VOLUME 08 * 2019

APPLIED FINANCE LETTERS

Applied Finance Letters is available on the website of the Auckland Centre for Financial Research
www.acfr.aut .ac.nz

Alternatively, copies can be ordered from the Auckland Centre for Financial Research at a cost. If you would like to receive a copy, please contact:

Auckland Centre for Financial Research
AUT Business School
Auckland University of Technology
Private Bag 92006
Auckland, 1142, New Zealand
✉ acfr@aut.ac.nz

Publishing in Applied Finance Letters

Applied Finance Letters publishes short empirical research with clear implications and relevance for the finance industry. The aim is to encourage discussions among academics, policy makers and financial practitioners.

Submitted articles go through a blind review process and may be returned to the authors for subsequent revision.

Please visit www.acfr.aut.ac.nz for author's guidelines and style requirements.

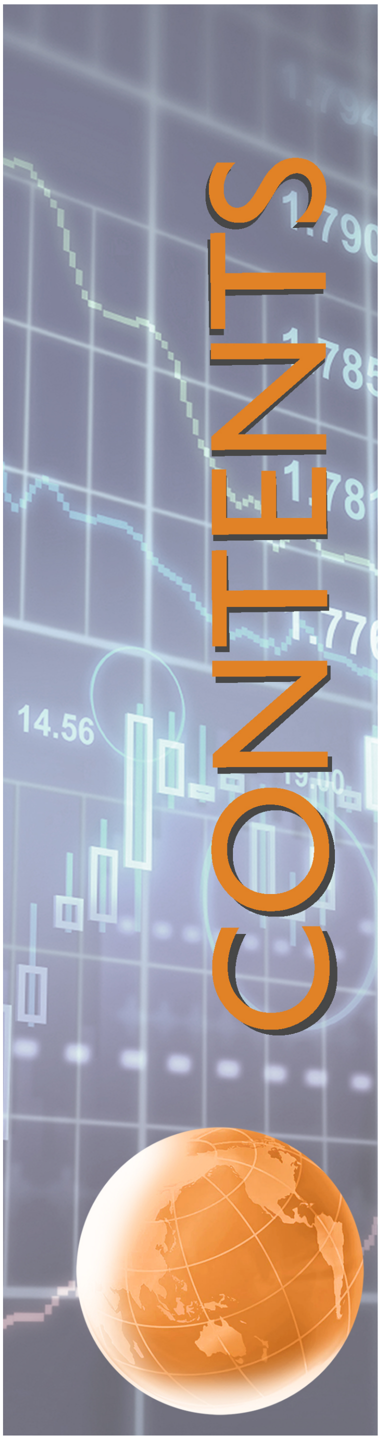
Copyright

ISSN 2253-5799 (Print)

ISSN 2253-5802 (Online)

AUCKLAND CENTRE FOR FINANCIAL RESEARCH

APPLIED FINANCE LETTERS



VOLUME 8, 2019

APPLIED FINANCE LETTERS

VOLUME 08, 2019

THE EFFECT OF RISK ON INVESTMENT: NEW EVIDENCE

Tingting Que

Corresponding Author: Tingting Que, College of Business, University of Alabama in Huntsville ☎ +256-824-6703
✉ tingting.que@uah.edu

Abstract: Previous results on the relation between risk and investment are mixed, partly due to endogeneity. To alleviate the effects of this bias, we adopt a generalized method of moments (GMM) dynamic panel estimator to investigate the relation. We find that the puzzling positive sensitivity of investment (i.e. firm's investment rate) to systematic risk as frequently documented in previous studies disappears. Further, we show that the more irreversible the firm's investments are, the more valuable is the option to delay investment when risk is high, which supports the model with irreversible investment.

Keywords: risk, investment, endogeneity, GMM, capital irreversibility

1. Introduction

The relation between risk and investment has been an important research topic for several decades. Despite an extensive theoretical literature predicting a negative impact of risk on investment (e.g., Bernanke (1983); Smith and Stulz (1985); McDonald and Siegel (1986); Froot et al. (1993); Dixit and Pindyck (1994); Abel and Eberly (1996)), empirical results are mixed. While some studies support the theoretical prediction (e.g., Leahy and Whited (1996); Gulen and Ion (2016); Julio and Yook (2012)), Bulan (2005) and Panousi and Papanikolaou (2012) show that idiosyncratic and systematic risk affect investment in different ways. In particular, they find that systematic risk actually encourages investment while idiosyncratic risk does the opposite. One goal of this paper is to reconcile these mixed empirical results.

We start by examining the average relation between risk and investment (i.e. firm's investment rate) using the conventional ordinary least squares (OLS) model with fixed effects. We confirm that an increase in idiosyncratic volatility depresses investment, while an increase in systematic volatility encourages investment.

Since systematic volatility depends on the firm's systematic risk exposure (beta) as well as market and industry risk, we further decompose systematic volatility into its individual components to examine the role of covariance and other components. Surprisingly, we find that a firm's exposure to systematic risk is positively correlated with investment. The result stands in contrast with the view that greater systematic risk tends to make investment less desirable.

We acknowledge that the OLS model may give rise to three potential sources of endogeneity. First, the OLS model ignores unobserved heterogeneity by assuming that neither the risk variables nor the control variables are correlated to unobserved firm characteristics. Second, OLS estimation relies on the assumption that none of the risk variables or the control variables is correlated with the error term. But if investment and risk are simultaneously determined, then this assumption is clearly violated, which leads to biased OLS estimates.

Lastly, although the fixed-effects estimation eliminates unobserved heterogeneity, it potentially introduces dynamic endogeneity. The fixed-effects estimation relies on a strict exogeneity assumption which requires that in our context, the risk variables we observe today are completely independent of any past, present and future investment. This assumption is likely to be violated if contemporaneous risk variables depend on past realization of investment, and thus fixed-effects estimation is likely to be inconsistent.

While there is substantial economic justification to suspect that the risk variables are not strictly exogenous, we confirm this with an econometric test of strict exogeneity suggested by Wooldridge (2010). The results show the risk variables are not strictly exogenous. Thus, dynamic endogeneity is likely to be a major source of bias in the baseline model.

Consistent estimation of the relation between risk and investment requires the use of an estimation technique which controls for unobserved heterogeneity and simultaneity while exploiting the dynamic association between risk and investment. A promising estimation technique is the GMM dynamic panel estimator. This estimator, first proposed by Holtz-Eakin et al. (1988) and Arellano and Bond (1991) and further developed by Arellano and Bover (1995) and Blundell and Bond (1998), provides an excellent econometric specification to deal with the abovementioned issues.

When the estimation is carried out using the GMM dynamic panel estimator, the puzzling positive sensitivity of investment to systematic risk disappears. We further apply the GMM estimator to estimate the relation between investment and various components of systematic risk and document a clear negative response of investment to systematic risk components. These results support the prediction that the greater the systematic risk the less the incentive to invest.

Both the traditional view (e.g. the capital asset pricing model (CAPM)) and the real option theory predict that greater uncertainty tends to make investment less desirable. However, the traditional view asserts that it is only the systematic risk that should matter for firm investment, while real option theory predicts that it is the total risk that should matter for investment. To this point, our GMM estimations show that both systematic and idiosyncratic risk have negative impacts on firm investment. Therefore, we take the next step to examine the predictions of real option theory in the context of capital irreversibility.

The ability to delay investment is valuable when the investment is irreversible, and the future is uncertain. The irreversibility of investment stems from capital specificity at the industry and/or at the firm level. We implement sample splits according to the firm's degree of irreversibility and re-estimate the relation between investment and risk using a GMM dynamic panel estimator in the two subsamples. We provide empirical support for the prediction of real options models that the more irreversible the firm's investments are, the more valuable is the option to delay investment when risk is high.

The main contribution of this paper is to reconcile mixed evidence on the relation between risk and investment. Prior studies documenting a puzzling positive sensitivity of investment to systematic volatility may have inadvertently relied on inconsistent estimation procedures. By implementing a GMM dynamic panel estimator that eliminates the major sources of endogeneity, we show that the positive sign is replaced by a negative relation between investment and systematic risk.

This paper also contributes to the literature that uses the GMM dynamic panel estimation in economics and finance where unobserved heterogeneity and dynamic endogeneity are prevalent and truly exogenous instruments are difficult to find. Examples of these studies include Caselli et al. (1996), Blundell and Bond (1998), Beck et al. (2000), Erickson and Whited (2000) and Wintoki et al. (2012). In estimating the relation between investment and risk, we apply the GMM dynamic panel estimator to control for unobserved heterogeneity, simultaneity and dynamic endogeneity.

In addition, this paper provides empirical support for the prediction of real options models. Our results show that the greater the degree of asset-specificity of capital (and hence the more irreversible the firm's investments are), the more valuable is the option to delay investment when uncertainty is high.

The remainder of the paper is organized as follows. Section 2 describes the sample. Section 3 provides empirical results. Section 4 concludes.

2. Sample

Following previous empirical work in the risk and investment literature, the sample includes all publicly traded firms in Compustat over the period 1970 to 2005, excluding firms in the financial (SIC code 6000−6999), utilities (SIC code 4900−4949), and government-regulated industries (SIC code > 9000). Firm-year observations with missing SIC codes, with missing values for investment, Tobin's Q, cash flows, size, leverage, stock returns, and with negative book values of capital are dropped. Firms with fewer than 40 weekly observations in that year are also excluded. The initial sample includes a total of 101,378 firm-year observations. Finally, data are winsorized by year at the 0.5% and 99.5% levels in all specifications. Descriptive statistics and correlations are presented in Table 1.

Table 1: Summary statistics

Investment rate ($Capx_t/A_{t-1}$) is defined as the ratio of capital expenditure to book assets. Our measure of idiosyncratic risk, $\log(\sigma_{i,t-1}^{idio})$, is constructed from a regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Systematic volatility ($\log(\sigma_{t-1}^{syst})$) is defined as the (log of the) square root of the difference between the firm's total variance and its idiosyncratic variance. Market volatility ($\log(\sigma_{t-1}^{mkt})$) is defined as the (log of the) square root of the variance of CRSP VW index. $\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$ are coefficient estimates from the regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Industry volatility ($\log(\sigma_{t-1}^{ind})$) defined as the (log of the) square root of the variance of VW industry portfolio. The sample period is 1970 to 2005.

Panel A: Descriptive Statistics (N=101,378)

Variables	Mean	Std Dev
$Capx_t/A_{t-1}$	0.071	0.078
$\log(\sigma_{i,t-1}^{idio})$	−0.959	0.545
$\log(\sigma_{t-1}^{syst})$	−1.949	0.735
$\log(\sigma_{t-1}^{mkt})$	−2.045	0.332
$\beta_{i,t-1}^{mkt}$	0.591	1.307
$\log(\sigma_{t-1}^{ind})$	−1.742	0.354
$\beta_{i,t-1}^{ind}$	0.374	0.998

Panel B: Sample Correlations (N=101,378)

Variables	$Capx_t/A_{t-1}$	$\log(\sigma_{i,t-1}^{idio})$	$\log(\sigma_{t-1}^{syst})$	$\log(\sigma_{t-1}^{mkt})$	$\beta_{i,t-1}^{mkt}$	$\log(\sigma_{t-1}^{ind})$	$\beta_{i,t-1}^{ind}$
$Capx_t/A_{t-1}$	1						
$\log(\sigma_{i,t-1}^{idio})$	-0.096	1					
$\log(\sigma_{t-1}^{syst})$	0.019	0.384	1				
$\log(\sigma_{t-1}^{mkt})$	-0.038	0.186	0.342	1			
$\beta_{i,t-1}^{mkt}$	0.002	0.114	0.217	0.005	1		
$\log(\sigma_{t-1}^{ind})$	-0.055	0.274	0.389	0.776	0.008	1	
$\beta_{i,t-1}^{ind}$	0.036	-0.051	0.214	-0.012	-0.767	0.016	1

3. Empirical Results

3.1. Baseline Model

In this section, we use the OLS model with fixed-effects as the baseline model to examine the response of investment to idiosyncratic and systematic volatility.

The baseline measure of idiosyncratic volatility is constructed using weekly data on stock returns from CRSP (Bulan (2005), Panousi and Papanikolaou (2012)). For every firm i and every year t, we regress the firm's return on the value-weighted market portfolio, R_{MKT}, and on the corresponding value-weighted industry portfolio, R_{IND}, based on the (Fama & French 1997) 30-industry classification, across the 52 weekly observations.

$$R_{i,\tau} = \alpha_{1,i} + \beta_i^{mkt} R_{MKT,\tau} + \beta_i^{ind} R_{IND,\tau} + \varepsilon_{i,\tau} \, , \tag{1}$$

where τ indexes weeks. Then idiosyncratic risk is the log volatility of the regression residuals

$$\log(\sigma_{i,t-1}^{idio}) = \log \sqrt{\textstyle\sum_{\tau \in t} \varepsilon_{i,\tau}^2} \, . \tag{2}$$

Systematic volatility is then defined as the (log of the) square root of the difference between the firm's total variance and its idiosyncratic variance.

The response of investment to idiosyncratic and systematic risk is estimated using the following equation:

$$Capx_t / A_{t-1} = \gamma_0 + \beta_1 \log(\sigma_{i,t-1}^{idio}) + \beta_2 \log(\sigma_{t-1}^{syst}) + \gamma_1 Z_{i,t-1} + \delta_i + \theta_t + \omega_{i,t} \, , \tag{3}$$

where the dependent variable is the firm's investment rate $(Capx_t/A_{t-1})$ and $Z_{i,t-1}$ is a vector of control variables: (i) log Tobin's Q; (ii) the ratio of cash flows to assets (CF_{t-1}/A_{t-2}); (iii) log firm size; (iv) the firm's own stock return (R_{t-1}); and (v) log firm leverage, measured as the ratio of equity to assets $(\log(E_{t-1}/A_{t-1}))$. Depending on the specification, we include firm (δ_i) or year dummies (θ_t). Finally, the errors $(\omega_{i,t})$ are clustered at the firm level.

The estimates of Equation (3) are reported in the first column of Table 2. The coefficient on idiosyncratic volatility is of −1% and statistically significant. The sign of the coefficient is consistent with Panousi and Papanikolaou (2012), but the magnitude is smaller due to the reason that we use book assets instead of replacement value of capital (see, Salinger and Summers (1983)) in the dependent variable in Equation (3). However, the coefficient on systematic volatility is positive and significant (0.2%). The positive sensitivity of investment to systematic volatility is puzzling. All else equal, an increase in systematic volatility increases the firms' cost of capital and therefore should decrease investment.

Since the measure of systematic volatility depends on the firm's systematic risk exposure (beta) as well as the amount of market and industry risk, we decide to decompose systematic volatility into individual components in an attempt to explain the positive response of investment to systematic volatility.

Table 2: Baseline OLS model of investment on risk

The table reports OLS estimation results of Equations (3) and (4), where the dependent variable is the investment rate ($Capx_t/A_{t-1}$). The idiosyncratic risk, $\log(\sigma_{i,t-1}^{idio})$, is constructed from a regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Systematic volatility ($\log(\sigma_{t-1}^{syst})$) defined as the (log of the) square root of the difference between the firm's total variance and its idiosyncratic variance. Market volatility ($\log(\sigma_{t-1}^{mkt})$) is defined as the (log of the) square root of the variance of CRSP VW index. $\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$ are coefficient estimates from the regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Industry volatility ($\log(\sigma_{t-1}^{ind})$) defined as the (log of the) square root of the variance of VW industry portfolio. Financial control variables include lagged values of: Tobin's Q $\log(Q_{t-1})$ defined as in Fazzari *et al.* (1988); operating cash flows (CF_{t-1}/A_{t-2}) defined as the ratio of operating income to book assets; the firm's size ($\log(A_{t-1})$) defined as the log value of book assets; the firm's stock return (R_{t-1}); leverage (E_{t-1}/A_{t-1}) defined as the ratio of book equity to book assets. The coefficients of these control variables are suppressed for brevity. The sample period is 1970 to 2005. *F, T* denotes firm and time fixed effects, and p-values are reported in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

$Capx_t/A_{t-1}$	1	2	3
$\log(\sigma_{i,t-1}^{idio})$	–0.010***	–0.002***	–0.009***
	(0.000)	(0.005)	(0.000)
$\log(\sigma_{t-1}^{syst})$	0.002***		
	(0.000)		
$\log(\sigma_{t-1}^{mkt})$		–0.001	–0.034***
		(0.589)	(0.000)
$\beta_{i,t-1}^{mkt}$		0.003***	0.001***
		(0.000)	(0.000)
$\log(\sigma_{t-1}^{ind})$		–0.011***	0.001
		(0.000)	(0.704)
$\beta_{i,t-1}^{ind}$		0.002***	0.001***
		(0.000)	(0.004)
Financial controls	Yes	No	Yes
Observations	101,378	101,378	101,378
R^2	0.555	0.494	0.555
Fixed effects	F,T	F	F,T

We estimate the response of investment to each component of systematic volatility and idiosyncratic risk using the following reduced-form equation:

$$Capx_t/A_{t-1} = \gamma_0 + \gamma_1 \log(\sigma_{i,t-1}^{idio}) + \gamma_2 \log(\sigma_{t-1}^{mkt}) + \gamma_3 \beta_{i,t-1}^{mkt} + \gamma_4 \log(\sigma_{t-1}^{ind}) + \gamma_5 \beta_{i,t-1}^{ind} + \gamma_6 Z_{i,t-1} + \delta_i + \theta_t + \omega_{i,t} \quad (4)$$

where four additional regressors are included: market volatility ($\log(\sigma_{t-1}^{mkt})$) defined as the (log of the) square root of the variance of CRSP VW index; $\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$ are coefficient estimates from the regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio in Equation (1); industry volatility ($\log(\sigma_{t-1}^{ind})$) defined as the (log of the) square root of the variance of VW industry portfolio.

The second and third column in Table 2 present estimates of Equation (4). In the second column, the coefficient on idiosyncratic volatility is negative and significant (–0.2%); the coefficients on systematic risk exposure ($\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$) are positive and significant whereas the coefficients on market and industry risk are negative (only the coefficient on industry risk is significant). The last column presents the results of the benchmark estimation for Equation (4). The coefficient on idiosyncratic volatility stays negative and significant. The coefficient on market risk is negative and significant whereas the coefficient on industry risk is positive but insignificant. Invariably, the coefficients on systematic risk exposure ($\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$) are positive and significant.

Our estimates are consistent with Bloom (2009), who finds a negative relation between investment and the volatility of the market portfolio. Nevertheless, the positive response of investment to a firm's exposure to systematic risk remains puzzling.

3.2. Testing for Strict Exogeneity

To investigate the puzzling positive sensitivity of investment to systematic risk, we need to realize that three potential sources of endogeneity may arise from estimating Equation (3) and (4) using the baseline model. Firstly, the OLS model ignores unobserved heterogeneity by assuming that neither the risk variables nor the control variables are correlated with unobserved firm characteristics. But it is quite easy to see that this assumption is likely to be violated when estimating the relation between investment and risk. For example, a firm' growth opportunity not only has a direct impact on investment but is also likely to be correlated with the firm's exposure to systematic risk. This suggests that OLS estimates are likely to be severely biased.

Aside from unobserved heterogeneity, OLS estimation relies on the assumption that neither the risk variables nor the control variables are correlated with the error term, $\omega_{i,t}$. If investment and risk are simultaneously determined, then this assumption is clearly violated, and OLS yields biased estimates.

Lastly, although the fixed-effects estimation employed in the previous section eliminates the unobserved heterogeneity, it potentially introduces dynamic endogeneity. The fixed-effects estimation relies on a strict exogeneity assumption which implies that, in our context the risk variables that we observe today is completely independent of any past, present and future investment. This assumption is likely to be violated if there is a dynamic relation between firms' investment and risk, and thus fixed-effects estimation is likely to be inconsistent.

While there is substantial economic justification to suspect that the risk variables are not strictly exogenous, we need to confirm this with an econometric test of strict exogeneity.

Wooldridge (2010) present a regression-based test for strict exogeneity that is relatively easy to implement. If $X_{i,t-1}$ contains the explanatory variables, a test of strict exogeneity is obtained by carrying out fixed-effects estimation on the equation:

$$Capx_t/A_{t-1} = \alpha + \beta X_{i,t-1} + \gamma Y_{i,t+1} + \delta_i + \theta_t + \omega_{i,t} \tag{5}$$

Where $Y_{i,t+1}$ is a forward subset of $X_{i,t-1}$. X includes idiosyncratic and systematic risk, market volatility ($\log(\sigma_{t-1}^{mkt})$), $\beta_{i,t-1}^{mkt}$, $\beta_{i,t-1}^{ind}$ and industry volatility ($\log(\sigma_{t-1}^{ind})$). Under the null hypothesis of strict exogeneity, $\gamma = 0$. Intuitively, if $\gamma \neq 0$, then current risk measures depend on past investment rate (or conversely, present investment affects firm's future risk). Thus, if we can reject the hypothesis, then fixed-effects estimation is likely to be biased by the presence of dynamic endogeneity and we are likely to obtain less biased and more consistent estimates using a dynamic estimation procedure.

Table 3 presents the results of estimating Equation (5), with different subsets of the risk variables, $Y_{i,t+1}$. In every specification in which they are included, the coefficient estimates for the forward values of idiosyncratic and systematic risk, $\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$ are significantly different from zero. This suggests that none of these risk variables are strictly exogenous and all of these variables adjust to firm investment.

Table 3: Tests of strict exogeneity

The table reports fixed-effects estimation results of Equation (5), where the dependent variable is the investment rate ($Capx_t/A_{t-1}$). Explanatory variables include idiosyncratic risk ($\log(\sigma_{i,t-1}^{idio})$), systematic risk ($\log(\sigma_{t-1}^{syst})$), market volatility ($\log(\sigma_{t-1}^{mkt})$), $\beta_{i,t-1}^{mkt}$, $\beta_{i,t-1}^{ind}$, industry volatility ($\log(\sigma_{t-1}^{ind})$) and forward values of these risk variables. Financial control variables include lagged values of: Tobin's Q $\log(Q_{t-1})$, operating cash flows (CF_{t-1}/A_{t-2}), the firm's size ($\log(A_{t-1})$), the firm's stock return (R_{t-1}) and leverage (E_{t-1}/A_{t-1}). The definitions of these variables are the same as in Table 2. The coefficients of these control variables are suppressed for brevity. The sample period is 1970 to 2005. *F, T* denotes firm and time fixed effects, and p-values are reported in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

$Capx_t/A_{t-1}$	1	2	3
$\log(\sigma_{i,t-1}^{idio})$	−0.010***	−0.004***	−0.009***
	(0.000)	(0.005)	(0.000)
$\log(\sigma_{t-1}^{syst})$	0.002***		
	(0.000)		
$\log(\sigma_{t-1}^{mkt})$		−0.002	−0.001
		(0.226)	(0.903)
$\beta_{i,t-1}^{mkt}$		0.003***	0.001***
		(0.000)	(0.002)
$\log(\sigma_{t-1}^{ind})$		−0.015***	−0.003*
		(0.000)	(0.091)
$\beta_{i,t-1}^{ind}$		0.003***	0.001***
		(0.000)	(0.009)
$\log(\sigma_{i,t+1}^{idio})$	−0.011***	−0.013***	−0.009***
	(0.000)	(0.000)	(0.000)
$\log(\sigma_{t+1}^{syst})$	0.006***		
	(0.000)		
$\beta_{i,t+1}^{mkt}$		0.006***	0.004***
		(0.000)	(0.000)
$\beta_{i,t+1}^{ind}$		0.008***	0.006***
		(0.000)	(0.000)
Financial controls	Yes	No	Yes
Observations	101,378	101,378	101,378
R^2	0.560	0.490	0.560
Fixed effects	F,T	F	F,T

Overall, the results from Table 3 suggest the risk variables are not strictly exogenous. Thus, dynamic endogeneity is likely to be a major source of bias in estimating the relation between investment and risk.

3.3. Estimating the Relation between Investment and Risk using a GMM Dynamic Panel Estimator

Consistent estimation of Equations (3) and (4) requires the use of an estimation technique which controls for unobserved heterogeneity and simultaneity while exploiting the dynamic association between risk and investment.

An appealing estimation technique is the GMM dynamic panel estimator. This estimator, first proposed by Holtz-Eakin et al. (1988) and Arellano and Bond (1991) and further developed by Arellano and Bover (1995) and Blundell and Bond (1998), provides an excellent econometric framework for dealing with the endogeneity issues. Moreover, as Nickell (1981) shows, when estimating a dynamic panel data, a bias arises in the "small T, large N" context. Our sample has a time dimension($T = 36$) and a large firm dimension ($N = 2{,}816$). GMM dynamic panel estimator is designed for small-T large-N panels.

The dynamic GMM estimator replaces the strict exogeneity assumption with a weaker form of exogeneity, sequential exogeneity. The sequential exogeneity assumption allows the risk variables to be determined by past and present realizations of investment, but not future values. This is a fairly reasonable assumption.

This assumption implies that the risk/investment relation should be treated as a dynamic unobserved effects model and Equations (2) and (3) should be estimated as:

$$Capx_t/A_{t-1} = \alpha\,(Capx_{t-1}/A_{t-2}) + \beta X_{i,t-1} + \gamma Z_{i,t-1} + \delta_i + \theta_t + \omega_{i,t} \qquad (6)$$

Where X includes idiosyncratic and systematic risk, market volatility ($\log(\sigma_{t-1}^{mkt})$), $\beta_{i,t-1}^{mkt}$, $\beta_{i,t-1}^{ind}$ and industry volatility ($\log(\sigma_{t-1}^{ind})$).

Arellano and Bond (1991) develop a first difference GMM estimator by transforming e
Equation (6) into a system of *T-1* equations in first differences:

$$\Delta\, Capx/A_i = \Delta\beta X_i + \Delta\omega_i \qquad (7)$$

Where X_i includes the risk variables, control variables and lagged investment rates. This step eliminates the unobserved heterogeneity and allows us to have a model where our risk variables can be arbitrarily correlated with any unobserved firm characteristics.

As Arellano and Bover (1995) and Blundell and Bond (1998) point out, we can improve the GMM estimator by including the equations in levels in the estimation procedure. We can use the first-differenced variables as instruments for the equations in levels. This will produce a system GMM estimator. The system GMM estimator enables us to obtain efficient estimates while maintaining all the essential elements of controlling for unobserved heterogeneity, simultaneity and dynamic endogeneity.

The basic steps underlying this estimation strategy is as follows. First, the regression equation of investment on risk is written as a dynamic model that includes lagged investment as an explanatory variable. Next, we can take first-difference and carry out GMM estimation using lagged values of the risk, as well as lagged values of investment as GMM instruments.

Table 4 presents the GMM dynamic panel estimator of investment on risk. We report the results in the same order as in Table 2. The first column reports the GMM estimation of investment on idiosyncratic and systematic risk. The coefficient on idiosyncratic volatility remains statistically negative (–0.5%). However, the positive sign on systematic volatility has disappeared, instead, the coefficient on systematic volatility is negative (–0.1%) and statistically significant. Columns 2 and 3 present the results of investment on various components of systematic volatility. The coefficients on systematic risk exposure ($\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$) are all negative: the coefficients on $\beta_{i,t-1}^{ind}$ are significant in both columns and the coefficients on $\beta_{i,t-1}^{mkt}$ statistically significant in column 3.

Table 4: GMM dynamic panel estimator of investment on risk

The table reports the GMM dynamic panel estimator of investment on risk, where the dependent variable is the investment rate ($Capx_t/A_{t-1}$). Our measure of idiosyncratic risk, $\log(\sigma_{i,t-1}^{idio})$, is constructed from a regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Systematic volatility ($\log(\sigma_{t-1}^{syst})$) is defined as the (log of the) square root of the difference between the firm's total variance and its idiosyncratic variance. Market volatility ($\log(\sigma_{t-1}^{mkt})$) is defined as the (log of the) square root of the variance of CRSP VW index. $\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$ are coefficient estimates from the regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Industry volatility ($\log(\sigma_{t-1}^{ind})$) defined as the (log of the) square root of the variance of VW industry portfolio. Financial control variables include lagged values of the investment rate ($Lag(Capx_t/A_{t-1})$), lagged values of: Tobin's Q $\log(Q_{t-1})$ defined as in Fazzari *et al.* (1988); operating cash flows (CF_{t-1}/A_{t-2}) defined as the ratio of operating income to book assets; the firm's size ($\log(A_{t-1})$) defined as the log value of book assets; the firm's stock return (R_{t-1}); leverage (E_{t-1}/A_{t-1}) defined as the ratio of book equity to book assets. The coefficients of these control variables are suppressed for brevity. The sample period is 1970 to 2005. *F* denotes firm fixed effects, *T* denotes time fixed effects. The standard errors are clustered at the firm-level, and p-values are reported in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

$Capx_t/A_{t-1}$	1	2	3
$\log(\sigma_{i,t-1}^{idio})$	–0.005***	–0.003**	–0.004***
	(0.000)	(0.012)	(0.000)
$\log(\sigma_{t-1}^{syst})$	–0.001**		
	(0.021)		
$\log(\sigma_{t-1}^{mkt})$		0.001	–0.005***
		(0.818)	(0.006)
$\beta_{i,t-1}^{mkt}$		–0.001	–0.002***
		(0.116)	(0.000)
$\log(\sigma_{t-1}^{ind})$		–0.008***	0.001
		(0.000)	(0.775)
$\beta_{i,t-1}^{ind}$		–0.002***	–0.003***
		(0.003)	(0.000)
Financial controls	Yes	No	Yes
Observations	83,687	91,686	83,687
Significance level (p-value)	0.000	0.000	0.000

The puzzling positive sensitivity of investment to systematic volatility documented in the baseline OLS model has been eliminated with the implementation of the GMM dynamic panel estimator. These results clearly support the hypothesis that the greater the systematic risk the less the incentive to invest. The application of the GMM dynamic panel estimator removes the major sources of endogeneity inherent in the estimation of the relation between risk and investment and thus enable us to reconcile the mixed results from prior studies.

3.4 Sample Splits

Although both assume a negative investment-risk relation, the traditional view (e.g. market uncertainty under the CAPM) states that it is only the systematic risk that should matter for firm investment; real option theory predicts, on the other hand, that it is total risk that should matter for firm investment. Our results in Table 4 show that both systematic and idiosyncratic risk matter for investment. Therefore, in this section we attempt to examine the predictions of real option models to differences in the irreversibility of capital.

Table 5: GMM dynamic panel estimator of investment on risk by asset specificity

The table reports the GMM dynamic panel estimator of investment on risk, where the dependent variable is the investment rate ($Capx_t/A_{t-1}$). The sample is split into high vs. low asset specificity subsamples. Asset specificity is the ratio of machinery and equipment to total assets. High (low) asset specificity subsamples are comprised of the firms whose asset specificity is above (below) the sample median at the three-digit SIC industry level. Our measure of idiosyncratic risk, $\log(\sigma_{i,t-1}^{idio})$, is constructed from a regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Systematic volatility ($\log(\sigma_{t-1}^{syst})$) is defined as the (log of the) square root of the difference between the firm's total variance and its idiosyncratic variance. Market volatility (σ_{t-1}^{mkt}) is defined as the square root of the variance of CRSP VW index. $\beta_{i,t-1}^{mkt}$ and $\beta_{i,t-1}^{ind}$ are coefficient estimates from the regression of weekly firm-level returns on the CRSP VW index and the corresponding industry portfolio. Industry volatility (σ_{t-1}^{ind}) defined as the square root of the variance of VW industry portfolio. Financial control variables include lagged values of the investment rate ($Lag(Capx_t/A_{t-1})$), lagged values of: Tobin's Q $\log(Q_{t-1})$ defined as in Fazzari *et al.* (1988); operating cash flows (CF_{t-1}/A_{t-2}) defined as the ratio of operating income to book assets; the firm's size ($\log(A_{t-1})$) defined as the log value of book assets; the firm's stock return (R_{t-1}); leverage (E_{t-1}/A_{t-1}) defined as the ratio of book equity to book assets. The coefficients of these control variables are suppressed for brevity. The sample period is 1970 to 2005. *F* denotes firm fixed effects, *T* denotes time fixed effects. The standard errors are clustered at the firm-level, and p-values are reported in parentheses. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	High asset specificity		**Low asset specificity**	
$Capx_t/A_{t-1}$	**1**	**2**	**3**	**4**
$\log(\sigma_{i,t-1}^{idio})$	–0.0024*	–0.0026**	–0.0010	–0.0016
	(0.056)	(0.026)	(0.545)	(0.300)
$\log(\sigma_{t-1}^{syst})$	–0.0003**		–0.0009	
	(0.047)		(0.260)	
$\beta_{i,t-1}^{mkt}\ \sigma_{t-1}^{mkt}$		–0.0005**		0.0005
		(0.019)		(0.247)
$\beta_{i,t-1}^{ind}\ \sigma_{t-1}^{ind}$		–0.0006**		0.0001
		(0.011)		(0.774)
Financial controls	Yes	Yes	Yes	Yes
Observations	13,684	13,684	14,240	14,240
Significance level (p-value)	0.000	0.000	0.000	0.000

The ability to delay investment is valuable when the investment is irreversible, and the future is uncertain. The irreversibility of investment expenditures stems from capital specificity at the industry and/or at the firm level. Dixit and Pindyck (1994) argue that the irreversibility of capital is more pronounced at the industry level because capital is industry-specific.

We split the sample according to the firm's degree of irreversibility and re-estimate the relation between investment and risk using a GMM dynamic panel estimator in two subsamples. We measure a firm's degree of irreversibility using the asset specificity. As in Klasa et al. (2018) and Valta (2012), we compute asset specificity as the ratio of machinery and equipment to book assets. Then the sample is split into high vs. low asset specificity subsamples, where high (low) asset specificity subsamples are the firms whose asset specificity is above (below) the sample median at the three-digit SIC industry level.

The results are presented in Table 5. The coefficients on both market and industry risk are significantly negative for irreversible (high asset specificity) firms while the coefficients are insignificant for reversible (low asset specificity) firms. These findings are consistent with real option behaviour when capital is industry-specific. The results on firm-specific risk show a similar pattern: the coefficients are significantly negative for the irreversible subsample but insignificant for the reversible sample. Overall, the main finding is that the greater the degree of asset-specificity of capital (and hence the more irreversible the firm's investments are), the more valuable is the option to delay investment when uncertainty is high.

4. Conclusion

Despite a vast theoretical literature that predicts an increase in risk should depress investment, the existing empirical results have been mixed. We recognize that three potential sources of inconsistency may arise when estimating the relation: unobserved heterogeneity, simultaneity and dynamic endogeneity. In an attempt to address these concerns, we use a GMM dynamic panel estimator. Results show that the puzzling positive sensitivity of investment to systematic risk documented in the OLS model has been replaced with a negative relation between investment and systematic risk which supports the hypothesis that the greater the systematic risk the less the incentive to invest. We also provide empirical support for the prediction of real options models with irreversible investment.

References

Abel, A.B., Eberly, J.C., 1996, Optimal investment with costly reversibility, Review of Economic Studies 63, 581-593.

Arellano, M., Bover, O., 1995, Another look at the instrumental variable estimation of error-components models, Journal of Econometrics 68, 29-51.

Beck, T., Levine, R., Loayza, N., 2000, Finance and the sources of growth, Journal of Financial Economics 58, 261-300.

Bernanke, B.S., 1983, Irreversibility, uncertainty, and cyclical investment, Quarterly Journal of Economics 98, 85-106.

Bloom, N., 2009, The impact of uncertainty shocks, Econometrica 77, 623-685.

Blundell, R., Bond, S., 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics 87, 115-143.

Bulan, L.T., 2005, Real options, irreversible investment and firm uncertainty: new evidence from US firms, Review of Financial Economics 14, 255-279.

Caselli, F., Esquivel, G., Lefort, F., 1996, Reopening the convergence debate: a new look at cross-country growth empirics, Journal of Economic Growth 1, 363-389.

Dixit, A.K., Pindyck, R.S., 1994. Investment under uncertainty. Princeton university press.

Erickson, T., Whited, T.M., 2000, Measurement error and the relationship between investment and q, Journal of Political Economy 108, 1027-1057.

Fama, E.F., French, K.R., 1997, Industry costs of equity, Journal of Financial Economics 43, 153-193.

Fazzari, S.M., Hubbard, R.G., Petersen, B.C., Blinder, A.S., Poterba, J.M., 1988, Financing constraints and corporate investment, Brookings Papers on Economic Activity 1988, 141-206.

Froot, K.A., Scharfstein, D.S., Stein, J.C., 1993, Risk management: Coordinating corporate investment and financing policies, Journal of Finance 48, 1629-1658.

Gulen, H., Ion, M., 2016, Policy uncertainty and corporate investment, Review of Financial Studies 29, 523-564.

Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988, Estimating vector auto-regressions with panel data, Econometrica, 1371-1395.

Julio, B., Yook, Y., 2012, Political uncertainty and corporate investment cycles, Journal of Finance 67, 45-83.

Klasa, S., Ortiz-Molina, H., Serfling, M., Srinivasan, S., 2018, Protection of trade secrets and capital structure decisions, Journal of Financial Economics 128, 266-286.

Leahy, J., Whited, T.M., 1996, The effects of uncertainty on investment: Some stylized facts, Journal of Money, Credit, and Banking 28, 64-83.

Nickell, S., 1981, Biases in dynamic models with fixed effects, Econometrica, 1417-1426.

Panousi, V., Papanikolaou, D., 2012, Investment, idiosyncratic risk, and ownership, Journal of Finance 67, 1113-1148.

Salinger, M., Summers, L.H., 1983, Tax reform and corporate investment: a microeconometric simulation study, In: Behavioral Simulation Methods in Tax Policy Analysis. University of Chicago Press, 247-288.

Valta, P., 2012, Competition and the cost of debt, Journal of Financial Economics 105, 661-682.

Wintoki, M.B., Linck, J.S., Netter, J.M., 2012 Endogeneity and the dynamics of internal corporate governance, Journal of Financial Economics 105, 581-606.

Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data. MIT Press, Cambridge, Massachusetts.

APPLIED FINANCE LETTERS

VOLUME 08, 2019

COINTEGRATION, PRICE-ADJUSTMENT DELAYS, AND OPTIMAL HEDGE RATIO IN THE PRECIOUS METAL MARKETS

Hirofumi Nishi

Corresponding Author: Hirofumi Nishi, Robbins College of Business and Entrepreneurship, Fort Hays State University, Hays, Kansas USA ☎ + (785) 628-5337 ✉ h_nishi@fhsu.edu

Abstract: Firms seeking to apply hedge accounting treatment under the Accounting Standards Codification Topic 815 must demonstrate higher hedge effectiveness, for which the regression analysis is commonly used as a testing method. An autoregressive distributed lag (ARDL) model is adopted in this article to examine the hedge effectiveness in the presence of a long-run cointegrating relationship between spot and futures prices while spot contracts are traded far less frequently. Using precious metal market data, our study empirically demonstrates that a hedge ratio estimated with a conventional OLS model tends to be downwardly biased. It is also shown that whether this omitted-variable bias is observable depends on the liquidity in a futures market.

Keywords: Commodity prices; Hedge accounting; Cointegration; Infrequent trading

1. Introduction

For both firms and investors, large earnings fluctuations associated with the derivative contracts used for hedging purposes are not desirable. In order to mitigate the earnings impact, the Accounting Standards Codification (ASC) Topic 815 issued by the Financial Accounting Standards Board (FASB) allows the gains and losses of derivatives to be combined in the same period with the gains and losses of the underlying hedged assets. Firms seeking to apply this special hedge accounting treatment must demonstrate the effectiveness of the hedge although the choice of the testing methodology is left to firms' discretion. One of the most commonly-used methods is the regression analysis, where the slope coefficient represents the optimal hedge ratio (i.e., the quantity of hedging instrument).[1] In this approach, the hedge effectiveness is generally considered high if the coefficient of determination, or R-squared, is 0.80 or greater.

Since Johnson (1960) introduced the portfolio theory into the study of hedging strategy, the optimal hedge ratio has been defined as the ratio of the covariance between the changes in spot and futures prices and the variance of the futures price changes. Ederington (1979), building on the foundation laid out by Johnson (1960), measures the effectiveness of a hedging strategy as the percent reduction in the variance between the hedged return and the unhedged return. More recent studies, however, have pointed out that the optimal hedge ratio estimation using conventional ordinary least squares (OLS) model does not necessarily incorporate all available information and therefore suffers from possible information inefficiency. While Myers and Thompson (1989) propose adding lagged spot and futures prices to a hedge ratio estimation, Viswanath (1993) shows that adding the spot-futures basis makes the hedge ratio estimate closer to unity than the conventional OLS estimate for most of the hedging

[1] Another commonly used approach is the dollar offset method with the 80%–125% rule.

durations.[2] Based on the analysis of the relationship between the basis and a futures contract maturity, Castelino (1992) shows that the minimum-variance hedge ratio increases as the hedge-lifting date approaches the contract maturity.

Information inefficiency in the OLS model is closely related to two econometric issues, one of which is the existence of a long-term relationship between spot and futures prices. The long-run equilibrium between two or more time series, the concept introduced by Granger (1981) and known as cointegration, often exists in the commodity markets. The effect on the hedge ratio estimate of the omission of a spot-future cointegration has been studied from various perspectives. Ghosh (1993) and Kroner and Sultan (1993) both suggest that incorporating the long-run cointegrating relationship between asset prices significantly improve the optimal hedge ratio estimate. Lien (2004) theoretically assesses the effect of a long-run relationship between spot and futures prices and suggests that omitting cointegration leads to a smaller hedge ratio. In contrast, Chen et al. (2004), based on the data of 25 different commodities, conclude that the conventional OLS hedge ratio estimate ultimately approaches unity if the hedge horizon is sufficiently long.[3] Likewise, Juhl et al. (2012) conclude that including the error-correction term to account for cointegration does not significantly improve the estimation performance for a long hedge horizon. The relationship between the hedge effectiveness and the hedge horizon is also discussed in section 3 of this article.

The second issue affecting the hedge effectiveness in the conventional OLS approach is infrequent trading, or low transaction volumes, of certain commodity contracts. The previous studies, thin trading has been investigated primarily in equity or bond markets. For example, Scholes and Williams (1977) and Dimson (1979) both examine the beta estimated by the capital asset pricing model (CAPM) in the presence of nonsynchronous trading and provide adjusted beta estimates. Lo and MacKinlay (1990a) utilize the data from various stocks grouped by firm size and generalize the models on thin trading by using stochastic intervals between trades. Using the data of thinly-traded stocks of Finnish firms, Luoma et al. (1993) conclude that the estimated error-correction term is highly dependent on the trading frequency of the underlying stock.[4] Wilkinson et al. (1999) analyse the New Zealand and Australian debt securities and argue that the lead-lag relations caused by infrequent trading do not impact hedge effectiveness.

The present article aims to expand the literature on commodity hedging by connecting two econometric issues: cointegration between spot and futures prices and infrequent trading in spot and futures markets. Despite a voluminous literature on liquidity in capital markets, little statistical work has been done in the context of commodity hedging and on how infrequent trading of a commodity contract affects the omitted-variable bias in the presence of spot-futures cointegration. Though seemingly unrelated, these issues are closely relevant to each other. The existence of a long-term equilibrium relationship between two prices implies that at least one of them must be pulled back to the equilibrium before deviating from it too far or too long. This means that, if the transaction frequency of a spot contract is significantly lower than that of a futures contract (i.e., a futures market is far more liquid than a spot market), the bias caused by ignoring their cointegrating relationship could be exacerbated. In order to test this effect, our study specifically utilizes the data from the precious metal markets where spot and futures contracts are traded at very different frequencies, yet their prices are still cointegrated. The autoregressive distributed lag (ARDL) model is used to estimate the optimal hedge ratio for cointegrated commodity price series while a long-run cointegrating relationship between the price series is tested with the ARDL bounds test procedure introduced by Pesaran et al. (2001).

---

[2] Viswanath (1993) notes that this approach is only correct if the spot-futures convergence is guaranteed. Such convergence often fails to take place in agricultural markets possibly because of the design of the exchange's delivery system (Aulerich et al., 2011).

[3] Also see Howard and D'Antonio (1991) and Benet (1992) for the studies on the relationship between the hedge period and hedging effectiveness.

[4] Luoma et al. (1993) and Chen et al. (2004) both utilize the model that only includes the contemporaneous and the one-period-lagged values of each variable (i.e., ARDL (1, 1)).

Briefly highlighting the results in this article, our study empirically demonstrates that an ARDL-adjusted estimate of a spot-futures hedge ratio tends to be higher than a conventional OLS estimate. This implies that an OLS estimate is downwardly biased when a cointegrating relationship between two prices is ignored. This is consistent to Lien's (2004) proposition. Our result also indicates that bias caused by the omission of a long-run equilibrium relation is associated with the liquidity in a futures market. We compare the daily transaction volumes of futures contracts across commodities, and it appears that the omitted-variable bias becomes observable only when a futures market is relatively active.

The remainder of this paper is organized as following. Section 2 describes the data used in this study and provides a brief description of the conventional OLS-based and the ARDL-adjusted hedge ratios. Various econometric biases in the hedge ratio estimation are also discussed. Section 3 presents the result of the ARDL bounds test for cointegration followed by the comparison of the OLS and the ARDL-based hedge ratios. Section 4 provides the summary.

2. Data and Methodologies

2.1 Data and Time-Series Structures

The daily spot price data for the period between January 2006 and June 2016 are collected from the London Bullion Market Association (LBMA). The LBMA gold auction takes place twice a day, at 10:30 a.m. and 3:00 p.m. London time, and administered by ICE Benchmark Administration (IBA). The LBMA silver auction takes place once a day at 12:00 noon and is operated by the CME Group and administered by Thomson Reuters. Lastly, the LBMA platinum and palladium auctions take place twice a day, at 9:45 a.m. and 2:00 p.m., and are administered by the London Metal Exchange (LME). For gold, platinum, and palladium, the prices observed in the afternoon session are used in this study.

The daily price settlement data of the Commodity Exchange Inc. (COMEX) gold and silver futures contracts and the New York Mercantile Exchange (NYMEX) palladium and platinum futures contracts during the sample period come from the Bloomberg Professional Service. The settlement prices, instead of intra-day futures prices, are used due to their transparency. Note that there are up to several hours of difference between the afternoon session of a LBMA auction (12:00 p.m., 2:00 p.m., or 3:00 p.m. London time) and the settlement of a COMEX/NYMEX futures contract (1:30 p.m. EST). In practice, however, this time lag has limited impact on a firm's hedging activities as a hedge horizon is typically set longer than one day. Moreover, the adjustment between spot and futures prices usually takes longer than several trading days, making several hours of difference relatively insignificant.[5]

Our study uses the prices of the second nearest-to-delivery futures contracts. That is, once the second month becomes the front month, the data rolls over to the next contract month. The second nearest-to-delivery contracts are used because the transaction volume of a metal futures contract tends to decline sharply after it becomes the nearest-to-delivery contract. This also implies that the second nearest-to-delivery contracts are most frequency traded. The weekly data are retrieved from every Tuesday; if Tuesday is not available in one or both the spot and futures markets (e.g., non-trading day), then Wednesday is used. If both Tuesday and Wednesday are unavailable, then Monday is used. This strategy ensures that each spot/futures price change pair is encompassed about seven days. The monthly contract data are retrieved on the first trading day of each month.[6] Table 1 presents the statistics of the daily spot prices and the daily second nearest-to-delivery futures contract settlement prices observed during the sample period.

[5] An unreported linear Granger-causality test suggests that, for all the commodities, the daily futures price returns lead the daily spot price returns for up to 12 trading days.

[6] The price series for one-month and one-week hedge periods are constructed as described here in order to avoid auto-correlated residuals caused by overlapping data bias.

Table 1: Daily Spot and Futures Precious Metal Prices: January 2006 - June 2016

A single asterisk (*) indicates the 5% level of significance and a double asterisk (**) means the 1% level of significance or better for the Jarque–Bera (JB) test on the hypothesis that the variable is normally distributed. The third column indicates the number of LBMA auctions held per day (spot market) and the average daily transaction volume of the second nearest futures contract during the sample period (futures market).

		Freq.	Price				Test for normality		
	Obs.	**/ Vol.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**JB test**
Gold									
LBMA Spot	2598	2/day	520.75	1891.00	1147.92	342.23	0.0056	2.0967	88.34**
COMEX Futures	2598	131,022	527.80	1889.00	1148.38	341.86	0.0064	2.0966	88.36**
Platinum									
LBMA Spot	2598	2/day	756.00	2276.00	1382.39	289.34	0.2023	2.6107	34.12**
NYMEX Futures	2598	7,748	760.00	2275.00	1381.43	288.68	0.2010	2.6334	32.03**
Palladium									
LBMA Spot	2598	2/day	168.00	901.00	541.07	198.80	-0.1422	1.6746	198.91**
NYMEX Futures	2598	3,740	167.00	900.00	539.80	195.80	-0.1441	1.6914	194.37**
Silver									
LBMA Spot	2598	1/day	8.83	48.70	19.74	7.83	1.0304	3.1453	461.97**
COMEX Futures	2598	43,433	8.79	47.53	19.72	7.79	1.0111	3.0772	443.31**

2.2 Conventional OLS and ARDL-Adjusted Hedge Ratios

Suppose that S_t and F_t represent the spot price of a certain commodity and the corresponding futures price, both at time *t*, respectively. Following Chen et al. (2004) and Juhl et al. (2012), our study applies the OLS regression model to first-differenced price series as following.

$$\Delta S_t = \alpha + \beta_{OLS}\Delta F_t + \varepsilon_t \tag{1}$$

where $\Delta S_t = S_t - S_{t-1}$ and $\Delta F_t = F_t - F_{t-1}$. β_{OLS} is the optimal hedge ratio between a spot contract and a futures contract and it is estimated as the ratio of the covariance between the changes in spot and futures prices to the variance of the futures price changes.

$$\hat{\beta}_{OLS} = \frac{Cov(\Delta S_t, \Delta F_t)}{Var(\Delta F_t)} \tag{2}$$

One statistical problem in this approach is that, especially in commodity markets, the transaction frequencies of spot contracts are often far less than those of the corresponding futures contracts. For example, the LBMA gold auction takes place twice a day at 10:30 a.m. and 3:00 p.m. in London time. Once the net volume of the buy and sell orders falls within the pre-determined tolerance level of imbalance, 10,000 ounces, then all the volume becomes tradeable at the initial price set by the chairperson.[7] The transaction frequencies in this system are clearly far less than those in the futures contracts at COMEX or NYMEX. In addition, the spot contracts typically have less transaction volumes than actively-traded futures contracts.[8] Infrequent trading in a spot market lowers the covariance estimate in Equation (2), which subsequently lowers the hedge ratio estimate.

[7] If the imbalance exceeds the tolerance level, the auction restarts with a revised auction price and continues until the equilibrium price is set.
[8] For example, 2.958 million COMEX gold futures contract (≈ 295.8 million troy ounces) were traded during January 2016 while the total volume transacted at the LBMA in the same period was 5.701 million troy ounces (bids and asks combined). This represents a difference of 52-to-l.

Perhaps, a more critical problem in the conventional OLS method is that it does not take into account the possibility that spot and futures prices are mutually cointegrated. As shown by the previous studies, the OLS approach is likely to suffer from omitted-variable bias if there exists a long-term equilibrium relationship between spot and futures prices. In contrast, the ARDL model allows the effect of an independent variable on the dependent variable to be distributed over time while lagged values of the dependent variable itself can also be additional regressors. In other words, there will be the immediate effect of a variable at time *t* followed by delayed effects taking place in later periods. Suppose the ARDL(p, q) model with an unrestricted intercept and time trend as following.

$$S_t = c_0 + c_1 t + \sum_{i=1}^{p} \Phi_i S_{t-i} + \sum_{i=0}^{q} \beta_i F_{t-i} + \mu_t \tag{3}$$

The lag orders for the dependent and independent variables in this model are equal to p and q, respectively. Following Pesaran and Shin (1998), the equation can be transformed into the error-correction form as following.

$$\Delta S_t = c_0 + c_1 t + \sum_{i=1}^{p-1} \lambda_{S,i} \Delta S_{t-i} + \sum_{i=0}^{q-1} \lambda_{F,i} \Delta F_{t-i} + \Phi(1)(S_{t-1} - \theta F_{t-1}) + \mu_t \tag{4}$$

where $$\theta = \frac{\sum_{i=0}^{q} \beta_i}{\Phi(1)}$$

and $$\Phi(1) = 1 - \sum_{i=1}^{p} \Phi_i$$

The set of $\lambda_{F,i}$ collectively reflects the effect of short-term changes in the futures price on the spot price change. $\Phi(1)(S_{t-1} - \theta F_{t-1})$ is referred to as the error-correction term; θ indicates the long-run equilibrium relation between the spot and futures prices while $\Phi(1)$ represents the speed of adjustment to a temporary deviation from such equilibrium.

The ECM derived from the ARDL approach has advantages relative to the two-step error correction model (ECM) shown by Engle and Granger (1987). First, the ARDL approach is a non-residuals-based method that yields the short-run and long-run parameter estimates in a single equation at a time. In addition, the ARDL bounds test procedure can be applied irrespective of whether the regressors are purely I(0), purely I(1), or mutually cointegrated. In contrast, conventional cointegration testing procedures, such as the Engle-Granger (1987) test and Johansen's (1988) rank test, require all the variables to be integrated of the same order and therefore pre-testing of unit roots is necessary.[9] The ARDL bounds test also provides robust results with a relatively small sample size (< 80) when using the critical value bounds calculated by Narayan (2005).[10]

[9] It is still advisable to do so in order to confirm that none of the variables is integrated of higher order.
[10] Given relatively large sample sizes, this study utilizes the critical value bounds provided by Pesaran et al. (2001).

3. Empirical Results and Discussion

3.1. ARDL Bounds Test for Cointegration

The first step of our analysis is to verify the existence of a long-run relationship between spot and futures prices in each of the precious metal markets. The ARDL bounds test is utilized for this purpose. As mentioned in sub-section 2.2, prior knowledge about the order of integration of each variable is not necessary in this approach. The optimal lag orders for spot and futures prices, denoted as p* and q* respectively, are determined using the Akaike Information Criterion (AIC) with the maximum lag order of 12.

Table 2: ARDL Bounds Test for Cointegration between Spot and Futures Prices

The optimal lag orders for the first-differenced spot price (= p*) and the first-differenced futures price (= q*) are determined with the Akaike Information Criterion (AIC). The F-test is performed against the joint hypothesis that the coefficients of the contemporaneous and lagged futures prices are collectively zero and the speed-of-adjustment coefficient is zero. The t-test is performed against the null hypothesis of zero speed-of-adjustment coefficient. The bounds on the critical values obtained from Pesaran et al. (2001). A single asterisk (*) and a double asterisk (**) indicate a significance level of 5% and 1%, respectively.

$$\Delta S_t = c_0 + c_1 t + \sum_{i=1}^{p-1} \lambda_{S,i} \Delta S_{t-i} + \sum_{i=0}^{q-1} \lambda_{F,i} \Delta F_{t-i} + \Phi(1)\,(S_{t-1} - \theta F_{t-1}) + \mu_t$$

Commodity	Data frequency	ARDL(p*, q*)	F-statistic	t-statistic
Gold	**Daily**	(11, 12)	77.351**	-12.438**
	Weekly	(12, 5)	74.405**	-12.199**
	Monthly	(1, 1)	70.284**	-11.855**
Platinum	**Daily**	(12, 12)	52.877**	-10.275**
	Weekly	(1, 1)	4179.755**	-91.054**
	Monthly	(1, 1)	3393.070**	-80.258**
Palladium	**Daily**	(12, 11)	47.225**	-9.717**
	Weekly	(8, 7)	27.690**	-7.441**
	Monthly	(2, 2)	47.478**	-9.727**
Silver	**Daily**	(11, 12)	61.001**	-11.044**
	Weekly	(6, 7)	55.696**	-10.544**
	Monthly	(10, 1)	1018.274**	-44.604**

The test is conducted by comparing the F-statistic and t-statistic against the lower and the upper bounds for the asymptotic critical values. The bounds F-test is performed against the joint null hypothesis that the coefficients of the contemporaneous and lagged futures prices are collectively zero (i.e., no long-run equilibrium relationship between spot and futures prices) and the speed-of-adjustment coefficient is zero.

$$H_0^F: \sum_{i=0}^{q} \beta_i = 0 \cap \Phi(1) = 0$$

If the F-statistic exceeds the upper bound, this indicates that there exists a long-run relationship between spot and futures prices; if the F-statistic is below the lower bound, there is no cointegration. If the F-statistic falls between the bounds, the test is inconclusive. As a supplemental test to confirm the validity of the abovementioned bounds F-test, a bounds t-test can be performed against the null hypothesis of zero speed-of-adjustment coefficient. If the test statistic is greater than the upper bound of the critical value, one concludes that there is a cointegrating relationship.

$$H_0^t: \Phi(1) = 0$$

3.2. Optimal Hedge Ratio Estimation

The primary objective of this study is to examine the impact of cointegration between spot and futures prices on the hedge ratio estimate. This section therefore involves two contrasting approaches: the model that takes into account a long-run equilibrium relationship between two prices, and the one that does not. First, the results with the conventional OLS regression models are shown in Table 3. Panels A, B, and C represent one-day, one-week, and one-month hedge horizons, respectively. The fourth column indicates the optimal hedge ratios estimated based on the first-differenced spot and futures prices (= $\hat{\beta}_{\mathrm{OLS}}$).

Table 3: OLS Hedge Ratio Estimations with 1-Day, 1-Week, and 1-Month Hedge Horizons

a is the constant. $\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}$ is the coefficient of the current change in the futures price and represents the short-run hedge ratio. A single asterisk (*) and a double asterisk (**) indicate a significance level of 5% and 1%, respectively, for the Wald tests on null hypotheses: a = 0 and $\boldsymbol{\beta}_{\mathbf{OLS}}$ = 1.

$$\Delta \mathrm{S}_t = \alpha + \beta_{\mathrm{OLS}} \Delta \mathrm{F}_t + \varepsilon_t$$

Panel A: One-Day Hedging				
Commodity	**Obs.**	**a**	$\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}$	**Adj. $\mathbf{R^2}$**
Gold	2597	0.1805	0.4155**	0.1745
Platinum	2597	-0.0014	0.6622**	0.3997
Palladium	2597	0.0638	0.4867**	0.2300
Silver	2597	0.0018	0.4765**	0.1977
Panel B: One-Week Hedging				
Commodity	**Obs.**	**a**	$\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}$	**Adj. $\mathbf{R^2}$**
Gold	547	0.1994	0.8719**	0.7754
Platinum	547	0.0052	0.9619*	0.8786
Palladium	547	0.0667	0.8917**	0.7675
Silver	547	0.0038	0.7418**	0.6928
Panel C: One-Month Hedging				
Commodity	**Obs.**	**a**	$\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}$	**Adj. $\mathbf{R^2}$**
Gold	125	0.4213	0.9441*	0.9132
Platinum	125	0.0445	0.9487**	0.9637
Palladium	125	0.0998	0.9752	0.9490
Silver	125	0.0043	0.9370*	0.8887

The Wald test conducted on the null hypothesis that the hedge ratio is equal to unity is rejected at a significance level of 0.05 or better for all the markets and hedge horizons, except for palladium with a one-month hedge period. Nevertheless, it is worthwhile to note that the OLS hedge ratio estimate, as well as adjusted R-squared, approaches unity as the length of the hedge period increases. This is consistent to the finding in Chen et al. (2004) based on different commodities and implies that, when more time for adjustments is given, the price-adjustment delays due to infrequent transactions become less prominent.

Table 4 presents a comparison between the OLD estimates of spot-futures hedge ratios and the ARDL-adjusted hedge ratios. Consistent with Table 3, Panels A, B, and C represent one-day, one-week, and one-month hedge horizons, respectively. The third column indicates the model specifications; the optimal lag orders for spot prices (= p*) and futures prices (= q*) are determined with the AIC. The OLS hedge ratios shown in the fifth column (= $\hat{\beta}_{\mathrm{OLS}}$) are compared to the ARDL-adjusted estimate in the sixth column (= $\hat{\lambda}_{\mathrm{F},0}$). The seventh and eighth columns indicate the long-run equilibrium relation between spot and futures prices (= θ) and the speed of adjustment toward the equilibrium (= $\Phi(1)$), respectively.

Table 4: ARDL-Adjusted Hedge Ratio Estimations with 1-Day, 1-Week, and 1-Month Horizons

ARDL (p*, q*) indicates the model specifications, where the optimal lag orders for the first-differenced spot price (= p*) and the first-differenced futures price (= q*) are determined with the Akaike Information Criterion (AIC). $\mathbf{c_0}$ is the constant. $\hat{\boldsymbol{\lambda}}_{\mathbf{F,0}}$ is the coefficient of the current change in the futures price and is considered the short-run hedge ratio. θ represents the long-run equilibrium relationship between spot and futures prices and Φ(1) is the speed of adjustment toward the equilibrium. A single asterisk (*) and a double asterisk (**) indicate a significance level of 5% and 1%, respectively, for the Wald tests on null hypotheses: $\mathbf{c_0}$ = 0, $\boldsymbol{\beta}_{\mathbf{OLS}}$ = 1, $\boldsymbol{\lambda}_{\mathbf{F,0}}$ = 1, and θ = 1.

$$\Delta S_t = c_0 + c_1 t + \sum_{i=1}^{p-1} \lambda_{S,i} \Delta S_{t-i} + \sum_{i=0}^{q-1} \lambda_{F,i} \Delta F_{t-i} + \Phi(1)\,(S_{t-1} - \theta F_{t-1}) + \mu_t$$

Commodity	Obs.	ARDL(p*, q*)	c_0	$\hat{\beta}_{OLS}$	$\hat{\lambda}_{F,0}$	$\hat{\theta}$	$\hat{\Phi}(1)$	Adj. R^2
Panel A: One-Day Hedging								
Gold	2586	(11, 12)	-0.6933	0.4155**	0.4020**	1.0007	-0.6654	0.7842
Platinum	2586	(12, 12)	-1.5159	0.6622**	0.5994**	1.0028	-0.5136	0.7607
Palladium	2586	(12, 11)	-2.7251**	0.4867**	0.4316**	1.0155**	-0.3877	0.7685
Silver	2586	(11, 12)	-0.0233	0.4765**	0.4799**	1.0052	-0.2963	0.6586
Panel B: One-Week Hedging								
Commodity	**Obs.**	**ARDL(p*, q*)**	c_0	$\hat{\beta}_{OLS}$	$\hat{\lambda}_{F,0}$	$\hat{\theta}$	$\hat{\Phi}(1)$	Adj. R^2
Gold	536	(12, 5)	-1.2254	0.8719**	0.8859**	1.0010	-1.2014	0.8995
Platinum	536	(1, 1)	-3.5989	0.9619*	0.9536**	1.0037	-0.9501	0.9398
Palladium	536	(8, 7)	-5.0654**	0.8917**	0.8907**	1.0153**	-0.7172	0.8919
Silver	536	(6, 7)	-0.0323	0.7418**	0.7993**	1.0023	-0.9839	0.8470
Panel C: One-Month Hedging								
Commodity	**Obs.**	**ARDL(p*, q*)**	c_0	$\hat{\beta}_{OLS}$	$\hat{\lambda}_{F,0}$	$\hat{\theta}$	$\hat{\Phi}(1)$	Adj. R^2
Gold	114	(1, 1)	0.4352	0.9441*	0.9666	0.9988	-1.1113	0.9610
Platinum	114	(1, 1)	4.1674	0.9487**	0.9426**	0.9968	-0.9456	0.9836
Palladium	114	(2, 2)	-7.0151**	0.9752	0.9583**	1.0120**	-1.2917	0.9774
Silver	114	(10, 1)	0.0296	0.9370*	1.0094	1.0002	-1.0092	0.9499

The ARDL-based model exhibits higher adjusted R-squared than the conventional OLS model does, regardless of the commodity or the hedge length. This is in line with expectation since an ARDL model includes additional variables (i.e., lagged futures prices). More interestingly, the comparison of the hedge ratio estimates highlights a clear contrast between actively-traded commodities and those that are not. For example, with respect to gold and silver, the ARDL-adjusted estimate is higher than that yielded in the corresponding OLS model with one-week or longer hedge periods. The ARDL-adjusted estimate of the one-week hedge ratio for gold (silver) is 0.8859 (0.7993), which is slightly higher than 0.8719 (0.7418) estimated with the OLS method. The difference between the hedge ratio estimates based on these two approaches is even greater with a one-month period; the ARDL estimate of gold (silver) hedge ratio, 0.9666 (1.0094), is substantially higher than the OLS estimate of 0.9441 (0.9370). Moreover, for both the commodities, the null hypothesis that the one-month ARDL hedge ratio is equal to unity cannot be rejected at a significance level of 0.05. Overall, our analysis on gold and silver markets empirically supports Lien's (2004) proposition that the omission of a long-run cointegrating relationship between two prices leads to a smaller hedge ratio estimate.

In contrast, the ARDL-adjusted hedge ratio estimate for platinum and palladium are almost equal to or smaller than the corresponding OLS estimates. Because this trend is consistent regardless of the hedge horizon, the variation in the results must be rather associated with differences in futures markets. While the LBMA auction is only held once or twice a day, futures transactions occur virtually on a continuous basis throughout a day. Nevertheless, futures trading volumes significantly vary across commodities. As shown in Table 1, the average numbers of gold, silver, platinum, and palladium futures contracts transacted per day during the sample period are in the ratios of 35:12:2:1. These ratios provide a rough proxy of the differences in trading frequencies across futures markets, and it is clearly indicated that the platinum and palladium are far less frequently traded than gold or silver. The contrast in our findings suggests that statistical bias created by ignoring spot-futures cointegration (i.e., a disadvantage of the OLS model) may be observable only if futures contracts are actively traded.

4. Conclusions

Risk managers seeking to apply special hedge accounting treatment under the ASC Topic 815 must demonstrate higher hedge effectiveness while the exact regression design for prospective and retrospective effectiveness testing has been debated over years. This article aims to simultaneously examine two econometric issues causing Information inefficiency in the conventional OLS hedge ratio: cointegration between spot and futures prices and infrequent trading in spot market. We specifically utilize the data from the precious metal markets, where spot and futures contracts are traded at very different frequencies while their prices are mutually cointegrated. The ARDL bounds test procedure is used to test a long-run equilibrium relationship between spot and futures prices.

The result in our study demonstrates that, in the gold and silver markets, a hedge ratio estimated with a conventional OLS model tends to be lower than an ARDL-adjusted estimate. Our finding empirically supports that the omission of a long-term spot-futures equilibrium relationship leads to a downwardly-biased hedge ratio estimate. On the other hand, such statistical bias is not observed with respect to the platinum or palladium markets. The contrast in our findings indicates that whether omitted-variable bias in the presence of spot-futures cointegration can be observed depends on the liquidity in a futures market.

Acknowledgements

My paper has greatly benefited from the insightful comments of an anonymous referee. I also thank Don MacDonald for his helpful suggestions. All errors are my own.

References

Aulerich, N.M., Fishe, R.P. and Harris, J.H., 2011. Why do expiring futures and cash prices diverge for grain markets?. *Journal of Futures Markets*, *31*(6), 503-533.

Benet, B.A., 1992. Hedge period length and Ex-ante futures hedging effectiveness: The case of foreign-exchange risk cross hedges. *Journal of Futures Markets*, *12*(2), 163-175.

Castelino, M.G., 1992. Hedge effectiveness: Basis risk and minimum-variance hedging. *Journal of Futures Markets*, *12*(2), 187-201.

Chen, S.S., Lee, C.F. and Shrestha, K., 2004. An empirical analysis of the relationship between the hedge ratio and hedging horizon: A simultaneous estimation of the short-and long-run hedge ratios. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *24*(4), 359-386.

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, *7*(2), 197-226.

Ederington, L.H., 1979. The hedging performance of the new futures markets. *The Journal of Finance*, *34*(1), 157-170.

Engle, R.F. and Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

Ghosh, A., 1993. Hedging with stock index futures: Estimation and forecasting with error correction model. *Journal of Futures Markets*, *13*(7), 743-752.

Granger, C.W., 1981. Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, *16*(1), 121-130.

Howard, C.T. and D'Antonio, L.J., 1991. Multiperiod hedging using futures: A risk minimization approach in the presence of autocorrelation. *Journal of Futures Markets*, *11*(6), 697-710.

Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, *12*(2-3), 231-254.

Johnson, L.L., 1960. The theory of hedging and speculation in commodity futures. *The Review of Economic Studies*, *27*(3), 139-151.

Juhl, T., Kawaller, I.G. and Koch, P.D., 2012. The effect of the hedge horizon on optimal hedge size and effectiveness when prices are cointegrated. *Journal of Futures Markets*, *32*(9), 837-876.

Kroner, K.F. and Sultan, J., 1993. Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of financial and quantitative analysis*, *28*(4), 535-551.

Lien, D., 2004. Cointegration and the optimal hedge ratio: the general case. *The Quarterly review of economics and finance*, *44*(5), 654-658.

Lo, A.W. and MacKinlay, A.C., 1990. An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, *45*(1-2), 181-211.

Luoma, M., Martikainen, T. and Perttunen, J., 1993. Thin trading and estimation of systematic risk: an application of an error-correction model. *Annals of Operations Research*, *45*(1), 297-305.

Myers, R.J. and Thompson, S.R., 1989. Generalized optimal hedge ratio estimation. *American Journal of Agricultural Economics*, *71*(4), 858-868.

Narayan, P.K., 2005. The saving and investment nexus for China: evidence from cointegration tests. *Applied economics*, *37*(17), 1979-1990.

Pesaran, M.H. and Shin, Y., 1998. An autoregressive distributed-lag modelling approach to cointegration analysis. *Econometric Society Monographs*, *31*, 371-413.

Pesaran, M.H., Shin, Y. and Smith, R.J., 2001. Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics*, *16*(3), 289-326.

Scholes, M. and Williams, J., 1977. Estimating betas from nonsynchronous data. *Journal of financial economics*, *5*(3), 309-327.

Viswanath, P.V., 1993. Efficient use of information, convergence adjustments, and regression estimates of hedge ratios. *Journal of Futures Markets*, *13*(1), 43-53.

Wilkinson, K.J., Rose, L.C. and Young, M.R., 1999. Comparing the effectiveness of traditional and time varying hedge ratios using New Zealand and Australian debt futures contracts. *Financial Review*, *34*(3), 79-94

APPLIED FINANCE LETTERS

VOLUME 08, 2019

DOES DEBT DIVERSIFICATION LEAD TO A DISCOUNT IN FIRM VALUE?

Nemiraja Jadiyappa[1], Namrata Saikia[2], Bhavik Parikh[3]

[1] Birla Institute of Technology and Science, Pilani. Hyderabad campus.
[2] Indiana University of Pennsylvania
[3] Saint Francis Xavier University

Corresponding Author: Bhavik Parikh, Assistant Professor, Department of Accounting and Finance, Gerald Schwartz School of Business, Saint Francis Xavier University, Office # 382, 3090 Martha Drive, Antigonish, Nova Scotia- B2G 2W5, Canada ☎ +1-902-968-4926 ✉ bparikh@stfx.ca

Abstract: Corporate firms access multiple sources of debt simultaneously. This study analyses the impact of debt diversification on firm value. We argue that, when firms diversify their debt sources, the monitoring role played by debt holders decreases as a result of the free rider problem. Hence, such firms should experience a value discount in the capital markets. Our empirical analysis provides evidence for the existence of a value discount in the capital markets for firms accessing multiple sources of debt. Our results remain robust for alternative measures of debt diversification.

Keywords: Debt Diversification, Debt Specialization, Firm Value, Free Rider, Agency Costs

1. Introduction

Rauh and Sufi (2010) and Colla et al. (2013) document the existence of debt diversification, i.e., accessing multiple sources of debt, among US corporate firms, as a common occurrence. Johnson (1997) reports that about 73% of US firms use more than one source of debt for a given level of debt. Surprisingly, despite the wide prevalence, the phenomenon of debt diversification is yet to be thoroughly examined. One such pertinent area is the impact of debt diversification on firm value. Managers of a firm are expected to take decisions that increase shareholders' wealth. This then leads us to the question whether the managerial decision to go for multiple sources of debt is a value-adding decision or not. In other words, do firms with diversified sources of debt command a better value in the capital markets?

The theoretical rationale for a relationship between debt diversification and firm value can be traced to Jaffee and Russell (1976) and Stiglitz and Weiss (1981) who argue the existence of credit rationing in the financial markets. Credit rationing limits the ability of firms to raise the required amount of debt from a single source or lender which in turn could potentially restrict managers from undertaking worthwhile projects. In such settings, debt diversification becomes an optimal strategy that managers can implement to overcome the constraints levied by credit rationing. This facilitating nature of the debt diversification decision, therefore, suggests a positive association between debt diversification and firm value. This positive association finds further theoretical ground through Harris and Raviv (1990) and Rajan (1992) who argue that debt plays an essential disciplining role by reducing the agency costs of equity. In the presence of moral hazard problems, debt holders typically tend to monitor the activities of the firm[1] and thereby, help alleviate the agency costs thereof. One would then expect this monitoring mechanism to be relatively more intense for a firm with multiple debt holders (i.e., greater debt diversification) relative to a firm with fewer debt holders. Consequently, a firm with greater debt diversification should experience lower agency costs and greater firm value, i.e., a positive association between debt diversification and firm value. This argument, however, finds a counter-hypothesis through findings in several studies (Krugman,1988; Carletti et al., 2007; Brunner and Krahnen, 2008) which argue the presence of multiple debt sources may lead to a drop in the efficiency of monitoring, due to the free rider problem and the lesser incentives for an individual debt holder to monitor the activities of the borrowing entity[2]. Carletti et al. (2007) in fact, propose a model in which the efficiency of monitoring is highest in the case of a single debt holder with substantial lending to the firm. This premise suggests a negative association between debt diversification and firm value. The objective of our paper is to therefore, examine whether debt diversification increases (via the financial constraints and the agency costs hypotheses) or erodes (via the free rider hypothesis) the value of the firm.

We employ Tobin's Q as a measure of firm value. Debt diversification is proxied using two measures. The first proxy is the total number of sources from which a firm has accessed debt (i.e., has an outstanding balance at the financial year-end). Since this measure does not account for the dispersion of debt within these sources, we use the normalized Herfindahl-Hirschman Index (HHI) as our second proxy of debt diversification. Our results using both measures support the free rider hypothesis – firms with diversified debt sources experience a value discount in the capital market. We further check the robustness of our results by splitting our sample firms into small and large groups based on the annual median sales. The negative association is observed for all firms irrespective of their size for both measurements of debt diversification.

The rest of the paper is organized as follows: data and methodological aspects are described in the second section, results are presented and discussed in the third section, and conclusions are presented in the last section.

2. Data and methodology

2.1 Data

Our sample period spans from 1962 to 2015, and the financial data for our analysis has been obtained from COMPUSTAT. We exclude financial, regulated and zero debt firms from our analysis[3]. The summary statistics for the variables used in our study are presented in Table 1. Our final sample consists of 149,938 firm-year observations.

As indicated by our first proxy for debt diversification, Debt Number, the average number of debt sources for our sample firms is about 2.6 with a median of 2. Since our sample comprises non-zero debt firms only, the minimum and the maximum debt

[1] This monitoring activity is aided by their access to private information of the firms, especially for banks. Bond holders on the other hand, typically form trusts to oversee the activities of the firm.

[2] Lender's rent is divided among many debt holders.

[3] We winsorized the data at 5% to limit the spurious effect of outlier and further restricted values of Tobin's Q, Tangibility, R&D ratio and DPR to non-negative values. Values of ROA less than negative one and of lagged leverage greater than one have also been dropped.

number are one and eight respectively. HHI which adjusts for the dispersion aspect of debt diversification is spread between a minimum and maximum of 0 and 0.957 respectively.

Table 1: Summary statistics

Tobin's Q is the ratio of the market value to the book value of the firm's total assets. Debt Number is the total number of debt sources that a firm has used. HHI is the dispersion-adjusted measure of debt diversification. Firm Size is log of firm sales. ROA is the return on total assets. Tangibility is the ratio of net investments in plant and machinery to total assets. Asset growth rate is the change in total assets over lagged total assets. R&D Ratio is the ratio of research and development expenditure to total assets. DPR is the ratio of total dividends to total assets. Leverage is the ratio of total debt to total assets. The values are rounded up to the nearest decimal.

Variable	Observations	Mean	Std. Dev.	Min	Max
Tobin's Q	149,938	1.468	1.281	0.442	6.170
Debt number	149,938	2.564	1.332	1.000	8.000
HHI	149,938	0.365	0.277	0.000	0.957
Firm Size	149,938	4.775	2.287	-0.038	8.689
Asset growth rate	149,938	0.142	0.324	-0.329	1.116
ROA	149,938	0.017	0.209	-0.779	0.231
Tangibility	149,938	0.302	0.207	0.026	0.762
R&D ratio	149,938	0.030	0.057	0.000	0.231
DPR	149,938	0.009	0.014	0.000	0.047
$Leverage_{t-1}$	149,938	0.293	0.209	0.015	0.895

2.2 Methodology

To estimate the marginal impact of debt diversification on firm value, we use Eq. (1) as our baseline model.

$$Y_{it} = \alpha_i + \beta_1\ Debt\ Diversification_{it} + \beta_2\ Firm\ Size_{it} + \beta_3\ ROA_{it} + \beta_4\ Tangibility_{it} + \beta_5\ Asset\ growth\ rate_{it} + \beta_6\ R\&D\ ratio_{it} + \beta_7\ DPR_{it} + + \beta_8\ Leverage_{it-1} + \varepsilon_{it} \tag{1}$$

Where, Y_{it}, the dependent variable, is the Tobin's Q calculated as the ratio of the market value to the book value of a firm's total assets. Of the two measures of debt diversification used, the first is Debt Number, which is the number of debt sources that a firm has accessed. We consider eight mutually exclusive debt sources in our study. They are: capitalized lease obligations (dclo); senior convertible debt (dcvsr); subordinated convertible debt (dcvsub); debt debentures (dd), debt notes (dn); subordinated debt (ds); notes payables (np) and other long-term debt (dlto)[4].

Our second measure of debt diversification is the HH index that accounts for dispersion of debt between the debt sources by assigning a higher weight to those sources with a higher proportion in the overall debt. This measure is computed using the same eight

[4] The variable codes used in the Compustat database are provided in parentheses.

sources of debt thus: we measure the concentration of debt, the Herfindahl-Hirschman scores, by summing the squared ratios of individual debt to total debt.

$$HHI_{it} = \sum_{i=1}^{8} \left(\frac{Debt\ Sources_i}{Sum\ of\ all\ debt\ sources} \right)^2 \tag{2}$$

The value obtained in (2) is then normalized using Eq. (3) to arrive at the normalized HH index.

$$HHI_{it} = \frac{HHI_{it} - (1/8)}{1 - (1/8)} \tag{3}$$

Greater *HHI* values indicate lesser debt diversification. The *HHI* variable, therefore, bears a negative correlation with debt diversification and with *Debt Number*. We subtract *HHI* (obtained in Eq.3) from one to make its interpretation consistent with that of *Debt Number* and use this modified proxy for the rest of the paper.

The control variables used in this study are: *Firm Size* is log of firm sales, *ROA* is return on total assets, *Tangibility* is the ratio of net investments in plant and machinery to total assets, *Asset growth rate* is the change in total assets over lagged total assets, *R&D ratio* is the ratio of research and development expenditure to total assets, *DPR* is the ratio of total dividends to total assets, and *Leverage* is the ratio of total debt to total assets. We use a one-year lagged value of *Leverage* to avert issues from a possible contemporaneous relationship between a firm's leverage and its debt number.

We use fixed effects estimator[5] to estimate the coefficients of our model presented in Eq. (1). This technique helps to control for unobserved time-invariant variables that might impact firm value proxied by Tobin's Q. The year-effects on firm value are controlled by using year dummies. We also use firm fixed effects to account for unobserved firm-level factors.

3. Results and discussion

The impact of debt diversification on firm value is examined by regressing *Tobin's Q* on our measure of debt diversification. The financial constraints and the agency costs hypotheses predict a positive association while the free rider hypothesis predicts a negative association. The results of the analysis using *Debt Number* are presented in Table 2.

The coefficient for *Debt Number* for the full sample analysis, presented in Model I, is negative and statistically significant at 1% significance level in support of the free rider hypothesis. This result is consistent with Carletti et al. (2007) who examined the impact of multiple banking relationships[6] on the value of Danish firms. To check the robustness of this negative association, we classify our sample firms into small and large firms based on the yearly median value of firm size (captured by the log of firm sales). Firms with below (above)-median firm size are classified as small (large) firms. We re-estimate Eq. (1) for these sub-samples separately. As presented in Model II and Model III, the coefficients of *Debt Number* for both small and large firms respectively are negative and significant. The

[5] Our data rejected the null of Hausman test at 1% confidence level.
[6] Their study is concerned with multiple banking relationships whereas our study is concerned with multiple sources of debt.

results, shown in Table 3, based on the normalized HHI as the measure of debt diversification, support the findings in Table 2, based on *Debt Number*. Overall, the analyses offer substantial evidence that firms which use diversified debt sources experience a value discount in capital markets.

Table 2: Regression analysis using Debt Number

Dependent Variable: Tobin's Q is the ratio of the market value to the book value of the firm's total assets. The main independent variable is Debt Number, which is the total number of debt sources that a firm has used. The control variables are defined as: Firm Size is log of firm sales, ROA is the return on total assets, Tangibility is the ratio of net investments in plant and machinery to total assets, Asset growth rate is the change in total assets over lagged total assets, R&D Ratio is the ratio of research and development expenditure to total assets, DPR is the ratio of total dividends to total assets, and Leverage is the ratio of total debt to total assets. The coefficients are estimated using fixed effects estimator, and heteroscedasticity-adjusted standard errors, clustered at the firm level, are presented in parentheses. ***, ** and * denotes significance at 1%, 5% and 10% respectively.

	Full Sample	Small Firms	Large Firms
VARIABLES	**Model I**	**Model II**	**Model III**
Debt number	-0.042***	-0.053***	-0.018***
	(0.004)	(0.006)	(0.004)
Firm Size	-0.183***	-0.233***	-0.135***
	(0.009)	(0.013)	(0.011)
ROA	0.066	-0.511***	3.140***
	(0.051)	(0.053)	(0.104)
Tangibility	-0.131**	-0.239***	0.110*
	(0.051)	(0.067)	(0.064)
Asset growth rate	0.586***	0.602***	0.350***
	(0.013)	(0.017)	(0.014)
R&D ratio	3.467***	2.991***	2.601***
	(0.231)	(0.250)	(0.485)
DPR	7.704***	3.439***	7.153***
	(0.528)	(0.741)	(0.625)
$Leverage_{t-1}$	0.494***	0.668***	0.141***
	(0.031)	(0.040)	(0.040)
Constant	1.929***	2.075***	1.493***
	(0.049)	(0.098)	(0.074)
Observations	149,938	74,959	74,959
R-squared	0.136	0.150	0.245
Number of firms	15,780	12,593	6,524
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

Table 3: Regression analysis using HHI

Dependent Variable: Tobin's Q is the ratio of the market value to the book value of the firm's total assets. The main independent variable is HHI which is the dispersion-adjusted measure of debt diversification. The control variables are: Firm Size is log of firm sales, ROA is the return on total assets, Tangibility is the ratio of net investments in plant and machinery to total assets, Asset growth rate is the change in total assets over lagged total assets, R&D Ratio is the ratio of research and development expenditure to total assets, DPR is the ratio of total dividends to total assets, and Leverage is the ratio of total debt to total assets. The coefficients are estimated using fixed effects estimator, and heteroscedasticity-adjusted standard errors, clustered at the firm level, are presented in parentheses. ***, ** and * denotes significance at 1%, 5% and 10% respectively.

	Full Sample	Small Firms	Large Firms
VARIABLES	**Model I**	**Model II**	**Model III**
HHI	-0.179***	-0.182***	-0.133***
	(0.016)	(0.024)	(0.018)
Firm Size	-0.184***	-0.236***	-0.132***
	(0.009)	(0.013)	(0.011)
ROA	0.076	-0.499***	3.141***
	(0.051)	(0.053)	(0.104)
Tangibility	-0.132**	-0.248***	0.116*
	(0.051)	(0.067)	(0.064)
Asset growth rate	0.582***	0.596***	0.351***
	(0.013)	(0.017)	(0.014)
R&D ratio	3.476***	2.996***	2.614***
	(0.231)	(0.250)	(0.483)
DPR	7.727***	3.495***	7.107***
	(0.528)	(0.741)	(0.625)
$Leverage_{t-1}$	0.492***	0.664***	0.148***
	(0.031)	(0.040)	(0.040)
Constant	1.893***	2.028***	1.452***
	(0.050)	(0.098)	(0.075)
Observations	149,938	74,959	74,959
R-squared	0.136	0.150	0.246
Number of firms	15,780	12,593	6,524
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

5. Conclusion

This study examines the impact of debt diversification on firm value. Our results suggest there is a negative impact of debt diversification on firm value. Prior studies maintain that the presence of debt in the capital structure tends to decrease agency costs; however, our results reveal that having debt from multiple sources tends to reduce that advantage. Thus, managers using diversified debt sources in an attempt to overcome financial constraints engendered by credit rationing might potentially erode shareholder wealth.

References

Brunner, A., Krahnen, J. P., 2008. Multiple lenders and corporate distress: Evidence on debt restructuring. The Review of Economic Studies 75(2), 415-442.

Carletti, E., Cerasi, V., Daltung, S., 2007. Multiple-bank lending: Diversification and free-riding in monitoring. Journal of Financial Intermediation 16(3), 425-451.

Colla, P., Ippolito, F., Li, K., 2013 Debt specialization. The Journal of Finance 68(5), 2117-2141.

Jaffee, D. M., Russell, T., 1976. Imperfect information, uncertainty, and credit rationing. The Quarterly Journal of Economics 90(4), 651-666.

Johnson, S. A., 1997. An empirical analysis of the determinants of corporate debt ownership structure. Journal of Financial and Quantitative Analysis 32(1), 47-69.

Harris, M., Raviv, A., 1990. Capital structure and the informational role of debt. The Journal of Finance 45(2), 321-349.

Krugman, P., 1998. Financing vs. forgiving a debt overhang. Journal of Development Economics 29(3), 253-268.

Rajan, R. G., 1992. Insiders and outsiders: The choice between informed and arm's-length debt. The Journal of Finance 47(4), 1367-1400.

Rauh, J. D., Sufi, A., 2010. Capital structure and debt structure. The Review of Financial Studies 23 (12), 4242-4280.

Stiglitz, J. E., Weiss, A., 1981. Credit rationing in markets with imperfect information. The American Economic Review, 71(3), 393-410.

APPLIED FINANCE LETTERS

VOLUME 8, 2019

THE SHIFT IN FIRMS' RELIANCE ON DEBT SOURCES

TU HO*

Auckland University of Technology, Auckland, New Zealand

* Corresponding Author: Tu Ho, Lecturer at Auckland University of Technology, New Zealand.
☎+ 09 921 9999 ext. 5826 ✉ ctho@aut.ac.nz

Abstract: Structural changes in capital market and information innovations have altered characteristics of debt sources, make them favourable to firms. This could possibly lead to a shift in firms' reliance on debt sources. Using a unique data set of debt mix of 1,100 U.S. non-financial firms, I conduct data analysis to reveal changes in firms' preference for different debt sources over a decade from 2004 to 2014. I find that bank debt remains the most common source of borrowing, followed by public debt and finally private placement debt. In addition, over time, firms have become more reliant on bank and public debt while less reliant on private placement debt. This pattern is consistent across different industries.

Keywords: Capital Markets, Financial Markets, bank debt, debt sources

1. Introduction

Firms can generally borrow from three main sources: debt issuance on financial markets, banks, and private lenders. These sources are distinct in various aspects that make them more or less desirable to firms depending on their needs and characteristics. Over the years, structural changes in capital markets and technology development have altered the distinctive characteristics of these debt sources (Boot and Thakor, 2000; Dinc, 2000; Gande and Sauders, 2012; Petersen and Rajan, 2002; Tracey and Carey, 2000). This raises an interesting question on how firms' preferences for debt sources have changed over time.

Observing changes in firms' reliance on different debt sources can partly reveal the answer to this question. This article uses a unique dataset of debt sources available to the U.S firms to carry out some data analysis on changes in the popularity of bank, public and private placement debt and in debt ownership structure over a ten-year period from 2004-2014. In general, I find that firms consistently rely the most on bank debt to finance their operations, followed by public debt and finally private placement debt. Over time, firms' reliance on bank and public debt tend to increase while the opposite pattern is observed for private placement debt.

2. Data

I hand collect debt source data for a sample of 1,100 randomly chosen U.S. non-financial firms in the three different years: 2004, 2009 and 2014. The random sample is based on the Compustat firm list from 2004. The dataset covers information on the three main borrowing sources that have been discussed in literature, namely bank debt, public debt, and private placement debt. The final dataset consists of 2,707 firm year observations for 1,100 US non-financial firms. The number of firms gradually decreases from 1,100 in 2004, to 894 in 2009 and finally to 713 in 2014.[1]

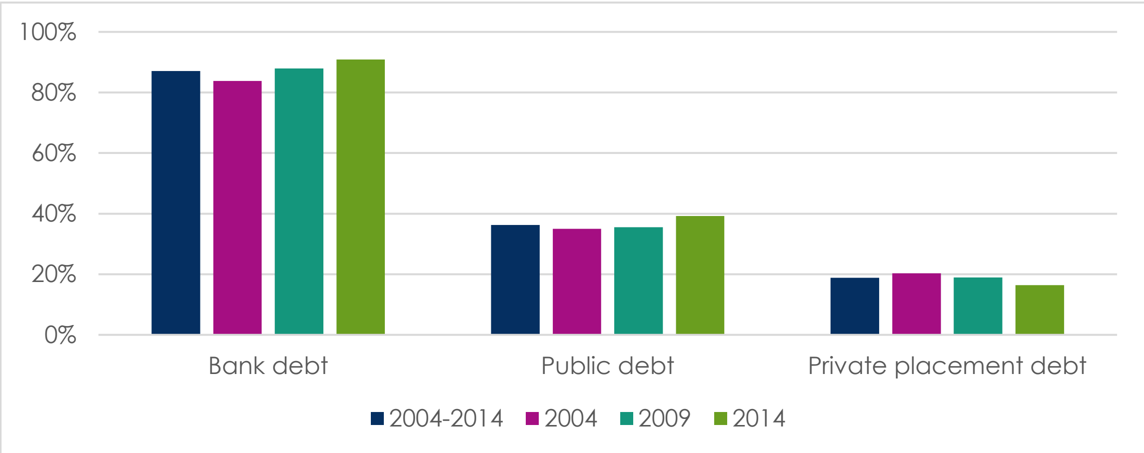
3. Analysis and findings

In this section, I conduct an analysis to reveal changes in firms' reliance on different debt sources over time. First, I provide an overall picture of how firms choose and rely on different types of lenders. To do so, I calculate the percentage of firms that has outstanding balance from a certain source to proxy its popularity and the proportion of that source to proxy how much firms rely on it. The second part focuses on three main debt sources only and their characteristics.

The popularity of bank, public and private placement debt sources

Graph 1 shows the percentage of firms that have outstanding balance of each debt source. Consistently, I find that bank debt remains the most popular, followed by public debt issuance and finally private placement. Among three main sources, bank debt is the most popular source of debt financing with more than 80% of the sample firms using or having financing agreements with banks. Public debt is the second with around 36% of firms having outstanding public bonds while private placement debt is the least popular with less than 20% of firms having outstanding balance from this source[2].

Figure 1: Percentage of observations that have outstanding debt of a given source



[1] To check the representativeness of my sample, I compare the mean and medians of some firm characteristics between my sample and the whole market, including all US nonfinancial firms. I find that firm size, firm age and market-to-book ratios are relatively similar in both the entire sample and yearly subsamples. I further carry out the difference in mean tests also confirm that my sample can be considered as representative of the market and any patterns found in my data analysis are likely applicable to out-of-sample nonfinancial firms in the US market.

[2] Since firms can simultaneously borrow from different debt sources, the sum of percentages of firms with outstanding balance from these sources can be greater than 100%.

Though the order of relative importance remains consistent over time, there is an upward trend in the number of firms using bank debt and issuing public debt on markets, in contrast to a decrease in that of private placement. Over the sample period, the popularity of bank and public debt has grown by 8.5% and 12.3%, respectively, and that of private placement debt significantly dropped by 19%.

The proportions of bank, public and private placement debt sources

Next, I analyse changes in each debt source by observing proportions of these sources used by firms over the sample period. Table 1 shows average and median proportions of different sources in total of outstanding debt for the full sample and the three subsamples. It can be observed that firms consistently rely the most heavily on bank debt (40.53%) and public debt (25.64%) to finance their business and the least on private placement debt (10.92%)[3]. Among all debt sources, only bank debt has median proportion greater than zero, confirming the fact that bank debt is the only source that is used by more than 50% of firms in the sample as was also shown in Figure 1.

Table 1: Debt ownership structure

This table presents proportions of debt sources for the full sample and the three subsamples. BankPercent, PubPercent, PriPercent, ProPercent, LeasePercent, FinPercent, PartyPercent and OtherPercent are proportions of bank debt, public debt and private placement debt, program debt, capital lease, financial company debt, third party debt and other unclassified debt respectively in total outstanding debt. Measurement unit of all variables is in %.

	2004-2014		2004		2009		2014	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
BankPercent	40.53	23.77	36.92	16.87	41.32	25.49	45.13	34.12
PubPercent	25.64	0	24.84	0	24.73	0	28.01	0
PriPercent	10.92	0	12.76	0	10.35	0	8.77	0

Moreover, the reliance of firms on both bank and public debt increases over time. Firms have 8.21% more bank debt and 3.17% more public debt in their debt ownership structure, equivalent to a growth of 22% and 12.8% in firms' reliance on these sources, respectively. Private placement debt, on the other hand, experienced a drop of 3.99% in proportion, which is converted to 31.3% decrease in firms' use of this source.

Since different industries with a distinctive business nature may prefer long- or short-term funding, they may have different preferences for bank, public and private placement debt. To make sure the observed pattern is not only present in certain industries, I split the sample into ten different industry groups and find consistent results in each of these groups. Moreover, I filter the samples into some subsamples with non-zero outstanding balance of each debt source to address the concern that averaging all numbers might not reflect the true picture. Consistent with the pattern found in the above section, I

[3] My dataset covers information of nine borrowing sources, namely bank debt, public debt, private placement debt, programme debt, government debt, capital lease, financial company loans, third- or related-party borrowing, and finally other debt. This paper only analyses the three main sources widely discussed in the literature, namely bank debt, private placement and public debt. Therefore, the proportions of these three sources do not add up to 100% in Table 1.

observe an upward trend in bank and public debt financing but a downward trend in private placement debt in each of these subsamples.

In general, banks remain the most important source of borrowing for the US firms, with the second-place public debt while private placement debt is the least important one in terms of both number of borrowers and the proportion in total debt. Moreover, over the ten-year period, firms tend to rely more on bank and public debt, and less on private placement debt. Since life insurance companies are dominant lenders in the private debt market (Pottier, 2007), a decrease in firms' preference to raise funds in this market might put these insurers in greater lending competition, which consequently can deteriorate credit quality and decrease bond yields.

4. Conclusions

This article uses a unique hand-collected dataset of debt sources to conduct some analysis on how firms' reliance on different sources of debt financing has changed among the US firms over time. The main finding is that among three main debt sources, bank debt remains the most important one, followed by public debt and lastly private placement debt. The difference between two ends of the scale of debt financing is getting wider in that firms are relying more and more on banks and less on private placement debt over the sample period. This finding is consistent across different industries and different subsamples of firms. I can see that there is a systematic shift in firms' choice of debt sources, and the interesting question is what factors are driving the trends. This systematic shift can hardly be explained by changes in macroeconomic factors since it is consistent through the pre- and post-global financial crisis periods. One possible explanation is that the structural changes in the debt markets and technology development have altered the distinction between debt sources, and thus made one more favourable than others as a borrowing source. It is well documented that firms with different level of information problem tend to seek fund from different debt markets. Those with the highest information asymmetry tend to rely on bank loans, while those with moderate informational problem rely more on private placement debt market and those with the lowest degree of information problem rely more on public debt. Recent innovations in information technology have allowed potential lenders to easily acquire information, which includes but is not limited to hard information about the credit quality of borrowers and access the data most of which was not available to public investors before (Petersen and Rajan, 2002; Tracey and Carey, 2000; DeYoung et al., 2011). This might have widened the entrance into public debt market and allowed more firms seeking long-term debt to access this source where they have more options at lower costs. In addition, the recent developments in the secondary loan market might contribute to the shift since it helps banks to share and reduce their credit risk and thus allowing firms to acquire bank debt more easily. Given these possible explanations, this shift is predicted to continue into near future. Further research on the link between capital market changes and the change in firms' debt mix is necessary to confirm the drivers of the shift. Understanding this link can be important to policy makers in regulating and implementing policies to ensure the sustainable and balanced development of the markets.

References

Boot, A. W., & Thakor, A. V. (2000). Can relationship banking survive competition? *Journal of Finance,* 55(2), 679-713.

Boyd, J., & Prescott, E. (1986). Financial intermediary-coalitions. *Journal of Financial Theory*, 38, 211-232.

Carey, M., Prowse, S., Rea, J., & Udell, G. (1993). The economics of the private placement market. *Washington: Federal Reserve System*.

Denis, D. J., & Mihov, V. T. (2003). The choice among bank debt, non-bank private debt, and public debt: evidence from new corporate borrowings. *Journal of Financial Economics*, 70, 3-28.

DeYoung, R., Frame, W. S., Glennon, D., & Nigro, P. (2011). The Information Revolution and Small Business Lending: The Missing Evidence. Journal of Financial Services Research, 39(1–2), 19–33.

Diamond, D. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51, 394-414.

Dinç, I. S. (2000). Bank reputation, bank commitment, and the effects of competition in credit markets. *The Review of Financial Studies*, (3). 781-812.

Fama, E (1985). What's different about banks? *Journal of Monetary Economics*, 15, 29–37.

Gande, A., & Sauders, A. (2012). Are banks still special when there is a secondary market for loans? *Journal of Finance*, 67(5), 1649-1684.

Leland, H., & Pyle, D. (1977). Informational asymmetries, financial structure and financial intermediation. *Journal of Finance*, 32, 371-387.

Petersen, M., & Rajan, R. (2002). Does distance still matter: the information revolution in small business lending. *Journal of Finance*, 57, 2533–2570.

Pottier, S. W. (2007). The Determinants of Private Debt Holdings: Evidence from the Life Insurance Industry. Journal of Risk & Insurance, 74(3), 591–612.

Rajan, R., & Winton, A. (1995). Covenants and collateral as incentives to monitor. *Journal of Finance*, 50, 1113-1146.

Tracy, W., & Carey, M. (2000). Credit risk rating systems at large U.S. banks. *Journal of Banking and Finance*, 24, 167–201.

Welch, I. (1997). Why is bank debt senior? A theory of asymmetry and claim priority based on influence costs. *Review of Financial Studies*, 10, 1203-1236.

APPLIED FINANCE LETTERS

VOLUME 08, 2019

PROFITABILITY, PRODUCT MARKET COMPETITION, AND STOCK RETURNS

SCOTT LI

California State University, Los Angeles

* Corresponding Author: Scott Li, PhD, MPA, Assistant Professor of Finance, College of Business and Economics, California State University, Los Angeles ✉ sli86@calstatela.edu

Abstract: This paper finds that product market competition level (measured by Herfindahl Hirschman Index using Fama French 48 industries) affects the performance of zero-cost investment strategies based on gross probability. From 1973 to 2017, the positive returns from such strategy mainly comes from the most competitive industry quintile while a strong reversal exists the second most competitive quintile. The same strategy does not generate any statistically significant returns in concentrated industry quintiles. Out of 25 dependently sorted portfolios on product market competition level and gross profitability, the top performing portfolio comes from the least profitable firms in the second most competitive industry quintile, where 65% of firms are from pharmaceutical and oil industries.

Keywords: Fama French, investment strategies, markets, stocks

1. Introduction

Sir John Templeton (1912-2008) is regarded by Money Magazine as "arguably the greatest global stock picker of the century" in 1999. During a one-on-one interview[1] with Tony Robbins on investing, he mentioned that one of the criteria he used to pick stocks was to buy firms with higher profitability compared to its direct competitors. There are two factors to this measure: profitability and product market competition. In a very competitive industry where many firms are competing for the same market, higher profitability demonstrates superior productivity and efficiency. However, if a firm resides in a highly concentrated industry where it possesses strong market power and majority market share, profitability becomes less comparable especially for industries that are monopolies. It also does not make much sense to compare two firms' profitability when they are from totally unrelated industries that have drastically different operation characteristics and cost structures. In other words, profitability, as a performance measure, is better used in a more comparable environment. Inspired by Sir John Templeton's stock picking criteria, the purpose of this paper is to test a zero-investment strategy that long stocks with the highest scaled profitability and short the ones with the lowest after controlling product market competition levels. This way firms are compared

[1] The video clip of the interview is part of Tony Robbins' Ultimate Edge Program

with their direct competitors in the same industry, and other industries that have similar product market competition levels.

In academia, on the one hand, there are many studies documenting the relationship between profitability and stock returns. Haugen and Baker (1996) and Cohen, Gompers, and Vuolteenaho (2002) find that firms with higher probabilities are associated with higher average stock returns. Fama and French (2008) find that, however, profitability produces a mixed picture, where higher positive profitability seems to be associated with higher abnormal returns, but there is no evidence that negative profitability leads to low abnormal return from 1963 to 2005. In their Table II (p.1660), it shows that only the small (versus micro and big) stock group generates positive high-minus-low (HML hereafter) value weighted returns with statistical significance (0.79% per month with a t-statistic of 2.87). Novy-Marx (2013) claims that profitability, measured by gross profits-to-asset ratio has roughly the same predicting power as book-to-market ratio (BM hereafter) in cross section of stock returns. His results in Table 6 (p.10) show that across all BM quintiles, the profitability HML raw returns are all positive and statistically significant, although the abnormal returns from Fama French three-factor regression are significant only in the 1st, 2nd, and 4th BM quintiles. Fama and French (2015) construct a five-factor model, adding profitability and investment factor to the existing three. Such model improved the explanatory power of their three-factor model (Fama and French, 1993), although they use operating profitability[2] instead of gross profitability to construct the new factor. At the same time, Ball at el. (2015) construct an alternative operating profitability measure[3] and claim that their measure displays a stronger link with stock returns than gross profit. Subsequently, Ball at el. (2016) use cash-based profitability measures and show that it subsumes predicting power of cross section of average returns by their previously used accruals based operating profitability measure. While scholars examine and argue the superiority of alternative profitability measure over others, it is reasonable to say that gross profitability is the cleanest measure, and it can be used to compare firms across industries because different industries have different cost structures by nature. I use scaled gross profit by book value of total assets to measure firm's profitability to form portfolios.

On the other hand, there are a few studies examining the relationship between product market competition level and stock returns with mixed results. For example, Hou and Robinson (2006) find that industry concentration level (measured by Herfindahl Hirschman Index using three-digit SIC code) is negatively related with stock returns from 1963 to 2001. Their study supports the creative destruction theory by Schumpeter (1912), which states that firms in competitive industries have more incentives to engage in innovation activities and are thus more likely to have higher future stock returns. However, their findings are challenged by Grullon et al. (2019), who show that during the last two decades, about 75% of industries (using three-digit NAICS[4] code) have become more and more concentrated. At the same time, firms in those concentrated industries display higher profits and stock returns. The difference is likely caused by using different sample periods because market structure evolves over time. It is also important

2 Operating profitability equals revenue minus cost of goods sold, interest expense, and SG&A expenses, and then divided by book equity (p.4). All accounting information is based fiscal year ending in t-1.

3 Their deflated operating profitability equals gross profit minus SG&A expenses (excluding research and development expenditure) and then divided the book value of total assets (p.240).

4 North America Industry Classification System (NAICS) was adopted in 1997 to replace Standard Industry Classification (SIC) code. Both NAICS and SIC codes are available in Compustat, while CRSP contains SIC code only. The SIC codes have substantial discrepancy between CRSP and Compustat database, documented by Kahle and Walking (1996).

to point out that these two studies use completely different industry classification systems.

This paper brings these two dimensions together and find that a value-weighted zero-investment strategy that longs stocks with high profitability and shorts the ones with low profitability works well only in the most competitive industries. More importantly, there seems to be a strong reversal in the second most competitive industry quintile, which is prominent during the past two decades. This phenomenon is likely caused by the evolution of market structure and the unique characteristics of industries in that group.

The rest of this paper is structed as follows. Section 2 describes the data and provides the summary of statistics. Section 3 examine the performance of the zero-investment strategy based on product market competition level and scaled gross profitability. Section 4 concludes.

2. Data and Statistics Summary

I use publicly held firms listed on the NYSE, AMEX, or NASDAQ from 1973 to 2017 and restrict the sample to firms issuing common shares. The monthly stock return file is from CRSP, and annual firm fundamentals come from Compustat. A firm must have non missing gross profit (GP hereafter), positive market equity (ME), total assets, book equity, and at least twelve consecutive monthly return observations prior to July of year t to be included in the sample of year t. After merging CRSP and Compustat datasets, firms without annual fundamentals are deleted from the sample. Following the common practice in the literature, financial industries (one-digit SIC as 6) are excluded from the sample. This produces 1,567,199 firm observations, with 2902 firms/year on average. I choose 1973 as the start year because it is the year that the inclusion of NASDAQ stocks takes place, although the results are consistent when extended to 1963. The number of firms in the sample continues to rise until it reaches its peak at 3926 in 1997. After that, the number starts to decline sharply. Likewise, the average industry concentration level also declines from 1973 to 1997, and then rises steadily after that. This is consistent with Grullon et al.'s (2019) study for the past two decades.

The scaled GP equals gross profit divided by book value of total assets (GP/AT). Because the unique characteristic of operation and cost structure, industries have different average GP/AT ratio. Among the 44 non-financial Fama French[5] industries, Retail has the highest average GP/AT ratio: 0.69, followed by Soda (0.59), Household (0.57), Clothing (0.54), Smoke (0.50), and Books (0.50). This is a huge contrast to Oil, Mines, Gold, Utility, and Pharmaceutical industry that all have GP/AT below 0.20 as Table 1 shows. It does not make much sense to compare firms' gross profit margins across industries that have very different nature.

I follow Hou and Robinson (2006) and measure the product market competition level (industry concentration level) using the Herfindahl–Hirschman Index (HHI hereafter) based on net sales:[6]

$$HHI_{j,t} = \sum_{i=1}^{I} product\ market\ share^2{}_{i,j,t} \tag{1}$$

[5] Fama French 48 industry classification is based on four-digit SIC. It can be downloaded from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_48_ind_port.html

[6] In the HHI equation, *i* stands for individual firm, *j* is the industry firm *i* belongs to, *t* stands for year *t*.

In each fiscal year, as equation (1) shows, the HHI is generated by calculating the sum of all sales in every industry grouped by Fama French 48 (FF 48 hereafter) industries. I use FF 48 instead of three-digit SIC industries because the latter produces highly concentrated industries that are consisted of mostly small single-firm industries, which take up the top 20% of all industries. Two-digit SIC industries mitigate such issue; but it still produces a heavily right-skewed distribution where only 3.86% and 8.32% of firm observations exist in the most and second-most concentrated industry quintiles. Meanwhile, over half of the sample (54.25%) are clustered in the most competitive industry quintile. This may not present an accurate description of market structure. In comparison, FF 48 classification produces a much more normalized distribution of firms across all industry quintiles. The market share of each firm in the industry is calculated by division of the firm's sales and industry total sales. I then square the market share of each firm and add all squared shares to compute the HHI of that industry, a value ranging from 0 to 1. If HHI equals 1, the industry is a monopoly. The bigger HHI is, the more concentrated the industry is.

Table 1: Average Gross Profitability by Industry 1973-2017

Note: The average gross profit over total assets ratio (GP/AT) is calculated for each of Fama French 48[7] industries (financial industries excluded) over 1973-2017. The sample includes public held firms with common shares traded on NYSE, AMEX and NASDAQ. A firm must have positive book equity, total assets, non-missing gross profit from Compustat to be included in the sample. For details of Fama French 48 industries please refer to the footnote.

FF IND	# of Firms	GP/AT	FF IND	# of Firms	GP/AT
RTAIL	177	0.69	HLTH	53	0.35
SODA	7	0.59	AUTOS	54	0.34
HSHLD	66	0.57	TXTLS	28	0.33
CLTHS	55	0.54	MEALS	61	0.32
SMOKE	3	0.50	FABPR	16	0.32
BOOKS	28	0.50	OTHER	58	0.31
TOYS	30	0.49	GUNS	7	0.30
FOOD	66	0.48	FUN	46	0.29
COMPS	120	0.47	BOXES	12	0.28
LABEQ	80	0.47	AERO	22	0.28
WHLSL	132	0.45	TELCM	69	0.27
BUSSV	312	0.45	STEEL	51	0.25
MEDEQ	102	0.43	SHIPS	8	0.25
PAPER	53	0.42	TRANS	75	0.25
PERSV	36	0.41	AGRIC	10	0.24
RUBBR	38	0.40	CNSTR	45	0.22
BEER	13	0.39	COAL	6	0.21
CHIPS	202	0.39	OIL	141	0.19
MACH	136	0.39	MINES	14	0.16
ELCEQ	56	0.38	GOLD	11	0.11
CHEM	67	0.38	UTIL	92	0.11
BLDMT	87	0.37	DRUGS	158	0.11

[7] Fama French 48 industry SIC codes are downloaded from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Following Fama and French (1992), at the end of June of year *t*, I form quintile portfolios based on firm's GP/AT ratio from fiscal year ending in year *t-1*, and then hold each quintile portfolio from July of year *t* to June of year *t+1* before rebalancing. The summary of average firm characteristics in each profitability quintile is presented in Table 2. It seems that, as profitability increases, the average monthly return increases monotonically from 1.08% to 1.67%. This supports the findings of previous studies. Accumulative momentum (average monthly returns from month *t-2* to month *t-12*) follows a similar pattern. The smallest average firm size belongs to the most profitable quintile (strong) while the biggest average firm size resides in the second weakest profitability quintile. Book to market ratio follows a similar pattern. Scaled R&D expenses display a strong U-shape pattern, where the firms with weakest GP/AT ratio having the highest R&D/AT ratio. However, R&D expenses is not recorded as part of Cost of Goods Sold (COGS) based on General Accepted Accounting Principles (GAAP), so it should not affect gross profitability. The inverted-U shape pattern of Sales seems to indicate that is no clear correlation between quantity of sales and gross profitability. At last, industry concentration level (HHI) seems to rise as the gross profitability rises, although not monotonically.

Table 2. Firm Level Summary Statistics by Gross Profitability 1973-2017

Note: This table presents the average firm level statistics from each profitability quintile portfolio. The portfolios are formed at the end of June at calendar year t. Firms are sorted by gross profit to total assets ratio (GP/AT) at the end of June into quintiles. The GP/AT ratio is from the report of last fiscal year ending in year *t-1*. Number of firms is the overall average number of firms per year in each GP/AT quintile from 1973 to 2017. Return (%) is the average monthly raw return calculated using the variable RET from CRSP dataset. Momentum (%) is the average monthly raw return of accumulative returns from month *t-2* to month *t-12* at the end of June each year. Log (ME) is the natural log of market equity (ME) calculated as the product of absolute value of price (PRC) and shares outstanding (SHROUT) from CRSP dataset. A firm must have positive ME to be included in the sample. BM is book to market ratio. Book equity is calculated from Compustat, market equity used in calculating BM is from December of year t-1. R&D/AT is scaled research and development expense by total assets. Log (sale) is the natural log of sales. HHI is Herfindahl Hirschman Index. For details on how to calculate HHI, please refer to the description in Section 2.

GP/AT Rank	# of firms	GP/AT	Return	Momentum	log(ME)	BM	R&D/AT	log(Sale)	HHI
Weak	580	0.024	1.082	0.910	4.760	0.593	0.078	3.167	0.072
2	580	0.232	1.186	1.149	5.100	0.612	0.018	3.411	0.080
3	580	0.349	1.348	1.388	4.917	0.556	0.029	3.252	0.084
4	580	0.482	1.459	1.637	4.872	0.504	0.042	3.171	0.085
Strong	580	0.789	1.666	1.931	4.682	0.454	0.052	3.185	0.081

3. Gross Profitability, Product Market Competition and Stock Returns

Previous studies have demonstrated that gross profitability is positively related to stock returns. A zero-investment strategy that longs stocks with the highest GP and shorts the ones with the lowest can generate positive and statistically significant returns. However, is this pattern consistent across all product market competition levels? To have higher GP than direct competitors demonstrate a firm's superior productivity and efficiency. However, if product market is highly concentrated, especially with one or only a few firms in the game, will profitability still be a relevant indicator to predict future stock returns? Fama and French (2008) find a mixed picture showing that there is lack of evidence that firms with negative profitability leads to lower abnormal returns. I hypothesize that the zero-investment strategy is prominent only in very competitive industries, where higher profitability in comparison with direct competitors can be used as a good signal for investors to pick stocks.

To analyse how product market competition level may affect the performance of the zero-investment strategy based on GP, at the end of June of year *t*, I sort all stocks into quintiles based on the industry concentration level (ICL hereafter) of the industry they belong to. With 44 non-financial industries in the sample, it gives us about 9 industries in each quintile. Within each quintile, I further sort the stocks into quintiles based on their GP/AT ratio from the fiscal year ending in year *t-1*. This produces 25 ICL & GP/AT portfolios. The zero-investment strategy is to long the stocks in the strongest GP/AT quintile and short the ones in the weakest GP/AT quintile, and then hold the Strong Minus Weak (SMW)[8] portfolios for 12 months before rebalancing.

First, without creating sections of stocks based on ICL, using the stock market as whole, the equal weighted SMW strategy generates 0.62% per month with a t-statistic of 4.92 from 1973-2017. This translates into a return of 7.44% per year. However, the value weighted SMW strategy generates much lower returns (0.24% per month) with no statistical significance during the same period. The comparison indicates that the higher returns of equal weighted SMW strategy may be driven by small stocks in the sample because small stocks outperform big stocks consistently over time. This is also known as the size effect identified by Fama and French (1992).

Table 3. Average Stock Returns of Portfolios by Gross Profitability, and by Gross Profitability & Industry Concentration Level 1973-2017

Note: In column "All", at the end of June of year *t* from 1973 to 2017, all firms are sorted by scaled gross profitability (gross profit/total assets) in quintiles. The zero-investment SMW strategy is to long stocks in the highest (Strong) GP/AT quintile and short the ones in the lowest (Weak) GP/AT quintile, and then held for 12 months from July of year *t* to June of year *t+1* before rebalancing. In column from "Low" to "High" as Industry Concentration Level (ICL), at the end of June of year *t*, all firms are sorted (by industry) first by the ICL into quintile. Then, within each ICL quintile, firms are sorted again based on GP/AT into quintiles. This creates 25 ICL & GP/AT dependently sorted portfolios. Within each ICL quintile, the zero-investment SMW strategy is to long stocks in the highest (Strong) GP/AT quintile and short the ones in the lowest (Weak) GP/AT quintile, and then held for 12 months from July of year *t* to June of year *t+1* before rebalancing. The average monthly stock raw returns (%), both equal weighted and value weighted, are presented in this table. T-statistics calculated with Newey-West adjusted standard errors are presented in parentheses.

Equal Weighted Portfolio Returns							Value Weighted Portfolio Returns						
		Industry Concentration level (FF48)							Industry Concentration level (FF48)				
GP	All	Low	2	3	4	High	GP	All	Low	2	3	4	High
Weak	1.14	0.96	1.38	1.10	0.89	1.07	Weak	1.47	1.22	2.44	1.86	1.43	1.57
	(4.01)	(3.58)	(3.13)	(3.12)	(2.29)	(2.88)		(6.98)	(6.24)	(6.67)	(5.90)	(5.46)	(5.46)
2	1.25	1.22	1.09	1.40	1.12	1.13	2	1.49	1.48	1.53	1.84	1.68	1.37
	(5.01)	(4.10)	(3.51)	(4.60)	(3.72)	(3.72)		(7.73)	(7.07)	(6.97)	(7.36)	(6.99)	(5.55)
3	1.43	1.38	1.21	1.45	1.37	1.39	3	1.69	1.67	1.45	1.68	1.53	1.62
	(5.73)	(4.42)	(4.15)	(4.68)	(4.30)	(4.36)		(7.91)	(6.69)	(7.17)	(6.23)	(6.96)	(6.87)
4	1.54	1.42	1.48	1.45	1.36	1.58	4	1.57	1.67	1.68	1.52	1.49	1.63
	(6.09)	(4.49)	(5.31)	(4.59)	(4.43)	(5.26)		(7.69)	(7.13)	(8.33)	(6.03)	(6.43)	(6.68)
Strong	1.76	1.55	1.64	1.79	1.76	1.78	Strong	1.71	1.95	1.77	1.81	1.80	1.37
	(6.81)	(4.52)	(5.40)	(5.44)	(5.35)	(5.96)		(8.28)	(8.12)	(8.03)	(6.07)	(7.97)	(5.87)
SMW	0.62	0.59	0.26	0.69	0.87	0.70	SMW	0.24	0.73	-0.67	-0.05	0.37	-0.21
	(4.92)	(3.54)	(1.01)	(3.81)	(5.19)	(3.02)		(1.56)	(4.03)	(-2.25)	(-0.21)	(1.61)	(-0.77)

Second, after controlling for ICL, on the left-hand side of Panel A in Table 3, the same equal-weighted strategy generates positive and significant returns four out of five ICL quintiles. In highly concentrated industry quintiles such as the 3rd, 4th, and the 5th, SMW

[8] SMW is named to differentiate from Fama and French's (2015) RMW factor, which is created based on operating profitability.

strategy generates higher returns than the most competitive quintile. The highest average monthly return belongs to the 4th ICL quintile: 0.87% per month with a t-statistic of 5.19. This equals 10.44% per year from 1973 to 2017. However, when using value-weighted strategy, the only quintile that still generates positive returns with statistical significance is the most competitive quintile: 0.73% per month with a t-statistic of 4.03. This supports my hypothesis. More importantly, the same SMW strategy in the 2nd ICL quintile produces a reversal return of -0.67% per month with a t-statistic of -2.25. Overall, out of the 25 ICL & GP/AT dependently sorted portfolios, the one with weakest GP/AT in the 2nd ICL quintile produces the highest monthly returns as 2.44% with a t-statistic of 6.67%. It is 0.81% higher than the average monthly returns of the 25 portfolios. This group stands out so much that it drives the reversal SMW returns in the 2nd ICL quintile. It turns out that this group is consisted of 29 industries over time, but with Pharmaceutical and Oil companies taking up more than 65% of the positions. Displayed in Table 1, these two industries are at the bottom of 44 FF industries in terms of average profitability. This suggests that gross profitability, may not be an efficient stock picking criterion when used to compare firms' performance across industries. The top 5 holding industries in this best performing portfolio are Pharmaceutical, Oil, Construction, Steel, and Business Services, together taking up 78.28% of the positions.

It is important to note that the market structure evolves over time. As mentioned earlier, the average industry concentration level declined sharply from 1973 to 1997 as more and more firms got listed on the three major exchanges. After 1997, the number of firms started to decline gradually, and 75% of industries become more and more concentrated over time. (Grullon et al., 2019). To test the robustness of SMW strategy over time, I split the sample into pre-1997 (industry expansion) and post-1997 (industry consolidation) period and calculate the equal weighted and value weighted returns. Results are presented in Table 4. Overall, the equal-weighted SMW strategies works consistently in both sub-sample periods, with the magnitude of returns slightly higher in industry expansion period. However, the 2nd ICL quintile still does not produce any statistically significant returns. In industry consolidation period (1997-2017), both 2nd and 5th (highest concentration) ICL quintiles produce SMW returns indifferent from zero. It is safe to state that zero-investment strategies based on GP works well in competitive industries overall.

Likewise, the value weighted SMW strategy displays consistent return patterns with only the most competitive industry quintiles generating statistically significant returns. During the industry expansion period (1973-1997), SMW strategy generates 0.79% per month with a t-statistic of 2.95 only in the most competitive industry quintile. The reversal in 2nd ICL quintile appears to be significant only during the industry consolidation period (1997-2017). The opposite performance between the most and second-most competitive industry quintiles is the reason that when using all stocks, the value weighted SMW strategy generates returns that are indifferent from zero. Indeed, product market competition affects the performance of zero-investment strategy based on GP/AT, especially when using the value-weighted scheme.

Fama and French (2015) develop two new factors: RMW and CMA. RMW is based on operating profitability. Although it is different from gross profitability, I suspect that RMW may explain most of the value weighted SMW returns. To test this hypothesis, I perform time series of regression of the monthly SMW returns in each ICL quintile using the FF four-factor (FF three-factor plus UMD[9]) and five-factor[10] model. The risk-adjusted returns

[9] UMD is the monthly premium of winners minus losers (Carhart, 1997).
[10] Fama French factors are downloaded from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

(alphas) are presented in Table 5. The two models are displayed as equation (2) and (3)[11] below.

Table 4. Subsample Period Average Stock Returns of Portfolios by Gross Profitability, and by Gross Profitability & Industry Concentration Level 1973-1996 and 1997-2017

Note: This table presents the subsample period results of the zero-investment strategy based on gross profit to total assets ratio (GP) alone, and dependently sorted portfolios by industry concentration level (ICL) and GP/AT, same with Table 3. Portfolios are formed at the end of June in year *t*, and then held from July of year *t* to June of year *t+1*. For portfolio formation, please refer to the detailed note for Table 3. Equal weighted raw returns (%) are presented on the left-hand side; value-weighted raw returns (%) are presented on the right-hand side. T-statistics calculated from Newey-West adjusted standard errors are presented in parentheses.

Panel A 1973-1996

Equal Weighted Portfolio Returns							Value Weighted Portfolio Returns						
	Industry Concentration level (FF48)							Industry Concentration level (FF48)					
GP	All	Low	2	3	4	High	GP	All	Low	2	3	4	High
Weak	1.27	0.97	1.37	1.39	0.98	1.01	Weak	1.56	1.33	1.84	2.07	1.6	1.76
	-4.01	-3.58	-3.13	-3.12	-2.29	-2.88		-6.98	-6.24	-6.67	-5.9	-5.46	-5.46
2	1.25	1.22	1.09	1.4	1.12	1.13	2	1.49	1.48	1.53	1.84	1.68	1.37
	-5.01	-4.1	-3.51	-4.6	-3.72	-3.72		-7.73	-7.07	-6.97	-7.36	-6.99	-5.55
3	1.43	1.38	1.21	1.45	1.37	1.39	3	1.69	1.67	1.45	1.68	1.53	1.62
	-5.73	-4.42	-4.15	-4.68	-4.3	-4.36		-7.91	-6.69	-7.17	-6.23	-6.96	-6.87
4	1.54	1.42	1.48	1.45	1.36	1.58	4	1.57	1.67	1.68	1.52	1.49	1.63
	-6.09	-4.49	-5.31	-4.59	-4.43	-5.26		-7.69	-7.13	-8.33	-6.03	-6.43	-6.68
Strong	1.76	1.55	1.64	1.79	1.76	1.78	Strong	1.71	1.95	1.77	1.81	1.8	1.37
	-6.81	-4.52	-5.4	-5.44	-5.35	-5.96		-8.28	-8.12	-8.03	-6.07	-7.97	-5.87
SMW	0.62	0.59	0.26	0.69	0.87	0.7	SMW	0.24	0.73	-0.67	-0.05	0.37	-0.21
	-4.92	-3.54	-1.01	-3.81	-5.19	-3.02		-1.56	-4.03	(-2.25)	(-0.21)	-1.61	(-0.77)

Panel B 1997 - 2017

Equal Weighted Portfolio Returns							Value Weighted Portfolio Returns						
	Industry Concentration level (FF48)							Industry Concentration level (FF48)					
GP	All	Low	2	3	4	High	GP	All	Low	2	3	4	High
Weak	0.99	0.95	1.39	0.77	0.78	1.15	Weak	1.37	1.09	3.15	1.61	1.23	1.35
	(2.17)	(2.06)	(1.84)	(1.34)	(1.11)	(1.80)		(3.76)	(3.93)	(4.91)	(2.91)	(3.02)	(3.20)
2	1.11	1.19	1.00	1.35	1.10	0.91	2	1.33	1.32	1.51	1.58	1.48	1.24
	(2.93)	(2.33)	(1.96)	(2.98)	(2.09)	(1.86)		(4.59)	(4.19)	(4.06)	(4.56)	(4.09)	(3.35)
3	1.30	1.39	1.16	1.29	1.31	1.37	3	1.63	1.71	1.34	1.36	1.56	1.61
	(3.65)	(2.72)	(2.56)	(2.95)	(2.69)	(2.82)		(4.70)	(3.87)	(4.34)	(3.41)	(4.55)	(4.45)
4	1.37	1.41	1.54	1.20	1.24	1.58	4	1.49	1.81	1.55	1.38	1.36	1.29
	(3.86)	(2.75)	(3.62)	(2.60)	(2.74)	(3.50)		(4.93)	(4.55)	(5.73)	(3.74)	(3.83)	(3.55)
Strong	1.57	1.65	1.53	1.64	1.70	1.63	Strong	1.61	1.75	1.68	1.81	1.56	0.94
	(4.29)	(3.04)	(3.12)	(3.44)	(3.35)	(3.81)		(6.04)	(5.29)	(5.34)	(4.46)	(4.98)	(3.91)
SMW	0.58	0.70	0.14	0.87	0.92	0.49	SMW	0.24	0.66	-1.47	0.20	0.34	-0.41
	(2.81)	(3.37)	(0.33)	(2.72)	(2.97)	(1.09)		(1.12)	(2.80)	(-3.09)	(0.55)	(0.91)	(-1.10)

[11] SMB (small minus big) is monthly premium of size factor; HML (high minus low) is the monthly premium of book to market factor; RMW (robust minus weak) is monthly premium of operating profitability factor; CMA (conservative minus aggressive) is monthly premium of investment factor.

$$EXRET^{12}{}_t = \alpha + \beta_1 EXMKT^{13}{}_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \varepsilon_t \quad (2)$$

$$EXRET_t = \alpha + \beta_1 EXMKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \varepsilon_t \quad (3)$$

Table 5. Risk Adjusted Returns of SMW Strategies by Industry Concentration Level Quintile 1973-2017

Note: This table presents the risk-adjusted monthly returns (%) (alpha) of Strong Minus Weak (SMW) zero-investment strategies based on 25 industry concentration level (ICL) and gross profitability (GP) portfolios from 1973 to 2017, and two subsample periods. The alphas are calculated as the intercepts from time series regressions of SMW returns on Fama French four- and five-factor models. Equal weighted SMW alphas are presented on the left-hand side; value weighted SMW alphas are presented on the right-hand side. T-statistics of the alphas are in parentheses.

	Equal Weighted SMW Portfolio					**Value Weighted SMW Portfolio**				
	Panel A 1973-2017									
	Industry Concentration Level					**Industry Concentration Level**				
	Low	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
FF4	0.53	0.34	0.74	0.90	0.83	0.66	-0.29	0.03	0.45	0.11
	(4.10)	(1.51)	(4.71)	(5.88)	(3.88)	(3.57)	(-1.11)	(0.14)	(1.93)	(0.47)
FF5	0.43	0.15	0.62	0.94	0.65	0.52	-0.64	-0.14	0.32	0.05
	(3.47)	(0.72)	(3.96)	(6.19)	(3.13)	(2.84)	(-2.68)	(-0.61)	(1.37)	(0.20)
	Panel B 1973-1996									
	Industry Concentration Level					**Industry Concentration Level**				
	Low	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
FF4	0.56	0.65	0.78	0.85	1.07	0.91	0.44	0.37	0.48	0.62
	(3.16)	(2.48)	(4.21)	(5.14)	(4.67)	(3.74)	(1.35)	(1.37)	(1.67)	(2.00)
FF5	0.14	0.27	0.71	0.93	0.79	0.19	-0.21	0.28	0.47	0.26
	(0.84)	(1.01)	(3.70)	(5.52)	(3.40)	(0.81)	(-0.67)	(0.98)	(1.60)	(0.81)
	Panel C 1997-2017									
	Industry Concentration Level					**Industry Concentration Level**				
	Low	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
FF4	0.56	0.26	0.89	1.10	0.71	0.46	-0.94	0.15	0.54	-0.30
	(3.08)	(0.72)	(3.48)	(4.44)	(1.94)	(1.65)	(-2.42)	(0.43)	(1.45)	(-0.79)
FF5	0.55	-0.03	0.67	1.15	0.53	0.37	-1.36	-0.05	0.35	-0.48
	(2.95)	(-0.08)	(2.62)	(4.36)	(1.43)	(1.32)	(-3.70)	(-0.13)	(0.92)	(-1.21)

On the left-hand side of all three panels in Table 5, the FF five-factor model seems to be able to explain the SMW returns in most competitive industry quintile during industry expansion period (1973-1996) and the returns in the most concentrated industry quintile during industry consolidation period (1997-2017). All the other quintiles except the 2nd ICL quintile still displays strong positive risk-adjusted returns. This is expected because FF factors are created using value-weighted scheme. To my surprise, on the right-hand side in Table 5 panel A, it seems the FF five-factor model cannot explain the SMW returns in the most competitive quintile and the SMW reversals in the 2nd ICL quintile. The FF four-factor model seems to fully explain the SMW reversal returns in the 2nd ICL quintile.

[12] EXRET stands for excess return, it is the stock/portfolio monthly return minus risk-free return.
[13] EXMKT stands for excess market return. It is the market return minus risk-free return.

However, as Panel C show, neither four- or five-factor model can explain the reversal returns during industry consolidation period (1997-2017).

In Panel B and C, the risk-adjusted returns of value-weighted SMW strategy from FF 5 factor model in the most competitive quintile is indifferent from zero, but the strong SMW reversal in the 2nd ICL quintile has a staggering -1.36% risk-adjusted return per month with a t-statistic of -3.70. It is safe to say that, using value weighted strategy, the most profitable one is to long the firms with lowest GP/AT ratio and short the ones with highest GP/AT ratio at the end of June each year, and then hold such portfolio for 12 months before rebalancing in the 2nd most competitive industry quintile. The evidence strongly suggests that product market competition level affects the performance of value-weighted zero-investment strategy based on gross profitability. While such strategy is profitable in the only most competitive industry quintile, a reversal strategy can generate much higher risk-adjusted returns in the 2nd most competitive industry quintile.

4. Conclusion

Using publicly held firms from NYSE, AMEX, and NASDAQ from 1973 to 2017, I find that product market competition level affects the performance of the zero-investment strategy based on firm's gross profitability. While the strategy with an equally weighted scheme produces positive and significant returns in four out of five industry concentration level quintiles, the value-weighted strategy appears to be profitable only in the most competitive industry quintile. The evidence seems to support one of Sir John Templeton's stock picking criteria, by which firms that have higher profitability than their direct competitors tend to have higher future stock returns.

The difference between the results using two weighting schemes suggests that small firms are driving the results. More importantly, in the 2nd most competitive industry quintile, the same value-weighted SMW strategy generates an astonishing risk-adjusted reversal return of -0.64% per month from 1973 to 2017, and an even stronger return of -1.36% per month during the past two decades (industry consolidation period: 1997-2017). This reversal return is driven by the top performing portfolio from the 25 ICL & GP/AT dependently sorted ones. I find that more than 65% of this portfolio consists of two industries (Pharmaceutical and Oil) at the bottom of gross profitability compared to other non-financial Fama French industries.

What is more valuable from these findings is that while using gross profitability as stock picking criteria is reasonable, investors should caution when comparing firms across industries with different natures of business and different product market competition levels. When the industry becomes highly concentrated, where only one or a few firms dominate the market, profitability becomes much less effective in the zero-investment strategy.

Acknowledgment

I am grateful to an anonymous referee for very helpful comments on earlier versions of the paper. All remaining errors are my own.

References

Ball, R., Gerakos, J., Linnainmaa, J., Nikolaev, V., 2015. Deflating profitability. *Journal of Financial Economics* 117 (2), 225–248.

Ball, R., Gerakos, J., Linnainmaa, J., Nikolaev, V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics* 121, 225–248.

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

Cohen, R., Polk, C., Vuolteenaho, T., 2003. The value spread. *Journal of Finance* 58 (2), 609–641

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns of bonds and stocks, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153-193.

Fama, Eugene F., and Kenneth R. French, 2008, Dissecting Anomalies, *Journal of Finance* 63, 1653-1678.

Fama, Eugene F. and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics*, 116, 1-22.

Grullon, G., Y. Larkin, and R. Michaely, 2019, Are US industries becoming more concentrated? *Review of Finance,* 1-47.

Haugen, Robert A., and Nardin L. Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401-439.

Hou, K. and D. Robinson, 2006, Industry concentration and average stock returns, *Journal of Finance* 10: 1927-1956.

Kahle, K. M. and R. A. Walking, 1996, The impact of industry classifications and financial research, *Journal of Financial and Quantitative Analysis* 31, 309-335.

Newey, W., and K. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1-28.

Schumpeter, J., 1912, *The Theory of Economic Development* (Harvard University Press, Cambridge, Mass).