JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 59, No. 1, Feb. 2024, pp. 307–338

doi:10.1017/S0022109023000169

The Impact of Uncertainty on Investment: Empirical Challenges and a New Estimator

Delong Li
University of Guelph Gordon S. Lang School of Business and Economics and Cornerstone Research
delong.li@cornerstone.com (corresponding author)

Yiguo Sun
University of Guelph Gordon S. Lang School of Business and Economics
yisun@uoguelph.ca

Abstract

This article proposes a new method for examining the impact on a firm's investment of uncertainty reflected in its stock-return volatility. We simultaneously address the endogeneity of uncertainty and mismeasurement in Tobin's Q, but earlier empirical work often neglects one of the two issues. Our nonparametric estimates further suggest that the relation between investment and uncertainty is significantly decreasing and strongly concave. This result contrasts with the existing literature that widely adopts linear regressions. Ignoring nonlinearity or measurement error in Q can lead to a substantial estimation bias. However, the bias due to the endogeneity of uncertainty is small.

I. Introduction

This article examines the impact of uncertainty measured by a firm's stock-return volatility on its investment through a novel empirical method. Our estimates simultaneously account for endogeneity in stock volatility and measurement error in Tobin's Q as a proxy for investment opportunities. We further allow the investment–uncertainty relation to be arbitrarily nonlinear. In contrast, earlier empirical work often leaves either the endogeneity or mismeasurement issue untreated and focuses on linear models (e.g., Panousi and Papanikolaou (2012), Alfaro, Bloom, and Lin (2022)). To what extent can these simplifications affect estimation results? We empirically find that neglecting measurement error in Q can introduce a substantial bias. However, the

We appreciate comments from seminar participants at the Econometric Society Asian meeting, China meeting, and North American summer meeting in 2019, Northern Finance Association 2019 conference (especially our discussant, Georgios Skoulakis), Bank of Finland, Shanghai University of Finance and Economics, and Wilfrid Laurier University. We give special thanks to Paul Malatesta (the editor) and Iulian Obreja (the referee) for detailed comments and helpful suggestions, and to the Social Sciences and Humanities Research Council of Canada (grant ID 430-2022-00119) for financial supports. In addition, Li coauthored this article before joining Cornerstone Research. The views expressed herein are solely those of the authors who are responsible for the content, and do not necessarily represent the views of Cornerstone Research.

bias caused by the endogeneity of uncertainty is small. Moreover, the investment–uncertainty relation is decreasing and strongly concave.

Another goal of this article is to establish a methodology to explicitly and jointly address unknown nonlinearities, regressor endogeneity, and mismeasurement. These problems are commonplace in empirical studies, but a comprehensive approach to address them all has not yet been developed. We propose a nonparametric method to fill the gap. To this end, our article extends the work of Erickson and Whited (2000) and Erickson, Jiang, and Whited (2014) that treat measurement error in Q in linear models without regressor endogeneity.

The theoretical literature shows that the relation between investment and uncertainty can have complex nonlinearities. For instance, Dixit and Pindyck (1994) document that heightened uncertainty decreases investment by increasing the real option value, and Sarkar (2000) and Wong (2007) demonstrate that this real option effect is nonlinear. The financial friction literature instead argues that high uncertainty can reduce investment by raising risk premiums, as in Bernanke, Gertler, and Gilchrist (1999), or due to managerial risk aversion, as in Panousi and Papanikolaou (2012). In this strand of literature, the relation between investment and uncertainty is typically concave. Moreover, greater uncertainty can also increase investment i) by inducing managers to overinvest (Eisdorfer (2008)); ii) if the increased uncertainty is due to "good news" (Segal, Shaliastovich, and Yaron (2015)); or iii) because of the Oi–Hartman–Abel effect (Oi (1961), Hartman (1972), and Abel (1983)). Given that all these effects can coexist, the combined effect of uncertainty on investment is expected to be nonlinear with a complex and unknown shape.

Due to such unknown nonlinearities, any parametric empirical model can lead to misspecification in estimating the impact of uncertainty on investment. This article instead proposes a nonparametric estimator based on penalized series approximation. We first approximate the investment–uncertainty relation by a linear combination of Hermite basis functions. Then, through a penalty method, we select the most relevant approximation terms relying on data to obtain a parsimonious model. We prove that this data-driven approach can recover the true investment–uncertainty relation under common regularity assumptions.

Moreover, there are regressor endogeneity and measurement error issues. On the one hand, the variable proxying for uncertainty at the firm level (such as stock-return volatility) can be affected by the firm's investment decisions. In linear models, this reverse causality is often addressed using the 2-stage least squares (2SLS) method with proper instrumental variables (see, e.g., Panousi and Papanikolaou (2012), Alfaro et al. (2022)). However, in nonparametric regressions, the 2SLS method faces an ill-posed inverse problem. Newey and Powell (2003) develop a nonparametric estimator analog to the 2SLS method to resolve this concern but assume that all variables are perfectly measured.

On the other hand, a firm's investment opportunities are unobservable, of which empiricists can obtain only imperfect measures, such as Tobin's Q. Since Q is often included in estimation as a control variable, its measurement error, if untreated, can lead to estimation bias of unknown directions to *all* coefficients, not only that of Q itself (Greene (2017)). It is also difficult to find suitable instrumental variables to deal with this mismeasurement.[1] Erickson and Whited ((2000), (2002), (2012)) and Erickson et al. (2014) develop a high-order moment approach that needs no such extraneous instruments (outside the model being considered) to correct for the measurement error bias. However, they focus on linear regressions of investment on Tobin's Q and an *exogenous* cash flow variable. Their method cannot account for unknown nonlinearities or endogenous regressors such as stock volatility. Meijer, Spierdijk, and Wansbeek (2017) propose another measurement error remedy in linear models that do not have additional regressor endogeneity problems.

To jointly address endogeneity and mismeasurement problems, we propose a novel control function approach combined with a nonparametric generalized method of moments (GMM) estimator. The control function approach is originally developed by Newey, Powell, and Vella (1999) for cross-sectional models, and ours is the first effort to extend it to panel data models with fixed effects. We demonstrate that it can address the endogeneity of uncertainty in the presence of mismeasured Tobin's Q. Then, we correct for the measurement error bias caused by Q through a GMM estimator in the spirit of Meijer et al. (2017) but in our nonparametric framework. Furthermore, in contrast to their method, our approach involves a penalty selection of moment conditions used in the GMM estimation, which greatly improves finite-sample performance. Last, we prove that the final estimator, which combines series approximation, control function approach, and measurement error remedy, is consistent and asymptotically normally distributed.

Applying our proposed methodology to U.S. firm-level data, we find that the investment–uncertainty relation is significantly decreasing and strongly concave. Thus, in net terms, uncertainty decreases investment, and this effect intensifies as uncertainty increases. The magnitude of the concavity is also considerable: Raising the level of uncertainty from the bottom to the top decile of our data more than doubles the marginal effect of uncertainty on investment. We further construct a point-by-point t-statistic to formally test the appropriateness of an otherwise similar linear regression model, demonstrating that the linear coefficient significantly overestimates the marginal effect when uncertainty is at low-to-median levels (accounting for more than 50% of our sample).

Due to methodological limitations, the existing literature often focuses on addressing either the endogeneity of uncertainty or mismeasurement in Q, but not both issues simultaneously. The endogeneity of uncertainty receives arguably more attention in the investment–uncertainty literature (discussed below), because it is intuitive to hypothesize that the mismeasured Q as a control variable should play a lesser role in affecting the coefficient of uncertainty. We find that this intuition is incorrect. To illustrate our point, we reestimate the model by leaving one of the two issues untreated and compare the resulting estimates with the estimates in which both issues are addressed. The results suggest that by not treating mismeasured Q, one can introduce a significant amplification bias in estimating the marginal impact of uncertainty on investment. The untreated measurement error can also lead to a misinterpretation of the shape of the investment–uncertainty

[1] Almeida, Campello, and Galvao (2010), among others, adopt lagged Tobin's Q's as instruments for the current Q. Erickson and Whited (2012) argue that lagged Q's are not valid instruments because the measurement error is likely to be serially correlated.

relation. However, the bias caused by the endogeneity of uncertainty is small and statistically insignificant.[2]

We further analyze subsamples and find evidence in line with the theoretical notion that uncertainty influences investment through various channels. This finding reaffirms the need for nonparametric estimation since the combined effect of uncertainty does not have an explicit functional form. Our analysis supports, specifically, the coexistence of the real option channel, risk premium channel, overinvestment channel, and good news principle that the uncertainty of good news affects investment differently than the uncertainty of bad news. We also demonstrate that some channels are difficult to detect using linear regressions.

Related Literature

Previous empirical studies examine the effects of uncertainty on investment in linear regression models, and most of them focus on dealing with the endogeneity of uncertainty. For instance, Eisdorfer (2008), Julio and Yook (2012), Gulen and Ion (2016), Kim and Kung (2017), and Doshi, Kumar, and Yerramilli (2018) choose to explore aggregate uncertainty fluctuations that are exogenous to firm-level investment. Alnahedh, Bhagat, and Obreja (2019) construct a backward-looking measure of cash flow uncertainty for individual firms, arguing that it is less prone to endogeneity. Bond and Cummins (2004) and Gilchrist, Sim, and Zakrajšek (2014) instead use stock-return volatility to measure uncertainty at the firm level and address endogeneity through a GMM method with lagged variables as instruments, following Arellano and Bond (1991). However, this approach can be complicated by weak-instrument and many-instrument problems (see discussions in Roodman (2009)). Panousi and Papanikolaou (2012) and Alfaro et al. (2022) also study individual firms' stock volatility but resolve endogeneity in a 2SLS framework with instrumental variables based on customer base concentration and industry-level heterogeneous exposure to aggregate uncertainty shocks, respectively.

In this literature, much less attention has been given to the mismeasurement of Tobin's Q. All the abovementioned work has included Q as a control variable (with slightly different definitions). Among them, Kim and Kung (2017) and Doshi et al. (2018) have explicitly corrected for this measurement error.[3] However, their uncertainty variables are aggregated, and therefore their conclusions cannot characterize the responses of investment to firm-specific uncertainty. Another branch of the literature constructs arguably better measures of Q (see, among others, Bond and Cummins (2004)).[4] However, since true investment opportunities are unobservable, it is still difficult to determine whether the measurement error contained in these alternative Q measures is negligible; thus, the mismeasurement concern is not fully addressed. Panousi and Papanikolaou (2012) and Gulen and Ion (2016) consider

[2]An important caveat is that these results apply only to our empirical sample. In econometric theory, the mismeasurement and endogeneity biases can be in any directions and of any magnitude (see more discussions in footnote 6).

[3]Both articles employ the high-order cumulant estimator developed in Erickson et al. (2014) to treat the measurement error. Kim and Kung (2017) also implement an alternative method that instruments the mismeasured Q with lagged Q and cash flow.

[4]Peters and Taylor (2017) also propose an alternative measure of Tobin's Q, but their work estimates the investment–Q relation rather than the investment–uncertainty relation. See also Philippon (2009) on the aggregate relation between investment and a Q measure constructed by bond prices.

both the endogeneity of uncertainty and measurement error in Q but separately address one of them at a time.[5] Doing so can still lead to biased estimates due to the issue that is left untreated.

II. Empirical Model and Challenges

We begin with a linear regression model to discuss the mismeasurement in Tobin's Q and endogeneity of uncertainty. We aim to illustrate that, even in this linear setup, it is difficult to simultaneously address the two problems. Then, we move to the more complicated nonparametric case and outline our approach.

The following regression (1) and its variations have been widely adopted by previous studies to examine how uncertainty affects investment at the firm level:

$$y_{i,t} = \mu_i + \lambda_t + \alpha_0 x^*_{i,t-1} + \beta_0 z_{i,t-1} + \gamma_0 s_{i,t-1} + u_{i,t}, \tag{1}$$

where $y_{i,t}$ denotes the investment of firm i in year t for $i = 1, \ldots, n$ and $t = 1, \ldots, T$; μ_i and λ_t represent firm and year fixed effects, respectively; $x^*_{i,t-1}$ denotes investment opportunities that are unobservable to researchers; $z_{i,t-1}$ is the cash flow representing internal funds; $s_{i,t-1}$ is a measure of uncertainty; and $u_{i,t}$ is the error term. All explanatory variables are lagged by 1 year to ensure that they are in the firm's information set when it invests.

In model (1), the investment opportunities $x^*_{i,t-1}$ are unobservable and often measured with error by Tobin's Q. To see how this measurement error affects the estimation results, following Erickson and Whited (2000), we assume that

$$x_{i,t-1} = x^*_{i,t-1} + e_{i,t-1}, \tag{2}$$

where $x_{i,t-1}$ is the measured Tobin's Q and $e_{i,t-1}$ is the measurement error. Substituting equation (2) into equation (1) yields

$$y_{i,t} = \mu_i + \lambda_t + \alpha_0 x_{i,t-1} + \beta_0 z_{i,t-1} + \gamma_0 s_{i,t-1} + \underbrace{u_{i,t} - \alpha_0 e_{i,t-1}}_{\varepsilon_{i,t}}, \tag{3}$$

where the regressor $x_{i,t-1}$ is correlated with the new error term $\varepsilon_{i,t}$, if $\alpha_0 \neq 0$, since they both contain the measurement error term $e_{i,t-1}$. That is, the measured Tobin's Q is an endogenous variable in model (3), thus introducing estimation bias to *all* the regression coefficients. In econometric theory, the directions and magnitude of the bias are also difficult to predict.[6]

[5]Both papers address the endogeneity issue via instrumental variable approaches assuming that there is no measurement error, and they correct for measurement error bias using the high-order moment/cumulant estimator developed in Erickson and Whited (2000) and Erickson et al. (2014), leaving the regressor endogeneity issue untreated.

[6]An exception is the estimates of α_0, where the bias is known as an attenuation bias. For other coefficients, however, the directions and magnitude of the bias are theoretically unknown (Greene (2017)). Erickson and Whited (2000) and Abel (2018) consider a special case in which the investment regression contains mismeasured Tobin's Q and cash flow as the only two regressors (i.e., these authors do not study the effects of uncertainty). In this simple two-variable case, a closed-form solution to the bias exists, suggesting that the coefficient of cash flow is biased upward if the correlation between the true Q and cash flow is positive. This correlation-based logic is no longer valid when there are more than two regressors, because the bias depends on the covariance between the true Q and other regressors in complex ways.

Correcting for the measurement error bias is challenging, partly because of a lack of proper instruments for Tobin's Q, and partly because of the potential endogeneity of uncertainty. For the former, the existing literature has proposed some unconventional measurement error remedies that require no such instruments (Erickson and Whited (2000), Erickson et al. (2014), and Meijer et al. (2017)). However, to achieve identification, all these methods assume that the regressors other than Tobin's Q are exogenous. This assumption does not hold in model (3) if the uncertainty variable, $s_{i,t}$, is endogenous, that is, correlated with $u_{i,t}$. One such case is when uncertainty is represented by individual stock volatility because the firm's stock returns are likely to be influenced by its investment. With such reverse causality, the previous literature has not yet developed a consistent estimator to the linear models (1)–(3).[7]

Moreover, as discussed previously, the investment–uncertainty relation is theoretically nonlinear with an unknown functional form. As a result, the linear investment model (1) is misspecified, as any specific parametric model can be. To account for this problem, we replace $\gamma_0 s_{i,t-1}$ with an unknown function $g_0(s_{i,t-1})$ in which we do not impose any specific assumptions on the functional form, except for smoothness, and aim to recover $g_0(\cdot)$ from data through a penalized series approximation approach (discussed in Section III). In addition, since previous studies suggest that the investment–cash flow relation is potentially nonlinear as well, we also replace $\beta_0 z_{i,t-1}$ with an unknown function $f_0(z_{i,t-1})$.[8] Our proposed nonparametric regression model is therefore defined as follows:

$$(4)\qquad y_{i,t}=\mu_i+\lambda_t+\alpha_0 x^*_{i,t-1}+f_0(z_{i,t-1})+g_0(s_{i,t-1})+u_{i,t}.$$

We implement a novel control function approach, accompanied by a nonparametric GMM estimator, to simultaneously deal with the endogeneity of uncertainty and measurement error in Tobin's Q in models (2) and (4). The control function approach decomposes the regression error term into two parts: one part that contains *all* the information that is correlated with the uncertainty variable (referred to as the "control function") and the other part that is uncorrelated with uncertainty:

$$(5)\quad y_{i,t}=\mu_i+\lambda_t+\alpha_0 x^*_{i,t-1}+f_0(z_{i,t-1})+g_0(s_{i,t-1})+\underbrace{\text{CTRLFUNC}}_{\text{correlated with } s}+\underbrace{u_{i,t}-\text{CTRLFUNC}}_{\text{uncorrelated with } s}.$$

In model (5), the part that is uncorrelated with uncertainty acts as the new error term, and, as a result, the uncertainty variable becomes exogenous. The control function is unknown, and we estimate it nonparametrically from data. Then, considering equations (2) and (5) together, since all the variables other than Tobin's Q are exogenous, we can correct for the measurement error bias through a GMM estimator in the spirit of Meijer et al. (2017) by extending their work to our nonparametric framework.

[7]We define the reverse causality issue rigorously in Section III.A. Note that using lagged uncertainty in model (3) cannot prevent this problem because, after the firm fixed effects are removed through the within-group transformation, the terms $\left[s_{i,t-1}-T^{-1}\Sigma_\tau s_{i,\tau}\right]$ and $\left[u_{i,t}-T^{-1}\Sigma_\tau u_{i,\tau}\right]$ are still correlated.

[8]Previous studies include, among others, Allayannis and Mozumdar (2004), Bhagat, Moyen, and Suh (2005), Cleary, Povel, and Raith (2007), and Firth, Malatesta, Xin, and Xu (2012).

Several caveats are worth nothing. First, we intentionally keep the term Q linear, that is, $\alpha_0 x^*_{i,t-1}$, so that our results provide a direct test of neoclassical theory. This linear assumption also simplifies our estimation. Schennach and Hu (2013) have considered a sieve maximum likelihood estimator to investigate the potential nonlinearity between investment and Q, finding that the nonlinear estimate of the investment–Q relation is close to the linear estimate once the measurement error in Q is corrected for. Therefore, we do not expect that this simplification would cause a significant bias to our estimation.

Second, methodologically, it is straightforward to extend model (4) to further account for interactions among Q, cash flow, and volatility by, for example, letting α_0 depend on z and s or considering a full nonparametric model $F\left(x^*_{i,t-1}, z_{i,t-1}, s_{i,t-1}\right)$. However, empirically, the estimation of such a model suffers from the curse of dimensionality, and the estimated marginal effects are also difficult to interpret. Thus, to obtain an estimate that is more comparable to the literature (i.e., model (1)), we adhere to the semiparametric partially linear additive setup and leave further investigation to future research.

Third, we assume that the cash flow and volatility variables are free of measurement error and that cash flow is exogenous.[9] These assumptions can be easily relaxed in our econometric theory but are empirically useful, because we need exogenous variation to correct for any mismeasurement bias. As we will discuss in Section III, moment conditions built on cash flow and volatility provide such variation within the model to treat measurement error in Tobin's Q. Further allowing cash flow and volatility to be mismeasured or cash flow to be endogenous requires finding additional instrumental variables to address the measurement error and endogeneity problems.

Last, even though our focus is on the investment–volatility relation g_0, we allow the cash flow effect f_0 and the control function to take general forms because Cleary et al. (2007) have predicted a nonlinear investment–cash flow relation. Accounting for nonlinearities therein helps test for their theory, also mitigating the concern that the nonlinearities can bias the estimates of g_0. That said, doing so complicates the estimation process. For our empirical sample, we conduct a robustness check by replacing f_0 and the control function with linear functions and find that the estimated shape of g_0 remains similar (see Appendix E of the Supplementary Material).

III. Identification and Estimator

A. Identification and Assumptions

We start with the endogeneity of uncertainty in model (4) to present our identification strategy. Specifically, the endogeneity issue we consider here is the

[9]Both assumptions are common in the empirical literature. Given that cash flow is from accounting data and volatility from the stock market, the first assumption implicitly assumes that there exists no accounting error and that the stock market is efficient enough to correctly reflect underlying uncertainty. The second assumption can be invalid if unobservable investment opportunities drive both investment and cash flow and are not well controlled for. We assume that this is not the case once measurement error in Tobin's Q is removed, so that controlling for Q is sufficient to capture investment opportunities.

reverse causality that an investment shock reflected by the error term $u_{i,t}$ can influence the changes in the firm's stock-return volatility, $\Delta s_{i,t}$, thereby leading to a contemporaneous correlation between $\Delta s_{i,t}$ and $u_{i,t}$. As a result, $s_{i,t} = s_{i,t-1} + \Delta s_{i,t}$ is an endogenous variable.[10] Suppose that $\mathbf{w}_{i,t}$ is a vector of strictly exogenous instruments for $\Delta s_{i,t}$ with the following first-stage regression:

$$\Delta s_{i,t} = \eta_0 + \mathbf{w}'_{i,t}\pi_0 + v_{i,t}, \tag{6}$$

where $\mathbf{w}_{i,t}$ satisfies the usual exclusion conditions as follows, and $v_{i,t}$ is an error term with zero mean and finite variance:

$$\mathbb{E}(u_{i,t}\mathbf{w}_{i,s}) = 0 \text{ and } \mathbb{E}(v_{i,t}\mathbf{w}_{i,s}) = 0, \ \forall(i,t,s). \tag{7}$$

Given the exclusion conditions, the correlation between $\Delta s_{i,t}$ and $u_{i,t}$ must be caused by the correlation between $v_{i,t}$ and $u_{i,t}$. Thus,

$$\mathbb{E}(u_{i,t}|v_{i,t}) \neq 0, \forall(i,t). \tag{8}$$

Let $r_0(v_{i,t}) = \mathbb{E}(u_{i,t}|v_{i,t})$ be the control function with an unknown functional form, and define

$$\omega_{i,t} = u_{i,t} - r_0(v_{i,t}). \tag{9}$$

Substituting (9) into (4), we obtain

$$y_{i,t} = \mu_i + \lambda_t + \alpha_0 x^*_{i,t-1} + f_0(z_{i,t-1}) + g_0(s_{i,t-1}) + \underbrace{r_0(v_{i,t}) + \omega_{i,t}}_{u_{i,t}}, \tag{10}$$

where $\omega_{i,t}$ is the new error term. Given that the control function $r_0(v_{i,t})$ already captures all information related to $v_{i,t}$ (and therefore to $\Delta s_{i,t}$), the new error term $\omega_{i,t}$ is uncorrelated with $v_{i,t}$. Thus, the endogeneity issue is addressed. We estimate $r_0(\cdot)$ directly from data, and because $v_{i,t}$ is unobservable, we replace it with the fitted error $\widehat{v}_{i,t}$ calculated from the first-stage regression (6).

We estimate the unknown functions $f_0(\cdot)$, $g_0(\cdot)$, and $r_0(\cdot)$ through a series approximation method by approximating each function with Hermite orthonormal basis functions:[11]

$$\begin{aligned} f_0(z) &\approx \vartheta^{(1)}_{z,0}P_1(z) + \vartheta^{(2)}_{z,0}P_2(z) + \cdots + \vartheta^{(k)}_{z,0}P_k(z) = \vartheta'_{z,0}\mathbf{P}^k(z), \\ g_0(s) &\approx \vartheta^{(1)}_{s,0}P_1(s) + \vartheta^{(2)}_{s,0}P_2(s) + \cdots + \vartheta^{(k)}_{s,0}P_k(s) = \vartheta_{s,0}\mathbf{P}^k(s), \\ r_0(v) &\approx \vartheta^{(1)}_{v,0}P_1(v) + \vartheta^{(2)}_{v,0}P_2(v) + \cdots + \vartheta^{(k)}_{v,0}P_k(v) = \vartheta'_{v,0}\mathbf{P}^k(v), \end{aligned} \tag{11}$$

[10]See Panousi and Papanikolaou (2012), Alfaro et al. (2018), and the references therein for more discussions about the cause of the reverse causality. We follow Alfaro et al. (2018) to treat volatility changes as the source of endogeneity because the lagged volatility $s_{i,t-1}$ is predetermined and less likely to be correlated with the error term $u_{i,t}$.

[11]The Hermite approximation is similar to Taylor expansions in polynomials but uses Hermite orthonormal basis functions instead. The literature has documented that the Hermite approximation produces smaller biases than the polynomial approximation.

where $\mathbf{P}^k(\cdot) = [P_1(\cdot), \ldots, P_k(\cdot)]'$ represents the up-to-kth-order Hermite functions; $\vartheta_{z,0} = \left[\vartheta_{z,0}^{(1)}, \ldots, \vartheta_{z,0}^{(k)}\right]'$, $\vartheta_{s,0} = \left[\vartheta_{s,0}^{(1)}, \ldots, \vartheta_{s,0}^{(k)}\right]'$, and $\vartheta_{v,0} = \left[\vartheta_{v,0}^{(1)}, \ldots, \vartheta_{v,0}^{(k)}\right]'$ are the expansion coefficients of $f_0(z)$, $g_0(s)$, and $r_0(v)$, respectively. In the series approximation literature, the highest order k is determined by the sample size nT. The larger nT is, the higher the k that can be used, and the more precise the approximation. However, this conventional approach has an unappealing feature: Once k is determined, *all* the expansion terms up to the kth order are used in the approximation. If some terms are redundant, then including them in the regression model decreases the estimation efficiency and can cause a many-regressor bias in finite samples.[12]

We propose the use of a penalty method to select the most relevant approximation terms among $\{P_1(\cdot), \ldots, P_k(\cdot)\}$ and remove the irrelevant terms based on the data. This approach decreases the number of regressors, thus improving the estimation efficiency and reducing finite-sample bias.[13] To simplify the notation, we continue to use equation (11) to represent the series approximation of $f_0(\cdot)$, $g_0(\cdot)$, and $r_0(\cdot)$ before formally introducing the penalty estimator in Section III.B. Substituting (11) into (10) yields

$$(12) \qquad y_{i,t} \approx \mu_i + \lambda_t + \alpha_0 x_{i,t-1}^* + \underbrace{\vartheta_{z,0}' \mathbf{P}^k(z_{i,t-1}) + \vartheta_{s,0}' \mathbf{P}^k(s_{i,t-1}) + \vartheta_{v,0}' \mathbf{P}^k(v_{i,t})}_{f_0(z_{i,t-1}) + g_0(s_{i,t-1}) + r_0(v_{i,t})} + \omega_{i,t}$$

as a *linear* regression model, where the regressors include a group of Hermite terms.[14] For compactness, we let $\vartheta_0 = \left[\vartheta_{z,0}', \vartheta_{s,0}', \vartheta_{v,0}'\right]'$ and $\mathbf{E}_{i,t} = \left[\mathbf{P}^k(z_{i,t-1})', \mathbf{P}^k(s_{i,t-1})', \mathbf{P}^k(v_{i,t})'\right]'$; then, the regression model can be rewritten as

$$(13) \qquad y_{i,t} \approx \mu_i + \lambda_t + \alpha_0 x_{i,t-1}^* + \mathbf{E}_{i,t}' \vartheta_0 + \omega_{i,t}.$$

Next, we substitute the measurement error equation (2) into model (13) and obtain

$$(14) \qquad y_{i,t} \approx \mu_i + \lambda_t + \alpha_0 x_{i,t-1} + \mathbf{E}_{i,t}' \vartheta_0 + \varepsilon_{i,t};$$

$$(15) \qquad \varepsilon_{i,t} = \omega_{i,t} - \alpha_0 e_{i,t-1},$$

where we treat $e_{i,t}$ as a *classical* measurement error following Erickson and Whited (2000); thus, $\mathbb{E}(u_{i,t} e_{i,s}) = 0$ for all i, t, and s. In model (14), $x_{i,t-1}$ (Tobin's Q) is an endogenous variable since it correlates with the error term $\varepsilon_{i,t}$ if α_0 is nonzero. In linear models, the literature has developed three different approaches to use internal variation to address classical measurement error when external instrumental

[12]Using polynomial approximation to illustrate this idea, when the true functional form is $h(a) = a^2$, the linear term is redundant in the approximation, and the same logic applies to Hermite approximation.

[13]The penalty estimator also allows $f_0(\cdot)$, $g_0(\cdot)$, and $r_0(\cdot)$ to consist of different approximation terms. In contrast, the conventional approach assigns the same k to all three approximations and thus implicitly assumes that the three unknown functions have the same highest order.

[14]The relation "$\approx$" indicates that approximation error exists in the Hermite expansions. This approximation error vanishes as $n \to \infty$ under Assumption 2(i), which we introduce later; thus, this error does not affect the consistency of our estimator.

variables for $x_{i,t-1}$ are difficult to find.[15] Erickson and Whited ((2000), (2002), (2012)) and Erickson et al. (2014) develop a high-order moment (cumulant) estimator. Meijer et al. (2017) document two additional approaches: one based on strongly exogenous regressors and another based on exploiting the error term's covariance matrix.[16] We extend the latter two remedies to fit into our nonparametric methodology and prove that the unified estimator is consistent.[17]

Specifically, we first cancel the firm fixed effect μ_i through within-group transformation and obtain the following transformed regression:

$$\tilde{y}_{i,t} \approx \tilde{\lambda}_t + \alpha_0 \tilde{x}_{i,t-1} + \tilde{\mathbf{E}}'_{i,t} \vartheta_0 + \tilde{\varepsilon}_{i,t}, \tag{16}$$

where $\tilde{a}_{i,t} = a_{i,t} - T^{-1}\sum_\tau a_{i,\tau}$ denotes the demeaned data. We estimate the transformed model through a GMM approach, where the orthogonal moment conditions are constructed based on two blocks of information. First, we utilize time dummies and regressors in $\left\{\tilde{\mathbf{E}}_{i,t}\right\}_{t=1}^{T}$ as instruments for the mismeasured Tobin's Q. Regarding the exclusion condition, time dummies represent macroeconomic fundamentals and are by default exogenous to individual firm's investment. Cash flow is assumed to be exogenous by following Erickson and Whited (2000) and Kim and Kung (2017). The measure of uncertainty and control function are exogenous by construction due to the step of the control function approach. In terms of the relevance condition, all these instruments are correlated with Tobin's Q following economic intuitions. The reason is that Q relies on the firm's stock market value, which is, in turn, associated with macroeconomic fundamentals, cash flow, and uncertainty. The second block of moments utilizes the *orthogonal complement* of the error terms' covariance matrix to construct instruments for Q. See Appendix B of the Supplementary Material for the technical details and (B.1) for a full list of instruments. We include both blocks of moments in the econometric theory for completeness, but the first block of moments is sufficient to identify model (16).[18]

[15]Ideally, measurement error in Q can be addressed by "external" instruments outside of the model that are uncorrelated with the regression error term (exclusion condition) but correlated with investment opportunities (relevance condition). For a small group of firms, like a specific industry, it is possible to identify external movements of investment opportunities to construct such instruments, for example, by exploiting exogenous changes in industry competition or regulation. However, it is difficult to find external instruments for a large group of firms and industries. As a result, in the existing literature, researchers instead utilize "internal" variation to correct for measurement error in Q.

[16]Consistent with the first methodology, Kim and Kung (2017) empirically use lagged Q and cash flow as internal instruments by assuming that these regressors are exogenous.

[17]In contrast, integrating the high-order moment (cumulant) method into our approach and proving consistency is theoretically demanding. We instead include some numerical analyses in Section IV.C.2.

[18]In our empirical analysis, we use only the first block of moments to reduce the total number of instruments, and among this block, we further select a smaller set of instruments that are most relevant to Tobin's Q based on the method developed by Belloni, Chen, Chernozhukov, and Hansen (2012). The purpose of these efforts is to mitigate the concern of many-instrument bias in finite samples. Notably, this instrument selection is independent of selecting relevant expansion terms in the series-approximation step. For our empirical sample, estimating a regression of the residual (of the main regression) on the selected instruments yields a very small F-statistic of 2.32 and an extremely low R^2 of 0.01, indicating that the instruments are statistically uncorrelated with the error term. In contrast, the first-stage regression of Tobin's Q on these instruments yields an F-statistic value of 32.11 and an R^2 of 0.16, suggesting that the instruments are sufficiently relevant to Q according to empirical rule of thumbs. In addition, we

https://doi.org/10.1017/S0022109023000169 Published online by Cambridge University Press

Identification is based on the next two assumptions. We introduce the following useful notation: $\boldsymbol{a}_i = [a_{i,1}, \ldots, a_{i,T}]'$ and $\boldsymbol{a}_{i,-1} = [a_{i,0}, \ldots, a_{i,T-1}]'$ are two $T \times 1$ vectors for any variable a; $\mathbf{W}_i = [\mathbf{w}_{i,1}, \ldots, \mathbf{w}_{i,T}]$ is a $d_w \times T$ matrix, where d_w is the number of instrumental variables for uncertainty (see equation (6)); and 0_T denotes a $T \times 1$ vector of zeros.

Assumption 1. (i) $\left\{ \left(y_{i,t}, x_{i,t}, z_{i,t}, \mathbf{w}'_{i,t}, s_{i,t}, u_{i,t}, v_{i,t}, e_{i,t} \right) \right\}$ is IID in index i and has a finite second-order moment; (ii) $\mathbb{E}(\mathbf{u}_i) = 0_T$, $\mathbb{E}(\mathbf{e}_i) = 0_T$, and $\mathbb{E}(\mathbf{v}_i) = 0_T$, and $\mathbb{E}\left(\mathbf{u}_i \mathbf{u}'_i\right) = \Sigma_u$, $\mathbb{E}\left(\mathbf{e}_i \mathbf{e}'_i\right) = \Sigma_e$, and $\mathbb{E}\left(\mathbf{v}_i \mathbf{v}'_i\right) = \Sigma_v$ are all nonsingular matrices; (iii) $(\mathbf{u}_i, \mathbf{v}_i)$, $\mathbf{e}_{i,-1}$, and $\left[\mathbf{x}^*_{i,-1}, \mathbf{z}_{i,-1}, \mathbf{W}'_i\right]'$ are mutually independent; and (iv) $\mathbb{E}[f_0(z_{i,t-1})] = \mathbb{E}[g_0(s_{i,t-1})] = 0$.

Assumption 1(i) assumes cross-sectional independence (see Erickson et al. (2014) for a similar assumption).[19] The assumption of identical distribution in index i is nonessential and can be relaxed with more complex mathematical expositions. Assumption 1(ii) assumes that all error terms have zero means, finite variances, and potentially serial correlations. Assumption 1(iii) is standard in the literature on classical measurement error. This assumption can be relaxed to $\mathbf{e}_{i,-1}$ and $\left[\mathbf{x}^*_{i,-1}, \mathbf{z}_{i,-1}, \mathbf{W}'_i\right]'$ being mutually uncorrelated and $(\mathbf{u}_i, \mathbf{v}_i)$ being independent of $\left[\mathbf{e}_{i,-1}, \mathbf{x}^*_{i,-1}, \mathbf{z}_{i,-1}, \mathbf{W}'_i\right]'$. Assumption 1(iv) is for separately identifying the functions $f_0(\cdot)$ and $g_0(\cdot)$ and firm fixed effects μ_i.

Assumption 2. (i) There exists $\vartheta_0 = \left[\vartheta'_{z,0}, \vartheta'_{s,0}, \vartheta'_{v,0}\right]'$ such that

$$\sup_{z \in \mathcal{S}_z} \left| f_0(z) - \vartheta'_{z,0} \mathbf{P}^k(z) \right| \le M_1 k^{-\zeta},$$
$$\sup_{s \in \mathcal{S}_s} \left| g_0(z) - \vartheta'_{s,0} \mathbf{P}^k(s) \right| \le M_2 k^{-\zeta},$$
$$\sup_{v \in \mathcal{S}_v} \left| r_0(v) - \vartheta'_{v,0} \mathbf{P}^k(v) \right| \le M_3 k^{-\zeta},$$

for $M_1 > 0$, $M_2 > 0$, $M_3 > 0$, and $\zeta > 2$ for all k, where $\mathcal{S}_z$, $\mathcal{S}_s$, and $\mathcal{S}_v$ represent the support set for z, s, and v, respectively; (ii) the orthonormal basis functions $\{P_j(\cdot), j = 1, 2, \ldots\}$ satisfy $\max_{x \in R} \max_j \left| P_j(x) \right| \le M < \infty$.

conduct sensitivity analysis by dropping each of the four groups of instruments at a time (i.e., those regarding time dummies, cash flow, uncertainty, and control function, respectively), finding that our results are robust (see Appendix E of the Supplementary Material).

[19]In practice, the error term $\varepsilon_{i,t}$ in the main regression model (14) may exhibit cross-sectional dependence due to, for instance, industry-cluster effects that violate Assumption 1(i). This violation is innocuous if the industry effects are time-invariant, since these effects are canceled out by the demeaning technique used in equation (16). If the cross-sectional dependence is caused by time-varying industry effects, the standard errors in estimation may need to be adjusted by using, for example, the nonparametric kernel heteroscedasticity autocorrelation spatial correlation (HACSC) robust standard errors. Driscoll and Kraay (1998) and Vogelsang (2012) study HACSC covariance estimators with arbitrary cross-sectional dependence in parametric panel data models with fixed effects. However, these estimators are inefficient in our case as we can observe to which industry a firm belongs. Thus, it would be beneficial to modify their estimators to directly employ the observed industry information, for example, by extending the HACSC estimator for cross-sectional data in Bester, Conley, Hansen, and Vogelsang (2016) to panel data cases. We leave this extension to future research.

This assumption is standard in the literature on series approximation. Assumption 2(i) requires that $f_0(\cdot)$, $g_0(\cdot)$, and $r_0(\cdot)$ are at least twice differentiable so that they can be approximated by typical orthonormal basis functions, and the approximation error vanishes as the sample size nT increases, where the highest order k is required to increase with nT. Assumption 2(ii) instead imposes requirements on the choice of the orthonormal basis functions used in the series approximation. The Hermite series satisfies this condition.

B. Penalized GMM Estimator

Let $\boldsymbol{\theta}_0 = \left[\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{T-1}, \alpha_0, \vartheta'\right]'$ include all unknown parameters to be estimated in model (16) and $\tilde{\mathbf{d}}_i$ denote a $T \times m_{iv}$ matrix of instruments used to correct for the measurement error in Tobin's Q, where m_{iv} is the number of instruments. Then, the following m_{iv} theoretical orthogonal moment conditions exist:

$$\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\varepsilon}_i\right) \approx 0_{m_{iv}}, \tag{17}$$

where " $\approx$ " is due to series approximation. Based on (17), we construct the following empirical moments as a function of the unknown parameters $\boldsymbol{\theta}$:

$$\bar{\mathbf{G}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\tilde{\mathbf{d}}}_i' \tilde{\hat{\varepsilon}}_i,$$

where $\tilde{\hat{\boldsymbol{a}}}$ equals $\tilde{\boldsymbol{a}}$ with $v_{i,t}$ replaced by $\widehat{v}_{i,t}$ calculated from fitting model (6).

The traditional GMM estimation minimizes the following objective function and uses the minimizer as the estimate of $\boldsymbol{\theta}_0$:

$$Q_n(\boldsymbol{\theta}) = \bar{\mathbf{G}}_n(\boldsymbol{\theta})'{}_n\bar{\mathbf{G}}_n(\boldsymbol{\theta}), \tag{18}$$

where $_n$ is a symmetric and nonsingular weight matrix.

We instead propose a penalized GMM estimator to select which regressors in model (16) are most relevant and remove the irrelevant regressors. Our objective function is as follows:

$$Q_n(\boldsymbol{\theta};\psi) = \bar{\mathbf{G}}_n(\boldsymbol{\theta})'{}_n\bar{\mathbf{G}}_n(\boldsymbol{\theta}) + \sum_{l=1}^{p} p_c(|\theta_l|, \psi), \tag{19}$$

where $p_c(\cdot, \psi)$ is a nonnegative penalty function with $c > 1$ controlling its concavity and $\psi > 0$ as the tuning parameter; a larger value of ψ indicates stronger penalization; p is the number of regressors in model (16); and $|\theta_l|$, for $l = 1, 2, \ldots, p$, is the absolute value of the lth coefficient.[20] We refer to the minimizer, $\widehat{\boldsymbol{\theta}}$, of (19) as the *penalty estimator* of $\boldsymbol{\theta}_0$.

[20]The total number of regressors is $p = T + 3k$, where $k = k_{nT}$ (increasing with the sample size nT) is the highest order of the series approximation of each of $f_0(\cdot)$, $g_0(\cdot)$, and $r_0(\cdot)$. Specifically, there exist $(T-1)$ time dummies, plus Tobin's Q coefficient α_0, and the Hermite expansion coefficients of $f_0(z)$, $g_0(s)$, and $r_0(v)$, each containing k elements. We set $k = 4$ in the estimation based on our sample size and check robustness for $k = 6$ (see Appendix E of the Supplementary Material).

This penalty approach is based on the belief that there are only a few important regressors in model (16), whereas the other regressors are redundant and do not contribute to investment. By construction, the penalty term $p_c\left(|\widehat{\theta}_l|,\psi\right)$ will be positive if $\widehat{\theta}_l \neq 0$. Thus, including a redundant regressor whose true coefficient is zero but the estimate $\widehat{\theta}_l$ is nonzero increases the objective function (19) and is disfavored by the penalty estimator. This approach improves estimation efficiency because it does not need to estimate the coefficients of redundant regressors.

Based on penalty selection, we further propose a *post-penalty estimator* $\tilde{\boldsymbol{\theta}}$ by reestimating the model by the traditional GMM approach after removing the redundant regressors chosen by the penalty estimator:

$$\text{(20)} \qquad \min_{\boldsymbol{\theta}} \bar{\mathbf{G}}_n(\boldsymbol{\theta})'{}_n\bar{\mathbf{G}}_n(\boldsymbol{\theta})$$

$$\text{s.t. } \theta_j = 0 \,\text{if}\, j \notin \text{supp}\left(\widehat{\boldsymbol{\theta}}\right),$$

where $\text{supp}\left(\widehat{\boldsymbol{\theta}}\right) = \left\{l \in (1,2,\ldots,p) : \widehat{\theta}_l \neq 0\right\}$ is the set of nonredundant regressors. An advantage of the post-penalty estimator is that the standard errors of the coefficients can be conveniently calculated using common procedures.

We use the minimax concave penalty (MCP) developed in Zhang (2010) as the penalty function $p_c(\cdot,\psi)$, where the tuning parameter ψ is chosen optimally using the Bayesian information criterion (BIC) and cross-validation, and the constant c is set to 3 by convention.[21] Under this choice, we derive the limiting results of both the penalty and post-penalty estimators in Appendix B of the Supplementary Material and find that $\widehat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ are asymptotically equivalent, converging to the true parameter values $\boldsymbol{\theta}_0$ as the cross-sectional sample size $n \to \infty$.

IV. Data and Estimation

A. Data and Variables

We collect annual data on U.S. firms from Compustat.[22] Following Erickson et al. (2014), we measure investment ($y_{i,t}$) as capital expenditures divided by capital stock, that is, gross plant, property, and equipment at the beginning of year t. Cash flow ($z_{i,t}$) is the income before extraordinary items plus depreciation and amortization, divided by capital stock. Tobin's Q ($x_{i,t}$) equals the sum of short-term and long-term debts, plus the market value of equity, minus the current assets, divided by the capital stock.

[21]See equation (F.17) for the definition of the MCP penalty function. Compared with another commonly used penalty function (the least absolute shrinkage and selection operator developed by Tibshirani (1996)), the MCP penalty has the advantage of being asymptotically unbiased. Moreover, we combine the BIC and cross-validation approaches to choose the tuning parameter by adopting the one that gives a smaller number of selected regressors.

[22]We consider only U.S. firms with ordinary common shares listed on the NYSE, American Stock Exchange, and Nasdaq Stock Market. Given that it is reasonable to assume that listed firms in the United States are less financially constrained than private firms and firms in emerging markets, and that it is commonly believed that financially constrained firms are often more vulnerable to uncertainty shocks, our estimates are likely to provide a lower bound of the effect of uncertainty on investment.

https://doi.org/10.1017/S0022109023000169 Published online by Cambridge University Press

The measure of uncertainty ($s_{i,t}$) and its instrumental variables ($\mathbf{w}_{i,t}$ in equation (6)) are constructed following Alfaro et al. (2022). Specifically, firm i's uncertainty is measured by its individual stock volatility, which equals the annualized standard deviation of the firm's daily stock returns in year t. The stock-return data are obtained from the CRSP.

The instrumental variables are at the *industry* level and constructed based on the by-industry heterogeneous responses of individual stock volatility to 10 different sources of aggregate uncertainty shocks: oil prices, U.S. 10-year Treasury yields, U.S. policy uncertainty, and the exchange rates between the U.S. dollar and seven major currencies worldwide.[23] In total, there are 10 instrumental variables. The idea is that firms have differential exposure to aggregate uncertainty shocks, and thus their individual stock volatility responds differently to these shocks. For example, when oil prices become more volatile, the stock volatility of companies in oil-related industries can increase more than that of companies in oil-unrelated industries. We first estimate industry-specific exposure to oil-price movements through the asset-pricing model in equation (D.1) in Appendix D of the Supplementary Material. We then multiply the resulting industry-specific exposure by oil uncertainty shocks to obtain the industry-level oil instrument. The same approach applies to the construction of all the 10 instrumental variables, corresponding to the 10 aggregate uncertainty shocks, respectively. It is worth noting that we choose to work with industry-level, rather than firm-level, instruments because they are more likely to be exogenous to investment of individual firms. Estimating the first-stage regression (6) leads to an $R^2 = 0.35$ and F-stat = 248.2.[24]

Our empirical sample covers the post-Great Recession period from 2010 to 2017. We aim to avoid including the Great Recession in this study, as it might cause structural breaks. We remove firms in the financial, utility, and public sectors and firms with less than 200 daily CRSP stock returns in any given year. We trim our sample at the top and bottom 0.5% of each variable to remove outliers. We further remove firms with individual stock volatility greater than 150% per annum, firms with a cash-flow-to-capital ratio greater than 1.5 or smaller than −1.5 in any year, and firms changing their major industry (according to the 3-digit SIC code) during the sample period.[25] Ultimately, we obtain a balanced panel of

[23]These currencies include the Australian dollar, British pound, Canadian dollar, Euro, Japanese yen, Swedish krona, and Swiss franc.

[24]The data of aggregate uncertainty are obtained from Bloomberg, with the U.S. policy uncertainty as an exception, for which the data are obtained from https://www.policyuncertainty.com. See Appendix D of the Supplementary Material and Alfaro et al. (2018) for more details. By convention, the first-stage 2SLS regression also includes all exogenous and perfectly measured regressors under our assumptions (i.e., cash flow and year dummies).

[25]These steps remove 2.2%, 7.8%, and 5.7% of firms, respectively, from our sample. We drop these firms not because they are necessarily outliers, but believe that they are different from most firms included in the study. Our baseline results are robust to a more generous selection criterion which removes firms with individual stock volatility greater than 200% per annum, firms with a cash-flow-to-capital ratio greater than 2 or smaller than −2 in any year, and also to a stricter criterion that removes firms with individual stock volatility greater than 100% per annum, firms with a cash-flow-to-capital ratio greater than 1 or smaller than −1 in any year (see Appendix E of the Supplementary Material).

TABLE 1
Summary Statistics

Table 1 includes the summary statistics of our sample with 8,200 observations, where Std. Dev. is the standard deviation, and p10, p25, p50, p75, and p90 represent the 10th, 25th, 50th, 75th, and 90th percentile values, respectively. All numbers are in decimal form.

	Mean	Std. Dev.	Skewness	p10	p25	p50	p75	p90
y: Investment/capital stock	0.10	0.09	3.96	0.03	0.05	0.08	0.13	0.20
x: Tobin's Q	3.29	5.35	5.79	0.17	0.63	1.54	3.89	8.28
z: Cash flow/capital stock	0.23	0.30	0.16	−0.01	0.09	0.18	0.35	0.61
s: Individual stock volatility	0.39	0.16	1.48	0.22	0.27	0.35	0.47	0.60

1,025 firms and 8,200 firm-year observations, and Table 1 presents the summary statistics.[26]

A remark is worth noting. Our estimator should better be considered for short panels if researchers are unsure about the persistence of regressors. The estimator can be applied to long panels only if all the variables are stationary. Some previous articles have considered longer sample periods, such as starting from 1970s or 1980s. Our data actually allow us to investigate an extended sample period between 1986 and 2017, because the U.S. policy uncertainty data date back to 1985. However, we find that in this extended sample, our key variables (investment, Tobin's Q, cash flow, and volatility) are likely to have unit roots for more than two-thirds of firms (see Table E.2 in Appendix E of the Supplementary Material for unit-root test results). Such time-series properties can play a significant role in a long panel of 32 years, though innocuous if the time period is short. Consequently, our proposed estimator and tests are no longer valid for long samples because the methodology has not accounted for the uncertain degree of persistence of variables. We leave extending our work to account for potential unit roots for future research.

B. Baseline Results

The estimates of the investment–uncertainty relation, $\widehat{g}(s)$, and its first-order derivative, $d\widehat{g}(s)/ds$, are plotted in Figure 1. As shown in Graph A, this relation is decreasing and concave in a majority range of volatility. Only when volatility is at high levels is the curve slightly convex. These patterns can be observed more clearly in Graph B, where the first-order derivative is significantly below 0 and decreases with volatility for most of the sample. The economic magnitude of the concavity is also large in that increasing the volatility from the bottom to top decile more than doubles the marginal impact on investment (from -0.03 to -0.07).[27]

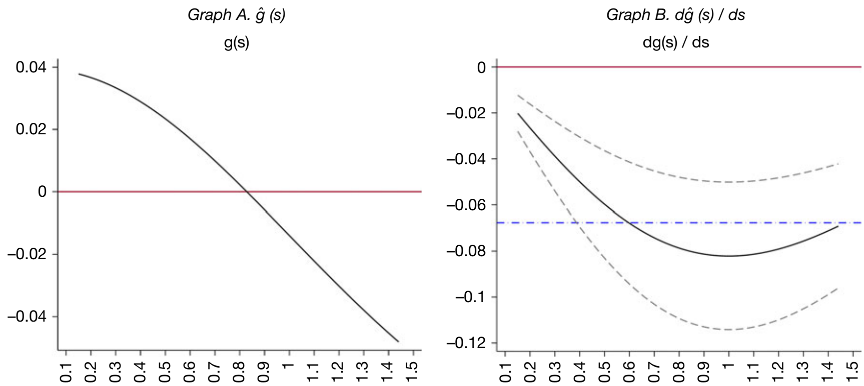
[26]We choose to study a balanced panel since firms with entry or exit during an 8-year period may have differential behavior from firms that stay active in this period. Nevertheless, our econometric theory also applies to unbalanced panels with more complicated mathematical notation, and our baseline results are robust to the following two unbalanced samples: i) an unbalanced panel between 2010 and 2017 that includes 1,598 firms and 11,324 observations and ii) an unbalanced panel between 2004 and 2017 that includes 2,427 firms and 22,122 observations (see Appendix E of the Supplementary Material).

[27]To put these numbers into perspective, a marginal effect equal to -0.07 indicates that a 1-standard-deviation increase in the individual stock volatility *reduces* the investment-to-capital rate by $0.07 \times 0.16 = 0.0112$. That is, the investment decreases by an amount equal to 1.12% of the capital stock.

FIGURE 1

Investment–Uncertainty Relation

Graph A of Figure 1 plots the estimate of the investment–uncertainty relation $\widehat{g}(s)$, where, for identification purposes, its sample mean is normalized to 0. Graph B plots its first-order derivative $d\widehat{g}(s)/ds$, where the dashed lines indicate the 95% confidence intervals. The horizontal dash-dotted line represents the linear estimate $\widehat{\gamma}$. All numbers are in decimal form.



In addition, when volatility is over 100% per annum, the first-order derivative starts to increase with volatility, suggesting that the investment–uncertainty relation is convex. However, this region only accounts for less than 1% of the sample.

We construct a joint Wald test for the null hypothesis of $g_0(s)=0$ for all s to formally test whether uncertainty has a meaningful impact on investment. The test significantly rejects the null with a p-value $=0.00$. The result implies that overall, the curve $g_0(s)$ is significantly different from 0, so uncertainty is a crucial determinant of investment.

A pointwise t-test is further introduced to test for the null hypothesis that $dg_0(s)/ds$ is constant. This test examines whether the investment–uncertainty relation can be characterized by a linear function $g_0(s)=\gamma_0 s$, thereby shedding light on the validity of linear regressions that are commonplace in the literature. The test procedure roughly consists of three steps: i) obtain a linear estimate $\widehat{\gamma}$; ii) compare $\widehat{\gamma}$ with the nonparametric estimates $d\widehat{g}(s)/ds$; and iii) reject the linear model if $\widehat{\gamma}$ is outside the 95% confidence intervals of the nonparametric estimates. The test results are shown in Graph B of Figure 1, where the horizontal dash-dotted line represents the linear estimate $\widehat{\gamma}$. The test rejects the linear model when volatility is below approximately 40% per annum. According to Table 1, this region accounts for more than half of the sample, in which the linear coefficient significantly overestimates the negative impact of uncertainty on investment. In contrast, when volatility exceeds this threshold, the linear model provides a good approximation of the investment–uncertainty relation, as the linear estimate falls within the 95% confidence intervals of the nonparametric estimates.[28]

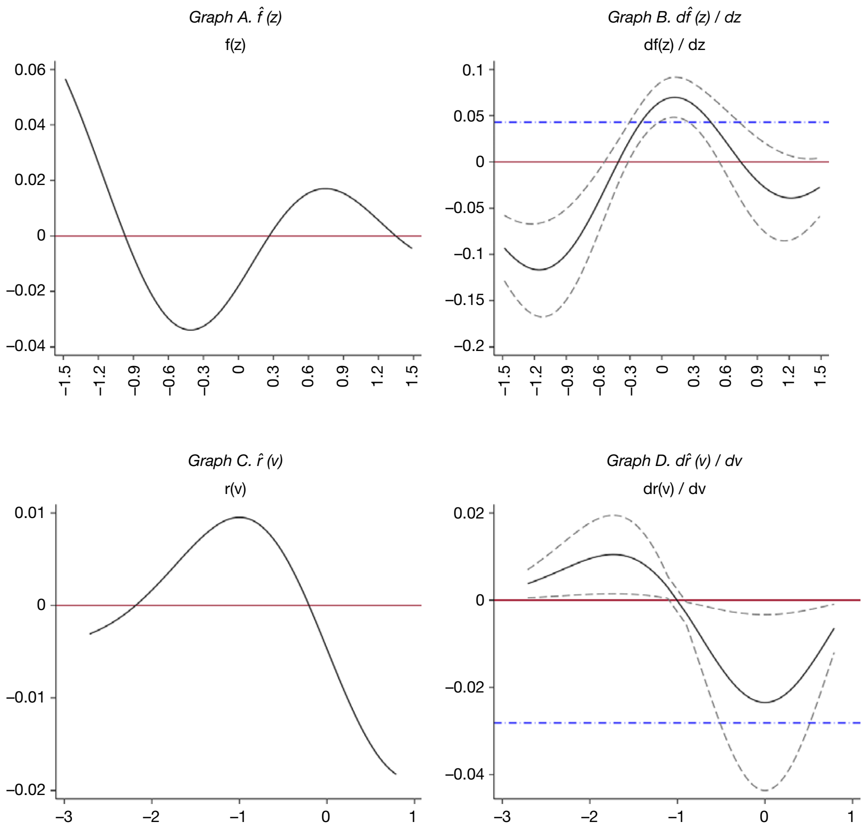
Moreover, Tobin's Q coefficient α_0 is estimated to be 0.014 and highly significant with a standard error of 0.002. In addition, Graphs A and B of Figure 2 plot the estimates of the investment–cash flow relation $\widehat{f}(z)$ and its first

[28]See Appendix C of the Supplementary Material for technical details of the two tests.

FIGURE 2

Investment–Cash Flow Relation and Control Function

Graphs A and C of Figure 2 plot the estimates of the investment–cash flow relation $\hat{f}(z)$ and the control function $\hat{r}(v)$, respectively, where, for identification purposes, their sample means are normalized to 0. Graphs B and D plot their first-order derivatives, where the dashed lines indicate the 95% confidence intervals. The horizontal dash-dotted line represents the corresponding linear estimates. All numbers are in decimal form.



derivative, showing that this relation is statistically significant and strongly non-linear. The joint Wald test rejects $f_0(z)=0$ with a p-value $=0.00$, and the pointwise t-test rejects linearity for 63% of the sample. The results demonstrate that cash flow contributes to investment in addition to Tobin's Q and volatility and therefore contrasts with neoclassical theory in which Q is a sufficient statistic for investment.[29] Furthermore, the investment–cash flow relation is first decreasing, then

[29]Our empirical results are also different from the results in Erickson and Whited (2000), and more recently Erickson et al. (2014), where cash flow does not contribute to investment once the measurement error in Tobin's Q is treated. The different results are likely due to differences in empirical samples rather than model specifications or estimation methods. To illustrate this point, we estimate the same linear regression (1) using our data by a linear version of our proposed estimator (i.e., without series approximation) and by the high-order cumulant estimator in Erickson et al. (2014), respectively (see Appendix E of the Supplementary Material). We find that the estimation results are close to each other and in both estimates, cash flow contributes to investment after treating the measurement error in Tobin's Q.

increasing, and finally flat (as the first-order derivative is insignificant from 0). This shape is consistent with the theoretical results in Cleary et al. (2007). Previous empirical studies have documented similar U-shapes by estimating linear regressions within negative and positive cash flow subsamples, finding that the linear slope is negative in the negative cash flow region and positive in the positive cash flow region.[30] Our nonparametric estimation does not need to impose such predetermined sample splits and further finds that the cash flow threshold at which the investment–cash flow dependence shifts from negative to positive should not be 0 but instead -0.5.

Graphs C and D of Figure 2 plot the estimated control function $\widehat{r}(v)$ and its first-order derivative. Most notably, the joint Wald test rejects the null of $r_0(v)=0$ with a p-value = 0.02.[31] This test result confirms that stock volatility is indeed an endogenous variable in model (4).

C. Further Discussions

1. Ignoring Measurement Error or Endogeneity

In econometric theory, neglecting the measurement error in Q or endogeneity of uncertainty can lead to inconsistent estimates. However, the magnitudes and directions of the bias rely on data in complex ways and are difficult to know a priori. We empirically examine the bias by reestimating our model and leaving either the mismeasurement or endogeneity issue untreated. Then, we compare the resulting biased estimates with the previous baseline estimates to evaluate the bias corresponding to the untreated issue.[32]

In Graph A of Figure 3, we compare the baseline estimates of the marginal effects of uncertainty on investment (solid line) to the estimates where measurement error in Q is untreated (dashed-dotted line). This comparison shows that measurement error bias can lead to misinterpretation of the shape of the investment–uncertainty relation. Specifically, the dash-dotted curve increases with volatility for most of our sample (see the kernel distribution of volatility from the dotted curve (using the right vertical axis)). That is, in the estimates with untreated measurement error, the investment–uncertainty relation is largely convex as opposed to concave. Moreover, ignoring measurement error leads to a significant overestimation of the negative impact of uncertainty on investment for over half of the observations. This

[30]See, among others, Bhagat et al. (2005), Cleary et al. (2007), and Firth et al. (2012). These papers also rely on quadratic and spline regressions to account for nonlinearity, but none of them have considered nonparametric estimation.

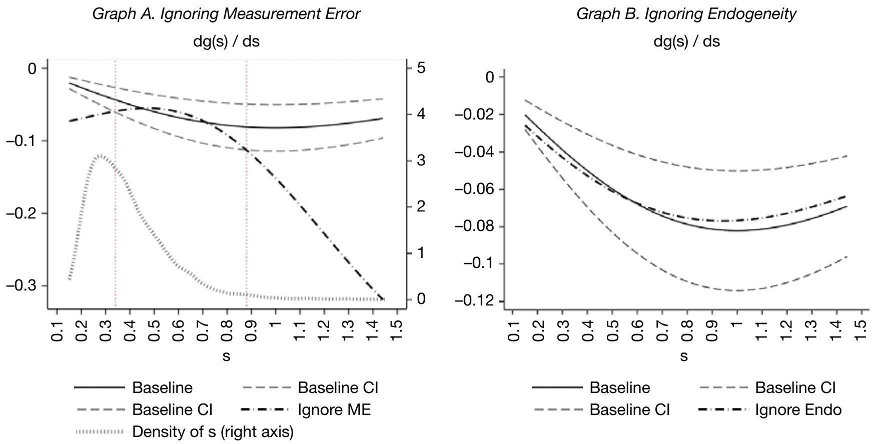
[31]The pointwise t-test rejects the linearity for only 0.4% of the observations, indicating that $r_0(v)$ is largely linear. Notably, this test result does not contradict the results shown in Graph D of Figure 2 because v is close to 0 for most observations, where the linear estimate (i.e., the horizontal dashed line) is within the 95% confidence interval of the nonparametric estimate.

[32]Note that in both cases, the moment conditions (17) still apply, and the same GMM approach can be used to estimate the model. In the case when the measurement error is ignored, the least squares method is more efficient than the GMM method, although both are consistent, because without measurement error, model (13) is a linear regression with all its variables being exogenous. We therefore adopt the least squares method to produce the Graph A of Figure 3 to achieve greater efficiency. Moreover, in the case when endogeneity of volatility is ignored, the terms of the control function are dropped, and the model is estimated by GMM to obtain the Graph B of Figure 3.

FIGURE 3

Ignoring Measurement Error or Endogeneity

Graph A of Figure 3 compares the marginal effects of uncertainty on investment, $d\widehat{g}(s)/ds$, between the baseline estimates and the estimates in which measurement error in Q is untreated. The solid line represents the baseline estimates, and the two dashed lines indicate the 95% confidence intervals. The dash-dotted line indicates the estimates where measurement error in Q is ignored. The two vertical dotted lines are at $s = 0.34$ and $s = 0.88$, respectively. The dotted curve represents the kernel density of s within our sample, where the values are on the right vertical axis. Graph B instead compares the baseline estimates and the estimates where the endogeneity of uncertainty is neglected. All numbers are in decimal form.



is when volatility is within the low-to-medium region ($s < 0.34$) and when volatility is high ($s > 0.88$), shown by the dash-dotted curve lying outside the 95% confidence intervals of the baseline estimates (dashed lines).

In Graph B of Figure 3, we compare the baseline estimates with the estimates where the endogeneity of uncertainty is neglected. The two curves are close to each other, indicating that the bias is small. Ignoring the endogeneity issue leads to slight overestimation of the impact of uncertainty on investment when uncertainty is low and a slight underestimation of the impact when uncertainty is high. However, since the dash-dotted curve lies within the 95% confidence intervals of the baseline estimates, the two estimates are statistically indifferent.

A crucial remark is worth mentioning here. The results that measurement error in Q can introduce a substantial bias into estimation but the endogeneity of uncertainty cannot are specific to our empirical sample. Our purpose is not to overly stress the role of measurement error, or to downplay the role of regressor endogeneity, but instead to provide a methodology to directly and jointly address both issues. Given the complex nature of the bias due to the two problems in data, such a method is necessary to obtain reliable empirical estimates.

2. High-Order Cumulant Estimator

Erickson and Whited ((2000), (2002), (2012)) introduce a high-order moment estimator to treat measurement error in Q. Based on the cumulant equations given in Geary (1941), Erickson et al. (2014) further develop a cumulant estimator that is more computationally efficient. In their setup, the regression model is linear, and all accurately measured regressors are exogenous. As a result, the high-order cumulant (or moment) estimator is not applicable to our model (4), since individual stock volatility is an endogenous variable and the regression features unknown nonlinearities.

However, after we address the endogeneity issue via the control function approach and the nonlinearities through penalized series approximation, our model (14) essentially satisfies the identification assumptions in Erickson et al. (2014), so their estimator becomes applicable. Even though deriving the limiting results of this *augmented high-order cumulant estimator* is theoretically demanding and beyond the scope of this article, we find that numerically, it yields estimation results similar to our baseline estimates. Specifically, we adopt the high-order cumulant method to estimate model (14), in which the series approximation terms in $\mathbf{E}_{i,t}$ are selected by our penalty estimator (i.e., equation (19)). The results are compared to the baseline estimates in Figure 4, where Graph A shows the level relation $\widehat{g}(s)$, and Graph B shows the marginal impact function $d\widehat{g}(s)/ds$. From this comparison, we observe that the estimates of marginal effects of uncertainty on investment obtained from the augmented high-order cumulant method (dash-dotted line) are slightly stronger but still lie in the 95% confidence intervals of the baseline estimates, and the two estimates of the level relation are also close to each other.

3. Robustness

The baseline results on the investment–uncertainty relation remain largely intact in the following robustness checks:[33] i) using pre-sample data (between 1998 and 2009) to construct the instrumental variables, $\mathbf{w}_{i,t}$, for individual stock volatility to ensure that they are exogenous to investment; ii) controlling for by-industry exposure to the *first*-moment movements of the aggregate uncertainty shocks that are used to construct $\mathbf{w}_{i,t}$ (Alfaro et al. (2022)); iii) including control variables as in Panousi and Papanikolaou (2012) and Kim and Kung (2017);[34] iv) estimating an unbalanced panel between 2010 and 2017 (1,598 firms and 11,324 observations) and a longer unbalanced panel between 2004 and 2017 (2,427 firms and 22,122 observations); v) replacing the realized stock volatility with the EGARCH-inferred volatility (Li, Magud, and Valencia (2020));[35] vi) considering the realized *idiosyncratic* volatility (Panousi and Papanikolaou (2012));[36] vii) replacing the cash flow variable with the measure constructed in Lewellen and Lewellen (2016);

[33]The results are shown in Appendix E of the Supplementary Material.

[34]Sales growth, stock returns, book leverage, and size of capital. See Appendix E of the Supplementary Material for more discussions.

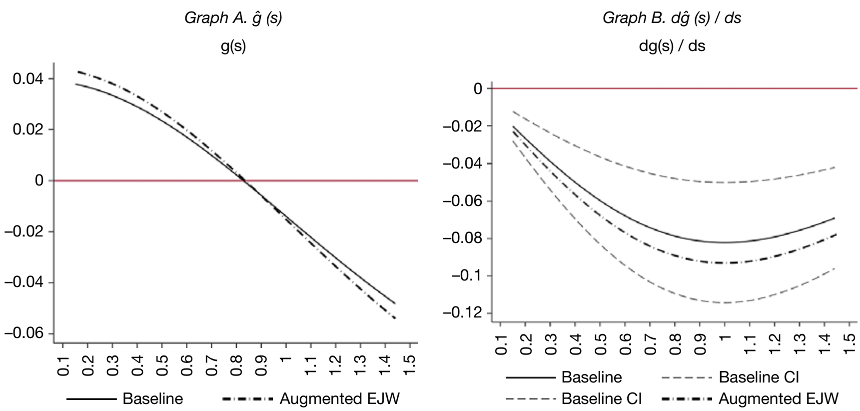
[35]We fit an EGARCH(p,q) model for each firm in each year using daily stock returns, where we choose the optimal lags based on the Akaike information criterion by setting the maximum number of lags to be 4 for both p and Q. The annual EGARCH volatility is calculated as the squared root of the sum of daily EGARCH-inferred variance divided by the number of days with non-missing data and multiplied by 252.

[36]The idiosyncratic stock returns are calculated based on the CAPM, where the exposure to market risk is partialled out. The first-stage regression of this idiosyncratic volatility on our instruments leads to an F-statistic = 168 and $R^2 = 0.27$. Following the usual convention, these high statistical values indicate that our instruments are relevant to the idiosyncratic volatility defined in this way. Moreover, because the CAPM-based idiosyncratic volatility does not contain market risk but can include both industry-level and firm-level risks, we expect it to be correlated with our industry-level instruments also from an economic perspective. In contrast, in Panousi and Papanikolaou (2012), the idiosyncratic volatility is defined in an alternative way. They have partialled out not only the exposure to market risk but also the exposure to industry risk. Then, we believe that our industry-level instruments can be less relevant.

FIGURE 4

Augmented Higher-Order Cumulant Estimates

Graph A of Figure 4 compares $\widehat{g}(s)$ between the baseline estimates and the estimates obtained from the high-order cumulant estimator developed by Erickson et al. (2014), with the latter augmented by the control function approach, series approximation, and penalty selection. The solid line shows the baseline estimates, and the dash-dotted line indicates the augmented high-order cumulant estimates. Graph B instead compares the estimates of $d\widehat{g}(s)/ds$ obtained from the two methods, where, additionally, the dashed lines indicate the 95% confidence intervals of the baseline estimates. All numbers are in decimal form.



viii) including R&D expenses as a part of the investment and cash flow (Chen, Goldstein, and Jiang (2007)); ix) replacing f_0 and r_0 with linear functions; x) dropping each of the four groups of instruments for Tobin's Q at a time (i.e., those regarding time dummies, cash flow, volatility, and control function, respectively); and xi) adopting the granular instrumental variables of Gabaix and Koijen (2020) for individual stock volatility.

V. Transmission Channels

Why does uncertainty affect investment beyond Q and cash flow? Theoretically, multiple channels exist: Heightened uncertainty may reduce investment due to the real option effect or by increasing risk premiums in the firm's borrowing; it may also enhance investment because of the overinvestment effect, the good news principle, or the Oi–Hartman–Abel effect.[37]

Understanding which effects drive the estimation results is of empirical interest. Among these mechanisms, the Oi–Hartman–Abel effect is less relevant to our results since typically, it is not very strong in the short run (Bloom (2014)). To investigate the remaining four channels, we divide our sample based on capital irreversibility, financial vulnerability, and size and further disentangle individual stock volatility into positive and negative semivariances to reestimate our model. We focus on the differences in the estimated investment–uncertainty relations across these specifications and find evidence that supports the coexistence of all four channels. Some of the evidence is difficult to obtain from linear regressions.

[37]See Bloom (2014) for a recent survey of these channels and the references therein; see Eisdorfer (2008) for the overinvestment channel.

A. Capital Irreversibility

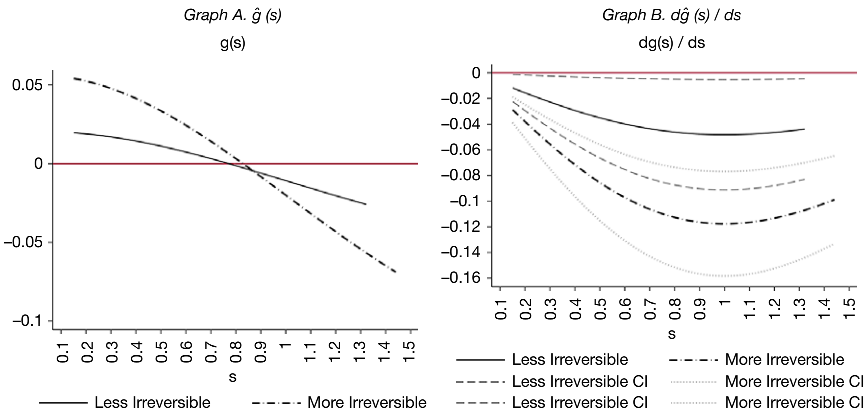
The real option theory predicts that firms investing in more irreversible capital reduce such investment to a greater extent in response to high uncertainty because irreversibility enhances the value of "wait-and-see" (Dixit and Pindyck (1994)). We test this hypothesis by dividing our sample firms into low-irreversibility versus high-irreversibility groups according to the industry-level index of capital resalability constructed in Balasubramanian and Sivadasan (2009) and Chirinko and Schaller (2009). Specifically, we collect data on industry-level total capital expenditures and expenditures on used capital from the Annual Capital Expenditures Survey (ACES) of the U.S. Census Bureau from 1994 to 2017. We use the time average of the survey results to construct an index of resalability, which is the share of used capital in total capital expenditures. A firm is assigned to the low-irreversibility group if its resalability index is *higher* than the median, and vice versa.[38]

The subsample estimates are compared in Figure 5. As shown in Graph A, the investment–uncertainty relations obtained from both subsamples are decreasing and concave, but the relation in the high-irreversibility subsample (dash-dotted line)

FIGURE 5

Subsample Estimates: Low Versus High Irreversibility

Graphs A and B of Figure 5 compare $\widehat{g}(s)$ and $d\widehat{g}(s)/ds$, respectively, between the estimates obtained from the low-irreversibility subsample and those from the high-irreversibility subsample. The degree of irreversibility is measured by an industry-level index of capital resalability constructed based on the share of used capital in aggregate industry capital expenditures, and the data are obtained from the ACES of the U.S. Census Bureau from 1994 to 2017 (Balasubramanian and Sivadasan (2009), Chirinko and Schaller (2009)). A firm is assigned to the low-irreversibility subsample if its resalability index is *higher* than the sample median and vice versa. The solid line represents the estimates obtained from the low-irreversibility subsample, and the two dashed lines indicate the 95% confidence intervals. The dash-dotted line represents the estimates obtained from the high-irreversibility subsample, whereas the two dotted lines represent the 95% confidence intervals. All numbers are in decimal form.



[38]We start from 1994 when the first survey is available, and we end in 2017 because this is the last year of our sample. The final resalability index varies by industry but not across time, and therefore we treat it as exogenous to individual firms' investment. The results in Figure 5 are also robust to the firm-level scale inflexibility index calculated from quarterly data in our sample period, following Gu, Hackbarth, and Johnson (2018) (see Appendix E of the Supplementary Material).

is steeper than that in the low-irreversibility subsample (solid line). This finding can be seen more clearly in Graph B. The marginal effects of uncertainty on investment are negative in both subsamples; however, the curve estimated for the high-irreversibility subsample is below that estimated for the low-irreversibility subsample. The two curves lie outside the 95% confidence intervals of each other, indicating that the differences between the two estimates are statistically significant. These results are consistent with the predictions from the real option channel.

B. Financial Distress or Constraints

The impact of financial distress on how uncertainty influences investment is theoretically mixed. First, the literature on default costs illustrates that high uncertainty can increase the likelihood of default by expanding the size of the left-tail default outcomes. Therefore, it raises the default-risk premium and the expected deadweight loss of bankruptcy (Bernanke et al. (1999)). Consequently, heightened uncertainty increases firms' borrowing costs and decreases investment. This effect is stronger when firms are closer to financial distress because an additional increase in the default probability can lead to more drastic changes in borrowing costs. Second, Eisdorfer (2008) documents an overinvestment channel that suggests the opposite. Given that greater uncertainty benefits shareholders of distressed firms at the expense of debtholders, it provides an incentive for managers who represent shareholders' interests to overinvest. Consequently, the negative relation between uncertainty and investment may be weakened or reversed among financially distressed firms.

We examine the roles of financial distress by dividing our sample according to *Z*-scores following Eisdorfer (2008). To mitigate the endogeneity concern that *Z*-scores may rely on investment, we use each firm's average *Z*-score over time throughout the sample period. A firm is assigned to the financially distressed group if its average *Z*-score is below 1.81, and vice versa.[39] The estimates obtained from the undistressed and distressed groups are compared in Figure 6. We observe that the estimates in the undistressed subsample (solid line) are similar to the baseline estimates, but those in the distressed subsample (dash-dotted line) are different. In Graph A, the investment–uncertainty relation in the distressed group displays a U-shape: Investment decreases with uncertainty for most observations but increases with uncertainty when uncertainty is high. In Graph B, the estimated marginal effects of uncertainty on investment for the distressed group increase with the level of uncertainty. These marginal effects are also more negative, lying outside the 95% confidence interval of the undistressed group, when the firm's stock volatility is lower than 60% per annum. Furthermore, when the stock volatility reaches 100% per year, the marginal effects of uncertainty estimated in the distressed group become positive.

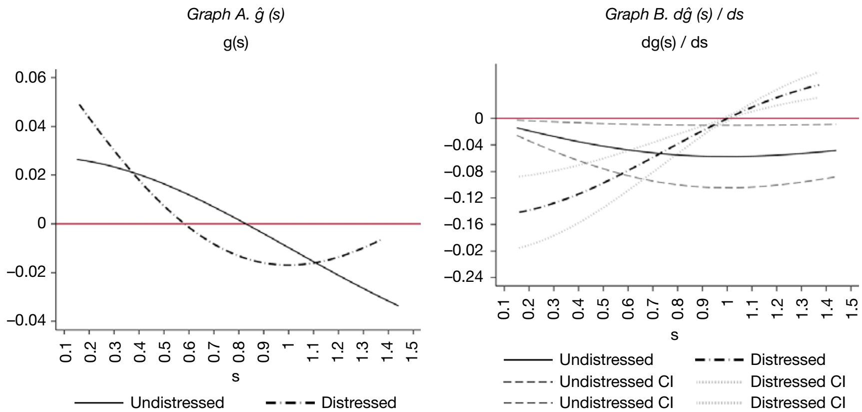
We interpret these results as evidence of the coexistence of the risk premium channel and the overinvestment channel. On the one hand, if uncertainty is not

[39]The definition of the *Z*-score and the criterion of 1.81 follow Eisdorfer (2008). The *Z*-score = 1.2 (working capital/total assets) + 1.4 (retained earnings/total assets) + 3.3 (earnings before interest and taxes/total assets) + 0.6 (market value of equity/book value of total liabilities) + 0.999 (sales/total assets). Approximately 17% of firms in our sample are assigned to the distressed group.

FIGURE 6

Subsample Estimates: Undistressed Versus Distressed

Graphs A and B of Figure 6 compare $\hat{g}(s)$ and $d\hat{g}(s)/ds$, respectively, between the estimates obtained for the undistressed group and those obtained for the distressed group. A firm is defined as financially distressed if the time average of its *Z*-scores (between 2010 and 2017) is below 1.81 following Eisdorfer (2008), and vice versa. The solid line represents the estimates obtained for the undistressed group, and the two dashed lines indicate the 95% confidence intervals. The dash-dotted lines are the estimates obtained for the distressed group, whereas the two dotted lines are the 95% confidence intervals. All numbers are in decimal form.



extremely high, then distressed firms reduce investment more aggressively under greater uncertainty than firms that are not distressed, supporting the risk premium channel. On the other hand, if uncertainty is sufficiently high, then investment increases with uncertainty for distressed firms because managers overinvest to transfer risk to debtholders. The two countervailing forces together produce a U-shaped investment–uncertainty relation.

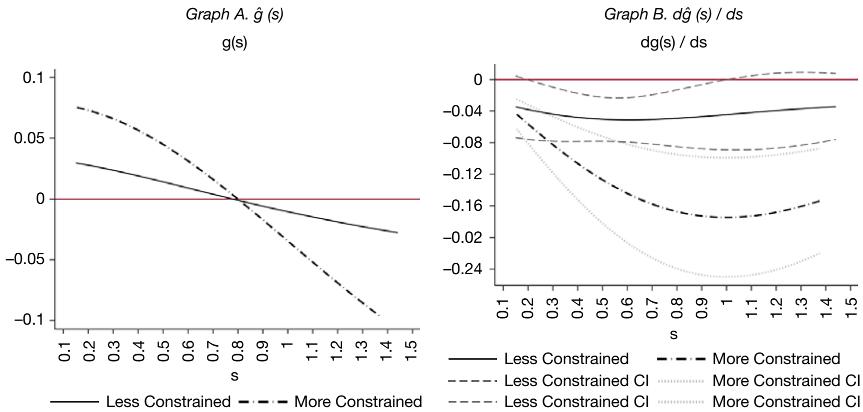
Nonparametric estimation is also important to this comparison. If we estimate linear regressions instead, then the coefficients of uncertainty are -0.05 and -0.11 in the undistressed and distressed subsamples, respectively. The linear regression results support the risk premium channel but cannot detect the overinvestment channel.

The above analysis is based on the assumption that the average *Z*-score over time throughout the study period is exogenous to investment. To further mitigate endogeneity concerns, we instead use the average *Z*-score over time in the *pre-sample* period from 1999 to 2009. However, using this approach, we need to change the interpretations of the resulting subsamples because a firm that was on average in financial distress between 1999 and 2009 was not necessarily in financial distress between 2010 and 2017. Therefore, this sample split is not suitable for testing the overinvestment channel. However, it can be reasonable to assume that firms experiencing financial distress during the pre-sample period would be "more financially constrained" during the sample period than firms that did not have that experience. This is because financial distress can hurt a firm's reputation, customer relationships, and human capital, and thus persistently damage their borrowing capacities. Lenders can also be more cautious in lending money to firms that were in financial distress in the recent past and therefore

FIGURE 7

Pre-Sample *Z*-Scores: Less Versus More Constrained

Graphs A and B of Figure 7 compare $\widehat{g}(s)$ and $d\widehat{g}(s)/ds$, respectively, between the estimates obtained for the less financially constrained subsample and those obtained for the more constrained subsample. A firm is defined as more constrained if the pre-sample average of its *Z*-scores (between 1999 and 2009) is below 1.81, that is, the firm was on average in financial distress prior to our sample period. The solid line shows the estimates obtained for the less constrained subsample, and the two dashed lines indicate the 95% confidence intervals. The dash-dotted line represents the estimates obtained for the more constrained subsample, whereas the two dotted lines are the 95% confidence intervals. All numbers are in decimal form.



request higher risk premiums.[40] According to the risk premium channel, the negative effects of uncertainty on investment in the more constrained group should be stronger than those in the less constrained group. As shown in Figure 7, this situation holds: For the more constrained group, in Graph A, the investment–uncertainty relation (dash-dotted line) is steeper, and in Graph B, the estimated marginal effects of uncertainty are further below 0. Moreover, for more than half of the observations, the estimates of the marginal effects of uncertainty obtained for the two groups are outside the 95% confidence intervals of each other, indicating that the differences are statistically significant.

C. Double Sorting: Irreversibility and Financial Constraints

To the extent that the real option channel and the risk premium channel can coexist, do they reinforce or undermine the impact of each other? We examine this question by double sorting our sample into four subsamples according to the resalability index and the criterion of financial constraints based on the pre-sample average *Z*-scores.

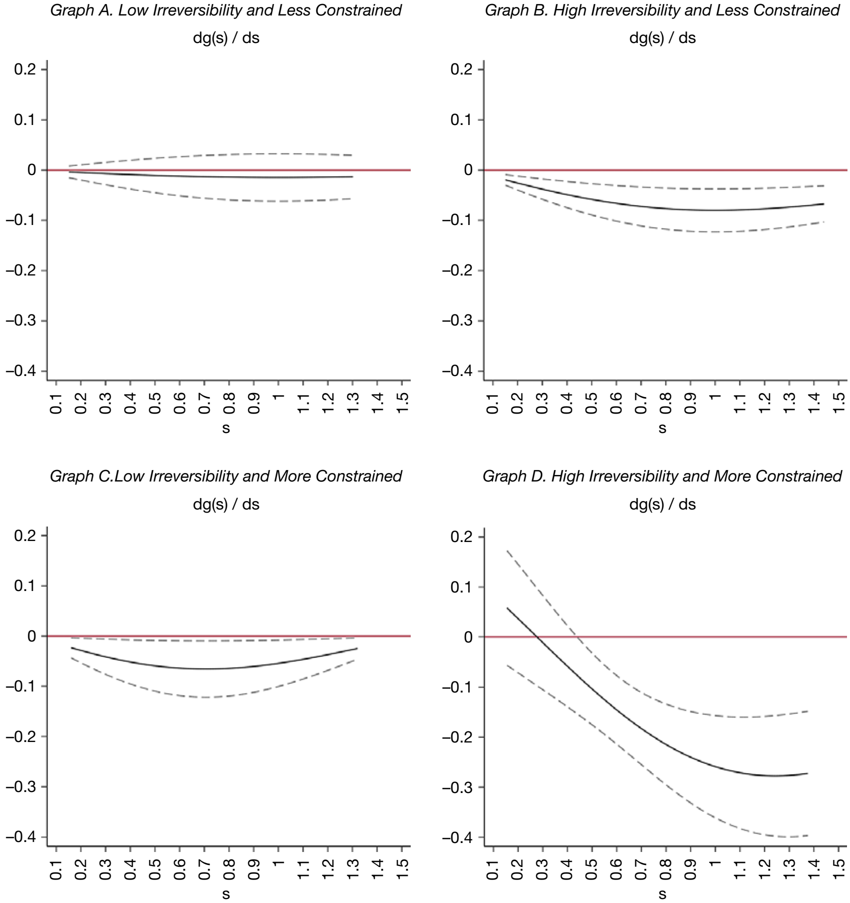
The estimates of the marginal effects of uncertainty on investment obtained in these four cases are compared in Figure 8. As shown in Graph A, the marginal effects are insignificant from 0 if firms are less financially constrained and have low

[40]An advantage of this backward-looking measure is that it is exogenous to the firm's current investment. From this perspective, this measure can be better than some commonly adopted indicators in the literature, such as the Cleary index, Kaplan–Zingales index, and Whited–Wu index. In fact, using a simulation-based method, Hennessy and Whited (2007) note that none of the latter three indexes is a good proxy for financial constraints due to endogeneity, but instead, firm size is a robust measure. We discuss the effects of firm size in Section V.D.

FIGURE 8

Double Sorting: Irreversibility and Financial Constraints

Figure 8 compares the estimates of $d\widehat{g}(s)/ds$ obtained for four subsamples as follows: Graph A presents the subsample of firms that have low irreversibility and are financially less constrained, Graph B presents the subsample of firms that have high irreversibility and are less constrained, Graph C presents the subsample of firms that have low irreversibility and are more constrained, and Graph D presents the subsample of firms that have high irreversibility and are more constrained. The degree of irreversibility is measured by the industry-level index of capital resalability as shown in Figure 5, and the degree of financial constraints is defined by the criterion as shown in Figure 7. All y-axes have the same scale, and all numbers are in decimal form.



capital irreversibility. In Graphs B and C, when firms either face high irreversibility or are more financially constrained, but not both, the marginal effects are significantly negative, and the magnitudes of the marginal effects in these two cases are similar. These results support the coexistence of the real option effect and the risk premium effect and suggest that the two isolated effects may be of a similar magnitude. In addition, when firms face tighter financial constraints and high capital irreversibility simultaneously, the negative effects of uncertainty on investment are the strongest, as shown in Graph D.[41] This finding indicates that the real

[41]Note that the subsample in Graph D includes only 90 firms. Thus, the results need to be interpreted with caution, as they can be sensitive to the inclusion of some particular firms.

https://doi.org/10.1017/S0022109023000169 Published online by Cambridge University Press

option channel and the risk premium channel may strengthen each other in reducing firms' investment in response to heightened uncertainty.

D. Firm Size

Next, we examine whether the investment of large firms and that of small firms are influenced differently by uncertainty. Firm size matters because, on the one hand, it reflects the degree of financial constraints (small firms are more constrained due to the more problematic asymmetric information; Hennessy and Whited (2007)). Following this logic, the investment of small firms should be more sensitive to uncertainty, according to the risk premium channel. On the other hand, in the industrial organization literature, large firms are often viewed as having more market power (Carlton, Perloff, and van 't Veld (1990)), which can intensify the real option effect, as documented by Caballero (1991). This mechanism leads to the opposite prediction that the investment of large firms is more responsive to uncertainty. Therefore, by dividing the sample based on firm size, we jointly test the following two predictions:

Prediction 1. The investment of large firms is affected more by uncertainty if size mainly represents market power AND the real option channel is at work.

Prediction 2. The investment of small firms is affected more by uncertainty if size mainly represents financial constraints AND the risk premium channel is at work.

We measure firm size by the book value of total assets and use the pre-sample average between 1999 and 2009 to mitigate endogeneity concerns. A firm is classified as a large firm if its pre-sample average of total assets is among the top 50% of all firms within our sample, and vice versa. The estimates obtained for the "large" and "small" subsamples are compared in Figure 9. We can observe that the negative effects of uncertainty on investment are actually stronger in the group of large firms given that the investment–uncertainty relation (solid line) is steeper in Graph A. Additionally, in Graph B, the estimates of the marginal effects of uncertainty among large firms are more negative than those among small firms, and the differences between the two groups of estimates are statistically significant. This empirical evidence supports *prediction 1* above; that is, size mainly measures market power, and the real option effect leads to the investment of larger firms reacting to heightened uncertainty more aggressively.

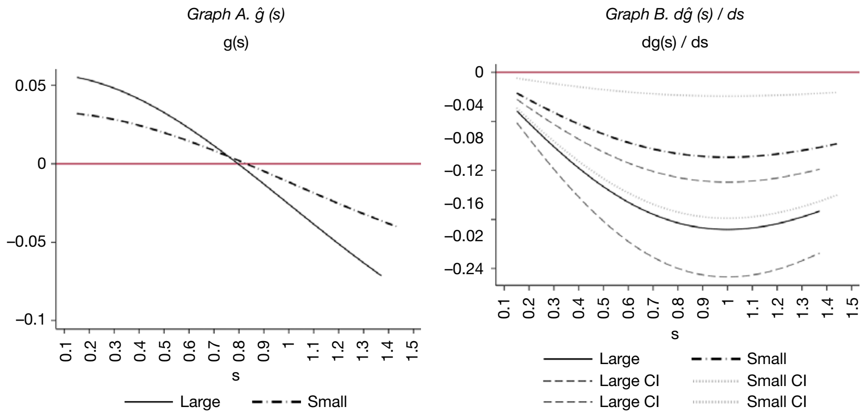
E. Good and Bad Uncertainty

A strand of literature regards uncertainty as having a "good news" component and a "bad news" component associated with positive and negative innovations, respectively, in stock prices and economic growth. For instance, Segal et al. (2015) decompose aggregate uncertainty into good and bad components and find that good uncertainty predicts an increase in economic activity (referred to as the "good news principle"), but bad uncertainty forecasts a decline in aggregate growth. Inspired by this idea, we divide the individual stock volatility into positive and negative realized semivariances, which represent good news and bad news uncertainty, denoted by $RS+$ and $RS-$, respectively, following Patton and Sheppard (2015). Our purpose is

FIGURE 9

Firm Size: Big Versus Small

Graphs A and B of Figure 9 compare $\widehat{g}(s)$ and $d\widehat{g}(s)/ds$, respectively, between the estimates obtained for the subsample of large firms and those obtained for the subsample of small firms. A firm is defined as a large firm if the pre-sample average of its total assets in book value (between 1999 and 2009) is above the median. The solid line represents the estimates obtained for the "large" subsample, and the two dashed lines indicate the 95% confidence intervals. The dash-dotted line represents the estimates obtained for the "small" subsample, whereas the two dotted lines are the 95% confidence intervals. All numbers are in decimal form.



to investigate whether these two components have different effects on firm investment.[42]

The comparison of the marginal effects on investment is shown in Figure 10, where Graph A is for good uncertainty and Graph B is for bad uncertainty. Notably, the negative effects of good uncertainty are smaller (in absolute values) than those of bad uncertainty at most uncertainty levels. We do not find direct evidence that good uncertainty can increase investment, but the magnitude of the negative effects indeed diminishes to 0 as good uncertainty rises. In contrast, the negative marginal effects of bad uncertainty become stronger as bad uncertainty heightens. These patterns are weak evidence of the good news principle, suggesting that the effects of good and bad uncertainty on investment are different.

VI. Concluding Remarks

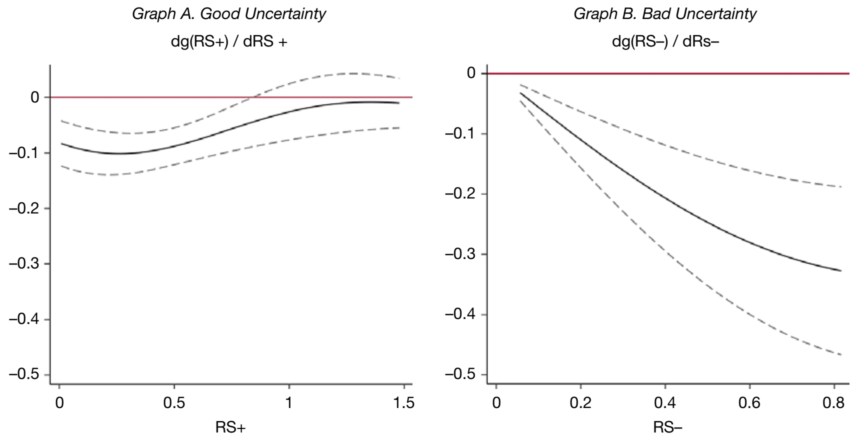
Examining U.S. firm-level data, this article estimates the relation between investment and uncertainty reflected in individual stock volatility. We propose a novel methodology to overcome three challenges in the data regarding regressor endogeneity of the stock volatility, mismeasurement of Tobin's Q, and the unknown function of the investment–uncertainty relation. We find that investment significantly decreases upon greater uncertainty and more so if uncertainty has already stood at a higher level. This nonlinear relation is likely due to a combination of the

[42]Specifically, for a given firm-year combination, we classify daily stock returns into a positive return group and a negative return group. The positive realized semivariance equals the sum of squared returns in the positive return group, and the negative realized semivariance equals the sum of squared returns in the negative return group.

FIGURE 10

Good Versus Bad Uncertainty

Graphs A and B of Figure 10 compare the marginal effects of uncertainty on investment between the estimates where uncertainty is "good uncertainty" and those where uncertainty is "bad uncertainty." Good and bad uncertainty are measured by the realized positive and negative semivariances, respectively, following Patton and Sheppard (2015). The solid line reflects the point estimates of the marginal effects, and the two dashed lines indicate the 95% confidence intervals. All numbers are in decimal form.



real option effect, risk premium effect, and overinvestment effect, and is different for good news uncertainty than for bad news uncertainty.

Our work contrasts with the existing empirical literature which ignores at least one or two of the three challenges due to limitations in estimation methods. In econometric theory, it is difficult to evaluate the resulting bias which relies on data in complicated ways. Our methodological framework can instead empirically assess the corresponding bias through "a controlled experiment" in which one of the three data issues is left untreated intentionally. We find that linear approximation can lead to significant overestimation of the marginal effect on investment of the stock volatility that is in the low region. Untreated measurement error in Q can instead result in overestimation of the marginal effect in both low and high volatility regions, and also incorrectly implies the shape of the investment–uncertainty relation. In contrast, neglecting the endogeneity of stock volatility would not cause significant bias in the estimation with our empirical sample.

The last result is to some extent surprising. Intuitively, one would expect that the regressor endogeneity of volatility has a more direct, and thus larger, impact on the estimation of the relation between investment and volatility than the mismeasurement of Tobin's Q does as a control variable. Possibly due to this intuition, previous empirical studies more often choose to address the regressor endogeneity problem than the measurement error issue. We find that this strategy can lead to significant bias in some situations, as in our empirical sample. However, we have no intention to generalize this finding to cases other than our empirical work, because the bias relies crucially on data in complex ways. This complexity further warrants the need for our estimation approach, since the bias is unpredictable prior to jointly addressing the measurement error and regressor endogeneity problems.

Finally, our proposed method can also be applied elsewhere since measurement error, regressor endogeneity, and nonlinearities are commonplace in empirical research. For example, a firm-level leverage regression that is similar to the one in Rajan and Zingales (1995) can be an ideal candidate—it is potentially nonlinear, where the *market-to-book ratio* may contain measurement error, and the *Tangibility* variable is likely an endogenous regressor. In econometric theory, future work may include considering nonclassical measurement error, cross-sectional dependence caused by time-varying industry effects, and a full nonparametric model $F\left(x^*_{i,t-1}, z_{i,t-1}, s_{i,t-1}\right)$ that allows for more interacting effects among regressors.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109023000169.

References

Abel, A. B. "Optimal Investment Under Uncertainty." *American Economic Review*, 73 (1983), 228–233.
Abel, A. B. "The Effects of *Q* and Cash Flow on Investment in the Presence of Measurement Error." *Journal of Financial Economics*, 128 (2018), 363–377.
Alfaro, I.; N. Bloom; and X. Lin. "The Finance Uncertainty Multiplier." NBER Working Paper No. 24571 (2022).
Allayannis, G., and A. Mozumdar. "The Impact of Negative Cash Flow and Influential Observations on Investment–Cash Flow Sensitivity Estimates." *Journal of Banking & Finance*, 28 (2004), 901–930.
Almeida, H.; M. Campello; and A. F. Galvao. "Measurement Errors in Investment Equations." *Review of Financial Studies*, 23 (2010), 3279–3328.
Alnahedh, S.; S. Bhagat; and L. Obreja. "Employment, Corporate Investment, and Cash-Flow Risk." *Journal of Financial and Quantitative Analysis*, 54 (2019), 1855–1898.
Arellano, M., and S. Bond. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58 (1991), 277–297.
Balasubramanian, N., and J. Sivadasan. "Capital Resalability, Productivity Dispersion, and Market Structure." *Review of Economics and Statistics*, 91 (2009), 547–557.
Belloni, A.; D. Chen; V. Chernozhukov; and C. Hansen. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica*, 80 (2012), 2369–2429.
Bernanke, B. S.; M. Gertler; and S. Gilchrist. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, Vol. 1. Amsterdam: Elsevier (1999), 1341–1393.
Bester, C. A.; T. G. Conley; C. B. Hansen; and T. J. Vogelsang. "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators." *Econometric Theory*, 32 (2016), 154–186.
Bhagat, S.; N. Moyen; and I. Suh. "Investment and Internal Funds of Distressed Firms." *Journal of Corporate Finance*, 11 (2005), 449–472.
Bloom, N. "Fluctuations in Uncertainty." *Journal of Economic Perspectives*, 28 (2014), 153–176.
Bond, S. R., and J. G. Cummins. "Uncertainty and Investment: An Empirical Investigation Using Data on Analysts' Profits Forecasts." Washington, D.C.: Federal Reserve Board (2004).
Caballero, R. J. "On the Sign of the Investment–Uncertainty Relationship." *American Economic Review*, 81 (1991), 279–288.
Carlton, D. W.; J. M. Perloff; and K. T. van 't Veld. *Modern Industrial Organization*. Glenview, IL: Scott, Foresman/Little, Brown Higher Education (1990).
Chen, Q.; I. Goldstein; and W. Jiang. "Price Informativeness and Investment Sensitivity to Stock Price." *Review of Financial Studies*, 20 (2007), 619–650.
Chirinko, R. S., and H. Schaller. "The Irreversibility Premium." *Journal of Monetary Economics*, 56 (2009), 390–408.
Cleary, S.; P. Povel; and M. Raith. "The U-Shaped Investment Curve: Theory and Evidence." *Journal of Financial and Quantitative Analysis*, 42 (2007), 1–39.

Dixit, A. K., and R. S. Pindyck, *Investment Under Uncertainty*. Princeton, NJ: Princeton University Press (1994).
Doshi, H.; P. Kumar; and V. Yerramilli. "Uncertainty, Capital Investment, and Risk Management." *Management Science*, 64 (2018), 5769–5786.
Driscoll, J. C., and A. C. Kraay. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80 (1998), 549–560.
Eisdorfer, A. "Empirical Evidence of Risk Shifting in Financially Distressed Firms." *Journal of Finance*, 63 (2008), 609–637.
Erickson, T.; C. H. Jiang; and T. M. Whited. "Minimum Distance Estimation of the Errors-in-Variables Model Using Linear Cumulant Equations." *Journal of Econometrics*, 183 (2014), 211–221.
Erickson, T., and T. M. Whited. "Measurement Error and the Relationship Between Investment and *Q*." *Journal of Political Economy*, 108 (2000), 1027–1057.
Erickson, T., and T. M. Whited. "Two-Step GMM Estimation of the Errors-in-Variables Model Using High-Order Moments." *Econometric Theory*, 18 (2002), 776–799.
Erickson, T., and T. M. Whited. "Treating Measurement Error in Tobin's *Q*." *Review of Financial Studies*, 25 (2012), 1286–1329.
Firth, M.; P. H. Malatesta; Q. Xin; and L. Xu. "Corporate Investment, Government Control, and Financing Channels: Evidence from China's Listed Companies." *Journal of Corporate Finance*, 18 (2012), 433–450.
Gabaix, X., and R. S. Koijen. "Granular Instrumental Variables." NBER Technical Report No. 28204 (2020).
Geary, R. C. "Inherent Relations Between Random Variables." *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, 47 (1941), 63–76.
Gilchrist, S.; J. W. Sim; and E. Zakrajšek. "Uncertainty, Financial Frictions, and Investment Dynamics." NBER Working Paper No. 20038 (2014).
Greene, W. H., *Econometric Analysis*. Upper Saddle River, NJ: Pearson (2017).
Gu, L.; D. Hackbarth; and T. Johnson. "Inflexibility and Stock Returns." *Review of Financial Studies*, 31 (2018), 278–321.
Gulen, H., and M. Ion. "Policy Uncertainty and Corporate Investment." *Review of Financial Studies*, 29 (2016), 523–564.
Hartman, R. "The Effects of Price and Cost Uncertainty on Investment." *Journal of Economic Theory*, 5 (1972), 258–266.
Hennessy, C. A., and T. M. Whited. "How Costly Is External Financing? Evidence from a Structural Estimation." *Journal of Finance*, 62 (2007), 1705–1745.
Julio, B., and Y. Yook. "Political Uncertainty and Corporate Investment Cycles." *Journal of Finance*, 67 (2012), 45–83.
Kim, H., and H. Kung. "The Asset Redeployability Channel: How Uncertainty Affects Corporate Investment." *Review of Financial Studies*, 30 (2017), 245–280.
Lewellen, J., and K. Lewellen. "Investment and Cash Flow: New Evidence." *Journal of Financial and Quantitative Analysis*, 51 (2016), 1135–1164.
Li, D.; N. E. Magud; and F. Valencia. "Financial Shocks and Corporate Investment in Emerging Markets." *Journal of Money, Credit and Banking*, 52 (2020), 613–644.
Meijer, E.; L. Spierdijk; and T. Wansbeek. "Consistent Estimation of Linear Panel Data Models with Measurement Error." *Journal of Econometrics*, 200 (2017), 169–180.
Newey, W. K., and J. L. Powell. "Instrumental Variable Estimation of Nonparametric Models." *Econometrica*, 71 (2003), 1565–1578.
Newey, W. K.; J. L. Powell; and F. Vella. "Nonparametric Estimation of Triangular Simultaneous Equations Models." *Econometrica*, 67 (1999), 565–603.
Oi, W. "The Desirability of Price Instability Under Price Uncertainty." *Econometrica*, 29 (1961), 58–64.
Panousi, V., and D. Papanikolaou. "Investment, Idiosyncratic Risk, and Ownership." *Journal of Finance*, 67 (2012), 1113–1148.
Patton, A. J., and K. Sheppard. "Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility." *Review of Economics and Statistics*, 97 (2015), 683–697.
Peters, R. H., and L. A. Taylor. "Intangible Capital and the Investment–*Q* Relation." *Journal of Financial Economics*, 123 (2017), 251–272.
Philippon, T. "The Bond market's *Q*." *Quarterly Journal of Economics*, 124 (2009), 1011–1056.
Rajan, R. G., and L. Zingales. "What Do We Know About Capital Structure? Some Evidence from International Data." *Journal of Finance*, 50 (1995), 1421–1460.
Roodman, D. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics*, 71 (2009), 135–158.
Sarkar, S. "On the Investment–Uncertainty Relationship in a Real Options Model." *Journal of Economic Dynamics and Control*, 24 (2000), 219–225.

https://doi.org/10.1017/S0022109023000169 Published online by Cambridge University Press

Schennach, S. M., and Y. Hu. “Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models Without Side Information.” *Journal of the American Statistical Association*, 108 (2013), 177–186.

Segal, G.; I. Shaliastovich; and A. Yaron. “Good and Bad Uncertainty: Macroeconomic and Financial Market Implications.” *Journal of Financial Economics*, 117 (2015), 369–397.

Tibshirani, R. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 58 (1996), 267–288.

Vogelsang, T. J. “Heteroskedasticity, Autocorrelation, and Spatial Correlation Robust Inference in Linear Panel Models with Fixed-Effects.” *Journal of Econometrics*, 166 (2012), 303–319.

Wong, K. P. “The Effect of Uncertainty on Investment Timing in a Real Options Model.” *Journal of Economic Dynamics and Control*, 31 (2007), 2152–2167.

Zhang, C.-H. “Nearly Unbiased Variable Selection Under Minimax Concave Penalty.” *Annals of Statistics*, 38 (2010), 894–942.