 2 to Theorem 1).

2. Since the rate of shrinking of the local alternative is now lower than before \((p = 1/4 < 1/2\) and \(q = 3/4 < 1\)), this means that \(LM^0_{\mu}\) is diverging if \(\mu_\beta \neq 0\) and/or \(\mu_\rho \neq 0\). Thus, although \(LM^0_{\mu}\) has non-negligible power against deviations that shrink towards the null at rate \(N^{-1/4}T^{-3/4}\) (Theorem 2), the power of \(LM^0_{\mu}\) is approaching one as \(N, T \to \infty\). Conversely, if \(p = 1/2\) and \(q = 1\), while the power of \(LM^0_{\mu}\) is non-negligible (and non-increasing), the power of \(LM^0_{\sigma_2}\) is now negligible. In an essence, when \(p = 1/2\) and \(q = 1\), the deviations from \(H_0\) (as measured by \(\beta_i \neq 0\)) are too small for \(LM^0_{\sigma_2}\) to be able to detect them.

3. It is interesting to compare the local power of the LM tests with that achievable using a time series test. Let us therefore consider the test of Lewellen (2004), which is asymptotically uniformly most powerful when \(p = 0\), \(q = 1\) and \(c_{\rho i} = 0\) (see Campbell and Yogo, 2006). The fact that \(LM^0_{\mu}\), and hence also \(LM^0\), have power within neighborhoods that shrink to the null at the rate \(N^{-1/2}T^{-1}\) means that, while \(T\) is relatively more important (because \(q = 1 > p = 1/2\)), a larger \(N\) leads to higher power in the sense that we can be even closer to the null (as measured by \(\beta_i \neq 0\)) and still have power. The test of Lewellen (2004) has power within \(T^{-1}\)-neighborhoods (corresponding to \(p = 0\) and \(q = 1\)). Hence, as expected, the power of this test is unaffected by \(N\). One implication of this is that since the rate of shrinking in terms of \(T\) is the same for the two test approaches \((q = 1)\), whenever \(N > 1\) \(LM^0_{\mu}\) will tend to dominate. The situation is quite different when considering \(LM^0_{\sigma_2}\). Indeed, since in this case the value of \(q\) for which power is negligible is given by \(q = 3/4 < 1\), this means that the time series test makes better use of the information contained in the time series dimension. However, this is compensated for in part by the fact that with \(LM^0_{\sigma_2}, q = 1/4 > 0\). In both cases \(p = q = 1\), suggesting that the relative power will have to depend on the relative expansion rate of \(N\) and \(T\). We have assumed that \(N/T \to 0\), which implies \(N^{-1/4}T^{-3/4} \to T^{-1}\). The times series test should therefore be more powerful, at least asymptotically.\(^3\) Of course, since these tests are not really designed to infer the same

\(^3\)In Section 5 we use Monte Carlo simulation to assess power in small samples.
hypothesis, the test of Lewellen (2004) (or indeed any other time series test) cannot be considered as a substitute for the panel tests developed here.\footnote{While one could in principle consider applying a unit-by-unit approach (see Section 1), which in the present context would amount to running cross-section-specific Lewellen (2004) tests, this would mean ignoring the multiplicity of the testing problem, which is in turn likely to result in too many rejections. An alternative that does not suffer from this problem is to use the so-called “Bonferroni inequality”; however, that will instead tend to make the test conservative. Hence, as usual, if the purpose is to conduct multiple hypothesis testing, then one should really consider a joint (panel) test.}

4. In contrast to $LM^0_{\mu}$, the power of $LM^0_{\sigma^2}$ is not only determined by $\mu_\beta$ but also by $\sigma^2_\beta$. Hence, the power of this test depends not only on the average $\beta_i$, but also on the heterogeneity of $\beta_i$, which is not the case for $LM^0_{\mu}$. Thus, unlike $LM^0_{\mu}$, $LM^0_{\sigma^2}$ has power against alternative hypotheses of the type $\mu_\beta = 0$ and $\sigma^2_\beta > 0$. Suppose, for example, that $LM^0_{\mu}$ is unable to reject. Then what is the correct conclusion to draw? Some researchers would probably take this as evidence in favor of $H_0$. However, since $\sigma^2_\beta$ might still be positive, this need not be the case. In other words, it is possible to have a situation in which there is predictability for each cross-sectional unit, but that positive and negative values of $\beta_i$ cancel out, causing $LM^0_{\mu}$ to accept the null. Only if $LM^0_{\sigma^2}$ also accepts can we say that there is no evidence against $H_0$.

5. Theorem 1 and Proposition 1 are based on an approximation that removes the dependence in higher moments of $c_\beta_i$. For this approximation to hold, we need $N/T \to 0$ as $N, T \to \infty$, which in practice means that $T >> N$. In Section 3.2 we consider feasible versions of the above statistics. In this case, $N/T \to 0$ is not only needed to ensure that the approximation holds, but also to eliminate the effect of the estimation of the parameters of the model.

### 3.2 The feasible test statistics

All results reported so far are based on the assumption that $\alpha_i, \gamma_i, \sigma^2_\gamma, \text{ and } \sigma^2_\xi$ are all known, which is of course not realistic. In this section we therefore consider replacing these parameters by their restricted ML estimators under $H_0$. The ML estimators of $\alpha$ (the constant part of $\alpha_i$) and $\gamma_i$ can be obtained by applying ordinary least squares (OLS) to the following auxiliary regression:

\[
y_{i,t} = \alpha + \gamma_i \Delta x_{i,t} + \text{error.} \tag{8}\]

While one could in principle consider applying a unit-by-unit approach (see Section 1), which in the present context would amount to running cross-section-specific Lewellen (2004) tests, this would mean ignoring the multiplicity of the testing problem, which is in turn likely to result in too many rejections. An alternative that does not suffer from this problem is to use the so-called “Bonferroni inequality”; however, that will instead tend to make the test conservative. Hence, as usual, if the purpose is to conduct multiple hypothesis testing, then one should really consider a joint (panel) test.
The ML estimators of $\sigma^2_{xi}$, $\sigma^2_{yi}$ and $\kappa_y$ are given by

$$\hat{\sigma}^2_{xi} = T^{-1} \sum_{t=2}^T (\Delta x_{it})^2, \quad \hat{\sigma}^2_{yi} = T^{-1} \sum_{t=2}^T f^2_{yi,t},$$

and $\hat{\kappa}_y = (NT)^{-1} \sum_{i=1}^N \sum_{t=2}^T \sigma^{-4}_{yi,t}$, respectively, where $f_{yi,t} = y_{i,t} - \hat{\alpha} - \hat{\gamma}_i \Delta x_{i,t}$. The feasible versions of $LM^0$, $LM^0_\mu$ and $LM^0_{\sigma^2}$ are given by

$$LM = LM_\mu + LM_{\sigma^2},$$

with

$$LM_\mu = \frac{A^2_\mu}{B_\mu},$$

$$LM_{\sigma^2} = \frac{12}{5(\hat{\kappa}_y - 1)} \frac{A^2_{\sigma^2}}{B_{\sigma^2}},$$

where $A_{\mu}$, $B_{\mu}$, $A_{\sigma^2}$ and $B_{\sigma^2}$ are $A^0_\mu$, $B^0_\mu$, $A^0_{\sigma^2}$ and $B^0_{\sigma^2}$, respectively, with $\alpha$, $\gamma_i$, $\sigma^2_{xi}$ and $\sigma^2_{yi}$ replaced by their corresponding ML estimates, and $\hat{r}_{yi,t}$ replaced by $\hat{r}_{yi,t}$. Theorem 2 shows that the effect of this replacement is negligible.

**Theorem 2.** Under the Assumptions 1 and 2, with $p = 1/2$ and $q = 1$ or $p = 1/4$ and $q = 3/4$, as $N, T \to \infty$ with $N/T \to 0$,

$$LM_\mu = LM^0_\mu + o_p(1),$$

$$LM_{\sigma^2} = LM^0_{\sigma^2} + o_p(1).$$

Theorem 2 shows that standard chi-squared inference is possible also in the case with unknown parameters. The problem is that, as already pointed out in remark 2 to Theorem 1, the asymptotic distributions of $LM_\mu$ and $LM_{\sigma^2}$, and therefore also that of $LM$, depend on $\hat{\gamma}$ and $\mu_{\rho r}$, which are unknown. Specifically, the problem is that while $\gamma_i$, and hence $\hat{\gamma}_i$, is consistently estimable, $\mu_{\rho r}$ is not. One can, of course, assume that $\mu_{\rho r} = 0$, but then one would no longer be testing the hypothesis of no predictability, but rather the joint hypothesis of no predictability and an average (exact) unit root predictor, which calls for careful interpretation of the test outcome in applied work. In particular, with a near unit root predictor under the null, researchers might incorrectly interpret a rejection as providing evidence of predictability, when in fact the predictor has no predictive ability at all.

Because of the dependence on $\mu_{\rho r}$, many studies begin by pretesting the predictor for a unit root, and the predictability test is then implemented conditional on the outcome of the pretest. Unfortunately, this means loosing control of the overall significance level of the joint test, which depends on the correlation between the two test statistics. Therefore, in order
to at least put an upper limit on the joint significance level, Cavanagh et al. (1995) have suggested the use of the Bonferroni inequality. Of course, being only a rough worst case approximation, it does not come as a surprise that tests based on Bonferroni critical values tend to be rather conservative.

Because of the pooling across the cross-section, our panel statistic has the advantage that it is asymptotically independent of any unit-specific unit root statistic that may be considered for the pretest. It also implies that the available information regarding \( \rho_i \) can be used in a relatively straightforward and uncomplicated fashion. Consider, for example, the test of Lewellen (2004), which can be seen as a bias-adjusted version of the conventional OLS \( t \)-statistic for testing \( \beta_i = 0 \) in regression \( i \). The idea is simple. Indeed, since the bias is given by \( \gamma_i(\hat{\rho}_i - \rho_i) \), where \( \hat{\rho}_i \) is the OLS estimator of \( \rho_i \), all that is needed in order to make the test operational is an estimate of \( \gamma_i \), and an educated “guess” of the value of \( \rho_i \). The obvious problem is that the guess might not be correct.

The next test that we consider has the advantage of being asymptotically invariant with respect to \( \gamma \) and \( \mu_\rho \) without requiring any assumptions regarding the values taken by these parameters. The idea is the following. Suppose again that all parameters except \( \beta_i \) are known. In this case it can be shown that the dependence on \( \gamma \) and \( \mu_\rho \) can be removed by simply adding \( \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i(\rho_i - 1)x_{i,t-1}^2 \) to the numerator of \( LM_0^\mu \). In view of this, the question naturally arises if the same holds true once \( \sigma_{yi}, \sigma_{xi}, \gamma_i \) and \( \rho_i \) have been replaced by estimates? It turns out that it does. Let us therefore consider the following modified version of \( LM_\mu \):

\[
LM^m_\mu = \frac{(A_\mu + NT\hat{\theta})^2}{(1 + \hat{\omega}^2)B_\mu},
\]

where

\[
\hat{\omega}^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_{xi}^2 \hat{\sigma}_{yi}^{-2} \hat{\gamma}_i^2,
\]

\[
\hat{\theta} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{\sigma}_{yi}^{-1} \hat{\sigma}_{xi}^{-1} \hat{\gamma}_i(\hat{\rho}_i - 1)x_{i,t-1}^2,
\]

with \( \hat{\rho}_i \) being the OLS estimator of \( \rho_i \). Note that this formula replicates the appropriately corrected version of \( LM_0^\mu \) in case of known parameters. The main difference is the scaling

---

5The idea here is to first find the minimum and maximum critical values for the predictability test for all possible values of \( \mu_\rho \), and then to reject if the value of the test statistic falls outside this range of critical values.
by \((1 + \omega^2)\), which is necessary because adding \(NT\hat{\theta}\) not only affects the mean of the test statistic but also the variance.

**Theorem 3.** Under the Assumptions 1 and 2, with \(p = 1/2\) and \(q = 1\), as \(N, T \to \infty\) with \(N/T \to 0\),

\[
LM^m_\mu \to_d \frac{\mu^2}{2(1 + \omega^2)} + \frac{\sqrt{2}\mu_\beta}{\sqrt{1 + \omega^2}} Z_1 + Z_1^2,
\]

where

\[
\omega^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma^2_x \rho_{yi}^2 \gamma_i^2.
\]

The beauty of this result is that, in contrast to the asymptotic distribution of \(LM^0_\mu\) (see Theorem 1), which also holds for \(LM_\mu\) (Theorem 2), here there is no dependence on \(\mu_\rho\), and the dependence on \(\gamma_i\) disappears under \(H_0\) (when \(\mu_\beta = 0\)). Thus, with this test there is no confusion about the interpretation of the test outcome in case of a rejection; if the test rejects it must be due to \(\mu_\beta \neq 0\). Of course, this does not mean that the level of power is also unaffected by \(\gamma_i\). In fact, as Theorem 3 makes clear, power is increasing in \(\mu_\beta\) and decreasing in \(\omega^2\), and hence also in \(\gamma_i\). The only case when power is unaffected by \(\omega^2\) is when \(\gamma_1 = \ldots = \gamma_N = 0\). Hence, while the power (and also the size) of \(LM^0_\mu\) and \(LM_\mu\) is expected to increase with \(\mu_\beta\) and \(\gamma_i\), here it is the other way around. Of course, in view of the usual power/efficiency–robustness/size trade-off, this is not totally unexpected.

### 3.3 Serial and cross-section dependence robust test statistics

One drawback with the above treatment is that it supposes that \(u_{i,t}\) (the composite error term in (3)) is iid, an assumption that is perhaps too strong to be held in applications. In this subsection we therefore generalize the analysis by allowing for more general error dynamics. In particular, following Campbell and Yogo (2004), instead of (2), we are going to assume that

\[
x_{i,t} = \delta_i(1 - \rho_i) + \rho_i x_{i,t-1} + \sum_{j=1}^{r} \phi_{j,i} \Delta x_{i,t-j} + \epsilon_{xi,t},
\]

thereby allowing for short-run serial correlation in \(x_{i,t}\). To also allow for some form of cross-section dependence, we are going to follow, for example, Forni et al. (2003), Stock and Watson (2002), and Ludvigson and Ng (2007), and consider the following factor augmented version of (1):

\[
y_{i,t} = \alpha_i + \beta_i x_{i,t-1} + \pi_i f_t + u_{i,t},
\]
where \( f_t \) is a stationary common factor with \( \pi_i \) being the associated factor loading. Hence, letting \( \lambda_{ij} = -\gamma_i \phi_{ij} \), the appropriate version of (8) to use in this case is given by

\[
y_{i,t} = \alpha + \pi_i f_t + \gamma_i \Delta x_{i,t} + \sum_{j=1}^{r} \lambda_{ij} \Delta x_{i,t-j} + \text{error.} \tag{11}
\]

By redefining \( r_{yi,t} = y_{i,t} - \alpha_i - \pi_i f_t - \gamma_i \Delta x_{i,t} - \sum_{j=1}^{r} \lambda_{ij} \Delta x_{i,t-j} \), and replacing \( \sigma^2_{xi} \) with \( \omega^2_{xi} = \sigma^2_{xi}(1 - \sum_{j=1}^{r} \phi_{ij})^{-2} \), the infeasible robust LM test statistic have exactly the same form as \( LM^0 \), and so does its asymptotic distribution.

The computation of the feasible robust LM test statistic depend on what is being assumed regarding \( f_t \). If \( f_t \) is known, then we begin by fitting (9) by OLS. This gives estimates \( \hat{\phi}_{1,i}, ..., \hat{\phi}_{r,i}, \hat{\sigma}^2_{xi} \) of \( \phi_{1,i}, ..., \phi_{r,i}, \sigma^2_{xi} \), which in turn can be used to obtain \( \hat{\sigma}^2_{xi} = \hat{\sigma}^2_{xi}(1 - \sum_{j=1}^{r} \hat{\phi}_{ij})^{-2} \). The only thing that is missing now is \( \hat{r}_{yi,t} \) (needed for computing \( \hat{\sigma}^2_{yi} \) and \( \hat{\kappa}_y \)), which can be obtained as the residual from the OLS fit of (11). The main problem with treating \( f_t \) as unknown is that now (11) is no longer feasible. Fortunately, there is a simple trick that can be used to circumvent this problem. We begin by taking cross-sectional averages and then solving (10) for \( f_t \), giving

\[
f_t = \frac{1}{\pi} \sum_{i=1}^{N} \left( y_{i,t} - \alpha_i - \gamma_i \Delta x_{i,t} - \sum_{j=1}^{r} \lambda_{ij} \Delta x_{i,t-j} \right) - \frac{1}{\pi} \overline{u}_t,
\]

where \( \pi = N^{-1} \sum_{i=1}^{N} \pi_i \) and a similar definition of \( \overline{u}_t \). The essential insight here, which is the same as in Pesaran (2006) (see also Hjalmarsson, 2010), is that since \( u_{ij} \) is mean zero and iid, we have \( \overline{u}_t = O_p(N^{-1/2}) \). This means that \( f_t \) can be approximated by (a linear combination of) the cross-sectional averages of \( y_{i,t} \) and \( \Delta x_{i,t}, ..., \Delta x_{i,t-r} \), which again makes (11) feasible after replacing \( f_t \) by these averages.

### 4 Tests of other hypotheses

As discussed in Section 3, the test statistics provided so far are quite flexible when it comes to the types of conclusions that can be drawn. If \( LM^\mu \) (or \( LM^\mu^w \)) and \( LM^\sigma^2 \) accepts, then there is no evidence against the no predictability null, whereas if at least one of the tests end up rejecting the null, then there is evidence to the contrary. If \( LM^\mu \) rejects while \( LM^\sigma^2 \) accepts, then the evidence is towards a common predictive slope coefficient. By contrast, if \( LM^\mu \) accepts while \( LM^\sigma^2 \) rejects, then there is evidence of predictability but not on average.

However, in some situations it might be interesting to test more general hypotheses regarding \( \beta_i \), and not just whether it is zero or not. The same is true for \( \alpha_i \) (the intercept in
the predictive regression). Suppose, for example, that $LM_\mu$ rejects. A natural question that arises is if the predictability is homogenous or not? Since $\mu_\beta \neq 0$, we can no longer use $LM_{\sigma^2}$ to test if $\sigma^2_\beta = 0$ (as this test statistic is also sensitive to $\mu_\beta$; see Theorem 2). In this section we therefore focus on inference regarding these parameters more generally. In so doing, we will relax the assumption that $\beta = 0$. One implication of this is that we can no longer rely on the restricted ML estimators of $\alpha$ and $\gamma_i$. This means that instead of defining $\hat{\alpha}$ and $\hat{\gamma}_i$ as the OLS estimators of $\alpha$ and $\gamma_i$ in (8) (or (11)), in this section we make use of the following unrestricted regression:

$$y_{i,t} = \alpha_i + \beta_i x_{i,t-1} + \gamma_i \Delta x_{i,t} + \text{error}. \quad (12)$$

Hence, in what follows $\hat{\alpha}_i, \hat{\beta}_i$ and $\hat{\gamma}_i$ refer to the OLS estimators of $\alpha_i, \beta_i$ and $\gamma_i$, respectively, in this regression.

In Appendix we show that the asymptotic distributions (as $N, T \to \infty$) of $\hat{\alpha}_i, \hat{\beta}_i$ and $\hat{\gamma}_i$ are normal. This result is very convenient because it means that hypotheses regarding the associated parameters can be tested in the usual fashion. It also provides a basis for deriving the limiting distributions of various test statistics. In this section we focus on inference regarding $\alpha_i$ and $\beta_i$. Let us therefore use $t_{\alpha i}(\alpha_0)$ ($t_{\beta i}(\beta_0)$) to denote the $t$-statistic for testing $H_0 : \alpha_i = \alpha_0$ ($H_0 : \beta_i = \beta_0$).

**Theorem 4.** Under the Assumptions 1 and 2, with $p = 0$ and $q = 1$, as $N, T \to \infty$,

$$t_{\alpha i}(\alpha) \to_d c_{\alpha i} \left(1 + \frac{\overline{W}_{xi}^2}{\int_0^1 (W_{xi}(s) - \overline{W}_{xi})^2 ds}\right)^{-1/2} + Z_1,$$

$$t_{\beta i}(\beta) \to_d c_{\beta i} \left(\int_0^1 (W_{xi}(s) - \overline{W}_{xi})^2 ds\right)^{1/2} + Z_2,$$

where $W_{xi}(s)$ is a standard Brownian motion and $\overline{W}_{xi} = \int_0^1 W_{xi}(s) ds$.

In analogy to the results reported for the LM test statistics, we see that the limiting distributions of $t_{\alpha i}(\alpha)$ and $t_{\beta i}(\beta)$ depend on the drift parameters $c_{\alpha i}$ and $c_{\beta i}$. However, since we are no longer dealing with pooled tests, the dependence is not on the mean and variance of $c_{\alpha i}$ and $c_{\beta i}$, but rather on the parameters themselves. If $H_0 : \alpha_i = \alpha$ ($H_0 : \beta_i = \beta$) is true, then $c_{\alpha i} = 0 (c_{\beta i} = 0)$, and therefore the asymptotic distribution of $t_{\alpha i}(\alpha)$ ($t_{\beta i}(\beta)$) reduces to $N(0, 1)$.

Theorem 4 can also be used to derive the limiting distributions of tests of poolability for the panel and a whole. Let us therefore denote by $\hat{\beta}$ the pooled OLS estimator of $\beta_i$ in (12),
and define $H_{\beta_i} = t_{\beta_i}(\hat{\beta})^2$, which is similar in spirit to the Hausman test statistic considered by Westerlund and Hess (2011). Under $H_0: \beta_i = \beta$ (for unit $i$), in view of Theorem 4, it is clear that $H_{\beta_i} \to_d \chi^2(1)$ as $N, T \to \infty$. This result can in turn be used to construct valid poolability tests for the panel as a whole. Thus, in this case we are interested in testing $H_0: \beta_1 = \ldots = \beta_N = \beta$. The test statistic that we will consider is the following normalized maximum:

$$H_{\hat{\beta}} = \max_{1 \leq i \leq N} \left( \frac{H_{\beta_i} - \tau_{2N}}{\tau_{1N}} \right),$$

where $\tau_{2N} = F^{-1}(1 - 1/N)$, $\tau_{1N} = F^{-1}(1 - 1/(Ne)) - \tau_{2N}$, and $F^{-1}(x)$ is the inverse of the chi-squared distribution with one degree of freedom. The asymptotic null distribution of $H_{\hat{\beta}}$ is given in the following corollary to Theorem 4.

**Corollary 1.** Under $H_0: \beta_1 = \ldots = \beta_N = \beta$, and Assumptions 1 and 2, with $p = 0$ and $q = 1$, as $N, T \to \infty$,

$$H_{\hat{\beta}} \to_d G(x),$$

where $x$ is any real number and $G(x) = \exp(-e^{-x})$ is the Gumbel distribution.

A similar result applies to the maximum Hausman statistic for testing $H_0: \alpha_1 = \ldots = \alpha_N = \alpha$, $H_{\alpha_i}$. The reason for taking an extremum statistic such as the maximum is that it allows for easy interpretation of the test outcome. If the null is accepted, then none of the individual statistics are large enough to cause a rejection, and therefore we conclude the panel can be pooled, at least at the desired level of significance. On the other hand, if the null rejected, then there is at least one unit $i$ for which the individual Hausman statistic is large enough for $\beta_i$ to be deemed different from $\hat{\beta}$, and therefore the panel cannot be pooled.

## 5 Monte Carlo simulations

In this section, we use Monte Carlo simulations to investigate the small-sample size and power of the new predictability tests. The data are generated from (1)–(5) with $\gamma_1 = \ldots = \gamma_N = \gamma$, $\alpha_i = \delta_i = \sigma_{\alpha_i}^2 = \sigma_{\delta_i}^2 = 1$, and $\epsilon_{it} \sim N(0, I_2)$. Moreover, since $\sigma_{\beta_i}^2$ should not affect the results, we set $c_{p1} = \ldots = c_{pN} = c_p$. This will also allow us to focus more on the

---

6Some simulation results for the Hausman poolability tests can be obtained from the corresponding author upon request.
drift parameter in (4), which is generated as $c_{\beta_i} \sim U(a, b)$, implying that $\mu_\beta = (a + b)/2$ and $\sigma^2_\beta = (b - a)^2/12$. Hence, by setting $a$ and $b$ we determine the values taken by $\mu_\beta$ and $\sigma^2_\beta$. The data are generated for 3,000 panels with $N$ cross-sectional and $T + 100$ time series observations, where the first 100 observations for each series are discarded in order to attenuate the effect of $x_{i,0}$, which is set to zero.

The results are reported not only for $LM, LM_\mu, LM_{\sigma^2}$ and $LM^m_\mu$, but also for the modified version of $LM$, defined by $LM^m = LM^m_\mu + LM_{\sigma^2}$. For comparison, the time series tests of Stambaugh (1999) and Lewellen (2004), henceforth denoted $t_{STA}$ and $t_{LEW}$, respectively, are also simulated. As already indicated, these tests are basically bias-adjusted versions of the conventional OLS $t$-test of $H_0: \beta_i = 0$ for unit $i$. All tests are carried out at the 5% significance level, and the rejection rates of the time series tests are averaged across the cross-section.

Consider first the size results reported in Table 1. We see that under the no predictability null and in absence of predictor endogeneity, all tests considered tend to perform well with sizes that are only marginally off the 5% nominal level. There are some distortions, though, especially for $LM, LM_{\sigma^2}$ and $LM^m$, which tend to be somewhat undersized when $T = 100$. However, things do improve as $T$ increases. Indeed, with $T = 400$, size accuracy is almost perfect. But this picture changes quite dramatically as endogeneity is introduced, especially for $LM$ and $LM_\mu$, and the distortions are particularly severe when $c_\rho = -10$, which is just as expected given our asymptotic results (see Remark 4 to Theorem 1). Another expected result is that $LM_{\sigma^2}, LM^m$ and $LM^m_\mu$ are almost unaffected by variations in $c_\rho$ and $\gamma$ (see Remark 1 to Proposition 1 and the discussion following Theorem 3).

Consider next the power results reported in Table 2, in which the tests are set up against the typical alternative with $p = 1/2$ and $q = 1$. The results are largely in accordance with our expectations. First, given the relatively fast rate of shrinking of the local alternative in this case, $LM_{\sigma^2}$ should not have any power beyond size (see Remark 2 to Proposition 1), and this is exactly what we see in the table. Second, we see that power is rather stable in $N$, which is consistent with the fact that theoretically there is no dependence on the sample size. On the other hand, there are some instances with $c_\rho = -10$ when power decreases quite substantially with $T$. However, this effect is mainly a reflection of the poorness of the asymptotic approximation when $T = 100$. Power is also flat in $\sigma^2_\beta$, corroborating the theoretical result that when $p = 1/2$ and $q = 1$ local power should only depend on $\mu_\beta$.

\footnote{Indeed, unreported results suggest that for $T \geq 200$ power is quite flat in the sample size.}
Third, while $LM_{\mu}$ and $LM_{m\mu}$ tend to perform very similarly when $\gamma = 0$, when $\gamma = -0.9$ the former test tends to dominate. Note in particular how the powers of $LM_{\mu}$ and $LM_{m\mu}$ seem to go in opposite directions; the power of $LM_{\mu}$ increases, while that of $LM_{m\mu}$ decreases. This is also in accordance with our expectations.

Fourth, except for $LM_{\sigma^2}$, the panel tests tend to dominate the pure time series tests, and the difference in power is increasing in $N$, suggesting that there are potentially large power gains to be made by exploring the cross-sectional dimension. This is in agreement with the discussion following Proposition 1 (see Remark 3).

Consider next the results reported in Table 3 when $a = -2$ and $b = 2$. As in Table 2, since $p = 1/2$ and $q = 1$ are the same as before, the power of $LM_{c2}$ should be negligible, which is also what we see. However, in contrast to Table 2, in this case the other panel tests also do not seem to rise much above size. The reason for this is that while $\sigma_{\beta}^2 > 0$, here $\mu_\beta = 0$, suggesting that $LM_{\mu}$ and $LM_{m\mu}$ should have negligible local power (see Remark 4 to Proposition 1). The fact that $LM$ and $LM_{\mu}$, and to some extent also $t_{LEW}$, reject more often when $c_\rho = -10$ and $\gamma = -0.9$ is to be expected given their size distortions under the null.

In Table 4 we again set $\mu_\beta = 0$, but this time $p = 1/4$ and $q = 3/4$, suggesting that $LM_{c2}$ should have non-negligible local power (see Proposition 1). Power should also be increasing in $\sigma_{\beta}^2$, although it should not depend on the values taken by $c_\rho$ or $\gamma$. Again, the results are quite suggestive of this. The other tests also have power, which might seem like a contradiction with theory. However, this is not necessarily the case, as the derivation of the asymptotic distributions of these tests are based on a relatively high rate of shrinking of the local alternative and any lower rate, such as the one considered here, will therefore tend to lead to divergence. The same is true for the two time series tests, which should have non-negligible power against alternatives that shrink towards the null hypothesis at rate $T^{-1}$. The rate $N^{-1/4}T^{-3/4}$ is slower in $T$ but faster in $N$. Thus, while decreasing in $N$, the power should tend to increase with $T$, and this is just what we see in the table.

The bulk of the simulation evidence reported above leads to the following practical guideline. First, while formally we require $N, T \to \infty$, in practice the tests seem to perform quite well already when $T = 100$ and $N = 10$. The requirement that $N/T \to 0$ means that while here $T = 100$ seems to be enough for good test performance, this is dependent on

---

8As discussed in Section 3, the power of $LM_{m\mu}$ is expected to go down with $\gamma$. Moreover, focusing on the mean effect, since in this case $(\mu_\beta - \gamma \mu_\rho)^2$ is much larger when $\gamma = -0.9$ than when $\gamma = 0$, the power of $LM_{\mu}$ should increase with $\gamma$. 

---
$N$ not being too large relative to $T$. In the simulations the largest value of $N$ considered is 20, although unreported results suggest that $N$ can be as large as 40 and still the tests perform well. Second, if the predictor is known to contain a unit root, given its superior power properties in the presence of endogeneity, $LM_{\mu}$ should be used. On the other hand, if there is uncertainty over the integratedness of the predictor, as there usually is, then $LM_{m\mu}$ should be used. Third, while $LM_{\mu}$ ($LM_{m\mu}$) is expected to lead to best power, in applications where $\mu_{\beta} = 0$ (and/or $\sigma_{\beta}^2 = 0$) cannot be ruled out a priori, inference should also be based on $LM_{m^2}$. A reasonable approach in such circumstances is therefore to focus on $LM$ ($LM_{m^2}$), which summarizes the evidence. On the one hand, if $LM$ ($LM_{m^2}$) rejects, then there is predictability, and in this case one may want to consider $LM_{\mu}$ ($LM_{m\mu}$) and $LM_{m^2}$ in order to investigate the cause of the rejection. On the other hand, if $LM$ ($LM_{m^2}$) accepts, then the predictability is absent, and therefore there is no point in looking at $LM_{\mu}$ ($LM_{m\mu}$) and $LM_{m^2}$.

6 Empirical results

6.1 Data

The empirical results reported in this section are based on firm-level data from the New York Stock Exchange. The data are sampled at a monthly frequency and cover the period August 1996–August 2010. The size of the cross-section is dictated by data availability. While there are several thousand firms listed at the New York Stock Exchange (NYSE), consistent time series data were available for only 1,559 firms. We extract data on six variables, namely, firm returns, share price, the book-to-market ratio (BM), the cash flow-to-price ratio (CFP), dividends and earnings per share. Dividends are 12-month moving sums of dividends paid on the NYSE index, and earnings are 12-month moving sums of earnings on the same index (see Welch and Goyal, 2008). We use these data to compute the dividend–price (DP) ratio, dividend yield (DY), the price–earnings ratio (PE), and dividend–payout (DE). DP is computed as the log difference between dividends and share price, DY is computed as log difference between dividends and the one period lagged share price, PE is computed as the log difference between earnings and share price, and DE is computed as the log difference between dividends and earnings.

All the data are downloaded from the Datastream database and are organized by sector. In particular, while $\beta_1, \ldots, \beta_N$ are not restricted to be equal, we do require that they are
drawn from the same distribution, which is unlikely to be the case when sampling from across the whole NYSE. One of the most natural splits along these lines is by sector. That is, $\beta_i$ is allowed to differ across firms, and then we also allow $\beta$, $\mu_\beta$ and $\sigma^2_\beta$ to differ by sector. In our sample there are no less than 15 sectors; banking, chemical, electricity, energy, engineering, real estate, technology hardware, household goods, mining, general retailers, software, telecom, transport, travel and leisure, and utilities. Retail is the largest sector and contains $N = 51$ firms. Thus, since $T = 169$, we have $T \gg N$, which is consistent with our theoretical requirement that $N/T \to 0$. The tests that we have developed should therefore be well-suited for the sample at hand.

For the chemical and software sectors we only have consistent time series data for two of the predictors, BM and CFP. Also, some of the predictors have missing observations within the sample range. In these cases, because the missing observations are very few and always single, we use the conventional approach of imputing the average of the two closest time series observations. Log dividends was replaced by zero whenever dividends turned out to be zero. Firms with no dividends are discarded in DP, DY and DE.

6.2 Preliminary results

Before we apply the new tests we need to know how to implement them, and this depends on the extent of the serial and cross-section dependence in the data. If there are no dependencies, the tests can be applied as described in Section 3.2, whereas if the data are dependent, then the test statistics need to be robustified, as discussed in Section 3.3.

In order to infer the significance of the cross-section correlation problem, we compute the pair-wise correlation coefficients of the returns. The simple average of these correlation coefficients across all pairs of stocks, together with the associated CD test discussed in Pesaran et al. (2008), are given in Table 5. The average correlation coefficient ranges between 0.27 and 0.48, and the CD statistic is highly significant for all sectors, which is suggestive of strong cross-section dependence. Thus, when testing the predictability hypothesis we focus on the cross-section dependence robust versions of our test statistics, although the results for the original tests are also reported for comparison.

As a second preliminary we test the variables for unit roots. However, because of the cross-correlations, we cannot use the conventional panel approach of just combining individual augmented Dickey–Fuller (ADF) unit root tests as if they were independent. For this
purpose, we employ the CIPS test of Pesaran (2007), which is based on a common factor-augmented version of the ADF test regression. The test is constructed with a common unit root under the null hypothesis and heterogeneous autoregressive roots under the alternative, suggesting that a rejection of the null should be taken as evidence in favor of stationarity for at least one unit. By contrast, if the null is accepted, we conclude that the panel is non-stationary as a whole. The order of the lag augmentation used to account for serial correlation is selected by the Bayesian information criterion (BIC) where the maximum lag length is allowed to increase with $T$ at the rate $\lfloor 4(T/100)^{2/9} \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of $x$. Also, since some of the predictors appear to be trending, the test regression is fitted with a constant and trend.

The test results reported in Table 6 suggest that the evidence against the unit root null is quite strong. Indeed, even if we look at the conservative 1% level, there is only a handful of cases when the null hypothesis is not rejected. As expected, the test values for returns are largest in absolute value with a preponderance of test values falling in the $(-14, -13)$ interval. The test values for the predictors are smaller but still significantly different from one, suggesting that there are at least one unit for which the autoregressive root is less than one. However, while significant, the estimated roots are still very close to one, suggesting that the predictors exhibit unit root-like behavior, which leads us to conclude that the local-to-unity model in (6) seems appropriate. It also implies that one cannot exclude the possibility that \( \mu \neq 0 \), which means that a rejection by \( LM_\mu \) need not be taken as evidence of predictability. For this reason, in what follows we focus on the modified test statistics.

### 6.3 Predictability test results

Having considered briefly the serial and cross-section correlation properties of the data, we now turn to the test for predictability. Since the predictors are persistent, the predictability tests are implemented using the lag augmentation approach discussed in Section 3.3, where the order of the augmentation is again selected by the BIC.\(^9\) As already mentioned, two versions of the tests are considered. While the first assumes that returns are cross-section uncorrelated, the other one does not, in which case the test regressions are further augmented.

---

\(^9\)As a robustness check, instead of the BIC we employed a sequential general-to-specific test rule based on the last lag. However, this did not have any effect on the conclusions.
with the cross-sectional averages of the observables (as discussed in Section 3.3).\textsuperscript{10}

The results are summarized in Table 7. The first thing to note is that, except possibly for the software, telecom and utilities sectors, for BM and CFP there is strong evidence against the no predictability null. Specifically, we see that while $LM_{\sigma^2}$ is generally highly significant, $LM_{\mu}^{m}$ is generally insignificant. Hence, while there is evidence of predictability at the level of the individual stocks, the predictive slopes average to zero. This means that existing tests for predictability based on estimates of the average predictive slope (see, for example, Hjalmarsson, 2008, 2010; Kauppi, 2000) are likely to be misleading in the sense that the no predictability null is unlikely to be rejected. In fact, it is not difficult to show how the local power of these tests is negligible in the direction of $\mu_\beta = 0$ and $\sigma_{\beta}^2 > 0$. The reason is that the power of $LM_\mu$ is negligible and $LM_\mu$ can be viewed as a squared $t$-statistic. Therefore, since the existing panel tests are all $t$-statistics, their local power properties are very similar to that of $LM_\mu$. That $LM_\mu$ suffers from poor power when the predictive slopes are close to zero is reflected in the results, and the application of the $t$-ration version of the same statistic did not alter the conclusions.\textsuperscript{11}

In view of the Monte Carlo results reported in Section 5, an alternative interpretation of the results for $LM_{\mu}^{m}$ is that there is predictability, even on average, but that the test is not powerful enough to detect it. While the truth is probably somewhere in between these explanations, there are reasons to believe that the averaging out story is more relevant. In Section 6.4 we take a look at the results obtained by applying the Hausman poolability tests of Section 4. Foreshadowing this, one of the findings of that section is that $\beta_i$ is close to zero on average, and that most of the “action” (if any) is coming from the variance. Of course, while certainly interesting, the main issue here is not so much which test that rejects, but if there is any evidence of predictability at all. To determine this we look at $LM^{m}$ (the joint statistic), whose results mirror those of $LM_{\sigma^2}^{m}$. That is, while there is ample evidence of predictability for BM and CFP, for DP, DY, EP and DE the evidence is much weaker.

The fact that BM and CFP appear to be able to predict returns is not surprising. In fact, several time series studies have found that BM predicts either returns or excess returns (see Lettau and Nieuwerburgh, 2008; Campbell and Thompson, 2008; Kothari and Shanken, \textsuperscript{10}In interest of robustness, different methods to eliminate the effect of the cross-section dependence were considered. In particular, as an alternative to using cross-sectional averages of the observables, we used estimated principal components factors. However, this difference only lead to minor differences in the results. \textsuperscript{11}The $t$-ratio results are available upon request, as are some confirmatory results based on the $t$-test of Hjalmarsson (2010).
1997; Lewellen, 2004; Pontiff and Schall, 1998), a finding that has been confirmed also when using cross-section data (see Desai et al., 2004; Pincus et al., 2007). The popularity of BM owes in large part to the findings of Fama and French (1992), who find that individual stocks have the ability to explain cross-sectional variations in stock returns. The main reason why BM appears to be one of the most successful predictors of returns is because, as many studies argue (see, for example, Ball, 1978), it is a ratio of cash flow proxy to the current price level. The price level changes with the discount rate, reflected in the change in BM (Pontiff and Schall, 1998).

As for CFP, while the relationship between cash flow (news) and stock returns has received much attention, the empirical evidence on whether or not CFP predicts returns is scarce. The relevance of cash flows to returns was first analyzed by Sloan (1996), who argued that investors tend to overweight accruals and underweight cash flows when forming future earnings expectations. Therefore, because accruals are less persistent than cash flows, high-accruals firms earn lower abnormal returns when compared to low-accruals firms.

Cohen et al. (2002) argue that cash flow can be seen as a measure of the change in the permanent component of stock prices, such that if expected returns do not change, then the change in stock returns will be the same as the change in the cash flow. Empirical evidence on the relationship between firm-level cash flow and returns tends to indicate that they are positively correlated (see, for example, Vuolteenaho, 2002; Cohen et al., 2002).

Looking next at DP and DY we see that the evidence against the no predictability null is generally very weak, which is in agreement with the results reported by, for example, Campbell and Togo (2006), and Torous et al. (2004), who find that DP does not predict monthly returns. However, our findings are inconsistent with Avramov and Chordia (2006), Lewellen (2004), and Kothari and Shanken (1997), who find that DY predicts returns. The evidence against the null is stronger for the transport sector where the null is rejected at the 10% level or better. The fact that all three tests lead to the same outcome suggests that for this sector the predictability does not cancel out at the aggregate panel level, and therefore that there is evidence of predictability at both the firm and panel levels. The returns for the hardware and retail sectors are also predictable, but here the evidence is more towards a homogenous nonzero predictive slope. Hence, the way that the predictability manifests itself for these sectors when DP and DY are used as predictors is very different from when BM and CFP are used.
If the evidence in favor of predictability was weak for DP and DY, it is even weaker for EP. In fact, even if one considers the liberal 10% level, there is only one rejection, for the hardware sector when using $LM^m_{ij}$ as a test statistic. The evidence of predictability is much stronger for DE. However, just as for DP and DY, we see that the evidence is mainly driven by the heterogeneity of the predictive slopes and not by their mean.

The observed “averaging out” phenomenon is in agreement with the results of Menzly et al. (2004), who use a general equilibrium model to study the cross-sectional differences in return predictability based on DY. According to their results, while time-varying risk preferences induce a positive relation between DY and expected returns, time-varying expected dividend growth induces a negative relation between them in equilibrium. These offsetting effects reduce the ability of DY to forecast future returns. Moreover, the extent of the offset depends on the properties of the asset’s cash flow process, thereby yielding different predictions across different portfolios, which is in agreement with our finding that the extent of predictability tend to vary by sector.

In the model of Menzly et al. (2004) agents have perfect knowledge, which is a rather strong assumption. If this assumption is relaxed, then there is also a possibility of learning over time, which is expected to lead to even more variability in the cross-sectional predictability of returns. This is in agreement with the (gradual) news diffusion models of, for example, Hong et al. (2007), and Rapach et al. (2013), implying that the extent of predictability across stocks is driven partly by information frictions. One source of information (flow) frictions is industry concentration. If industry concentration is high it suggests that investors are most likely to have complete information on just few firms in the market. In light of this, we proxy information flow frictions with industry concentration and investigate whether industry concentration helps explain the sectoral differences in the predictability results. We focus on BM and CFP, which stand out as the most successful of predictors. Following Jiang et al. (2009) industry concentration is calculated as the sum of the earnings share (in %) associated with the eight largest firms in a particular industry. The relationship between return predictability and industry concentration is motivated by Hoberg and

---

12 See also Hjalmarsson (2010), and Rapach et al. (2013) for some confirmatory empirical results based on country-level data.

13 Diffusion of news has been used previously in the literature to explain sector-level differences in predictability. For example, while Hirshleifer et al. (2009) use it to explain differences in the predictive ability of cash flows, in Narayan and Sharma (2011) it is used to explain how returns from different sectors respond differently to oil price shocks.
Phillips (2010), who find that less concentrated industries have more predictable average returns. Our findings are generally consistent with both Jiang et al. (2009) and Hoberg and Phillips (2010). For software, telecom, and utility, where there is no or limited evidence of predictability, industry concentration is in the 60%–65% range. In the other sectors, where the evidence of predictability is relatively high, industry concentration is lower, between 30% and 57%. To make this relationship more clear, we run time series OLS regressions of the form \( \hat{y}_{i,t} = \delta_1 + \delta_2 IC_t + e_{i,t} \), where \( \hat{y}_{i,t} \) is the time-\( t \) return forecast for stock \( i \) belonging to a particular sector, \( IC_t \) is industry concentration for the same sector, and \( e_{i,t} \) is an error term. The results are untabulated to conserve space but are available upon request. Our findings can be summarized as follows. In eight of the 12 sectors where predictability based on BM is strong, industry concentration has a statistically significant (at conventional levels of significance) effect on the forecasted returns, whereas when returns are based on CFP in seven sectors industry concentration has a statistically negative effect. By comparison, when we consider the three sectors where evidence of predictability was weak industry concentration has a statistically significant and positive effect regardless of the predictor used. From these results, it seems that at least for some sectors industry concentration helps explain predictability.

### 6.4 Poolability test results

Table 8 summarizes the poolability test results for the intercept and slope of the predictive regressions. As expected given the predictability test results, we see that in the case of BM and CFP the evidence is mostly against the poolability null. The evidence is strongest with BM in which case we count no less than 12 rejections at the 10% level when using the robust version of our tests. This is in agreement with the models of Menzly et al. (2004), Hong et al. (2007), and Rapach et al. (2013), suggesting that the predictive slopes should tend to vary across stocks. We also see that for BM and CFP the average of the individual slope estimates tend to be close to zero, which is in agreement with the predictability test results for these variables.

The slope coefficients are generally positive, as expected. The main exception is returns when DE is used as a predictor (the only other exception is for BM in the chemical sector). In this case, five sectors appear with a negative slope, three (banking, household, and to some extent also mining) of which are significant according to the predictability test results.
Hence, for these sectors expected returns are high when dividend payout is low. Kothari et al. (2006) find that earnings surprises, contrary to predictions of behavioral models, negatively predict returns (see Hjalmarsson, 2010; Rapach et al., 2013, for similar findings). They make the point that earnings are positively related to inflation and interest rates in that earnings may contain information about future inflation. Dividend payout can also have a similar relationship with inflation. If this is true, a negative slope may be reflecting the negative reaction to inflation of the type shown in the work of Fama and Schwert (1977). Moreover, Campbell and Thompson (2008) argue that it is not uncommon to find negative slope coefficients. They explain that: “A regression estimated over a short sample period can easily generate perverse results, such as a negative coefficient when theory suggests that the coefficient should be positive” (page 1516). They argue that one way to remedy this theoretically inconsistent sign on the slope coefficient is to apply a “sensible restriction” which they define as setting the coefficient to zero whenever it appears with a wrong (negative) sign. The motivation for such a restriction has roots in the idea that typically an investor will not use a perverse coefficient for the purpose of forecasting returns (Campbell and Thompson, 2008). They find that a restricted sign-based model outperforms the benchmark historical average in out-of-sample forecasting evaluations.

As for remaining predictors, DP, DY, EP and DE, we see that the evidence against the null is much weaker than for BM and CFP. We also see that the averages of the individual slope estimates tend to be much smaller in absolute value than before, suggesting that the homogenous slope is also equal to zero. The fact that the null hypothesis of homogenous intercepts also does not meet much resistance implies that expected return does not vary much across stocks, which is in contrast to studies such as Jorion and Goetzman (1999) showing that the equity premium varies across countries. On the other hand, given the great deal of similarity that exists within a sector, the finding that the predictability characteristics also seem to be very similar is maybe not that surprising.

6.5 Economic significance

In this section, we assess the economic significance of our panel predictability test results by using forecasts of returns based on each of the predictors to evaluate trading strategies. In so doing, we adopt the approach of Markowitz (1952), which assumes the existence of a mean-variance investor, whose utility function is given by $E(y_{i,t+1}|I_t) - \tau \var(y_{i,t+1}|I_t)/2$, where $E(y_{i,t+1}|I_t)$ is the expected return on stock $i$ at time $t$, $\tau$ is the investor's coefficient of relative risk aversion, and $\var(y_{i,t+1}|I_t)$ is the variance of the expected return on stock $i$ at time $t$.
where $y_{i,t}$ is again returns for firm $i$ in time period $t$, $I_t$ is the information set available in the same period, and $\tau$ is the coefficient of relative risk aversion. The investor invests in two assets, one is risky, while the other is risk-free, as measured by the three-month US Treasury bill rate. Using $\mathbb{P}_t$ to denote market returns, the proportion invested in the risky asset is set optimally to $E(y_{i,t+1}|I_t) / \tau \text{var}(y_{t+1}|I_t)$ (Marquering and Verbeek, 2004).

Two sets of return forecasts are generated; one is static, while the other is dynamic. The forecasts are generated as follows. The dynamic forecast is simply the one step-ahead forecast obtained fitting (1) by OLS. With this approach we allow an investor to rebalance his portfolio once a month (since we use monthly data). The coefficients of the forecasting model are re-estimated at the end of each month when new information becomes available, so that each month the investor revises his beliefs about expected returns and volatility. The static forecast is the dynamic forecast but with $\beta_1, \ldots, \beta_N$ set to zero, which is usually referred to as the “constant returns” forecast. These forecasts are used to proxy $E(y_{i,t+1}|I_t)$. The next issue is how to proxy $\text{var}(y_{i,t+1}|I_t)$ and $\text{var}(\mathbb{P}_{t+1}|I_t)$. For the first variance we use the estimated variance of the return forecasts, whereas for the second we follow, for example, Marquering and Verbeek (2004), and Westerlund and Narayan (2012), and use a 12-month rolling variance of the return on the NYSE. We set $\tau = 6$, so that the investor is “moderately” risk-averse. Limited borrowing and short selling is permitted by constraining the portfolio weights to lie in $[0, 1.5]$. In forecasting returns under each of the trading strategies, following Marquering and Verbeek (2004), we use a transaction cost of 0.1%, which is deducted from profits. Also, since poolability is not always supported (see Section 6.4), all forecasts are generated using unrestricted OLS.

The results are reported in Table 9. The first thing to note is that the investor utilities associated with the using dynamic trading strategy are generally higher than those associated with the static trading strategy. As for the relative performance of the predictors (under the dynamic strategy), CFP stands out as the overall best predictor leading to the highest return in all but four sectors, namely, banking, chemical, real estate and travel, for which BM or DP (as in the case of real estate) stands out as the best performing predictor. We also see that the returns for BM are much more volatile across sectors than for the other predictors. For example, while the (cross-section) range of the BM returns is equal to $[0.01, 8.22]$, the range of the CFP returns is $[0.23, 4.62]$. The gain from the dynamic strategy is therefore largely idiosyncratic, especially for BM, which is in agreement with the results of Narayan.
Looking next at the estimated utilities we see that they vary substantially across sectors. This is again in agreement with results for returns. Interestingly, while the two measures tend to follow each other very closely for CFP, this is not the case for the other predictors where low (high) returns can be are accompanied by either low or high utilities. Since the main difference between these measures is that the utility not only accounts for the level but also the variance of the returns, this suggests that variance, or risk, is relatively unimportant for CFP. Indeed, except for the banking sector, the utility–return relationship is almost perfectly linear.

The previous section revealed that not all predictors are equally useful. Indeed, while BM and CFP turned out to be quite effective, DE, EP, DY and DP did not. A reasonable conclusion from these results is that investors should be less interested in tracking DE, EP, DY and DP. However, in a recent study, Cenesizoglu and Timmermann (2012) find that some of the predictors that were statistically insignificant nevertheless turned out to be economically significant. In our case, considering only the insignificant predictors, we find that while most insignificant predictors do indeed allow investors to earn significant profits, investor utility turns out to be negative. Therefore, our analysis suggests at best partial evidence that insignificant predictors lead to economically significant outcomes for investors.

7 Concluding remarks

This paper develops a new procedure for testing the null hypothesis of no predictability in panels where the heterogeneity of the predictive slope $\beta_i$ can be assumed to be random across the cross-section. This is quite important since in most, if not all, related work, whenever heterogeneity is allowed, it is assumed to be non-random. This means that each individual coefficient has to be fitted separately, leading to multiple estimation errors in the test procedure. The purpose of the current paper is to device a test that exploits the information that under the null hypothesis the predictability is absent, when a random approach is used, $\beta_i$ has zero mean and zero variance. This leads naturally to the consideration of the LM principle, from which three test statistics are derived. The first is designed to test the joint restriction that $\beta_i$ has zero mean and zero variance, while the other two are designed to test the mean and variance restrictions separately.
The asymptotic distributions of the test statistics are derived and verified in small samples using Monte Carlo simulations. In the empirical part of the paper, based on a large panel comprising 1,559 firms between 1996 and 2010, we find that while CFP and BM are able to predict returns, this is not the case for the other predictors considered.
References


Appendix: Proofs

Lemma A.1. Under Assumptions 1 and 2,

\[ T^{-1/2} x_{i,t} = \frac{1}{\sqrt{T}} \left( x_{0i,t} + N^{-1/2} \sigma_y \sigma_x^{-1} c_{ip} \sum_{s=1}^{t-1} x_{0i,s} \right) + O_p(T^{-1/2}), \]

where \( x_{0i,t} = \sum_{s=1}^{t} \epsilon_{xi,s} \).

Proof of Lemma A.1.

By repeated substitution into (2),

\[ x_{i,t} = \sum_{s=1}^{t} \rho_{i}^{t-s} \delta_{i} (1 - \rho_{i}) + \rho_{i}^{t} x_{i,0} + \sum_{s=1}^{t} \rho_{i}^{t-s} \epsilon_{xi,s}, \]

which, via a first-order Taylor expansion and insertion of \( \rho_{i} = 1 + N^{-1/2} \sigma_y \sigma_x^{-1} c_{ip} \) (Assumption 1), can be rewritten as

\[
T^{-1/2} x_{i,t} = \frac{1}{\sqrt{T}} \sum_{s=1}^{t} \epsilon_{xi,s} + T^{-1/2} x_{i,0} + (NT)^{-1/2} \sigma_y \sigma_x^{-1} c_{ip} \left( t(x_{i,0} + \delta_{i}) + \sum_{s=1}^{t} (t-s) \epsilon_{xi,s} \right) + N^{-1} T^{-3/2} \sigma_y^2 \sigma_x^{-2} c_{ip} \sum_{s=1}^{t} (t-s) \epsilon_{xi,s} + O_p(T^{-1/2}).
\]

The last equality follows from assuming \( x_{i,0} = O_p(1) \), which makes \( T^{-1/2} x_{i,0} \) the leading term. The proof is completed by noting that

\[
\sum_{s=1}^{t} (t-s) \epsilon_{xi,s} = \sum_{s=1}^{l-1} \sum_{j=1}^{s} \epsilon_{xi,j} = \sum_{s=1}^{t-1} x_{0i,s}.
\]

Proof of Theorem 1.

Consider the numerator of \( LM_{1,0} A_{0} \). Since

\[ r_{yi,t} = y_{i,t} - \alpha_i - \gamma_i \Delta x_{i,t} = \epsilon_{yi,t} - \gamma_i \delta_i (\rho_i - 1) + [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1}, \]
we can show that

\[ N^{-1/2}T^{-1}A_{\mu}^{0} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi} \sigma_{xi}^{-1} y_{i,t} x_{i,t-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} y_{i,t} x_{i,t-1} - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} \gamma_{i} (\rho_{i} - 1) x_{i,t-1} \]

\[ + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi} \sigma_{xi}^{-1} [\beta_{i} - \gamma_{i} (\rho_{i} - 1)] x_{i,t-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} y_{i,t} x_{i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} [\beta_{i} - \gamma_{i} (\rho_{i} - 1)] x_{i,t-1} + O_{p}(T^{-1/2}), \]  

(A1)

where the last equality holds, because \( T^{-3/2} \sum_{t=2}^{T} x_{i,t-1} = O_{p}(1) \), and therefore

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} \gamma_{i} \delta_{i} (\rho_{i} - 1) x_{i,t-1} = \frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sigma_{xi}^{-2} \gamma_{i} \delta_{i} c_{pi} \frac{1}{T^{3/2}} \sum_{t=2}^{T} x_{i,t-1} = O_{p}(T^{-1/2}). \]

We similarly have \( T^{-3/2} \sum_{s=2}^{l-2} x_{0i,s} = O_{p}(1) \), where \( x_{0i,l} = \sum_{t=2}^{T} \epsilon_{xi,s} \) is as in Lemma A.1. Since \( \sum_{s=2}^{l-2} \epsilon_{yi,t} x_{0i,s} \) is mean zero and independent across \( i \),

\[ \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=2}^{l-2} \sigma_{xi}^{-2} c_{pi} \epsilon_{yi,t} x_{0i,s} = O_{p}(N^{-1/2}). \]

By using this and Lemma A.1, we obtain

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} \epsilon_{yi,t} x_{i,t-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} \epsilon_{yi,t} x_{0i,t-1} + \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=2}^{l-2} \sigma_{xi}^{-2} c_{pi} \epsilon_{yi,t} x_{0i,s} + O_{p}(\sqrt{NT}^{-1/2}) \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} \epsilon_{yi,t} x_{0i,t-1} + O_{p}(\sqrt{NT}^{-1/2}), \]

and by similar arguments,

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi} \sigma_{xi}^{-1} [\beta_{i} - \gamma_{i} (\rho_{i} - 1)] x_{i,t-1}^{2} \]

\[ = \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{xi}^{-2} (c_{pi} - \gamma_{i} c_{pi}) x_{i,t-1}^{2} = \frac{1}{NT^{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{xi}^{-2} (c_{pi} - \gamma_{i} c_{pi}) x_{0i,t-1}^{2} + O_{p}(T^{-1/2}). \]

Thus, by adding the results,

\[ N^{-1/2}T^{-1}A_{\mu}^{0} = A_{\mu}^{0} + A_{2\mu}^{0} + O_{p}(\sqrt{NT}^{-1/2}), \]

(A2)

38
where
\[
A_{1\mu}^0 = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi} \sigma_{xi}^{-1} \epsilon_{yi,t} x_{0i,t-1},
\]
\[
A_{2\mu}^0 = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \sigma_{xi}^{-2} (c_{\beta i} - \gamma_i c_{\rho i}) x_{0i,t-1}^2.
\]

By a functional central limit theorem, \(T^{-1/2} x_{0i,\lfloor rT \rfloor} \to_d \sigma_{xi} W_{xi}(r)\) and \(T^{-1/2} \sum_{i=2}^T \epsilon_{yi,s} \to_d \sigma_{yi} W_{yi}(r)\) as \(T \to \infty\), where \(\lfloor x \rfloor\) is the integer part of \(x\), and \(W_{xi}(r)\) and \(W_{yi}(r)\) are two standard Brownian motion that are independent of each other. It follows that, by the continuous mapping theorem,
\[
\sigma_{xi}^{-2} \frac{1}{T^2} \sum_{t=2}^T x_{0i,t-1}^2 \to_d \int_0^1 W_{xi}(r)^2 dr,
\]
\[
\sigma_{yi}^{-1} \frac{1}{T} \sum_{t=2}^T x_{0i,t-1} \epsilon_{yi,t} \to_d \int_0^1 W_{xi}(r) dW_{yi}(r).
\]
Hence, by the properties of Brownian motion,
\[
E \left( \sigma_{xi}^{-2} \frac{1}{T^2} \sum_{t=2}^T (c_{\beta i} - \gamma_i c_{\rho i}) x_{0i,t-1}^2 \right) \to (c_{\beta i} - \gamma_i c_{\rho i}) \int_0^1 E[W_{xi}(r)^2 | c_{\beta i}, c_{\rho i}] dr
\]
\[
= \frac{(c_{\beta i} - \gamma_i c_{\rho i})}{2}.
\]
as \(T \to \infty\), and therefore we obtain the following unconditional expectation:
\[
E \left( \sigma_{xi}^{-2} \frac{1}{T^2} \sum_{t=2}^T (c_{\beta i} - \gamma_i c_{\rho i}) x_{0i,t-1}^2 \right) \to \frac{(\mu_{\beta} - \gamma_i \mu_{\rho})}{2}.
\]
The conditions of the law of large numbers of Phillips and Moon (1999, Corollary 1) are satisfied (details are available upon request). It follows that
\[
A_{2\mu}^0 \to_p \frac{(\mu_{\beta} - \gamma_i \mu_{\rho})}{2}
\] (A3)
as \(N, T \to \infty\), where \(\gamma = \lim_{N \to \infty} N^{-1} \sum_{i=1}^N \gamma_i\).

Consider \(A_{1\mu}^0\). Since \(\epsilon_{yi,t}\) and \(x_{0i,t-1}\), and hence also \(dW_{yi}(r)\) and \(W_{xi}(r)\), are independent of each other and across \(i\), it is clear that \(E(A_{1\mu}^0) = 0\). Moreover,
\[
\text{var}(A_{1\mu}^0) = E[(A_{1\mu}^0)^2] \to E \left[ \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N \int_0^1 W_{xi}(r) dW_{yi}(r) \right)^2 \right] = \frac{1}{N} \sum_{i=1}^N \int_0^1 E(W_{xi}(r)^2) dr = \frac{1}{2}
\]
as \(T \to \infty\). The conditions of the central limit theorem of Phillips and Moon (1999, Theorem 2) are satisfied (details are again available upon request). Hence,
\[
A_{1\mu}^0 \to_d \frac{1}{\sqrt{2}} Z_1
\] (A4)
as \( N, T \to \infty \), where \( Z_1 \sim N(0,1) \).

Consider next the denominator of \( LM^\mu_\nu, B^0_\mu \). By Lemma A.1,

\[
N^{-1}T^{-2}B^0_\mu = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} x_{i,t-1}^2
\]

\[
= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} x_{0i,t-1}^2 + O_p(T^{-1/2})
\]

\[
= B^0_1 + O_p(T^{-1/2}) \quad \text{(A5)}
\]

where, by another application of Corollary 1 of Phillips and Moon (1999),

\[
B^0_1 = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} x_{0i,t-1}^2 \overset{p}{\to} \frac{1}{2} \quad \text{(A6)}
\]

as \( N, T \to \infty \). Thus, by putting everything together,

\[
LM^\mu_\nu = \frac{(N^{-1/2}T^{-1}A^0_\mu)^2}{N^{-1}T^{-2}B^0_\mu} = \frac{(A^0_{1 \mu})^2}{B^0_1} + \frac{2A^0_{1 \mu}A^0_{2 \mu}}{B^0_1} + \frac{(A^0_{2 \mu})^2}{B^0_1} + O_p(\sqrt{NT}^{-1/2}), \quad \text{(A7)}
\]

where

\[
\frac{(A^0_{1 \mu})^2}{B^0_1} \overset{d}{\to} Z^2 ,
\]

\[
\frac{2A^0_{1 \mu}A^0_{2 \mu}}{B^0_1} \overset{d}{\to} \sqrt{2}(\mu_\beta - \overline{\gamma}_\mu_\beta)Z_1 ,
\]

\[
\frac{(A^0_{2 \mu})^2}{B^0_1} \overset{p}{\to} \frac{(\mu_\beta - \overline{\gamma}_\mu_\beta)^2}{2}
\]

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

Next, consider \( LM^\mu_{\nu \nu} \). We begin by noting that

\[
\gamma^2_{yi,t} = (\epsilon_{yi,t} - \gamma_i \delta_i (\rho_i - 1) + [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1})^2
\]

\[
= \epsilon^2_{yi,t} + \gamma^2_i \delta^2_i (\rho_i - 1)^2 + [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1}^2 - 2\epsilon_{yi,t} \gamma_i \delta_i (\rho_i - 1) - 2\epsilon_{yi,t} [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1} - 2\gamma_i \delta_i (\rho_i - 1) [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1} \quad \text{(A8)}
\]

The effects of \( [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1}^2 \) and \( \epsilon_{yi,t} [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1} \), which are the leading remainder terms in this expansion, can be deduced from noting that

\[
\frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} [\beta_i - \gamma_i (\rho_i - 1)]^2 x_{i,t-1}^4 = \frac{1}{N^{3/2}T^{7/2}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-4} (\epsilon_{yi,i} - \gamma_i \epsilon_{yi})^2 x_{i,t-1}^4 = O_p((NT)^{-1/2})
\]
\[
\frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2} \sigma_{yi,t}^{-2} \epsilon_{yi,t} [\beta_i - \gamma_i (\rho_i - 1)] x_{i,t-1}^2 = \frac{1}{NT^{5/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-3} \sigma_{yi,t}^{-1} (\epsilon_{yi,t} - \gamma_i \epsilon_{yi}) x_{i,t-1}^3 = O_p((NT)^{-1/2}).
\]

By using this and Lemma A.1,
\[
N^{-1/2} T^{-3/2} A_{11i}^0 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2} \sigma_{yi,t}^{-2} (\epsilon_{yi,t} - \sigma_{yi}^2) x_{i,t-1}^2 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2} \sigma_{yi,t}^{-2} (\epsilon_{yi,t} - \sigma_{yi}^2) x_{i,t-1}^2 + O_p((NT)^{-1/2}) = A_{11i}^0 + O_p((NT)^{-1/2}),
\]

where
\[
A_{11i}^0 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2} \sigma_{yi,t}^{-2} (\epsilon_{yi,t} - \sigma_{yi}^2) x_{i,t-1}^2 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2} \sigma_{yi,t}^{-2} (\epsilon_{yi,t} - \sigma_{yi}^2) x_{i,t-1}^2 + o_p(1),
\]

where the last equality follows by taking deviations from the mean. Clearly, since \(E(\epsilon_{yi,t}^2 - \sigma_{yi}^2) = 0\), we have \(E(A_{11i}^0) = 0\). For the variance, we use the fact that
\[
E \left[ \left( \frac{1}{\sqrt{T}} \sum_{s=2}^{T} \sigma_{yi,s}^{-2} (\epsilon_{yi,s} - \sigma_{yi}^2) \right)^2 \right] = \frac{1}{T} \sum_{s=2}^{T} [\sigma_{yi}^{-4} E(\epsilon_{yi,s}^4 - 1)] = \kappa_y - 1,
\]

where \(\kappa_y = \sigma_{yi}^{-4} E(\epsilon_{yi,s}^4)\), suggesting
\[
\frac{1}{\sqrt{T}} \sum_{s=2}^{T} \sigma_{yi,s}^{-2} (\epsilon_{yi,s} - \sigma_{yi}^2) \rightarrow_{d} \sqrt{\kappa_y - 1} V_i(r)
\]
as \(T \rightarrow \infty\), where \(V_i(r)\) is a standard Brownian motion that is independent of \(W_{xi}(r)\) and \(W_{yi}(r)\) (see McCabe and Tremayne, 1995, Lemma 1). It follows that
\[
\text{var}(A_{11i}^0) = E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sqrt{\kappa_y - 1} \int_{0}^{1} \left( W_{xi}(r)^2 - \int_{0}^{1} W_{xi}(s)^2 ds \right) dV_i(r) \right)^2 \right] = (\kappa_y - 1) \frac{1}{N} \sum_{i=1}^{N} \int_{0}^{1} E \left[ \left( W_{xi}(r)^2 - \int_{0}^{1} W_{xi}(s)^2 ds \right)^2 \right] dr
\]
as \(T \rightarrow \infty\), where
\[
\int_{0}^{1} E \left[ \left( W_{xi}(r)^2 - \int_{0}^{1} W_{xi}(s)^2 ds \right)^2 \right] dr = \int_{0}^{1} E[W_{xi}(r)^4] dr - E \left[ \left( \int_{0}^{1} W_{xi}(r)^2 dr \right)^2 \right].
\]
By using the moments of Brownian motion,
\[
\int_0^1 E[W_{xi}(r)^4]dr = 3 \int_0^1 r^2 dr = 1,
\]
\[
\int_0^1 \int_0^1 E[W_{xi}(r)^2W_{xi}(s)^2]drds = 2 \int_0^1 \int_0^1 (rs + 2s^2)drds = \frac{7}{3} \int_0^1 r^3 dr = \frac{7}{12},
\]
and therefore
\[
\text{var}(A_{1v^2}) \to \frac{5(\kappa - 1)}{12}
\]
as \(N, T \to \infty\), which, via Theorem 2 of Phillips and Moon (1999), yields
\[
A_{1v^2} \to_d \frac{\sqrt{5(\kappa_y - 1)}}{\sqrt{12}} Z_2 \quad \text{(A10)}
\]
as \(N, T \to \infty\), where \(Z_2 \sim N(0, 1)\) is independent of \(Z_1\), as follows from the fact that \(W_{xi}(r)\) and \(V_{ji}(r)\) are independent for all \((i, j)\).

As for the denominator of \(LM_{v^2}^0\), by using Lemma 1 and the fact that \(A_{1v^2} = O_p(1)\),
\[
N^{-1}T^{-3}B_{v^2}
\]
\[
= \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi}^2 \sigma_{xi}^{-4} (2\sigma_{yi,t}^2 - \sigma_{yi}^2)x_{0,i,t-1}^4
\]
\[
= \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi}^2 \sigma_{xi}^{-4} \epsilon_{yi,t}^2 x_{0,i,t-1}^4 - \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi}^2 \sigma_{xi}^{-4} (\epsilon_{yi,t}^2 - \sigma_{yi}^2)x_{0,i,t-1}^4 + O_p(T^{-1/2})
\]
\[
= B_{v^2}^0 + O_p(T^{-1/2}),
\]
where, via Corollary 1 of Phillips and Moon (1999),
\[
B_{v^2}^0 = \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=2}^T \sigma_{yi}^2 \sigma_{xi}^{-4} \epsilon_{yi,t}^2 x_{0,i,t-1}^4 \to_p \int_0^1 E[W_{xi}(r)^4]dr = 1 \quad \text{(A11)}
\]
as \(N, T \to \infty\). It follows that
\[
LM_{v^2}^0 = \frac{12}{5(\kappa_y - 1)} \frac{(N^{-1/2}T^{-3/2}A_{1v^2})^2}{N^{-1}T^{-3}B_{v^2}} = \frac{12}{5(\kappa_y - 1)} \frac{(A_{1v^2})^2}{B_{v^2}^0} + O_p(T^{-1/2}), \quad \text{(A12)}
\]
where
\[
\frac{(A_{1v^2})^2}{B_{v^2}^0} \to_d \frac{5(\kappa_y - 1)}{12} Z_2^2
\]
as \(N, T \to \infty\). This completes the proof. \(\blacksquare\)

**Proof of Proposition 1.**
The proof of Proposition 1 follows by simple manipulations of the proof of Theorem 1, and hence only essential details will be provided. We begin by considering $A_{12}^0$. With $p = 1/4$ and $q = 3/4$, all terms in the expansion of $r_{yi,t}^2$ (see the proof of Theorem 1) are negligible, except for $c_{yi,t}^2$ and $\beta_i^2 x_{t,-1}^2$, suggesting that, via Lemma A.1,

$$N^{-1/2} T^{-3/2} A_{12}^0 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi,t}^2 (r_{yi,t}^2 - \sigma_{yi,t}^2) x_{t,-1}^2$$

$$= \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi,t}^2 (c_{yi,t}^2 - \sigma_{yi,t}^2) x_{t,-1}^2 + \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi,t}^2 (c_{yi,t}^2 - \sigma_{yi,t}^2) x_{t,-1}^2 + o_p(1)$$

$$= A_{12}^0 + A_{22}^0 + o_p(N),$$

where $A_{12}^0$ is as in the proof of Theorem 1. Moreover, since $E(c_{yi,t}^2) = \mu_\beta^2 + \sigma_\beta^2$,

$$A_{22}^0 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi,t}^2 (c_{yi,t}^2 - \sigma_{yi,t}^2) x_{t,-1}^2$$

$$\rightarrow_d E(c_{yi,t}^2) \int_0^1 E[W_{xi}(r)^4]dr = \mu_\beta^2 + \sigma_\beta^2$$

as $N, T \rightarrow \infty$, suggesting that

$$(N^{-1/2} T^{-3/2} A_{12}^0)^2 \rightarrow_d \frac{\sqrt{5(k_y - 1)}}{\sqrt{12}} Z_2 + \mu_\beta^2 + \sigma_\beta^2. \tag{A13}$$

Thus, since the denominator of $LM_{12}^0$ is unaffected by the change in $(p, q)$ (when compared to the case considered in Theorem 1), we have

$$LM_{12}^0 = \frac{12}{5(k_y - 1)} \frac{(A_{12}^0 + A_{22}^0)^2}{B_{12}} + o_p(1)$$

$$= \frac{12}{5(k_y - 1)} \left( \frac{(A_{12}^0)^2}{B_{12}} + \frac{(A_{22}^0)^2}{B_{22}} + \frac{2A_{12}^0 A_{22}^0}{B_{12}} \right) + o_p(1), \tag{A14}$$

where the first term on the right-hand side is as before and

$$\frac{A_{12}^0 A_{22}^0}{B_{12}} \rightarrow_d \frac{\mu_\beta^2 + \sigma_\beta^2 \sqrt{5(k_y - 1)}}{\sqrt{12}} Z_2,$$

$$\frac{(A_{22}^0)^2}{B_{22}} \rightarrow_p (\mu_\beta^2 + \sigma_\beta^2)^2$$

as $N, T \rightarrow \infty$, and so the proof is complete.

**Proof of Theorem 2.**
We begin by considering \( \hat{\alpha} \), which, via \( \alpha_i = \alpha + T^{-1/2} \sigma_{yi} c_{ai} \), can be expanded in the following way:

\[
\hat{\alpha} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (y_{i,t} - \hat{\gamma}_i \Delta x_{i,t})
\]

\[
= \alpha + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (\beta_i x_{i,t-1} + \gamma_i \epsilon_{xi,t} + \epsilon_{yi,t} - \hat{\gamma}_i \Delta x_{i,t}) + \frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sigma_{yi} c_{ai}
\]

\[
= \alpha + R_1 + R_2,
\]

where

\[
R_2 = \frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sigma_{yi} c_{ai} = O_p((NT)^{-1/2}).
\]

\[R_1\] can be expanded in the following way:

\[
R_1 = \frac{1}{N} \sum_{i=1}^{N} [(\beta_i - \gamma_i (\rho_i - 1)) \bar{x}_{i,-1} + \bar{\epsilon}_{yi} - (\hat{\gamma}_i - \gamma_i) \bar{\Delta x}_i - \gamma_i \delta_i (\rho_i - 1)]
\]

with \( \bar{x}_{i,-1} = T^{-1} \sum_{t=2}^{T} x_{i,t-1} \) and similar definitions of \( \bar{\epsilon}_{yi} \) and \( \bar{\Delta x}_i \). \( \epsilon_{yi,t} \) is independent across both \( i \) and \( t \). Therefore, \( \bar{\epsilon}_{yi} = N^{-1} \sum_{i=1}^{N} \bar{\epsilon}_{yi} = O_p((NT)^{-1/2}) \), and, by the cross-section independence of \( c_{pi} \), \( N^{-1} \sum_{i=1}^{N} \gamma_i \delta_i (\rho_i - 1) = N^{-3/2} T^{-1} \sum_{i=1}^{N} \gamma_i \delta_i \sigma_{yi} \sigma_{xi}^{-1} c_{pi} = O_p(N^{-1/2} T^{-1}) \).

As for the first term in \( R_1 \), we have

\[
\frac{1}{N} \sum_{i=1}^{N} [\beta_i - \gamma_i (\rho_i - 1)] \bar{x}_{i,-1} = \frac{1}{N^{3/2} T} \sum_{i=1}^{N} \sigma_{yi} \sigma_{xi}^{-1} (\epsilon_{pi} - \gamma_i \epsilon_{pi}) \bar{x}_{i,-1} = O_p(N T^{-1/2}).
\]

Finally, as for the third term, since \( \Delta \bar{x}_i = \delta_i (1 - \rho_i) + (\rho_i - 1) \bar{x}_{i,-1} + \bar{\epsilon}_{xi} \),

\[
\frac{1}{N} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \Delta \bar{x}_i = -\frac{1}{N^{3/2} T} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \delta_i \sigma_{yi} \sigma_{xi}^{-1} c_{pi} + \frac{1}{N^{3/2} T} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \sigma_{yi} \sigma_{xi}^{-1} c_{pi} \bar{x}_{i,-1} + \frac{1}{N} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \bar{\epsilon}_{xi}.
\]

In the proof of Theorem 3 we show that \( \sqrt{T} (\hat{\gamma}_i - \gamma_i) = O_p(1) \). By using this and the Cauchy–Schwarz inequality, we obtain

\[
\frac{1}{N^{3/2} T} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \delta_i \sigma_{yi} \sigma_{xi}^{-1} c_{pi} \leq \frac{1}{\sqrt{NT^{3/2}}} \left( \frac{1}{N} \sum_{i=1}^{N} |\sqrt{T} (\hat{\gamma}_i - \gamma_i)|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} (\delta_i \sigma_{yi} \sigma_{xi}^{-1} c_{pi})^2 \right)^{1/2} = O_p(N^{-1/2} T^{-3/2}),
\]

and by the same argument,

\[
\frac{1}{N^{3/2} T} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \sigma_{yi} \sigma_{xi}^{-1} c_{pi} \bar{x}_{i,-1} = O_p(N^{-1/2} T^{-1}),
\]

\[
\frac{1}{N} \sum_{i=1}^{N} (\hat{\gamma}_i - \gamma_i) \bar{\epsilon}_{xi} = O_p(N^{-1/2} T^{-1}).
\]
Under our assumption that \( N/T \to 0 \), we have \( O_p(N^{-1/2}T^{-1/2}) > O_p(N^{-1/2}T^{-1}) \). The leading terms in \( R_1 \) are therefore given by the cross-sectional averages of \( [\beta_i - \gamma_i(\rho_i - 1)]\bar{x}_{i,t-1} \) and \( \bar{e}_{yi} \). It follows that
\[
R_1 = O_p((NT)^{-1/2}) + O_p(N^{-1}T^{-1/2}).
\] (A17)

Let us now consider \( \hat{r}_{yi,t} \). By adding and subtracting \( \alpha_i \) and \( \gamma_i\Delta x_{i,t} \),
\[
\hat{r}_{yi,t} = yi,t - \hat{\alpha} - \hat{\gamma}_i\Delta x_{i,t} = r_{yi,t} - (\hat{\alpha} - \alpha_i) - (\hat{\gamma}_i - \gamma_i)\Delta x_{i,t}.
\]

Insertion of \( r_{yi,t} = e_{yi,t} - \gamma_i(\rho_i - 1) + [\beta_i - \gamma_i(\rho_i - 1)]x_{i,t-1} \) now yields
\[
\hat{r}_{yi,t} = r_{yi,t} - (\hat{\alpha} - \alpha) + T^{-1/2}\sigma_{yi}c_{ai} - (\hat{\gamma}_i - \gamma_i)\Delta x_{i,t}
\]
\[
= r_{yi,t} - (\hat{\gamma}_i - \gamma_i)\Delta x_{i,t} - (R_1 + R_2) + T^{-1/2}\sigma_{yi}c_{ai}
\]
\[
= e_{yi,t} - \gamma_i\delta_i(\rho_i - 1) + [\beta_i - \gamma_i(\rho_i - 1)]x_{i,t-1} - (\hat{\gamma}_i - \gamma_i)\Delta x_{i,t} - (R_1 + R_2)
\]
\[
+ T^{-1/2}\sigma_{yi}c_{ai},
\]

which can be used to show that (as is done in the proof of Theorem 4)
\[
\sigma_{yi}^2 = \sigma_{yi}^2 + O_p(T^{-1/2}),
\]
and similarly, \( \sigma_{xi}^2 = \sigma_{xi}^2 + O_p(T^{-1/2}) \) and \( \bar{e}_y = \kappa_y + O_p(T^{-1/2}) \). Thus, by using Taylor expansion of the inverse square root and then substitution for \( \hat{r}_{yi,t} \),
\[
N^{-1/2}T^{-1}A_{\mu} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{T} \frac{\sigma_{yi}^{-1}\sigma_{xi}^{-1}}{\sigma_{yi}^{-1}\sigma_{xi}^{-1}} \hat{r}_{yi,l}x_{l,t-1}
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{T} \sigma_{yi}^{-1}\sigma_{xi}^{-1} \hat{r}_{yi,l}x_{l,t-1} + O_p(\sqrt{NT^{-1/2}})
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{T} \sigma_{yi}^{-1}\sigma_{xi}^{-1} \hat{r}_{yi,l}x_{l,t-1} - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{T} \sigma_{yi}^{-1}\sigma_{xi}^{-1}\gamma_i\delta_i(\rho_i - 1)x_{i,t-1}
\]
\[
+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{T} \sigma_{yi}^{-1}\sigma_{xi}^{-1}[\beta_i - \gamma_i(\rho_i - 1)]x_{i,t-1}^2
\]
\[
- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{l=1}^{T} \sigma_{yi}^{-1}\sigma_{xi}^{-1}(\hat{\gamma}_i - \gamma_i)\Delta x_{i,t}x_{i,t-1} - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{yi}^{-1}\sigma_{xi}^{-1}\bar{x}_{i,-1}(R_1 + R_2)
\]
\[
+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{yi}^{-1}c_{ai} + O_p(\sqrt{NT^{-1/2}}).
\] (A18)

There are five terms to consider, of which the first and third are the same as in the proof of
Theorem 1. As for the second term, by the Cauchy–Schwarz inequality,
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i \delta_i (\rho_i - 1) x_{i,t-1} \\
= \frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sigma_{xi}^{-2} \gamma_i \delta_i c_{\rho_i} \frac{1}{T^{3/2}} \sum_{t=2}^{T} x_{i,t-1} \\
\leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{xi}^{-4} \gamma_i \delta_i c_{\rho_i}^2 \right)^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T^{3/2}} \sum_{t=2}^{T} x_{i,t-1} \right)^2 \right]^{1/2} = O_p(T^{-1/2}),
\]
and by the same inequality the order of the fourth term is given by
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} (\hat{\gamma}_i - \gamma_i) \Delta x_{i,t} x_{i,t-1} \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \sqrt{T} (\hat{\gamma}_i - \gamma_i) \frac{1}{T} \sum_{t=2}^{T} \Delta x_{i,t} x_{i,t-1} \\
\leq \sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-2} T (\hat{\gamma}_i - \gamma_i)^2 \right)^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=2}^{T} \Delta x_{i,t} x_{i,t-1} \right)^2 \right]^{1/2} = O_p(\sqrt{NT^{-1/2}}).
\]
Next, consider the fifth term. From \( \bar{\epsilon}_y = N^{-1} \sum_{i=1}^{N} \bar{\epsilon}_{yi} = O_p((NT)^{-1/2}) \) and
\[
\bar{\epsilon}_{i,-1} = \frac{1}{N} \sum_{i=1}^{N} \bar{\epsilon}_{i,-1} = \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} T^{-1/2} \bar{\epsilon}_{i,-1} = O_p(\sqrt{TN^{-1/2}}),
\]
we obtain
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \bar{\epsilon}_{i,-1} \frac{1}{N} \sum_{j=1}^{N} [\beta_j - \gamma_j (\rho_j - 1)] \bar{x}_{j,-1} \\
= \frac{1}{NT} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \bar{\epsilon}_{i,-1} \frac{1}{N} \sum_{j=1}^{N} \sigma_{yi} \sigma_{xi}^{-1} (\epsilon \beta_j - \gamma_j c_{\rho_j}) \bar{x}_{j,-1} = O_p(N^{-1}),
\]
and
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \bar{\epsilon}_{i,-1} \bar{\epsilon}_y = \frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \bar{\epsilon}_{i,-1} \sqrt{NT} \bar{\epsilon}_y = O_p(N^{-1/2}).
\]
Hence, by leading term approximation,
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \bar{\epsilon}_{i,-1} R_1 = O_p(N^{-1/2}).
\]
The term involving \( R_2 \) is of the same order, as seen by writing
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \bar{\epsilon}_{i,-1} R_2 = \frac{1}{N^{3/2} \sqrt{T}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \sum_{j=1}^{N} \sigma_{yi} \sigma_{xj} \bar{\epsilon}_{k} = O_p(N^{-1/2}).
\]
The sixth and final term is
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma^{-1}_{x_i} c_{ai} = O_p(T^{-1/2}).
\]
Thus, by putting everything together,
\[
N^{-1/2}T^{-1} A_{\mu} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-1}_{y_i} \sigma^{-1}_{x_i} r_{y_i,t} x_{i,t-1}
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-1}_{y_i} \sigma^{-1}_{x_i} c_{y_i,t} x_{i,t-1}
\]
\[
+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-1}_{y_i} \sigma^{-1}_{x_i} [\beta_i - \gamma_i \rho_i (\rho_i - 1)] x_{i,t-1}^2 + O_p(\sqrt{NT}^{-1/2})
\]
\[
= A^0_{1\mu} + A^0_{2\mu} + O_p(\sqrt{NT}^{-1/2}),
\]
where \(A^0_{1\mu}\) and \(A^0_{2\mu}\) are the same as in the proof of Theorem 1, and where the last equality holds because of Lemma A.1. Similarly, since
\[
N^{-1}T^{-2}B_{\mu} = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma^{-2}_{x_i} x_{i,t-1}^2 = B^0_{1\mu} + O_p(T^{-1/2}),
\]
where \(B^0_{1\mu}\) is again as in the proof of Theorem 1, we can show that
\[
LM_{\mu} = LM^0_{\mu} + O_p(\sqrt{NT}^{-1/2}).
\]
Consider \(LM_{\mu^2}\). In particular, consider
\[
N^{-1/2}T^{-3/2} A_{\mu^2} = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-2}_{x_i} \sigma^{-2}_{y_i} (r^2_{y_i,t} - \sigma^2_{y_i}) x_{i,t-1}^2
\]
\[
= \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-2}_{x_i} \sigma^{-2}_{y_i} (r^2_{y_i,t} - \sigma^2_{y_i}) x_{i,t-1}^2 + O_p(\sqrt{NT^{-1/2}}),
\]
where
\[
r^2_{y_i,t} = [r_{y_i,t} (\hat{\alpha} - \alpha_i) - (\hat{\gamma}_i - \gamma_i) \Delta x_{i,t}]^2
\]
\[
= r^2_{y_i,t} - 2r_{y_i,t} (\hat{\alpha} - \alpha_i) - 2r_{y_i,t} (\hat{\gamma}_i - \gamma_i) \Delta x_{i,t} + (\hat{\alpha} - \alpha_i)^2 + 2(\hat{\alpha} - \alpha_i)(\hat{\gamma}_i - \gamma_i) \Delta x_{i,t}
\]
\[
+ (\hat{\gamma}_i - \gamma_i)^2 (\Delta x_{i,t})^2.
\]
Hence, there are six terms to consider in \(N^{-1/2}T^{-3/2} A_{\mu^2}\). Since \(\hat{\alpha} - \alpha_i = \hat{\alpha} - \alpha - T^{-1/2} \sigma_{y_i} c_{ai} = R_1 + R_2 - T^{-1/2} \sigma_{y_i} c_{ai}\), by the Cauchy–Schwarz inequality, the term involving \(r_{y_i,t} (\hat{\alpha} - \alpha_i)\) can be written as
\[
\frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-2}_{x_i} \sigma^{-2}_{y_i} r_{y_i,t} (\hat{\alpha} - \alpha_i) x_{i,t-1}^2
\]
\[
= (R_1 + R_2) \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-2}_{x_i} \sigma^{-2}_{y_i} r_{y_i,t} x_{i,t-1}^2 - \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma^{-2}_{x_i} \sigma^{-2}_{y_i} c_{ai} r_{y_i,t} x_{i,t-1}^2.
\]
By the same arguments used in the proof of Theorem 1, we have

\[
\frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} r_{yi,t}^2 x_{i,t-1}^2
\]

\[
= \frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} e_{yi,t} x_{i,t-1}^2 - \frac{1}{N T^{5/2}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-3} \sigma_{yi}^{-1} \gamma_i \delta_i c_{pi} x_{i,t-1}^2
\]

\[
+ \frac{1}{N T^{5/2}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-3} \sigma_{yi}^{-1} (c_{pi} - \gamma_i c_{pi}) x_{i,t-1}^3,
\]

where

\[
\frac{1}{T^{3/2}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-1} e_{yi,t} x_{i,t-1}^2 \to_d \int_0^1 W_{xi}(r)^2 dW_{yi}(r),
\]

as \( T \to \infty \), which is clearly mean zero and independent across \( i \). The first term in the expansion of \( N^{-1/2} T^{-3/2} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} r_{yi,t} x_{i,t-1}^2 \) is therefore \( O_p(1) \). The second and third terms are \( O_p((NT)^{-1/2}) \) and \( O_p(N^{-1/2}) \), respectively, and therefore

\[
\frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} r_{yi,t} x_{i,t-1}^2 = O_p(1).
\]

Similarly,

\[
\frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-1} c_{ai} r_{yi,t} x_{i,t-1}^2 = O_p(T^{-1/2}).
\]

Hence, since \( R_1 = O_p((NT)^{-1/2}) + O_p(N^{-1}T^{-1/2}) \),

\[
\frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} r_{yi,t} (\delta - \gamma_i) x_{i,t-1}^2 = O_p(T^{-1/2}).
\]

The term involving \((\hat{\gamma}_i - \gamma_i)^2(\Delta x_{i,t})^2\) can be expanded as

\[
\frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} (\hat{\gamma}_i - \gamma_i)^2 (\Delta x_{i,t})^2 x_{i,t-1}^2
\]

\[
= \frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-2} (\hat{\gamma}_i - \gamma_i)^2 x_{i,t-1}^2
\]

\[
+ \frac{1}{\sqrt{N T^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{xi}^{-2} \sigma_{yi}^{-2} (\hat{\gamma}_i - \gamma_i)^2 [(\Delta x_{i,t})^2 - \sigma_{xi}^2] x_{i,t-1}^2.
\]

By the same arguments used in the proof of Theorem 1, we have \( T^{-1/2} \sum_{i=2}^{t} (\epsilon_{xi,t}^2 - \sigma_{xi}^2) = O_p(1) \), which can in turn be used to show that

\[
\frac{1}{T^{3/2}} \sum_{i=2}^{T} [(\Delta x_{i,t})^2 - \sigma_{xi}^2] x_{i,t-1}^2 = O_p(1).
\]
Hence,
\[
\frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2}(\gamma_i - \gamma) \left[ (\Delta x_{i,t})^2 - \sigma_{xi}^2 \right] x_{i,t-1}^2 \\
\leq \frac{\sqrt{N}}{T} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sigma_{yi}^{-2}(\gamma_i - \gamma) \left[ (\Delta x_{i,t})^2 - \sigma_{xi}^2 \right] \right)^{1/2} \\
\leq O_p(\sqrt{NT^{-1}}).
\]

We similarly have
\[
\frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-2}(\gamma_i - \gamma) x_{i,t-1}^2 \\
\leq \frac{\sqrt{N}}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T^{3/2}} \sum_{t=2}^{T} \left[ (\Delta x_{i,t})^2 - \sigma_{xi}^2 \right] \right) \right)^{1/2} = O_p(\sqrt{NT^{-1/2}}).
\]

The effect of the remaining terms in the expansion of \( \hat{r}_{yi,t}^2 \) is of lower order than this. Hence,
\[
N^{-1/2}T^{-3/2} A_{\nu^2}^0 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-2}(\gamma_i - \gamma) (\hat{r}_{yi,t}^2 - \sigma_{yi}^2) x_{i,t-1}^2 + O_p(\sqrt{NT^{-1/2}}) \\
= A_{1\nu^2}^0 + O_p(\sqrt{NT^{-1/2}}), \tag{A24}
\]

where \( A_{1\nu^2}^0 \) is as in the proof of Theorem 1. Similarly,
\[
N^{-1}T^{-3} B_{\nu^2} = \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-2} \hat{\gamma}_i^{-4} (\hat{r}_{yi,t}^2 - \sigma_{yi}^2) x_{i,t-1}^4 = B_{1\nu^2}^0 + O_p(T^{-1/2}), \tag{A25}
\]

where \( B_{1\nu^2}^0 \) is again the same as in the proof of Theorem 1. Hence,
\[
LM_{\nu^2} = \frac{12}{5(k_y - 1)} A_{\nu^2}^0 B_{\nu^2} = \frac{12}{5(k_y - 1)} \frac{A_{1\nu}^0}{B_{1\nu}^0} + O_p(\sqrt{NT^{-1/2}}) = LM_{\nu^2}^0 + O_p(\sqrt{NT^{-1/2}}), \tag{A26}
\]

which completes the proof.

\[\Box\]

**Proof of Theorem 3.**

The inclusion of a common intercept in \( \hat{\theta} \) and \( \hat{\rho}_i \) does not affect the results. In what follows we therefore disregard the effect of the intercept.

From the proof of Theorem 2,
\[
\hat{r}_{yi,t} - [\hat{\beta}_i - \gamma_i(\hat{\rho}_1 - 1)] x_{i,t-1} = e_{yi,t} - \gamma_i \delta_i (\hat{\rho}_1 - 1) - (\gamma_i - \gamma) \Delta x_{i,t} - (R_1 + R_2) + T^{-1/2} \sigma_{yi} e_{ai}.
\]
which in turn implies

\[ N^{-1/2}T^{-1}A_\mu - \sqrt{N}\hat{\theta} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \sigma_{yi,i}^{-1} \rho_{yi,i} x_{i,t-1} + \sqrt{N}\hat{\theta} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \rho_{yi,i} x_{i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \rho_{yi,i} x_{i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} \]

where the first two terms can be analyzed in the same way as in the proof of Theorem 2. The third term can be written as

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \rho_{yi,i} x_{i,t-1} + O_p(T^{-1/2}) \]

where the second equality makes use of the Cauchy–Schwarz inequality, from which it follows that

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \gamma_i (\hat{\rho}_i - 1) x_{i,t-1} \]

\[ \leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \gamma_i \right)^2 \right)^{1/2} \leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{t=2}^{T} \sigma_{yi,i}^{-1} \gamma_i \right)^2 \right)^{1/2} = O_p(T^{-1/2}) \]

As for the third equality, note that

\[ \hat{\rho}_i = \frac{\sum_{t=2}^{T} x_{i,t-1} x_{i,t}}{\sum_{t=2}^{T} x_{i,t-1}^2} = \rho_i + \delta_i (1 - \rho_i) \frac{\sum_{t=2}^{T} x_{i,t-1} x_{i,t}}{\sum_{t=2}^{T} x_{i,t-1}^2} + \frac{\sum_{t=2}^{T} x_{i,t-1}^2 e_{i,t}}{\sum_{t=2}^{T} x_{i,t-1}^2} \]
simplifying that

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i (\hat{\rho}_i - \rho_i) x_{i,t-1}^2
\]

\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i \left( \delta_i (1 - \rho_i) \frac{\sum_{t=2}^{T} x_{i,t-1}^2}{\sum_{t=2}^{T} x_{i,t-1}^2} + \frac{\sum_{i=2}^{T} x_{i,t-1} \epsilon_{x_i,t}}{\sum_{t=2}^{T} x_{i,t-1}^2} \right) \sum_{t=2}^{T} x_{i,t-1}^2
\]

\[
= - \frac{1}{NT^2} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i \delta_i \epsilon_{pi} \sum_{t=2}^{T} x_{i,t-1}^2 + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i \sum_{t=2}^{T} x_{i,t-1} \epsilon_{x_i,t}
\]

Moreover, by Lemma A.1,

\[
\sigma_{xi}^{-1} \frac{1}{T} \sum_{i=1}^{T} \epsilon_{xi,t} x_{i,t-1} = \sigma_{xi}^{-1} \frac{1}{T} \sum_{i=1}^{T} \epsilon_{xi,t} x_{0i,t-1} + \sigma_{yi} \sigma_{xi}^{-1} \frac{cpi}{\sqrt{NT^2}} \sum_{i=1}^{T} \epsilon_{xi,t} \sum_{s=2}^{T} x_{0i,s}
\]

\[
\rightarrow^d \int_0^1 W_{xi}(r) dW_{xi}(r) + \sigma_{yi} \sigma_{xi}^{-1} \frac{cpi}{\sqrt{N}} \int_0^1 \int_0^r W_{xi}(s) ds dW_{xi}(r)
\]

as \( T \to \infty \). The mean and variance of the first of the two limiting distributions are equal to zero and \( 1/2 \), respectively. The mean of the second distribution is zero too. It follows that

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i \epsilon_{x_i,t} x_{i,t-1}
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{xi} \sigma_{yi}^{-1} \gamma_i \sigma_{xi}^{-1} \frac{1}{T} \sum_{i=1}^{T} \epsilon_{x_i,t} x_{0i,t-1} + \frac{1}{N} \sum_{i=1}^{N} \sigma_{xi} \gamma_i \sigma_{xi}^{-1} \frac{1}{T} \sum_{i=1}^{T} \epsilon_{x_i,t} \sum_{s=2}^{T} x_{0i,s}
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_{xi} \sigma_{yi}^{-1} \gamma_i \sigma_{xi}^{-1} \frac{1}{T} \sum_{i=1}^{T} \epsilon_{x_i,t} x_{0i,t-1} + O_p \left( N^{-1/2} \right) \rightarrow^d \frac{\omega^2}{\sqrt{2}} Z_4
\]

as \( N, T \to \infty \), where \( \omega^2 = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sigma_{xi}^2 \sigma_{yi}^{-2} \gamma_i^2 \) and \( Z_4 \sim N(0,1) \) is independent of \( Z_1 \) and \( Z_2 \). Making use of this and the results provided in the proof of Theorem 2,

\[\begin{align*}
N^{-1/2}T^{-1} \Lambda_{\mu} - \sqrt{N} \theta &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \epsilon_{yi,i} x_{0i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \beta_i x_{0i,t-1} \\
&+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \sigma_{yi}^{-1} \sigma_{xi}^{-1} \gamma_i \epsilon_{xi,t} x_{0i,t-1} + O_p \left( \sqrt{NT}^{-1/2} \right) \\
&\rightarrow^d \frac{\mu^\theta}{2} + \frac{1}{\sqrt{2}} Z_1 + \frac{\omega}{\sqrt{2}} Z_4
\end{align*}\]

(A28)

as \( N, T \to \infty \). Moreover, since \( Z_1 \) and \( Z_4 \) are independent, we have that \( Z_1 / \sqrt{2} + \omega Z_4 / \sqrt{2} \sim \sqrt{1 + \omega^2} Z_3 / \sqrt{2} \), where \( Z_3 \sim N(0,1) \). By using this and \( \omega^2 - \omega^2 = o_p(1) \),

\[
\frac{\sqrt{2}}{\sqrt{1 + \omega^2}} (N^{-1/2}T^{-1} \Lambda_{\mu} - \sqrt{N}) \rightarrow^d \frac{\mu^\theta}{\sqrt{2}(1 + \omega^2)} + Z_3
\]

(A29)
which in turn implies
\[
LM^m_\mu \rightarrow_d \frac{\mu^2_\beta}{2(1 + \omega^2)} + \frac{\sqrt{2\mu_\beta}}{\sqrt{1 + \omega^2}}Z_3 + Z^2_3,
\]
and so the proof is complete.

**Proof of Theorem 4.**

It is convenient to rewrite (1) and (3) as
\[
y_{i,t} = \alpha_i + \beta_i x_{i,t-1} + \gamma_i (\epsilon_{xi,t} - \Delta x_{i,t-1}) + \epsilon_{yi,t}
\]
\[
= \alpha_i + \beta_i x_{i,t-1} + \gamma_i \Delta x_{i,t} + v_{i,t}
\]
\[
= x_{i,t}^\prime \beta_i + v_{i,t}
\]
where \( x_{i,t} = (1, x_{i,t-1}, \Delta x_{i,t})\)'s, \( \beta_i = (\alpha_i, \beta_i, \gamma_i)' \) and \( v_{i,t} = \gamma_i (\epsilon_{xi,t} - \Delta x_{i,t-1}) + \epsilon_{yi,t} \). In what follows we will use \( \beta_i = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i)' \) to denote the OLS estimator of \( \beta_i \) and \( \sigma^2_{yi} = T^{-1} \sum_{t=2}^T (y_{i,t} - x_{i,t}^\prime \hat{\beta}_i)^2 \). Let \( D_T = \text{diag}(\sqrt{T}, T, \sqrt{T}) \), such that
\[
D_T^{-1}(\hat{\beta}_i - \beta_i) = \left(D_T^{-1} \sum_{t=2}^T x_{i,t}^\prime x_{i,t} D_T^{-1}\right)^{-1} D_T^{-1} \sum_{t=2}^T x_{i,t} v_{i,t}.
\]
By using the fact that \( (\rho_i - 1) = O_p(N^{-1/2}T^{-1}) \), we obtain
\[
\frac{1}{\sqrt{T}} \sum_{t=2}^T v_{i,t} = \gamma_i \frac{1}{\sqrt{T}} \sum_{t=2}^T (\epsilon_{xi,t} - \Delta x_{i,t}) + \sqrt{T} \epsilon_{yi}
\]
\[
= -\sqrt{T} \gamma_i \delta_i (1 - \rho_i) - \gamma_i (\rho_i - 1) \sqrt{T} x_{i,t-1} + \sqrt{T} \epsilon_{yi}
\]
\[
= \sqrt{T} \epsilon_{yi} + O_p(N^{-1/2}),
\]
where \( x_{i,t-1} = T^{-1} \sum_{t=2}^T x_{i,t-1} \) with an analogous definition of \( \epsilon_{yi} \). We similarly have
\[
\frac{1}{T} \sum_{t=2}^T x_{i,t-1} v_{i,t} = -\gamma_i \delta_i (1 - \rho_i) x_{i,t-1} - \gamma_i (\rho_i - 1) \frac{1}{T} \sum_{t=2}^T x_{i,t-1}^2 + \frac{1}{T} \sum_{t=2}^T x_{i,t-1} \epsilon_{yi,t}
\]
\[
= \frac{1}{T} \sum_{t=2}^T x_{i,t-1} \epsilon_{yi,t} + O_p(N^{-1/2}),
\]
\[
\frac{1}{\sqrt{T}} \sum_{t=2}^T \Delta x_{i,t-1} v_{i,t} = -\gamma_i \delta_i (1 - \rho_i) \sqrt{T} \Delta x_i - \gamma_i (\rho_i - 1) \frac{1}{\sqrt{T}} \sum_{t=2}^T \Delta x_{i,t} x_{i,t-1} + \frac{1}{T} \sum_{t=2}^T \Delta x_{i,t} \epsilon_{yi,t}
\]
\[
= \frac{1}{T} \sum_{t=2}^T \Delta x_{i,t} \epsilon_{yi,t} + O_p(N^{-1/2}),
\]

52
from which it follows that
\[ D_T^{-1} \sum_{t=2}^{T} x_{i,t} \epsilon_{y,t} = D_T^{-1} \sum_{t=2}^{T} x_{i,t} \epsilon_{y,t} + O_p(N^{-1/2}). \]

Moreover, since \( T^{-3/2} \sum_{t=2}^{T} x_{i,t-1} \Delta x_{i,t-1} \) and \( \Delta x_i \) are both \( O_p(T^{-1/2}), \)
\[ D_T^{-1} \sum_{t=2}^{T} x_{i,t} x_{i,t}^T D_T^{-1} = D_T^{-1} \sum_{t=2}^{T} \begin{bmatrix} 1 & x_{i,t-1} & 0 \\ x_{i,t-1}^2 & x_{i,t-1}^2 & 0 \\ 0 & 0 & (\Delta x_{i,t})^2 \end{bmatrix} D_T^{-1} + O_p(T^{-1/2}). \]

The first two elements of \( D_T^{-1} (\hat{\beta}_i - \beta_i) \) are therefore asymptotically uncorrelated with the third, suggesting that we do not lose generality by analyzing \( \sqrt{T}(\hat{\alpha}_i - \alpha_i) \) and \( T(\hat{\beta}_i - \beta_i) \) separately. In so doing, it is not difficult to see that
\[ \sqrt{T}(\hat{\alpha}_i - \alpha_i) = \sqrt{T} \epsilon_{yi} - T(\hat{\beta}_i - \beta_i) T^{-1/2} \bar{x}_{i-1} + O_p(T^{-1/2}) + O_p(N^{-1/2}), \]
\[ T(\hat{\beta}_i - \beta_i) = \frac{T^{-1} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i-1}) \epsilon_{yi,t}}{T^{-2} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i-1})^2} + O_p(T^{-1/2}) + O_p(N^{-1/2}). \]

The asymptotic distribution of the second element is given by
\[ T(\hat{\beta}_i - \beta_i) \to_d \sigma_{yi}^{-1} \left( \frac{\int_0^1 (W_{xi}(r) - \bar{W}_{xi}) dW_{yi}(r)}{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr} \right) Z_1, \]
where \( Z_1 \sim N(0,1) \). As for the first element in \( D_T^{-1} (\hat{\beta}_i - \beta_i), \)
\[ \sqrt{T}(\hat{\alpha}_i - \alpha_i) \to_d \sigma_{yi} \left( W_{yi}(1) - \frac{\int_0^1 (W_{xi}(r) - \bar{W}_{xi}) dW_{yi}(r)}{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr} \bar{W}_{xi} \right), \]
where
\[ \left( W_{yi}(1) - \frac{\int_0^1 (W_{xi}(r) - \bar{W}_{xi}) dW_{yi}(r)}{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr} \bar{W}_{xi} \right) |F \sim Z_2 - \frac{\bar{W}_{xi}}{\sqrt{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr}} Z_1, \]
with \( Z_2 \sim N(0,1) \). The (conditional) covariance between \( Z_1 \) and \( Z_2 \) is given by
\[ \int_0^1 E[(W_{xi}(r) - \bar{W}_{xi})dW_{yi}(r)W_{yi}(1)|F] = \int_0^1 (W_{xi}(r) - \bar{W}_{xi})E[dW_{yi}(r)W_{yi}(1)|F] = 0, \]
from which it follows that
\[ \sigma_{yi} \left( Z_2 - \frac{\bar{W}_{xi}}{\sqrt{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr}} Z_1 \right) |F \sim \sigma_{yi} \left( 1 + \frac{\bar{W}_{xi}^2}{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr} \right)^{1/2} N(0,1). \]
Let us now consider \( \hat{\sigma}^2_{yi}(D_T^{-1} \sum_{t=2}^T x_{i,t} x_{i,t}^T D_T^{-1})^{-1} \). The limits of the first and second diagonal elements of this matrix are given by

\[
\frac{\hat{\sigma}^2_{yi}}{T^{-2} \sum_{t=2}^T (x_{i,t-1} - x_{i,t-1})^2} \xrightarrow{d} \frac{\sigma^2_{yi}}{f_0^1(W_{xi}(r) - \bar{W}_{xi})^2 dr},
\]

as \( T \to \infty \), where we have made use of the fact that

\[
\frac{1}{T} \sum_{t=2}^T x_{i,t}^2 = \frac{1}{T} \sum_{t=2}^T [\gamma_i (\epsilon_{x_i,t} - \Delta x_i,t) + \epsilon_{y_i,t}]^2
\]

\[
= \gamma_i^2 \frac{1}{T} \sum_{t=2}^T (\epsilon_{x_i,t} - \Delta x_i,t)^2 + \gamma_i \frac{1}{T} \sum_{t=2}^T (\epsilon_{x_i,t} - \Delta x_i,t) \epsilon_{y_i,t} + \frac{1}{T} \sum_{t=2}^T \epsilon_{y_i,t}^2
\]

\[
= \gamma_i^2 \frac{1}{NT^2} \sum_{t=2}^T \sum_{t=2}^T (\epsilon_{y_i,t} - \Delta x_i,t)^2 + \gamma_i \frac{1}{NT^2} \sum_{t=2}^T \sum_{t=2}^T (\epsilon_{y_i,t} - \Delta x_i,t) \epsilon_{y_i,t} + \frac{1}{T} \sum_{t=2}^T \epsilon_{y_i,t}^2
\]

\[
= \frac{1}{T} \sum_{t=2}^T \epsilon_{y_i,t}^2 + O_p(N^{-1/2}T^{-1}),
\]

which in turn implies

\[
\hat{\sigma}^2_{yi} = \frac{1}{T} \sum_{t=2}^T (y_{i,t} - x_{i,t}^T \hat{\beta}_i)^2 = \frac{1}{T} \sum_{t=2}^T (v_{i,t} - x_{i,t}^T (\hat{\beta}_i - \beta_i))^2
\]

\[
= \frac{1}{T} \sum_{t=2}^T v_{i,t}^2 - 2 \frac{1}{T} \sum_{t=2}^T v_{i,t} x_{i,t}^T (\hat{\beta}_i - \beta_i) + (\hat{\beta}_i - \beta_i) \frac{1}{T} \sum_{t=2}^T x_{i,t} x_{i,t}^T (\hat{\beta}_i - \beta_i)
\]

\[
= \frac{1}{T} \sum_{t=2}^T v_{i,t}^2 + O_p(T^{-1/2}) = \frac{1}{T} \sum_{t=2}^T \epsilon_{y_i,t}^2 + O_p(T^{-1/2}) \xrightarrow{p} \sigma^2_{yi} \quad \text{(A34)}
\]

as \( T \to \infty \). We can therefore show that

\[
t_{\hat{\beta}_i}(\beta_i) = \frac{T(\hat{\beta}_i - \beta_i)}{\sigma_{yi} / \sqrt{T^{-2} \sum_{t=2}^T (x_{i,t-1} - x_{i,t-1})^2}} \xrightarrow{d} \frac{f_0^1(W_{xi}(r) - \bar{W}_{xi})dW_{yi}(r)}{\sqrt{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr}},
\]

where

\[
\frac{f_0^1(W_{xi}(r) - \bar{W}_{xi})dW_{yi}(r)}{\sqrt{\int_0^1 (W_{xi}(r) - \bar{W}_{xi})^2 dr}} \mid F \sim Z_1
\]

Moreover, because \( Z_1 \) is independent of \( \{W_{xi}(s)\}_{0}^r \), it is also the asymptotic unconditional distribution of \( t_{\hat{\beta}_i}(\beta_i) \). The same steps can be used to show that the asymptotic distribution of \( t_{\hat{\alpha}_i}(\alpha_i) \) is also \( N(0, 1) \).
Let us now consider $t_{bi}(\beta)$. Under the condition that $p = 0$ and $q = 1$, we have $\hat{\beta}_i = \beta + T^{-1} \sigma_{yi} \sigma^{-1}_{xi} c_{yi}$, and therefore
\[
t_{bi}(\beta) = \frac{T(\hat{\beta}_i - \beta)}{\sigma_{yi} / \sqrt{T^{-2} \sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{i-1})^2}} + c_{yi} \sigma_{yi} \sigma^{-1}_{yi}\left(\frac{\sigma_{xi}^{-2} T^{-1} \sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{i-1})^2}{1}\right)^{1/2}
\rightarrow_d Z_1 + c_{yi} \left(\int_0^1 (W_{xi}(s) - \bar{W}_{xi})^2 ds\right)^{1/2}
\] (A35)
as $N, T \rightarrow \infty$. We similarly have
\[
t_{ai}(\alpha) = \frac{\sqrt{T}(\hat{\alpha}_i - \alpha_i)}{\sigma_{yi} / \sqrt{T^{-2} \sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{i-1})^2}} + c_{ai} \sigma_{yi} \sigma^{-1}_{yi}\left(\frac{\sigma_{xi}^{-2} T^{-1} \sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{i-1})^2}{1}\right)^{1/2}
\rightarrow_d N(0, 1) + c_{ai} \left(\int_0^1 (W_{xi}(s) - \bar{W}_{xi})^2 ds\right)^{1/2}
\] (A36)
The asymptotic normality of the $t$-statistic for testing $\gamma_i$ follows from standard arguments for stationary processes, and is therefore omitted.

**Proof of Corollary 1.**

Let $\hat{\beta}$ denote the pooled OLS estimator of $\beta_i$ in (8). Under $H_0 : \beta_1 = \ldots = \beta_N = \beta$, we have that
\[
T(\hat{\beta}_i - \hat{\beta}) = T(\hat{\beta}_i - \beta) - T(\hat{\beta} - \beta) = T(\hat{\beta}_i - \beta) + O_p(N^{-1/2}),
\]
where the last equality follows from the fact that $(\hat{\beta} - \beta) = O_p(N^{-1/2} T^{-1})$ (details are available upon request). This result, together with Theorem 3, implies
\[
H_{bi} = t_{bi}(\hat{\beta})^2 = \frac{T^2(\hat{\beta}_i - \hat{\beta})^2}{\sigma_{yi}^2} \sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{i-1})^2
\]
\[= \frac{T^2(\hat{\beta}_i - \beta)^2}{\sigma_{yi}^2} \sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{i-1})^2 + O_p(N^{-1/2})
\]
\[= t_{bi}(\beta)^2 + O_p(N^{-1/2}) \rightarrow_d Z_1^2
\] (A37)
as $N, T \rightarrow \infty$. The rest of the proof is almost identical to that of Theorem 1 in Westerlund and Hess (2011). It is therefore omitted.
<table>
<thead>
<tr>
<th>$c_\rho$</th>
<th>$T$</th>
<th>$N$</th>
<th>$LM$</th>
<th>$LM_{\mu}$</th>
<th>$LM_{\mu}^2$</th>
<th>$LM_{\mu}^{n}$</th>
<th>$t_{STA}$</th>
<th>$t_{LEW}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>1.8</td>
<td>4.4</td>
<td>0.8</td>
<td>1.8</td>
<td>4.1</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>1.7</td>
<td>4.2</td>
<td>0.4</td>
<td>1.8</td>
<td>4.3</td>
<td>5.8</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>3.2</td>
<td>4.2</td>
<td>3.3</td>
<td>3.3</td>
<td>4.1</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>3.0</td>
<td>4.2</td>
<td>2.7</td>
<td>3.0</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td>−10</td>
<td>100</td>
<td>10</td>
<td>5.8</td>
<td>4.3</td>
<td>6.9</td>
<td>5.7</td>
<td>4.4</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>4.7</td>
<td>3.8</td>
<td>5.7</td>
<td>5.0</td>
<td>3.8</td>
<td>5.6</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>6.3</td>
<td>4.8</td>
<td>7.1</td>
<td>6.2</td>
<td>4.8</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>5.2</td>
<td>4.6</td>
<td>6.1</td>
<td>5.0</td>
<td>4.6</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>1.8</td>
<td>4.4</td>
<td>0.8</td>
<td>2.2</td>
<td>4.4</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>1.7</td>
<td>4.2</td>
<td>0.4</td>
<td>1.9</td>
<td>4.8</td>
<td>9.3</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>3.2</td>
<td>4.2</td>
<td>3.3</td>
<td>3.2</td>
<td>5.0</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>3.0</td>
<td>4.7</td>
<td>2.7</td>
<td>3.1</td>
<td>4.7</td>
<td>8.8</td>
</tr>
<tr>
<td>−10</td>
<td>100</td>
<td>10</td>
<td>84.6</td>
<td>89.8</td>
<td>7.4</td>
<td>6.3</td>
<td>4.6</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>95.5</td>
<td>97.7</td>
<td>6.8</td>
<td>5.6</td>
<td>4.3</td>
<td>8.4</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>84.6</td>
<td>89.6</td>
<td>7.6</td>
<td>6.4</td>
<td>4.7</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>94.6</td>
<td>97.4</td>
<td>6.7</td>
<td>5.5</td>
<td>4.5</td>
<td>7.8</td>
</tr>
</tbody>
</table>

**Notes:** $\gamma$ refers to the covariance between the errors in the predictive and predictor equations and $c_\rho$ is such that $\rho = 1 + N^{-1/2}T^{-1}c_\rho$, where $\rho$ is the autoregressive coefficient of the predictor, $T$ and $N$ are the number of time series and cross-section units, respectively. $t_{STA}$ and $t_{LEW}$ refer to the tests of Stambaugh (1999) and Lewellen (2004), respectively. See Section 3 for a description of the LM tests.
Table 2: Power at the 5% level when $p = \frac{1}{2}$ and $q = 1$.

<table>
<thead>
<tr>
<th>$c_p$</th>
<th>$T$</th>
<th>$N$</th>
<th>$LM$</th>
<th>$LM_{\mu}$</th>
<th>$LM_{c^2}$</th>
<th>$LM^{w}_{\mu}$</th>
<th>$t_{STA}$</th>
<th>$t_{LEW}$</th>
<th>$LM$</th>
<th>$LM_{\mu}$</th>
<th>$LM_{c^2}$</th>
<th>$LM^{w}_{\mu}$</th>
<th>$t_{STA}$</th>
<th>$t_{LEW}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No endogeneity: $\gamma = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>42.9</td>
<td>54.8</td>
<td>0.7</td>
<td>43.2</td>
<td>55.5</td>
<td>7.0</td>
<td>7.2</td>
<td>42.9</td>
<td>57.5</td>
<td>0.7</td>
<td>43.1</td>
<td>57.8</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>46.9</td>
<td>61.4</td>
<td>0.4</td>
<td>47.6</td>
<td>61.5</td>
<td>6.4</td>
<td>6.6</td>
<td>47.0</td>
<td>61.2</td>
<td>0.4</td>
<td>48.0</td>
<td>62.0</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>27.5</td>
<td>36.7</td>
<td>3.3</td>
<td>27.3</td>
<td>36.6</td>
<td>6.4</td>
<td>6.5</td>
<td>25.8</td>
<td>36.4</td>
<td>3.1</td>
<td>25.6</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>27.3</td>
<td>36.6</td>
<td>2.6</td>
<td>27.5</td>
<td>36.9</td>
<td>5.7</td>
<td>5.8</td>
<td>26.4</td>
<td>36.3</td>
<td>2.6</td>
<td>26.5</td>
<td>36.5</td>
</tr>
<tr>
<td>$-10$</td>
<td>100</td>
<td>10</td>
<td>10.7</td>
<td>10.7</td>
<td>6.7</td>
<td>10.9</td>
<td>10.9</td>
<td>6.4</td>
<td>6.8</td>
<td>10.5</td>
<td>10.1</td>
<td>6.6</td>
<td>10.6</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>11.4</td>
<td>13.5</td>
<td>5.8</td>
<td>11.4</td>
<td>13.8</td>
<td>6.0</td>
<td>6.3</td>
<td>11.0</td>
<td>12.9</td>
<td>5.8</td>
<td>10.7</td>
<td>13.5</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>10.6</td>
<td>11.0</td>
<td>6.9</td>
<td>10.7</td>
<td>10.9</td>
<td>5.8</td>
<td>5.9</td>
<td>10.3</td>
<td>10.8</td>
<td>7.1</td>
<td>10.5</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>11.5</td>
<td>13.1</td>
<td>6.0</td>
<td>11.4</td>
<td>13.3</td>
<td>5.4</td>
<td>5.6</td>
<td>11.3</td>
<td>13.1</td>
<td>6.0</td>
<td>11.4</td>
<td>13.3</td>
</tr>
<tr>
<td>High endogeneity: $\gamma = -0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>42.9</td>
<td>54.8</td>
<td>0.7</td>
<td>23.5</td>
<td>32.8</td>
<td>10.8</td>
<td>7.2</td>
<td>42.9</td>
<td>57.5</td>
<td>0.7</td>
<td>22.1</td>
<td>33.1</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>46.9</td>
<td>61.4</td>
<td>0.4</td>
<td>25.1</td>
<td>36.0</td>
<td>10.3</td>
<td>6.5</td>
<td>47.0</td>
<td>61.2</td>
<td>0.4</td>
<td>24.9</td>
<td>35.6</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>27.5</td>
<td>36.7</td>
<td>3.3</td>
<td>14.3</td>
<td>19.3</td>
<td>10.8</td>
<td>6.5</td>
<td>25.8</td>
<td>36.4</td>
<td>3.1</td>
<td>13.2</td>
<td>18.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>27.3</td>
<td>36.6</td>
<td>2.6</td>
<td>13.2</td>
<td>19.8</td>
<td>9.9</td>
<td>5.7</td>
<td>26.4</td>
<td>36.3</td>
<td>2.6</td>
<td>12.9</td>
<td>19.0</td>
</tr>
<tr>
<td>$-10$</td>
<td>100</td>
<td>10</td>
<td>94.2</td>
<td>96.9</td>
<td>8.4</td>
<td>9.1</td>
<td>6.6</td>
<td>9.1</td>
<td>19.7</td>
<td>94.5</td>
<td>97.0</td>
<td>8.5</td>
<td>8.9</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>99.1</td>
<td>99.6</td>
<td>8.7</td>
<td>9.6</td>
<td>8.7</td>
<td>9.3</td>
<td>14.0</td>
<td>99.2</td>
<td>99.7</td>
<td>8.6</td>
<td>9.3</td>
<td>8.3</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>94.1</td>
<td>96.9</td>
<td>7.6</td>
<td>8.1</td>
<td>7.2</td>
<td>9.0</td>
<td>17.8</td>
<td>94.2</td>
<td>97.1</td>
<td>7.6</td>
<td>8.2</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>99.2</td>
<td>99.8</td>
<td>7.0</td>
<td>8.1</td>
<td>7.5</td>
<td>8.6</td>
<td>12.3</td>
<td>99.2</td>
<td>99.7</td>
<td>6.9</td>
<td>8.0</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Notes: $a$, $b$, $p$ and $q$ are such that $\beta_i = N^{-p}T^{-q}c_{\beta_i}$, where $c_{\beta_i} \sim U(a, b)$ is a drift term and $\beta_i$ is the predictive slope. See Table 1 for an explanation of the rest.
Table 3: Power at the 5% level when $p = 1/2$, $q = 1$, $a = -2$ and $b = 2$.

<table>
<thead>
<tr>
<th>$c_\rho$</th>
<th>$T$</th>
<th>$N$</th>
<th>$LM$</th>
<th>$LM_{p^1}$</th>
<th>$LM_{p^2}$</th>
<th>$LM^{*\mu}$</th>
<th>$LM^{*\mu}_p$</th>
<th>$t_{STA}$</th>
<th>$t_{STA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No endogeneity: $\gamma = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>4.5</td>
<td>9.1</td>
<td>0.7</td>
<td>4.5</td>
<td>8.9</td>
<td>6.1</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>3.3</td>
<td>6.8</td>
<td>0.4</td>
<td>3.3</td>
<td>6.6</td>
<td>5.9</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>10</td>
<td>4.2</td>
<td>5.7</td>
<td>3.3</td>
<td>4.2</td>
<td>5.8</td>
<td>5.7</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>3.8</td>
<td>5.7</td>
<td>2.7</td>
<td>3.9</td>
<td>5.9</td>
<td>5.4</td>
<td>5.5</td>
</tr>
<tr>
<td>−10</td>
<td>100</td>
<td>10</td>
<td>5.8</td>
<td>4.8</td>
<td>6.8</td>
<td>5.9</td>
<td>4.6</td>
<td>5.8</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>5.2</td>
<td>4.2</td>
<td>5.5</td>
<td>5.2</td>
<td>4.2</td>
<td>5.8</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>10</td>
<td>6.5</td>
<td>4.8</td>
<td>6.9</td>
<td>6.3</td>
<td>5.0</td>
<td>5.3</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>5.2</td>
<td>4.8</td>
<td>6.2</td>
<td>5.2</td>
<td>4.7</td>
<td>5.2</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>High endogeneity: $\gamma = -0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>4.5</td>
<td>9.1</td>
<td>0.7</td>
<td>3.3</td>
<td>6.8</td>
<td>9.6</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>3.3</td>
<td>6.8</td>
<td>0.4</td>
<td>3.1</td>
<td>6.3</td>
<td>9.3</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>10</td>
<td>4.2</td>
<td>5.7</td>
<td>3.3</td>
<td>4.0</td>
<td>5.7</td>
<td>9.3</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>3.8</td>
<td>5.7</td>
<td>2.7</td>
<td>3.3</td>
<td>5.4</td>
<td>8.9</td>
<td>5.5</td>
</tr>
<tr>
<td>−10</td>
<td>100</td>
<td>10</td>
<td>84.0</td>
<td>89.5</td>
<td>7.3</td>
<td>6.5</td>
<td>4.8</td>
<td>8.0</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>95.3</td>
<td>97.5</td>
<td>6.7</td>
<td>5.9</td>
<td>4.5</td>
<td>8.5</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>10</td>
<td>84.7</td>
<td>89.5</td>
<td>7.4</td>
<td>6.7</td>
<td>4.8</td>
<td>7.9</td>
<td>13.8</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>94.6</td>
<td>97.4</td>
<td>6.6</td>
<td>5.6</td>
<td>4.4</td>
<td>7.8</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Notes: See Table 1 for an explanation.
Table 4: Power at the 5% level when $p = 1/4$ and $q = 3/4$.

<table>
<thead>
<tr>
<th>$c_p$</th>
<th>$T$</th>
<th>$N$</th>
<th>$a = -4, b = 4$</th>
<th>$a = -6, b = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$LM$</td>
<td>$LM_{\mu}$</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>10</td>
<td>72.4</td>
<td>71.7</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>71.5</td>
<td>67.2</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>78.9</td>
<td>75.1</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>80.4</td>
<td>72.6</td>
</tr>
<tr>
<td>$-10$</td>
<td>100</td>
<td>10</td>
<td>32.6</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>38.0</td>
<td>35.7</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>45.7</td>
<td>46.8</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>50.2</td>
<td>49.2</td>
</tr>
<tr>
<td>$-10$</td>
<td>100</td>
<td>10</td>
<td>72.4</td>
<td>71.7</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>71.5</td>
<td>67.2</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>78.9</td>
<td>75.1</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>80.4</td>
<td>72.6</td>
</tr>
<tr>
<td>$-10$</td>
<td>100</td>
<td>10</td>
<td>69.4</td>
<td>71.5</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>80.2</td>
<td>80.8</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td></td>
<td>68.3</td>
<td>69.9</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>77.0</td>
<td>76.4</td>
</tr>
</tbody>
</table>

Notes: See Tables 1 and 2 for an explanation.
Table 5: Cross-correlation test results for returns.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Correlation</th>
<th>CD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banking</td>
<td>0.454</td>
<td>148.089</td>
<td>0.000</td>
</tr>
<tr>
<td>Chemical</td>
<td>0.314</td>
<td>99.677</td>
<td>0.000</td>
</tr>
<tr>
<td>Electricity</td>
<td>0.398</td>
<td>137.162</td>
<td>0.000</td>
</tr>
<tr>
<td>Energy</td>
<td>0.479</td>
<td>94.592</td>
<td>0.000</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.375</td>
<td>84.388</td>
<td>0.000</td>
</tr>
<tr>
<td>Real estate</td>
<td>0.470</td>
<td>166.398</td>
<td>0.000</td>
</tr>
<tr>
<td>Hardware</td>
<td>0.350</td>
<td>72.389</td>
<td>0.000</td>
</tr>
<tr>
<td>Household</td>
<td>0.344</td>
<td>109.234</td>
<td>0.000</td>
</tr>
<tr>
<td>Mining</td>
<td>0.284</td>
<td>16.912</td>
<td>0.000</td>
</tr>
<tr>
<td>Retail</td>
<td>0.267</td>
<td>124.078</td>
<td>0.000</td>
</tr>
<tr>
<td>Software</td>
<td>0.293</td>
<td>22.833</td>
<td>0.000</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.339</td>
<td>23.335</td>
<td>0.000</td>
</tr>
<tr>
<td>Transport</td>
<td>0.329</td>
<td>77.006</td>
<td>0.000</td>
</tr>
<tr>
<td>Travel</td>
<td>0.300</td>
<td>70.353</td>
<td>0.000</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.323</td>
<td>84.699</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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.
### Table 6: Unit root test results.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Return</th>
<th>BM</th>
<th>CFP</th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>Critical values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1%</td>
</tr>
</tbody>
</table>

*Notes:* The results reported in the table refer to the CIPS panel unit root test of Pesaran (2007). The lag lengths are selected using the BIC and the appropriate critical values are taken from Pesaran (2007). The test regression is fitted with both a constant and trend.
Table 7: Predictability test results.

<table>
<thead>
<tr>
<th>Sector</th>
<th>No factor</th>
<th>Factor</th>
<th>No factor</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$LM^m$</td>
<td>$LM^u$</td>
<td>$LM_{c2}$</td>
<td>$LM^m$</td>
</tr>
<tr>
<td>Banking</td>
<td>0.000</td>
<td>0.047</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Chemical</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Electricity</td>
<td>0.006</td>
<td>0.111</td>
<td>0.006</td>
<td>0.063</td>
</tr>
<tr>
<td>Energy</td>
<td>0.000</td>
<td>0.125</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.000</td>
<td>0.004</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Real estate</td>
<td>0.000</td>
<td>0.080</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Hardware</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>Household</td>
<td>0.000</td>
<td>0.011</td>
<td>0.000</td>
<td>0.020</td>
</tr>
<tr>
<td>Mining</td>
<td>0.123</td>
<td>0.246</td>
<td>0.092</td>
<td>0.059</td>
</tr>
<tr>
<td>Retail</td>
<td>0.004</td>
<td>0.491</td>
<td>0.001</td>
<td>0.056</td>
</tr>
<tr>
<td>Software</td>
<td>0.075</td>
<td>0.567</td>
<td>0.028</td>
<td>0.603</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.022</td>
<td>0.236</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>Transport</td>
<td>0.000</td>
<td>0.069</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Travel</td>
<td>0.000</td>
<td>0.153</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.217</td>
<td>0.171</td>
<td>0.277</td>
<td>0.687</td>
</tr>
</tbody>
</table>

Continued overleaf
Table 7: Continued.

<table>
<thead>
<tr>
<th>Sector</th>
<th>No factor</th>
<th>Factor</th>
<th>No factor</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( LM^m )</td>
<td>( LM^m )</td>
<td>( LM_{\mu} )</td>
<td>( LM_{\mu} )</td>
</tr>
<tr>
<td></td>
<td>( \mu )</td>
<td>( \mu )</td>
<td>( \sigma^2 )</td>
<td>( \sigma^2 )</td>
</tr>
<tr>
<td>Banking</td>
<td>0.220 0.316 0.155</td>
<td>0.220 0.247 0.194</td>
<td>0.406 0.354 0.331</td>
<td>0.343 0.250 0.365</td>
</tr>
<tr>
<td>Electricity</td>
<td>0.654 0.361 0.899</td>
<td>0.773 0.476 0.928</td>
<td>0.648 0.356 0.898</td>
<td>0.779 0.480 0.998</td>
</tr>
<tr>
<td>Energy</td>
<td>0.253 0.100 0.846</td>
<td>0.316 0.129 0.971</td>
<td>0.350 0.147 0.995</td>
<td>0.394 0.174 0.890</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.463 0.358 0.404</td>
<td>0.748 0.459 0.852</td>
<td>0.638 0.411 0.637</td>
<td>0.741 0.452 0.853</td>
</tr>
<tr>
<td>Real estate</td>
<td>0.218 0.176 0.270</td>
<td>0.249 0.299 0.193</td>
<td>0.212 0.221 0.205</td>
<td>0.185 0.293 0.132</td>
</tr>
<tr>
<td>Hardware</td>
<td>0.173 0.062 0.879</td>
<td>0.176 0.062 0.966</td>
<td>0.190 0.070 0.825</td>
<td>0.227 0.085 0.917</td>
</tr>
<tr>
<td>Household</td>
<td>0.384 0.197 0.616</td>
<td>0.524 0.262 0.851</td>
<td>0.550 0.274 0.992</td>
<td>0.498 0.314 0.537</td>
</tr>
<tr>
<td>Mining</td>
<td>0.946 0.978 0.739</td>
<td>0.981 0.979 0.846</td>
<td>0.976 0.990 0.826</td>
<td>0.957 0.960 0.771</td>
</tr>
<tr>
<td>Retail</td>
<td>0.097 0.031 0.845</td>
<td>0.095 0.037 0.541</td>
<td>0.054 0.016 0.921</td>
<td>0.093 0.032 0.683</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.901 0.994 0.648</td>
<td>0.832 0.953 0.546</td>
<td>0.770 0.928 0.473</td>
<td>0.672 0.997 0.372</td>
</tr>
<tr>
<td>Transport</td>
<td>0.003 0.021 0.012</td>
<td>0.013 0.028 0.050</td>
<td>0.008 0.036 0.022</td>
<td>0.019 0.023 0.094</td>
</tr>
<tr>
<td>Travel</td>
<td>0.680 0.380 0.977</td>
<td>0.793 0.496 0.996</td>
<td>0.634 0.340 0.991</td>
<td>0.743 0.441 0.997</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.756 0.456 0.961</td>
<td>0.778 0.481 0.945</td>
<td>0.702 0.401 0.951</td>
<td>0.710 0.412 0.908</td>
</tr>
</tbody>
</table>

Notes: The number in the table are the test \( p \)-values. The lag lengths are selected using the BIC. While the “No factor” test results assume that the cross-section units are independent, the “Factor” results allow for cross-section dependence in the form of a common factor; see Section 3.3 for more details.
Table 8: Poolability test results.

<table>
<thead>
<tr>
<th>Sector</th>
<th>No factor</th>
<th>Factor</th>
<th>No factor</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \bar{\alpha} )</td>
<td>( H_\alpha )</td>
<td>( \bar{\beta} )</td>
<td>( H_\beta )</td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banking</td>
<td>-0.017</td>
<td>0.000</td>
<td>0.037</td>
<td>0.000</td>
</tr>
<tr>
<td>Chemical</td>
<td>0.031</td>
<td>0.092</td>
<td>-0.002</td>
<td>0.030</td>
</tr>
<tr>
<td>Electricity</td>
<td>-0.047</td>
<td>0.001</td>
<td>0.081</td>
<td>0.001</td>
</tr>
<tr>
<td>Energy</td>
<td>-0.051</td>
<td>0.181</td>
<td>0.121</td>
<td>0.054</td>
</tr>
<tr>
<td>Engineering</td>
<td>-0.051</td>
<td>0.020</td>
<td>0.132</td>
<td>0.002</td>
</tr>
<tr>
<td>Real estate</td>
<td>-0.030</td>
<td>0.000</td>
<td>0.060</td>
<td>0.000</td>
</tr>
<tr>
<td>Hardware</td>
<td>-0.042</td>
<td>0.197</td>
<td>0.140</td>
<td>0.147</td>
</tr>
<tr>
<td>Household</td>
<td>-0.049</td>
<td>0.001</td>
<td>0.101</td>
<td>0.000</td>
</tr>
<tr>
<td>Mining</td>
<td>-0.028</td>
<td>0.002</td>
<td>0.091</td>
<td>0.003</td>
</tr>
<tr>
<td>Retail</td>
<td>-0.040</td>
<td>0.000</td>
<td>0.128</td>
<td>0.000</td>
</tr>
<tr>
<td>Software</td>
<td>-0.022</td>
<td>0.044</td>
<td>0.101</td>
<td>0.009</td>
</tr>
<tr>
<td>Telecom</td>
<td>-0.012</td>
<td>0.000</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>Transport</td>
<td>-0.044</td>
<td>0.043</td>
<td>0.096</td>
<td>0.002</td>
</tr>
<tr>
<td>Travel</td>
<td>-0.043</td>
<td>0.000</td>
<td>0.114</td>
<td>0.000</td>
</tr>
<tr>
<td>Utilities</td>
<td>-0.061</td>
<td>0.006</td>
<td>0.120</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Continued overleaf
Table 8: Continued.

<table>
<thead>
<tr>
<th>Sector</th>
<th>No factor</th>
<th>Factor</th>
<th>No factor</th>
<th>Factor</th>
<th>No factor</th>
<th>Factor</th>
<th>No factor</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{\pi}$</td>
<td>$H_\alpha$</td>
<td>$\bar{\beta}$</td>
<td>$H_\beta$</td>
<td>$\bar{\pi}$</td>
<td>$H_\alpha$</td>
<td>$\bar{\beta}$</td>
<td>$H_\beta$</td>
</tr>
<tr>
<td>Banking</td>
<td>0.010</td>
<td>0.012</td>
<td>0.002</td>
<td>0.018</td>
<td>0.031</td>
<td>0.023</td>
<td>0.007</td>
<td>0.051</td>
</tr>
<tr>
<td>Electricity</td>
<td>0.101</td>
<td>0.162</td>
<td>0.032</td>
<td>0.044</td>
<td>0.065</td>
<td>0.343</td>
<td>0.020</td>
<td>0.302</td>
</tr>
<tr>
<td>Energy</td>
<td>0.079</td>
<td>0.000</td>
<td>0.020</td>
<td>0.000</td>
<td>0.039</td>
<td>0.086</td>
<td>0.009</td>
<td>0.088</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.134</td>
<td>0.244</td>
<td>0.030</td>
<td>0.265</td>
<td>0.098</td>
<td>0.597</td>
<td>0.022</td>
<td>0.584</td>
</tr>
<tr>
<td>Real estate</td>
<td>0.096</td>
<td>0.002</td>
<td>0.033</td>
<td>0.002</td>
<td>0.094</td>
<td>0.000</td>
<td>0.032</td>
<td>0.000</td>
</tr>
<tr>
<td>Hardware</td>
<td>0.039</td>
<td>0.933</td>
<td>0.008</td>
<td>0.945</td>
<td>0.031</td>
<td>0.843</td>
<td>0.006</td>
<td>0.794</td>
</tr>
<tr>
<td>Household</td>
<td>0.063</td>
<td>0.006</td>
<td>0.016</td>
<td>0.005</td>
<td>0.040</td>
<td>0.469</td>
<td>0.009</td>
<td>0.391</td>
</tr>
<tr>
<td>Mining</td>
<td>0.074</td>
<td>0.665</td>
<td>0.013</td>
<td>0.646</td>
<td>0.089</td>
<td>0.457</td>
<td>0.016</td>
<td>0.368</td>
</tr>
<tr>
<td>Retail</td>
<td>0.093</td>
<td>0.000</td>
<td>0.021</td>
<td>0.000</td>
<td>0.069</td>
<td>0.000</td>
<td>0.015</td>
<td>0.000</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.039</td>
<td>0.701</td>
<td>0.008</td>
<td>0.686</td>
<td>0.050</td>
<td>0.154</td>
<td>0.014</td>
<td>0.143</td>
</tr>
<tr>
<td>Transport</td>
<td>0.109</td>
<td>0.012</td>
<td>0.026</td>
<td>0.013</td>
<td>0.063</td>
<td>0.109</td>
<td>0.014</td>
<td>0.109</td>
</tr>
<tr>
<td>Travel</td>
<td>0.037</td>
<td>0.906</td>
<td>0.007</td>
<td>0.867</td>
<td>0.018</td>
<td>0.336</td>
<td>0.001</td>
<td>0.197</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.113</td>
<td>0.056</td>
<td>0.033</td>
<td>0.050</td>
<td>0.084</td>
<td>0.104</td>
<td>0.024</td>
<td>0.093</td>
</tr>
</tbody>
</table>

Notes: $\bar{\pi}$ and $\bar{\beta}$ refer to the average of the estimated intercept and slope coefficient, respectively. The results reported in the columns labelled as $H_\alpha$ and $H_\beta$ are the $p$-values for the maximum of the individual Hausman tests of the null hypothesis of poolability; see Section 3.3. See Table 7 for an explanation of the rest.

Table 9: Economic significance results.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Constant-only</th>
<th>BM</th>
<th>CFP</th>
<th>DE</th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Utility</td>
<td>Mean</td>
<td>SD</td>
<td>Utility</td>
<td>Mean</td>
</tr>
<tr>
<td>Banking</td>
<td>0.259</td>
<td>0.004</td>
<td>−0.013</td>
<td>1.358</td>
<td>1.264</td>
<td>1.450</td>
<td>0.582</td>
</tr>
<tr>
<td>Chemical</td>
<td>0.259</td>
<td>0.001</td>
<td>−0.007</td>
<td>0.288</td>
<td>0.230</td>
<td>−0.158</td>
<td>0.231</td>
</tr>
<tr>
<td>Electricity</td>
<td>0.259</td>
<td>0.000</td>
<td>−0.008</td>
<td>1.417</td>
<td>1.535</td>
<td>1.241</td>
<td>4.044</td>
</tr>
<tr>
<td>Energy</td>
<td>0.259</td>
<td>0.004</td>
<td>−0.017</td>
<td>0.406</td>
<td>0.541</td>
<td>1.061</td>
<td>2.385</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.259</td>
<td>0.001</td>
<td>−0.012</td>
<td>1.401</td>
<td>2.079</td>
<td>13.946</td>
<td>3.723</td>
</tr>
<tr>
<td>Real Estate</td>
<td>0.259</td>
<td>0.000</td>
<td>−0.008</td>
<td>3.629</td>
<td>0.350</td>
<td>3.341</td>
<td>0.314</td>
</tr>
<tr>
<td>Hardware</td>
<td>0.259</td>
<td>0.003</td>
<td>−0.021</td>
<td>1.247</td>
<td>1.624</td>
<td>5.002</td>
<td>1.564</td>
</tr>
<tr>
<td>Household</td>
<td>0.259</td>
<td>0.005</td>
<td>−0.018</td>
<td>1.074</td>
<td>1.401</td>
<td>1.127</td>
<td>1.902</td>
</tr>
<tr>
<td>Mining</td>
<td>0.259</td>
<td>0.001</td>
<td>0.003</td>
<td>0.283</td>
<td>0.227</td>
<td>0.010</td>
<td>0.408</td>
</tr>
<tr>
<td>Retail</td>
<td>0.259</td>
<td>0.002</td>
<td>0.003</td>
<td>0.130</td>
<td>0.271</td>
<td>12.233</td>
<td>2.336</td>
</tr>
<tr>
<td>Software</td>
<td>0.260</td>
<td>0.002</td>
<td>−0.015</td>
<td>0.283</td>
<td>0.284</td>
<td>0.022</td>
<td>0.226</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.293</td>
<td>0.857</td>
<td>−0.199</td>
<td>0.270</td>
<td>0.215</td>
<td>0.941</td>
<td>4.441</td>
</tr>
<tr>
<td>Transport</td>
<td>0.259</td>
<td>0.001</td>
<td>−0.012</td>
<td>1.057</td>
<td>3.020</td>
<td>10.527</td>
<td>2.425</td>
</tr>
<tr>
<td>Travel</td>
<td>0.259</td>
<td>0.002</td>
<td>−0.016</td>
<td>8.215</td>
<td>2.011</td>
<td>8.186</td>
<td>4.623</td>
</tr>
<tr>
<td>Utility</td>
<td>0.259</td>
<td>0.004</td>
<td>−0.013</td>
<td>0.009</td>
<td>0.011</td>
<td>−0.193</td>
<td>1.603</td>
</tr>
</tbody>
</table>

Notes: Mean and SD refer to the average and standard deviation of the forecasted returns. The estimated utilities are based on the forecasted returns.