Volume Title: NBER Macroeconomics Annual 1998, volume 13

Volume Author/Editor: Ben S. Bernanke and Julio Rotemberg, editors

Volume Publisher: MIT Press

Volume ISBN: 0-262-52271-3

Volume URL: http://www.nber.org/books/bern99-1

Publication Date: January 1999

Chapter Title: Investment: Fundamentals and Finance

Chapter Author: Simon Gilchrist, Charles Himmelberg

Chapter URL: http://www.nber.org/chapters/c11248

Chapter pages in book: (p. 223 - 274)

Simon Gilchrist and Charles Himmelberg
BOSTON UNIVERSITY AND NBER, AND COLUMBIA UNIVERSITY

Investment: Fundamentals and Finance

1. *Introduction*

It is well recognized that financial variables such as cash flow and cash stocks are robust and quantitatively important explanatory variables for investment in reduced-form equations estimated with firm-level data. Following the seminal work of Fazzari, Hubbard, and Petersen (1988), a large body of recent empirical work attributes these findings to capital-market imperfections [see the extensive survey by Hubbard (1998)]. This literature argues that when access to external debt and equity is costly, internally available funds provide a cheaper source of financing, thus increasing the desired level of investment. Such cost premiums for external finance are generally explained by appealing to models with asymmetric information and agency problems.

Despite the volume of empirical work in this literature, financing-based interpretations of the explanatory power of cash flow and other financial variables in investment equations remain controversial.[1] Even among economists who agree that firms face some degree of financial frictions, there remains substantial disagreement over the magnitude of such frictions and whether they are large enough to affect investment behavior. The controversy can be traced to two distinct but related problems with identification. The first problem is that financial variables may contain information about future returns to capital. In a forward-looking

We thank Charles Calomiris, Russel Cooper, Jan Eberly, Bernd Fitzenberger, Bill Gentry, Mark Gertler, Bob Hodrick, Glenn Hubbard, Cornelia Kullman, Chris Mayer, John Shea, and Kristen Willard for helpful comments and suggestions and the NSF for financial support. We are especially grateful for excellent comments from Ben Bernanke, David Gross, Julio Rotemberg, and Ken West. We are also grateful to seminar participants at Boston University, the CEPR/DFG/ZEW Conference on Industrial Structure and Input Markets, Columbia, CREST, Georgetown, the Federal Reserve Bank of Boston, the Federal Reserve Bank of New York, and the 1998 NBER Macro Annual Conference.

1. See, for example, Kaplan and Zingales (1997) and Fazzari, Hubbard, and Petersen (1996).

model, investment depends on marginal Q, the present value of expected future marginal returns to capital. This present value is the "fundamental" to which investment should respond, even in the absence of capital-market imperfections. Any variable that helps predict marginal Q should appear as a state variable in the firm's decision rule for investment, and should therefore have explanatory power for investment. The ratio of cash flow to capital is obviously closely related to the return on capital, so from the perspective of models based on forward-looking fundamentals, if other variables in the regression (like Tobin's Q) do not fully specify the expected marginal value of capital, it is not surprising that cash flow appears in reduced-form regression models. The same logic makes it difficult to interpret the role of other financial variables such as cash stocks and leverage as well.

The second identification problem relates to the distinction between the *marginal* return to capital and the *average* return to capital. In the absence of financial-market imperfections, the present value of expected future marginal profitability of capital (MPK) should be the sole determinant of investment at the firm level. Lacking good measures of the marginal return to capital, the empirical investment literature often relies on the *average* return to capital—the ratio of profits to capital—as a proxy for the marginal return. Unfortunately, this proxy also provides a good measure of the financial health of the firm, which, in the presence of financial-market imperfections, should also influence investment. By not carefully distinguishing between the present value of marginal and average returns, the existing empirical investment literature potentially confounds the influence of investment fundamentals with the effect of financial factors that reflect premiums on external finance.

In Gilchrist and Himmelberg (1995), we attempted to resolve the first of these identification problems by using a vector autoregression to model the forward-looking role of cash flow in a structural model for investment. Using firm-level data, we confirmed that the predictive power of cash flow for future MPK in a model with perfect capital markets could account for a significant portion of the overall explanatory power of cash flow for investment. But we also found evidence against the model. Like previous studies that used Tobin's Q to control for the expected return to investment, we found that investment is "excessively sensitive" to cash flow, that is, more sensitive to cash flow than the neoclassical model of investment information would predict. We concluded that financial-market imperfections were a likely source of the model's rejection, but our modeling framework was not sufficiently general to assign a structural interpretation to investment's excess sensitivity to cash flow.

In this paper, we attempt to resolve the second identification problem by extending the empirical framework used in our previous work to a (linearized) structural model of investment that explicitly incorporates financial frictions. Like our previous work, this empirical framework uses panel-data vector autoregressions (VARs) to construct expectations of the future marginal profitability of capital. Unlike our previous work, however, we introduce financial frictions into the model, and we develop improved measures of the marginal profitability of capital (MPK) that sharpen the distinction between MPK and financial factors. By combining better measures of MPK with our extended model, we substantially improve our ability to identify and quantify the influence of financial factors on investment decisions.

Although panel-data VARs have not been widely used by previous researchers to describe investment behavior, we believe they can be a useful tool for summarizing the data and testing structural model assumptions.[2] We consider two strategies for using VARs to model investment. First, we use VARs to summarize the dynamic relationship among investment, MPK, and cash flow. By imposing a recursive structure on the contemporaneous shocks of the model (a standard identification technique in VAR analysis), we identify shocks to cash flow that are orthogonal to MPK. The impulse response functions for this model show that the orthogonalized shocks to cash flow elicit a substantial and prolonged response from investment. Moreover, the cash-flow shock predicts either zero or negative response to MPK. This result implies that the response of investment to cash flow cannot be attributed to revisions in the expected return to capital. Indeed, the negative response of MPK implies (counterfactually) that investment should fall rather than rise in response to the cash-flow shock. This evidence is difficult to reconcile with a model in which cash flow's influence on investment is entirely attributable to nonfinancial fundamentals.

Our second strategy uses panel-data VARs to impose structural restrictions on the investment equation derived from a model with costly external finance. The use of VARs to estimate structural investment models was introduced into the empirical investment literature by Abel and Blanchard (1986), and was subsequently applied to panel data by Gilchrist and Himmelberg (1995). The modeling contribution in this paper is to show that putting financial frictions in the model introduces a state-dependent discount factor that depends on the firm's balance-sheet condition. Because it is not possible to solve this model analyti-

2. The two applications of panel-data VARs to firm-level investment of which we are aware are Whited (1992) and Himmelberg (1990).

cally, we work with a linearized version that is amenable to VAR methods. This structure allows us to identify the sensitivity of investment to changes in the expected marginal value of capital. With financial frictions in the model, we show that investment should also display excess sensitivity to the present value of financial variables because these variables influence the future shadow cost of funds used to discount future MPK.

In our empirical results, we find that investment is responsive to both fundamental and financial factors, as predicted by the existence of financial frictions. This response is both statistically and economically significant—for the average firm in our sample, our estimates show that financial factors increase the overall response of investment to an expansionary shock by 25% over the first few years following the initial impulse.

Although the average firm in our sample shows a quantitatively significant response to financial factors, we also find that financial factors play little, if any, role in determining the investment behavior of bond-rated firms. Because bond-rated firms account for a large fraction of overall investment activity (on the order of 50% in manufacturing), this reduces the role of financial factors for aggregate investment, at least during normal times. While non-bond-rated firms are quantitatively less important for aggregate investment, they are more labor-intensive and are influential in the determination of inventory dynamics.[3] To the extent that non-bond-rated firms rely on external funds to finance both labor inputs and inventory investment, our evidence that such firms do indeed face capital market imperfections suggests that financial factors will have important influences through these channels as well.[4]

2. *Investment, MPK, and Cash Flow: Simple VAR Evidence*

In this section we begin by discussing the importance of measuring MPK as accurately as possible. We then briefly discuss the estimation of panel-data vector autoregressions and argue the merits of using VARs as summary statistics that provide a full, dynamic description of the relationship among investment, MPK, and financial variables. Finally, we suggest a recursive ordering of the VAR that allows us to identify the component of cash-flow innovations that is orthogonal to the MPK shock. We report

3. See Carpenter, Fazzari, and Peterson (1997).
4. Sharpe (1993) shows the importance of financial constraints for employment dynamics. Kashyap, Lamont, and Stein (1994), Carpenter, Fazzari, and Peterson (1997), and Gertler and Gilchrist (1994) provide evidence for inventory dynamics, and Himmelberg and Peterson (1994) estimate the effect of financial frictions on R&D spending.

impulse response functions for investment based on this ordering, and argue that the results provide evidence of a financing role for cash flow. These results motivate a more structural econometric investigation, which we provide in Section 3.

2.1 MEASURING MPK

Suppose a firm has a Cobb–Douglas production function $y = Ak^{\alpha_k}n^{\alpha_n}x^{\alpha_x}$, where A is the total factor productivity, y is output, k and n are quasifixed capital stocks, and x is a variable factor input. We allow for nonconstant returns to scale by assuming $\alpha_k + \alpha_n + \alpha_x = 1 + \gamma$, where γ is the return-to-scale parameter. We allow for multiple quasifixed factors because we are concerned about the empirical implications of ignoring omitted quasifixed factors. The idea here is that k represents the stock of fixed property, plant, and equipment, while n represents R&D capital and other intangible assets. The assumption of a single variable input is without loss of generality. Assuming that the firm faces an inverse demand curve $p(y)$, variable factor prices w, and fixed costs F, the profit function is defined by

$$\begin{aligned}\pi(k, n, w, F) &= \max_{x>0} p(y)y - wx - F \\ \text{s.t.} \quad y &= Ak^{\alpha k}n^{\alpha n}x^{\alpha x}.\end{aligned} \tag{2.1}$$

This specification of the profit function allows fixed costs F to be time-varying. For example, if n represents the stock of the firm's R&D workers, which are quasifixed factors due to hiring and firing costs, the F could represent the wages paid to these workers.

By applying the envelope theorem, the marginal profitability of fixed capital, denoted by MPK, is readily shown to be

$$\text{MPK} \equiv \frac{\partial\pi}{\partial k} = \theta\left(\frac{s}{k}\right), \tag{2.2}$$

where $\theta = (1 + \eta^{-1})\ \alpha_k$, $\eta \equiv (\partial y/\partial p)p/y < -1$ is the (firm-level) price elasticity of demand,[5] α_k is the capital share of output from the Cobb–Douglas specification, and $s = py$ is the firm's sales. Equation (2.2) shows that, up to a scale parameter,[6] the ratio of sales to capital measures the

5. Note that if firms are profit maximizers, they will produce on the elastic portion of the demand curve, so that $\eta < -1$.

6. If the effect of corporate taxes is included, then the tax-adjusted expression for MPK takes the form MPK = $(1 - \tau)\theta(s/k)$, where τ is the corporate tax rate on profits. Our estimates of θ allow variation in τ over industries but not over time. Time variation in tax rates would, to some degree, be captured by our year dummies. We plan to explore the effects of taxes in more detail in future work.

Table 1 TWO-DIGIT SIC ESTIMATES OF $\hat{\theta}_j$

		$\hat{\theta}_j$				$\hat{\theta}_j$	
SIC	*Obs.*	*Sales*	*OI*	*SIC*	*Obs.*	*Sales*	*OI*
20	1112	0.036	0.387	30	670	0.040	0.373
21	34	0.027	0.171	31	153	0.017	0.233
22	549	0.035	0.376	32	420	0.069	0.571
23	332	0.017	0.185	33	821	0.063	0.612
24	298	0.044	0.489	34	958	0.040	0.375
25	373	0.031	0.330	35	2161	0.036	0.328
26	562	0.077	0.598	36	2123	0.039	0.304
27	700	0.042	0.300	37	1062	0.037	0.353
28	1504	0.051	0.334	38	1411	0.036	0.313
29	469	0.097	0.722	39	398	0.032	0.301

marginal profitability of fixed capital.[7]

Because it is unreasonable to assume that manufacturing firms in different industries face the same price elasticity of demand, η, or the same capital share of sales, α_k, we construct industry-level estimates of θ. We assume that firms are, on average, at their equilibrium capital stocks. Ignoring adjustment costs, this says the marginal profitability of capital should roughly equal the cost of capital, that is, $\text{MPK}_{it} = r_{it} + \delta_{it}$, where r_{it} and δ_{it} are the risk-adjusted discount rate and depreciation rate of capital, respectively. Substituting $\theta_j(s/k)_{it}$ for MPK_{it} and averaging over all firms $i \in I(j)$ and years $t \in T(i)$ in industry j suggests that a reasonable estimate of θ_j is given by

$$\hat{\theta}_j = \left(\frac{1}{N_j} \sum_{i \in I(j)} \sum_{t \in T(i)} (s/k)_{it} \right)^{-1} \frac{1}{N_j} \sum_{i \in I(j)} \sum_{t \in T(i)} (r_{it} + \delta_{it}),$$

where N_j is the number of firm–year observations for industry j. In practice, we assume that $(1/NT) \sum_{i \in I(j)} \sum_{t \in T(i)} (r_{it} + \delta_{it}) = 0.18$ for all industries.[8]

To show the degree to which θ varies across two-digit industries, columns 3 and 7 in Table 1 report the values of $\hat{\theta}_j$. The table shows that the value of $\hat{\theta}_j$ ranges from .017 to 0.097. The assumptions $\alpha_k = 0.06$ and $\eta = -4.0$ imply a value of $(1 + \eta^{-1})\alpha_k = 0.045$. These values seem

7. This derivation ignores the difference between production and sales. For the smaller subset of firms in Compustat that report finished goods, the correlation between production-to-capital ratio and sales-to-capital ratio exceeds 0.99. In light of this fact and because of the limited availability of data on finished goods, we opted to measure MPK using the sales-to-capital ratio.

8. We experimented with different values for $(1/NT) \sum_{i \in I(j)} \sum_{t \in T(i)} (r_{it} + \delta_{it})$, including the calculation of industry-specific depreciation rates. In practice, this adds very little variation to the estimated value of θ_j, but this is probably an issue that future work could profitably explore in more depth.

plausible, suggesting that our estimates of $\hat{\theta}_j$ reported in Table 1 are reasonable. We therefore construct estimates of the marginal profit using

$$\mathrm{MPK1}_{it} = \hat{\theta}_j \frac{s_{it}}{k_{it}}.$$

In the results reported in the paper, this is our preferred measure of MPK, which we refer to as MPK1. The summary statistics reported in Table 2 indicate that MPK1 has a mean of 0.200, with an interquartile range of 0.121 to 0.240.

2.2 WHY IT IS LESS DESIRABLE TO MEASURE MPK USING OPERATING INCOME

Previous authors have measured MPK using the ratio of operating income to capital. For example, using aggregate data for U.S. manufactur-

Table 2 VARIABLES: ACRONYMS, DEFINITIONS, AND SUMMARY STATISTICS

Acronym	*Description/Compustat Definition*	*Mean (S.D.)*	*Percentiles* *Min*	*25%*	*50%*	*75%*	*Max*
MPK1	Sales-based marginal profitability of capital (see text)	0.200 (0.125)	0.019	0.121	0.170	0.241	1.37
MPK2	Operating-income-based marginal profitability of capital (see text)	0.164 (0.131)	−0.410	0.090	0.1521	0.241	2.04
S/K	Sales/capital $=x_{12}/k_8(t-1)$	5.03 (3.36)	0.518	2.86	4.27	6.21	30.4
OI/K	(Operating income)/capital $=x_{13}/k_8(t-1)$	0.467 (0.392)	−1.23	0.237	0.419	0.637	3.86
CF/K	(Cash flow)/capital $=(x_{18}+x_{14})/x_8(t-1)$	0.291 (0.291)	−0.976	0.148	0.274	0.427	2.44
I/K	(Gross investment)/capital $=x_{30}/k_8(t-1)$	0.227 (0.175)	0.013	0.119	0.185	0.280	1.56
CE/K	(Cash and equivalents)/capital $=x_1/k_8(t-1)$	0.271 (0.372)	0.00	0.045	0.129	0.344	4.13
FW/K	(Financial working capital)/capital $=(x_4-x_5+x_3)/x_8(t-1)$	0.218 (0.548)	−1.00	−0.092	0.122	0.441	2.98
TD/K	(Total debt)/capital $=(x_{34}+x_9)/x_8(t-1)$	1.01 (0.825)	0.001	0.481	0.764	1.27	7.64
TQ	Tobin's Q $=(x_{25}x_{199}+10x_{19}+x_{181})/x_6$	1.47 (0.955)	0.170	0.972	1.17	1.48	12.0

The notation x_{99} refers to Compustat data item 99, etc.

ing, Abel and Blanchard (1986) used the average profitability of capital to measure MPK. In Gilchrist and Himmelberg (1995), we constructed a similar measure with firm-level data by using the ratio of operating income to capital. In hindsight, we think the assumptions necessary to make this approximation—zero fixed costs and perfect competition—are unreasonable at the firm level.[9] The sales-based measure described in the previous section is our preferred measure, for reasons which we explain in this section.

Working from the firm's objective function in equation (2.1), an alternative representation of the marginal profit is

$$\frac{\partial \pi}{\partial k} = \varphi \left(\frac{\pi}{k} + \frac{F}{k} + \eta^{-1} \frac{py}{k} \right), \tag{2.3}$$

where $\varphi = \alpha_k/(\alpha_k + \alpha_n - \gamma)$.[10] With accounting data, we observe π/k and py/k, but not F. Hence, to use equation (2.3), we must assume fixed costs are zero, so that $F = 0$. Moreover, η and φ cannot be separately identified without access to additional data, so it is also necessary to at least assume that η^{-1} is constant across industries; more conventionally, perfect competition is assumed, so that $\eta^{-1} = 0$. Under these assumptions, a measure of MPK based on operating income is given by

$$\frac{\partial \pi}{\partial k} = \varphi \frac{\pi}{k} . \tag{2.4}$$

Just as we used industry estimates of θ_j to adjust the sales-to-capital ratio, we implement equation (2.4) using industry estimates of the capital share of variable profits, φ_j. Thus, a second measure of MPK is given by

$$\mathrm{MPK2}_{it} = \hat{\varphi}_j \frac{\mathrm{oi}_{it}}{k_{it}},$$

9. In defense of Abel and Blanchard (1986), one advantage of using aggregate data is the availability of prices and wages, which make it possible to construct variable costs. At the firm level, however, only total costs are available, so variable costs are unmeasurable unless we assume fixed costs are zero.

10. The derivation of equation (2.3) follows from the first-order condition for variable inputs, $(1 + \eta^{-1})\alpha_x py = wx$. The returns-to-scale parameter γ is defined so that the factor shares sum to $1 + \gamma$, i.e., $\alpha_x + \alpha_k + \alpha_n = 1 + \gamma$, so that constant returns would imply $\gamma = 0$. Substituting for α_x in the first-order condition and rearranging, we find $(1 + \eta^{-1})(\alpha_k + \alpha_n - \gamma)py = (1 + \eta^{-1})py - wx$. Using $\pi + F = py - wx$, this can be written

$$(1 + \eta^{-1})\alpha_k py = \frac{\alpha_k}{\alpha_k + \alpha_n - \gamma}(\pi + F + \eta^{-1}py).$$

Dividing both sides by k gives the desired result. Note that if $\gamma = 0$, then $\varphi = \alpha_k/(\alpha_k + \alpha_n)$ is simply the capital share of quasifixed inputs.

where oi_{it} denotes operating income.

It is important to stress that for our purposes, MPK2 is less desirable than MPK1. This is because the accuracy of MPK2 requires the added assumptions of zero fixed costs and perfect competition, whereas MPK1 does not. In other words, MPK2 is a noisier measure of MPK. But the most important shortcoming of MPK2 is that the noise component is correlated with cash flow, and thus MPK2 could spuriously attribute cash-flow fluctuations to changes in MPK. This distinction is obviously important, because MPK is what matters for fundamental explanations, whereas cash flow is more likely to matter for financial reasons. The empirical results in this paper exploit this difference.

2.3 MEASURING CASH FLOW

Our accounting definition of cash flow is net income before extraordinary items plus depreciation. Equivalently, cash flow is operating income before depreciation and minus taxes, minus interest payments, plus nonoperating income, plus special items. To provide a feel for relative magnitudes, Table 3 reports the aggregate income sheet for the Compustat universe of manufacturing firms in 1988.

With respect to the terms in Equation (2.3), py corresponds to sales, while $wx + F$ corresponds to cost of goods sold plus selling, general, and administrative expenses. It is therefore not possible with accounting data to disentangle variable and fixed costs. This is one of the reasons we gave in the previous section for preferring MPK1 over MPK2.

In equation 2.3, the difference between marginal and average profits introduced scope for identifying changes in cash flow distinct from changes in MPK. Our definition of cash flow provides additional sources of independent variation from MPK, because it treats taxes payable and

Table 3 AGGREGATE INCOME STATEMENT IN 1988 (PERCENTAGE OF SALES)

Sales	
−Cost of goods sold	67.6
−Selling, general, and administrative expenses	17.6
Operating income before depreciation	14.9
−taxes payable	3.6
−interest payments	2.2
+(−) Nonoperating income	1.8
+(−) Special items	0.0
Cash flow	10.9
−depreciation	5.0
Net income before extraordinary items	5.9

interest payments as fixed charges.[11] In addition, as the table shows, many firms generate internal funds from financial investments and other nonoperating assets. These funds provide a third source of cash-flow variation, which is distinct from MPK variation.

Our definition of cash flow is only partly correlated with operating income, which in turn is only partly correlated with MPK. This is an important empirical distinction which previous authors (including Gilchrist and Himmelberg, 1995) have failed to exploit. We exploit this difference below to distinguish the investment response to pure cash-flow shocks from the response to mere MPK shocks.

2.4 PANEL-DATA VECTOR AUTOREGRESSIONS

It is uncommon to see VARs estimated with panel data, and VARs have not been widely used in the investment literature, so we provide a brief discussion of the (minimal) econometric assumptions for their estimation with panel data. Without loss of generality, consider the following VAR(1) with fixed firm effects and year effects:

$$y_{it} = Ay_{it-1} + f_i + d_t + u_{it},$$

where A is a $k \times k$ matrix of slope coefficients, f_i is a $k \times 1$ vector of (unobserved) firm effects, and d_t is $k \times 1$ vector of year effects (to be estimated). In this paper, y_{it} will generally consist of a $k \times 1$ vector of firm-level state variables and decision variables that will include variables like investment, MPK, and cash flow. More generally, this notation will be used to describe the companion form of a VAR(p) model for y_{it}. In either case, the matrix of parameters A is redefined accordingly.

A VAR model provides a surprisingly flexible framework for describing the dynamic relationship among firm-level panel data. For one, the inclusion of the time effects d_t accommodates aggregate shocks to y_{it} that are common across firms. Thus, to the extent that there may be common movements to interest rates or other macroeconomic conditions that are not captured by lagged y_{it}, these factors will be captured by time dummies. In addition, under the assumption that $E(u_{it}) = 0$ and $E(u_{it}u_{it}') = \Omega_i$ (and conditional on $\{d_t\}_{t=1}^{T}$, where the d_t's are parameters that will be estimated), y_{it} has unconditional mean and variance given by $E(y_{it}) = (I - A)^{-1} f_i$ and $\mathrm{Var}(y_{it}) = (I - A)^{-1}\Omega_i(I - A)^{-1}$. Thus, while the model imposes the same slope coefficients A across firms, it imposes no restrictions on the unconditional mean and variance of y_{it}. This is an important feature

11. We do not need to assume that current interest and tax payments are strictly predetermined. Rather, we only assume that to the extent they are endogenous, they are determined by factors independently of the decision to invest.

of the model, since the unconditional means and variances of most firm-level variables display substantial cross-sectional heterogeneity.

The estimation of panel-data VARs has been discussed by Holtz-Eakin, Newey, and Rosen (1988), among others, and they show that panel data pose no particular problem for the estimation of VARs. In fact, asymptotic results are, if anything, easier to derive for panel data than for time series. We mention this because it is still common to encounter confusion (usually among macroeconomists) over the feasibility of estimating a time-series model (such as a VAR) using only a few years of data. Because the sampling properties depend on the number of cross-sectional observations, not the number of time-series observations, it is technically possible, for example, to estimate an AR1 on a panel with as few as 3 years of data, although it is preferable to have panels with 5 or more years (because this increases the availability of instruments required for the estimation technique described below). All that is required is that the slope coefficients be the same across observations in the cross section. Estimation does not require homogeneity of the intercepts or the variances of the error terms. More details on the econometrics are included in the appendix.

2.5 THE DATA

Our data set is a firm-level panel of annual data on firms drawn from the Compustat universe of manufacturing firms from 1980 to 1993. We sampled every available firm–year observation during this time period without regard to whether the firm was in existence for the length of the time period; that is, we did *not* require a balanced panel. We then removed observations for which the data required to construct the variables in Table 2 were not available. We also imposed outlier rules on the Table 2 variables by removing observations that fell below the first or above the 99th percentile. Rules of this sort are both common and necessary when working with large panels, because some firms have very small (measured) capital stocks, and these cause large outliers when capital appears in the denominator as a scaling variable.[12]

12. To deal with large discrete changes in firm identity due to large mergers, acquisitions, and divestitures, we deleted observations which had large outliers in the amount by which the percentage change in the capital stock differed from the gross investment rate net of depreciation. For robustness issues, when estimating structural models, we considered financial variables that were ratios of both capital and debt. Because these financial variables have more dispersion in the tails, the forecasting equations used in our structural estimates were in some cases less precise without more stringent outlier rules for these variables. We therefore imposed the additional requirements that the ratios CE/debt and FW/K be within (−1,3). These rules are approximately equivalent to trimming the tails of these variables at the 2% level.

2.6 IDENTIFICATION USING RECURSIVELY ORDERED VECTOR AUTOREGRESSIONS

When interpreting the effect of cash flow on investment, the primary identification problem is to distinguish the information revealed about future MPKs from the information revealed about the financial condition of the firm. One way to make this distinction is by using a *structural* VAR, which imposes restrictions on the contemporaneous shocks but not on the coefficients of lagged variables. In our empirical specification, we estimate a three-variable, two-lag VAR that controls for fixed firm and year effects. The VAR variables are I/K, the ratio of gross investment to capital; MPK, the marginal profit of capital (based on sales as described in the previous section); and CF/K, the ratio of cash flow to capital. In the context of this VAR system, there are two issues that affect the interpretation of cash flow in the investment equation as evidence of a financing effect.

The first issue, of which the literature has long been aware, is that even after conditioning on lagged investment and MPK, lagged cash flow can still contain information about the future marginal profitability of capital. In this case, the responsiveness of investment to cash flow simply reflects the fact that we are estimating a forward-looking decision rule, and that CF/K belongs in the information set. The second issue is that it is difficult to identify the effects of contemporaneous cash-flow shocks on investment. To deal with this issue, we postulate a causal relationship among contemporaneous shocks that is obtained from a standard Cholesky decomposition using the ordering I/K, MPK, CF/K. This ordering allows for the possibility that I/K shocks contemporaneously cause movements in cash flow and MPK, but assumes there is no feedback (contemporaneously) from MPK shocks to I/K, or from cash flow to MPK.

This ordering is particularly interesting for investigating the effect of cash flow's financing role because orthogonal cash-flow shocks, by construction, contain no information about current MPK. While this represents progress toward identifying pure cash-flow effects, it does not confront the first issue. That is, while our orthogonal cash-flow shocks are uncorrelated with current MPK, they may nonetheless be correlated with *future* MPK. Thus, when using the impulse response functions to interpret the dynamic response of investment to orthogonal cash-flow shocks, it is important to inspect the dynamic response of MPK for evidence that the cash-flow shock predicts future marginal profits.

We report the impulse response of investment to both MPK shocks and cash-flow shocks, where the residuals are orthogonalized using the

Table 4 SELECTED IMPULSE RESPONSE FUNCTIONS

		Response						
Shock	*Variable*	*T=0*	*T=1*	*T=2*	*T=3*	*T=4*	*T=5*	*T=6*
MPK	$(I/K)_{it}$	0.00	0.021	0.01	0.004	0.002	0.001	0.00
	$MPK1_{it}$	0.041	0.031	0.02	0.012	0.008	0.005	0.003
	$(CF/K)_{it}$	0.079	0.054	0.028	0.016	0.009	0.005	0.003
Cash flow	$(I/K)_{it}$	0.00	0.034	0.02	0.011	0.006	0.003	0.001
	$MPK1_{it}$	0.00	0.003	−0.001	−0.003	−0.003	−0.003	−0.002
	$(CF/K)_{it}$	0.184	0.074	0.034	0.014	0.005	0.001	0.00

Impulse response functions based on a two-lag VAR for investment, MPK, and cash flow. Impulse response functions show the response to a one-standard-deviation shock.

decomposition described above. The top part of Table 4 reports the impulse response of all three variables to the MPK shock. As expected, investment, MPK and cash flow all rise in response to such a shock, with the effect persisting over a two- to three-year horizon before returning slowly to steady state.

The bottom part of Table 4 reports the response of investment to a cash-flow shock that is orthogonal to MPK. In this case investment responds positively to cash flow (the magnitude of response here is actually slightly larger than for the MPK shock), despite the fact that the marginal profitability of capital falls in response to such a shock. Thus, while fundamentals are falling, investment is rising. These results suggest that the positive investment response to cash flow is not caused by the predictive content of cash flow for future investment opportunities. Indeed, the negative response of future MPK implies that the impulse response for investment understates the full magnitude of the financing effect.[13]

In summary, reduced-form VAR analysis shows that investment re-

13. It is possible that, in addition to financial factors, cash flow also captures information about cost shocks that are not reflected in our sales-based fundamental. Under the assumption of Cobb–Douglas production, our measure of MPK captures the influence of both cost shocks and demand shocks on the marginal profitability of capital. If there are large deviations from Cobb–Douglas production, however, then cost shocks may be an issue. To investigate this possibility, we augmented our VAR framework by adding the ratio of cost of goods sold to capital (COG/K) as another variable in the VAR. We then considered a shock to cash flow that was orthogonal to I/K, MPK, and COG/K. We still obtain the result that investment responds positively to cash flow even though fundamentals are falling. Indeed, the quantitative results from this exercise are very close to those reported in Table 4. Thus, it seems unlikely that unmeasured variations in costs are driving this result.

sponds to both fundamentals (as measured by MPK) and financial factors (as measured by cash flow). The positive investment response to cash-flow shocks cannot be attributed to rising profit opportunities using our measure of the marginal profitability of capital, and is therefore most likely to be explained by financial frictions that generate excess sensitivity of investment to cash-flow shocks. While these results suggest that financial factors influence investment at the firm level, this exercise is limited in its ability to provide an economic description of the exact channel through capital-market imperfections influence investment dynamics. To say more, it is necessary to consider the more structural approach provided in the next section.

3. *A Model of Investment with Financial Frictions*

In this section we develop a model of investment with financial frictions that is similar to models that have been explored in the literature. The goal here is not to show how financial frictions can be integrated into the standard investment model, but to show how the resulting model, which is nonlinear, can be linearized to obtain a tractable dynamic system of equations that describe the joint evolution of investment, MPK, and financial variables. This framework includes the standard Q model of investment as a special case.

Let $\Pi(K_t, \xi_t)$ denote the maximized value of current profits taking as given the beginning-of-period capital stock, K_t, and a profitability shock, ξ_t. For the time being, we make no assumptions regarding the nature of returns to scale or competition in the product and factor markets, other than to assume that the profit function is concave and bounded. The time to build and install one unit of capital is one period,[14] where δ is the rate of capital depreciation and I_t is the investment expenditure, so that the capital stock evolves according to the equation $K_{t+1} = (1 - \delta)K_t + I_t$. Finally, as is common in the literature, we assume that $C(I_t,K_t)$ is the resource cost of installing I_t units of capital.[15] For simplicity, the numeraire is the price of capital.

14. The true time to build is probably somewhere closer to six months, for which we have no corresponding assumption using annual data. In the absence of a strong empirical motive, a good theoretical reason for assuming one-period time to build is that it simplifies the inversion of the marginal adjustment cost function.
15. Future research could investigate alternative adjustment cost technologies designed to deal with asymmetries and nonconvexities, such as those developed by Abel and Eberly (1994, 1996) and Caballero (1997). Under one such alternative specification of adjustment costs, Caballero and Leahy (1996) show why average Q may be theoretically more effective than marginal Q for explaining investment. Recent papers by Goolsbee and Gross (1997) and Caballero and Engel (1998) provide empirical evidence on the importance of such factors.

A simple way to incorporate financial frictions is to assume that the marginal source of external finance is debt, and to assume that risk-neutral debt holders demand an external finance premium, $\eta_t = \eta(K_t, B_t, \xi_t)$, which in general depends on the entire state vector of the firm, and is increasing in the amount borrowed ($\partial\eta/\partial B > 0$). The idea is that highly leveraged firms have to pay an additional premium to compensate debt holders for increased costs due to information problems (e.g., *ex post* monitoring costs and/or moral hazard costs). While previous researchers have derived this premium in equilibrium,[16] it is sufficient for our purposes to postulate the existence of such a function, and to assume that this function is increasing in the debt level. Hence, we assume that the gross required rate of return on debt is $(1 + r_t)$ $[1 + \eta(K_t, B_t, \xi_t)]$, where r_t is the risk-free rate of return.

We have in mind that B_t summarizes the firm's net financial liabilities (bank debt, trade debt, cash holdings, etc.). This is the simplest possible model of financial assets and liabilities. In our empirical work, we consider several alternative definitions of B_t, one measure being long-term debt minus the net short-term financial assets of the firm, i.e., long-term debt minus financial working capital. Alternative specifications of B_t and $\eta(K_t, B_t, \xi_t)$ could be easily investigated.[17]

To guarantee that debt (and not equity) is the firm's marginal source of finance, we need either to assume a binding non-negativity constraint on dividends, or to assume that equity holders prefer to have dividends paid out rather than reinvested. One way to make this operational is to assume a utility function for dividends (e.g. Gross, 1997). This assumption is particularly useful when constructing numerical solutions to the model, because it avoids corner solutions. For our purposes, however, it is sufficient to display a model that generates a *shadow cost* of equity, and the simplest way to this is to assume that dividends cannot be negative (i.e., that marginal equity is prohibitively expensive).

For simplicity, assume a constant price of new capital goods, normalized at unity, and let $(1 + r_t)^{-1}$ be the *ex ante* one-period discount factor used to value period-$t + 1$ dividends at time t. Then the manager's problem is

$$V(K_t, B_t, \xi_t) = \max_{\{I_{t+s}, B_{t+s+1}\}_{s=0}^{\infty}} D_t + E_t \sum_{s=1}^{\infty} \left(\prod_{k=1}^{s} (1 + r_{t+k})^{-1} \right) D_{t+s}$$

subject to

16. For example, Moyen (1997) derives an equilibrium debt premium generated by default costs.

17. Future research on the underlying sources of capital-market frictions could usefully guide future empirical work by suggesting appropriate functional forms for η and λ.

$$D_t = \Pi(K_t, \xi_t) - C(I_t, K_t) - I_t + B_{t+1} - (1 + r_t)[1 + \eta(B_t, K_t, \xi_t)]B_t,$$
$$K_{t+1} = (1 - \delta)K_t + I_t,$$
$$D_t \geq 0,$$

where E_t is the exceptions operator conditional on the time-t information set Ω_t.

To see the effect of financial frictions, let λ_t be the Lagrange multiplier for the non-negativity constraint on dividends. The multiplier λ_t indicates the shadow value of paying a negative dividend, and can thus be interpreted economically as the shadow cost of internally generated funds. The role of this shadow cost in the firm's investment decision is exposed by deriving the Euler equation for investment[18]:

$$1 + \frac{\partial C(I_t, K_t)}{\partial I_t} = E_t \left\{ \frac{1}{1 + r_t} \left(\frac{1 + \lambda_{t+1}}{1 + \lambda_t} \right) \left[\frac{\partial D_{t+1}}{\partial K_{t+1}} + (1 - \delta) \left(1 + \frac{\partial C(I_{t+1}, K_{t+1})}{\partial I_{t+1}} \right) \right] \right\}. \quad (3.1)$$

If $\lambda_{t+1} = \lambda_t = 0$ and $\eta_t = 0$, then the shadow cost of internal funds is one, and the Euler equation is identical to the one provided by the perfect-capital-markets model. In the presence of financial market imperfections, however, $\lambda_t = \lambda(K_t, B_t, \xi_t)$ and $\eta_t = \eta(K_t, B_t, \xi_t)$ are state-dependent and time-varying.[19] The first-order condition for debt requires that

$$E_t \left[\frac{1 + \lambda_{t+1}}{1 + \lambda_t} \left(1 + \eta_{t+1} + \frac{\partial \eta_{t+1}}{\partial B_{t+1}} B_{t+1} \right) \right] = 1.$$

The marginal cost of debt determines the shadow cost of funds today vs. tomorrow (i.e., λ_t vs. λ_{t+1}), and hence provides a time-varying discount factor that depends on the level of net financial liabilities, B_t (among

18. A number of papers in the literature estimate this Euler equation directly by assuming a parametric form for the shadow cost term: Himmelberg (1990), Whited (1992), Hubbard and Kashyap (1992), Hubbard, Kashyap, and Whited (1995), and Jaramillo, Schiantarelli, and Weiss (1996).

19. While it is not necessary to resolve such issues for our empirical specification, it is interesting to ask under what conditions the premium on external funds is likely to be stationary. For simplicity, suppose η_{t+1} doesn't depend on θ_{t+1}, so that we can ignore the expectations operator (B_{t+1} and K_{t+1} are known at time t). Then in steady state, a constrained firm would have $\lambda_t = \lambda_{t+1}$, which implies $(\partial\eta/\partial B)B + \eta = 0$. Since $\partial\eta/\partial B > 0$, we would observe $B > 0$ only if $\eta < 0$. This is possible if, for example, the premium η is net of tax advantages or agency benefits. That is, despite the positive marginal premium on debt, the average premium might be negative. In a more general model, a steady-state equilibrium with $\partial\eta/\partial B > 0$ could be maintained by modeling managers as "impatient." In Bernanke, Gertler, and Gilchrist (1998), for example, exogenous firm "failure" generates this behavior.

other state variables). This point is general and does not depend in any specific way on our particular dividend assumption.

3.1 A LINEARIZED EMPIRICAL FRAMEWORK

Let $c(I_t, K_t)$ denote the *marginal* adjustment cost function, and let MPK_t denote the marginal profit function net of adjustment costs and financing costs.[20] For simplicity, assume the discount rate r_t is constant over time and over firms (in the discussion below, we explain how this assumption could easily be relaxed). Then the first-order conditions for the above model with financial frictions can be written

$$\begin{aligned} 1 + c(I_t, K_t) &= E_t \sum_{s=1}^{\infty} \left[\prod_{k=1}^{s} \frac{1-\delta}{1+r} \left(\frac{1+\lambda_{t+k}}{1+\lambda_{t+k-1}} \right) \right] \mathrm{MPK}_{t+s} \\ &= E_t \sum_{s=1}^{\infty} \left(\frac{1-\delta}{1+r} \right)^s \left(\prod_{k=1}^{s} \frac{1+\lambda_{t+k}}{1+\lambda_{t+k-1}} \right) \mathrm{MPK}_{t+s} \\ &= E_t \sum_{s=1}^{\infty} \beta^s \Theta_{t,t+s} \mathrm{MPK}_{t+s} \end{aligned}$$

where the discount factor has been factored into a deterministic component, $\beta = (1-\delta)/(1+r)$, times a stochastic component $\Theta_{t,t+s} \Pi_{k=1}^{s} (1+\lambda_{t+k})/(1+\lambda_{t+k-1})$, which in general will be a function of firm-level variables.

Since the mean of $\Theta_{t,t+s}$ should be near one, we can use a first-order Taylor approximation around $E(\Theta_{t,t+s}) \simeq 1$ and $E(\mathrm{MPK}_{t+s}) \simeq \gamma$ to write

$$\Theta_{t,t+s}\mathrm{MPK}_{t+s} \simeq \gamma_0 + \gamma\Theta_{t,t+s} + \mathrm{MPK}_{t+s}.$$

Furthermore, we can approximate the expression for $\Theta_{t,t+s}$ to get

$$\begin{aligned} \Theta_{t,t+s} &= \prod_{k=1}^{s} \frac{1+\lambda_{t+k}}{1+\lambda_{t+k-1}} \\ &\simeq 1 + \sum_{k=1}^{s} \frac{\lambda_{t+k} - \lambda_{t+k-1}}{1+\lambda_{t+k-1}} \\ &\simeq \text{const} + \sum_{k=1}^{s} \phi \mathrm{FIN}_{t+k}, \end{aligned}$$

where we have assumed that $(\lambda_{t+k} - \lambda_{t+k-1})/(1+\lambda_{t+k-1}) = \phi_0 + \phi \mathrm{FIN}_{t+k}$ is a linear approximation representing the dependence of the shadow discount term on a financial state variable represented by FIN_{t+k}. This functional form assumption for $\Theta_{t,t+s}$ obviously allows us to specify FIN either as net financial liabilities (i.e., B_t), in which case the predicted sign of ϕ is negative, or as net financial assets (i.e., $-B_t$), in which case the predicted

20. In our empirical work, we ignore the marginal reduction of financing costs in our construction of MPK because it is a small effect relative to $\partial\Pi/\partial K$.

sign of ϕ is positive. In our empirical work, we prefer to work with net financial assets.

Note that with additional notation, we could have allowed the stochastic component of the discount factor, $\Theta_{t,t+s}$, to include a time-varying discount factor, r_t. Then the above linearization would include an additional term capturing the effect of r_t. In our empirical work, the inclusion of time dummies in our panel-data regressions accommodates time-varying discount rates. By the same logic, allowing for firm fixed effects accommodates firm-specific discount rates attributable to differences in the average firm-level "beta" as well as differences in the average level of the firm's external finance premium.

It is useful at this point to briefly consider what would constitute a plausible range of values of ϕ for our model. One way to do this is to consider a plausible range of variation for the premium on external funds across firms. Letting σ_r represent the standard deviation of the net external finance premium, our model suggests that $\sigma_r \simeq \phi\sigma_{\text{Fin}}$. Calomiris and Himmelberg (1998) report that the standard deviation for underwriting spreads for seasoned equity issues is 5.8%. For annual data, the measured premium on average loan rates can easily vary by 5 percentage points across firms, or over time for a given firm. Thus a range of 5% to 10% seems reasonable for the *marginal* premium on external funds. In our empirical work below, we use the ratio of cash and equivalents to capital as one measure of FIN_t. This variable has a standard deviation of 0.37, implying that a ballpark figure for ϕ is on the order of 0.1 to 0.3.

Substituting the above approximations for $\Theta_{t,t+s}\text{MPK}_{t+s}$ and $\Theta_{t,t+s}$ into the present value and collecting constant terms yields

$$\begin{aligned}
1 + c(I_t, K_t) &= E_t \sum_{s=1}^{\infty} \beta^s \Theta_{t,t+s} \text{MPK}_{t+s} \\
&= \text{const} + \gamma E_t \sum_{s=1}^{\infty} \beta^s\, \Theta_{t,t+s} + E_t \sum_{s=1}^{\infty} \beta^s \text{MPK}_{t+s} \\
&= \text{const} + \gamma\phi E_t \sum_{s=1}^{\infty} \sum_{k=1}^{s} \beta^s\, \text{FIN}_{t+k} + E_t \sum_{s=1}^{\infty} \beta^s \text{MPK}_{t+s}.
\end{aligned}$$

Estimation requires a functional form for adjustment costs. Following standard practice, we assume that $C(I_t, K_t)$ is quadratic in I_t/K_t, so that marginal adjustment costs are linear in I_t/K_t. We also extend the specification to include a technology shock ω_t. Thus, the marginal adjustment cost function is assumed to be

$$c(I_t, K_t) = \text{const} + \alpha^{-1} (I/K)_t - \omega_t.$$

Under this specification of the adjustment cost technology, the relationship between investment, the present value of future FIN_t, and the present value of future MPK_t is given by

$$(I/K)_t = \text{const} + \alpha\gamma\phi E_t \sum_{s=1}^{\infty} \sum_{k=1}^{s} \beta^s \, \mathrm{FIN}_{t+k} + \alpha E_t \sum_{s=1}^{\infty} \beta^s \, \mathrm{MPK}_{t+s} + \omega_t. \tag{3.2}$$

The standard Q-model of investment is a special case of the above model where $\phi = 0$, and the model is typically estimated using Tobin's Q as a proxy for the present value of future marginal profits, i.e., $Q_t = E_t \sum_{s=1}^{\infty} \beta^s \mathrm{MPK}_{t+s}$. With financial frictions, however, Tobin's Q-values not only future MPK, but also changes in the expected financial status of the firm, $E_t \sum_{s=1}^{\infty} \sum_{k=1}^{s} \beta^s \, \mathrm{FIN}_{t+k}$. Thus Tobin's Q would appear to be a poor choice for estimating investment models when the goal is to identify financial frictions.[21] We elaborate on this point in Section 4. As an alternative to using Tobin's Q, we propose the method used by Abel and Blanchard (1986) and Gilchrist and Himmelberg (1995), which constructs present-value terms by estimating a VAR for the vector of state variables that help to forecast MPK_t and FIN_t.

3.2 THE EXPECTED PRESENT VALUE OF MPK AND FINANCIAL FACTORS USING VAR FORECASTS

In our notation, we now add the subscript i to index firm-level variables. To construct this expectation using a VAR model of the firm's state vector, let x_{it} be a vector containing current and lagged values of MPK_{it}, FIN_{it}, and any other variables containing information that can be used to forecast the future marginal profitability of investment.[22] This information $x_{it-1} \subseteq \Omega_t$ is available at time t when the firm i makes its investment decision. We assume that these variables follow an autoregressive process, and to simplify notation, we write this VAR in companion form as

$$x_{it} = Ax_{it-1} + u_{it},$$

21. Under some specifications of the external finance premium, it is possible to show that Tobin's Q remains a sufficient statistic for investment (see Chirinko, 1993). This further shows why Tobin's Q is a poor choice for estimating investment models when the goal is to detect and quantify the importance of financing constraints. For further evidence on this point, see the simulation results reported by Gomes (1997).

22. The variables included in the forecast VAR should not include lagged investment. In theory, it is feasible and even desirable to include lagged investment in the forecast VAR, but doing so makes it much more difficult to impose the cross-equation restrictions. This is a difficult methodological issue on which we are currently working and which we hope to explore in a future paper.

and we assume $E(u_{it}|x_{it-1}) = 0$. By recursive substitution, the conditional expectation of x_{it+s} given x_{it-1} is easily seen to be

$$E[x_{it+s}|x_{it-1}] = A^{s+1}x_{it-1}.$$

Let MPK_{it} be the first element of x_{it}, and let FIN_{it} be the second element. If we let c_j denote a vector of zeros with a one in the jth position, then $\mathrm{MPK}_{it} = c_1'x_{it}$ and $\mathrm{FIN}_{it} = c_2'x_{it}$. Using this notation, the expected present value of MPK is given by

$$\begin{aligned}\mathrm{PV}_{it}^{\mathrm{MPK}} &= E_{it}\sum_{s=1}^{\infty}\beta^s\,\mathrm{MPK}_{it+s}\\ &= \sum_{s=1}^{\infty}\beta^s E[\mathrm{MPK}_{it+s}|x_{it-1}]\\ &= c_1'\sum_{s=1}^{\infty}\beta^s A^{s+1}x_{it-1}\\ &= c_1'(I-\beta A)^{-1}\beta A^2 x_{it-1}.\end{aligned}$$

Analogously, using our notation $\mathrm{FIN}_{it} = c_2'x_{it}$, the expected present value of financial factors is given by[23]

$$\begin{aligned}\mathrm{PV}_{it}^{\mathrm{FIN}} &= E_t\sum_{s=1}^{\infty}\sum_{k=1}^{s}\beta^s\,\mathrm{FIN}_{t+k}\\ &= \sum_{s=1}^{\infty}\sum_{k=1}^{s}\beta^s\,E[\mathrm{FIN}_{it+k}|x_{it-1}]\\ &= c_2'\sum_{s=1}^{\infty}\sum_{k=1}^{s}\beta^s A^{k+1}x_{it-1}\\ &= c_2'(1-\beta)^{-1}(I-\beta A)^{-1}\beta a^2 x_{it-1}.\end{aligned}$$

These present-value formulae allow us to specify a structural reduced-form model of investment that is linear in x_t:

$$(\mathrm{I/K})_{it} = \mathrm{const} + \alpha(\mathrm{PV}_{it}^{\mathrm{MPK}}) + \alpha\gamma\phi(\mathrm{PV}_{it}^{\mathrm{FIN}}) + f_i + d_t + \omega_{it}, \tag{3.3}$$

The terms f_i and d_t represent fixed firm and year effects that are controlled for in the estimation. The residual satisfies the moment condition $E[\omega_{it}x_{it-s}] = 0$ for all s, so all lagged values of x_{it} are valid for estimation.

23. Here we make use of the result that $\Sigma_{s=1}^{\infty}\beta^s\,\Sigma_{k=1}^{s}A^k = (1-\beta)^{-1}(I-\beta A)^{-1}\beta A$.

4. *Model Implications and a Discussion of the Recent Investment Literature on Financing Constraints*

The empirical framework in the previous section shows that in a (linearized) model with financial frictions, investment is a function of both (1) the expected present value of future MPKs, or *fundamental Q*, and (2) the expected present value of future financial state variables of the firm, or *financial Q*. That is,

$$(I/K)_t = \underbrace{\alpha E_t \sum_{s=1}^{\infty} \beta^s \mathrm{MPK}_{t+s}}_{\text{fundamental } Q} + \underbrace{\alpha\gamma\phi E_t \sum_{s=1}^{\infty}\sum_{k=1}^{s} \beta^s \mathrm{FIN}_{t+k}}_{\text{financial } Q}.$$

Although the above equation has not been used in past research, it nevertheless explains the intuition underlying many of the empirical specifications in the literature surveyed by Hubbard (1998). Specifically, it shows that investment equations based only on fundamental Q contain an omitted variable in the error term, so that investment will appear to be excessively sensitive to any explanatory variable (e.g., cash flow) that helps to predict current or future values of FIN_t. This equation also shows that investment can be excessively sensitive even to nonfinancial variables such as sales growth, provided such variables help to forecast future financial conditions.

While it is easy in theory to see why investment should be excessively sensitive to variables that are correlated with financial Q, it is difficult in practice to assign econometric interpretations to the explanatory variables in a reduced-form regression. The interpretation of cash flow, for example, is not obvious, because it could predict fundamental Q as well as financial Q. By the same logic, the role of Tobin's Q is theoretically ambiguous, because Tobin's Q measures the average value of capital, and this is closely related to financial Q in some theoretical models. In Gertler (1992), for example, the firm's net worth determines the degree to which external investors can write contracts that reduce moral hazard on the part of insiders, and thus determines the severity of financing constraints. In his model, or in any model where the specification of the financial friction depends on net worth, Tobin's Q will contain information about both fundamental Q and financial Q, and will therefore be difficult to interpret in the absence of more model structure.[24]

24. There are additional problems in the empirical literature that are usefully viewed in the context of the above model. First, as we argued in Gilchrist and Himmelberg (1995), there are reasons to believe that Tobin's Q is a poor proxy for fundamental Q. For example, if firms enjoy market power, or if firms employ multiple quasifixed factor inputs, or if Tobin's Q is measured with noise, then Tobin's Q will not be a sufficient

A recent paper by Cummins, Hassett, and Oliner (1997) provides a useful illustration of this problem. In their paper, the idea is to use analysts' earnings forecasts to construct fundamental Q. They use data obtained from IBES, which provides forecasts at one-year and two-year horizons, $f1_{it}$ and $f2_{it}$, as well as forecasts of the expected annual growth rate, g_{it}, of earnings in years 3 through 5. Assuming a discount rate β, they approximate the present value of earnings by assuming that earnings continue to grow at the rate g_{it} for 10 years. Thus, they assume the present value of earnings is well approximated by[25]

$$PV_{it}^{EARN} = \beta\, f1_{it} + \beta^2 \left(\sum_{s=0}^{8} \beta^s (1 + g_{it})^s \right) f2_{it}.$$

Defining "fundamental Q" as PV_{it}^{EARN}/K_{it}, they regress investment on this measure of fundamental Q and current earnings and report that investment displays no "excess sensitivity" to current earnings.[26] But does PV_{it}^{EARN}/K_{it} measure fundamental Q or financial Q? Our model makes it clear that identification requires two separate terms; at best, PV_{it}^{EARN}/K_{it} combines these two Q-variables into one term. Indeed, PV_{it}^{EARN}/K_{it} is conceivably a better measure of financial Q than of fundamental Q,

---

statistic for fundamental Q, and investment will display excess sensitivity to any variable (including financial variables) that contains forecast information for the future marginal profitability of capital. In other words, poor proxies for fundamental Q may spuriously give rise to excess cash-flow sensitivity and thus overstate the importance of financial Q. For this reason, the literature has generally placed more emphasis on the fact that cash-flow sensitivity tends to be greater for firms that are more likely to face financial frictions on the basis of some *a priori* measure (such as size, dividend payout, or access to public debt markets). Even so, the interpretation of excess cash-flow sensitivities remains controversial. See the critique by Kaplan and Zingales (1997) and the rebuttal by Fazzari, Hubbard, and Petersen (1996).

25. Strictly speaking, they do not report this equation, but this is what we infer based on our reading of the verbal description in their paper.

26. In addition to their conceptual failure to distinguish between fundamental Q and financial Q, we have some doubts about the robustness of their empirical results. We have examined the IBES data ourselves, and in contrast to the claims made by Cummins, Hassett, and Oliner, we do not find that the inclusion of earnings forecasts eliminates the explanatory power of cash flow and other balance-sheet variables. We have not yet been able to trace the sources of this discrepancy, but one possible explanation may be the fact that Cummins, Hassett, and Oliner use IBES earnings per share (EPS) and shares outstanding to construct current total earnings. This is problematic, because the number of shares reported by IBES corresponds to the number outstanding on the date of the *forecast* and does not correspond to the number outstanding on the date of the *fiscal year end*. Stock splits and other share adjustments make this calculation impossible. When we use Compustat cash flow and earnings measures instead of IBES earnings, we find no evidence to suggest that earnings forecasts are sufficient statistics for investment. In the end, however, the robustness of their results is a red herring. Our main objection is conceptual, because the present value of earnings is a good measure of financial Q.

because the only difference between earnings and cash flow is depreciation, whereas it is a long way from earnings to MPK1 (see the discussion in Section 2.1.1, and the income sheet reported in Section 2.2).

5. *Empirical Results on the Structural Model*

In this section of the paper we explore the extent to which investment responds to fundamental Q versus financial Q in our structural model. We begin our analysis using the full sample. We then look at how our results vary across subsamples where the data are split based on indicators that capture a firm's likely degree of access to finance.

The estimates of equation (3.3) described in the previous section are constructed as follows. First, a VAR(2) is specified with the following vector of variables: MPK1, MPK2, and the state variable FIN_{it} measuring the firm's financial status. Because MPK1 depends on sales, and MPK2 depends on operating income, the VAR system includes information on both revenues and profits as well as financial factors. The instrument set includes lags one and two each of the variables used in the forecasting system. For the regressions that do not include a PDV of financial factors, we include the cash and equivalents to capital ratio in the forecasting system. Second, this VAR is used to construct the fundamental Q, $\text{PV}_{it}^{\text{MPK}}$, and the financial Q, $\text{PV}_{it}^{\text{FIN}}$:

$$\text{PV}_{it}^{\text{MPK}} = c_2'(I - \beta A)^{-1}\beta A^2 x_{it-1},$$
$$\text{PV}_{it}^{\text{FIN}} = c_3'(1 - \beta)^{-1}(I - \beta A)^{-1}\beta A^2 x_{it-1}.$$

Finally, investment is regressed on $\text{PV}_{it}^{\text{MPK}}$ and $\text{PV}_{it}^{\text{FIN}}$ using the same set of instrumental variables as those used when estimating the VAR. All regressions are run using the forward mean-differencing transformation described in the appendix.

We consider two alternative definitions of the state variable measuring financial Q: the ratio of cash and equivalents to capital, $(\text{CE}/\text{K})_{it}$, and the ratio of financial working capital minus long-term debt to capital, $(\text{FW}/\text{K} - \text{LD}/\text{K})_{it}$.[27] The first definition captures the short-term liquid asset position of the firm. It thus reflects the amount of savings inside the firm. It also reflects the share of assets that is most easily used as collateral. The second variable measures the leverage position of the firm, net of current liquid assets. A distinct advantage of both of these variables is that they measure financial stocks rather than financial flows. Because finan-

27. We define "financial working capital" as current assets minus current liabilities plus inventories. Exact definitions are provided in Table 2.

Table 5 FULL-SAMPLE RESULTS

	Estimated Parameter Value					
Variable	*Sales-Based MPK*			*OI-Based MPK*		
PV_{it}^{MPK}	1.48 (0.261)	1.16 (0.229)	1.27 (0.237)	1.22 (0.233)	1.07 (0.218)	1.1 (0.216)
$PV_{it}^{CE/K}$	— —	0.056 (0.008)	— —	— —	0.05 (0.011)	— —
$PV_{it}^{FW/K}$	— —	— —	0.048 (0.008)	— —	— —	0.035 (0.01)
R^2	0.356	0.385	0.394	0.377	0.401	0.398
P	0.00	0.00	0.00	0.007	0.118	0.009
N_{obs}	8520	8520	8520	8520	8520	8520

Adjusted standard errors in parentheses (see appendix).

cial stocks are less directly linked to the marginal profitability of capital, their present values are more likely orthogonal to PV_{it}^{MPK} than are present values that are constructed from financial flows.[28]

5.1 FULL-SAMPLE RESULTS

Table 5 reports estimates of parameters corresponding to α and $\alpha\gamma\phi$ in equation (3.3). For comparison purposes, we report results using the two alternative definitions of PV_{it}^{MPK} based on the two alternative measures of MPK. As described in Section 2, the first definition is based on the ratio of sales to capital, while the second definition is based on the ratio of operating income to capital. These results are estimated under the assumption that the time to build is one period, and the information set used by the firm is based on time-$(t-1)$ information. The first three columns contain results using the sales-based MPK, and the last three

28. We also investigated a third specification of financial Q using cash flow, $(CF/K)_{it}$, as the financial state variable. This specification of financial Q was a robust explanatory variable in all of the specifications that we tried. Unfortunately, it was also highly correlated with our measure of fundamental Q, just as one might have anticipated from our theoretical discussion of MPK in Section 2.2. As a consequence of this collinearity, the coefficient on fundamental Q in these regressions was typically insignificant and occasionally even negative. We obviously do not view this as evidence that adjustment costs are negative. Rather, we view this as evidence of model misspecification caused by the fact that financial Q was picking up information about fundamental Q. As we explain in the text, when we used stock measures of the firm's financial status, this was not a problem. We consider additional theoretical work on the financial side of the model to be an important direction for future research because this could help resolve this specification choice.

columns contain results using MPK based on operating income. This table reveals one of the major results in the paper: Fundamental Q does very well in explaining the investment data. In particular, the coefficients on PV_{it}^{MPK} suggest rapid adjustment speeds and hence reasonable adjustment costs.

Despite the success of the sales-based measure of fundamental Q in explaining investment, investment is still highly responsive to financial factors. In all cases, financial Q is an important determinant of investment. With standard errors adjusted for the fact that the present-value terms are generated from previous regressions, the t-statistics are on the order of 3–8 for all three variables reported in the first three columns of Table 5.[29]

Besides reporting coefficient values, Table 5 also reports two diagnostic statistics: the P-value from a chi-squared test of orthogonality between error terms and instruments, and the R^2 from the regression. The orthogonality tests reject the model overwhelmingly and suggest model misspecification, even when financial factors are included. As we show below, this model misspecification is due to firms that are most likely to face severe financial constraints.

We now consider the results using fundamental Q constructed from the operating-income-based measure of MPK. In the full-sample results (Table 5, columns 4–6), it appears to make little difference whether our measure of MPK is based on sales or operating income. We obtain similar coefficients for adjustment costs, and approximately the same R^2. The fact that the coefficients on fundamental Q are fairly close across both measures suggests that we are using the correct normalizations of sales to capital and operating income to capital ratios when constructing MPK measures.

5.2 RESULTS BASED ON SAMPLE SPLITS

We now consider how our results vary across subsamples of firms when the subsamples are designed to sort firms by their ability to access financial markets. The traditional argument for performing subsample splits in the literature is that not all firms have the same degree of access to financial markets. The response of investment by firms with costly access is more likely to be sensitive to financial factors than that of firms with cheap access to external financial markets. Sample splitting thus provides a way to test for the presence of financial factors, even with imperfect measures of investment fundamentals. For example, large

29. The standard-error correction that results from generated regressors raised the standard errors by approximately 75–100%.

firms and firms that have issued public debt or have established commercial-paper programs are likely to have established lines of credit that may be drawn down during periods of low profitability. As a result, the investment policy of such firms may not be responsive to swings in balance-sheet conditions. By not taking such differences into account, we may not obtain an accurate description of the importance of financial factors in investment. Also, to the extent that we can identify a subset of firms that do not face financial frictions, and for whom the baseline investment model without financial frictions fits well, we can be more confident that our underlying investment model is correct. Such a result would imply that the presence of financial factors does not simply capture an undetermined source of model misspecification.

When splitting the sample, we consider three alternative criteria. The first criterion sorts firms according to whether or not they have an S&P bond rating. Because most firms that issue public debt obtain a bond rating, this effectively sorts the full sample into firms that have issued public debt in the past, versus those that have not. Calomiris, Himmelberg, and Wachtel (1995) argue that public-debt issuance is a good indication that a firm has low-cost access to capital markets, because firms with serious adverse selection or moral hazard problems are forced to rely on intermediated finance such as bank debt and private placements. Because the population of public-debt issuers is relatively stable over time, this selection criterion has the advantage of being relatively exogenous with respect to the time-series variation in the data. It has the disadvantage of only capturing a subset of the best-quality firms.

The other two criteria that we use to split the sample are the dividend payout ratio and firm size. The dividend payout ratio was originally used by Fazzari, Hubbard, and Petersen (1988) and has been employed in a number of additional studies. The size split has also been used extensively to distinguish between *constrained* and *unconstrained* firms (Gertler and Gilchrist, 1994; Carpenter, Fazzari, and Petersen 1996). The rationale for splitting the sample according to dividend policy is that when firms declare dividends, they endogenously reveal that they have a low shadow value of internal funds. For a number of reasons, firm size is another common way of identifying firms with low external-financing premiums. For one, it is plausible that costs of obtaining funds contain a significant fixed-cost component. The presence of such increasing returns suggests that small firms face higher costs of obtaining external funds than large firms. In addition, size is a proxy for age and other unobservable firm attributes that affect the degree to which public information about the firm's investment projects is available. Among publicly traded firms, smaller, newer firms are less likely to be tracked by analysts and less likely to have been through multiple equity or debt offerings that result in substantial production of public information.

While the bond-rating categorization is based on a zero–one variable (rating versus no rating), both the dividend payout ratio and size are continuous. Because we wish to distinguish firms with cheap access to credit from firms that face potential credit frictions whose investment will be responsive to financial state variables, we divide the sample conservatively and classify firms who are in the top one-third of the dividend payout or size distribution as likely to be unconstrained.[30]

For each sample split, we allow the VAR forecasting system to vary across the constrained and unconstrained subsamples. By allowing the VAR forecasting system to vary across subsamples we correct for any systematic differences in forecasting properties that may bias results. We report the results using both the sales-based MPK and the operating-income-based MPK, and consider the cash-and-equivalents variable as our financial state variable. The regression results for this exercise are reported in Tables 6 and 7.

The results from the sample-splitting exercise provide strong evidence that financial factors are important determinants of investment, principally for firms classified as constrained. Table 6 reports the results for the bond-rating split, using both the sales-based and operating-income-based measures of MPK. Using either definition of MPK, firms with a bond rating show no sensitivity of investment to financial factors. Thus all of the contribution of financial factors in explaining investment comes through firms without bond ratings. The orthogonality conditions for the baseline investment model are not rejected for firms classified as unconstrained. This result implies that the underlying investment model does well at explaining the data, in the absence of financial frictions. In addition, adding the financial factor adds very little in terms of explanatory power as measured by R^2 for unconstrained firms. For firms without bond ratings, the coefficients on financial factors increase by 40% relative to the full sample results.

Table 6 also shows that, for bond-rated firms, the sales-based MPK measure does a much better job explaining investment than the operating-income-based MPK. In particular, the coefficient on fundamental Q is much higher and the model is not rejected when using a sales-based

30. Because we compute the 66th percentile for the dividend payout and size variables before dropping firms because of missing values, we end up with slightly different sample sizes than the one-third–two-thirds split of the original sample. The actual values used are a ratio of common dividends to capital greater or less than 0.05 and real sales greater or less than $364 million. Real sales were constructed using the GDP deflator. This cutoff for real sales is close to the value of $250 million used by Gertler and Gilchrist (1994) in their study of small versus large manufacturing firms.

Table 6 BOND RATED VERSUS NON-BOND-RATED FIRMS

Variable	Estimated Parameter Value					
	Bond Rating			No Bond Rating		
	Sales-Based MPK					
PV_{it}^{MPK}	1.32	1.26	1.21	1.55	1.24	1.32
	(0.603)	(0.622)	(0.536)	(0.399)	(0.353)	(0.36)
$PV_{it}^{CE/K}$	—	0.003	—	—	0.07	—
	—	(0.01)	—	—	(0.015)	—
$PV_{it}^{FW/K}$	—	—	0.006	—	—	0.049
	—	—	(0.021)	—	—	(0.013)
R^2	0.419	0.428	0.426	0.318	0.342	0.358
P	0.889	0.789	0.743	0.00	0.00	0.00
N_{obs}	1720	1720	1720	4420	4420	4420
	Operating-Income-Based MPK					
PV_{it}^{MPK}	0.318	0.254	0.357	1.22	1.02	0.993
	(0.205)	(0.175)	(0.218)	(0.364)	(0.34)	(0.313)
$PV_{it}^{CE/K}$	—	0.013	—	—	0.063	—
	—	(0.009)	—	—	(0.018)	—
$PV_{it}^{FW/K}$	—	—	−0.006	—	—	0.04
	—	—	(0.021)	—	—	(0.014)
R^2	0.418	0.45	0.436	0.339	0.357	0.354
P	0.025	0.018	0.034	0.058	0.082	0.01
N_{obs}	1720	1720	1720	4420	4420	4420

Adjusted standard errors in parentheses (see appendix).

MPK. These findings suggest that our sales-based MPK captures most of the information about fundamentals in the absence of credit frictions.

Table 7 reports results for alternative sample splits using the sales-based measure of fundamentals (similar conclusions are reached using MPK based on operating income). Small firms are clearly more responsive to financial factors than large firms. For the dividend split, the differences across subsamples are not so obvious. There is less of a difference in estimates of ϕ for low- vs. high-dividend firms, once one corrects for the fact that the coefficient estimate on fundamental Q is much lower for high-dividend firms.

5.3 GOODNESS OF FIT AND ROBUSTNESS EXERCISES

We conducted a variety of robustness exercises that are not reported in the tables. First, we investigated the robustness of our results across industries. While it is not possible to estimate separate investment equations for each industry, it is possible to consider a more homogeneous sample than the full set of manufacturing firms considered above. For robustness, we reestimated all the regressions reported in Table 5 for a sample that is limited to durable-goods industries only (two-digit SICs between 3200 and 3999). The argument for doing this exercise is that the durable-goods industries are much more homogeneous than the nondu-

Table 7 ALTERNATIVE SAMPLE SPLIT CRITERIA

	Estimated Parameter Value					
Variable	*High Dividend Payout*			*Low Dividend Poyout*		
PV_{it}^{MPK}	0.422	0.215	0.516	1.84	1.44	1.57
	(0.2)	(0.117)	(0.234)	(0.428)	(0.384)	(0.362)
$PV_{it}^{CE/K}$	—	0.038	—	—	0.085	—
	—	(0.007)	—	—	(0.014)	—
$PV_{it}^{FW/K}$	—	—	0.031	—	—	0.062
	—	—	(0.013)	—	—	(0.011)
R^2	0.31	0.385	0.363	0.37	0.401	0.41
P	0.00	0.022	0.023	0.00	0.00	0.00
N_{obs}	2900	2900	2900	5240	5240	5240
	Large Firm			*Small Firm*		
PV_{it}^{MPK}	0.714	0.513	0.616	1.35	1.28	1.39
	(0.174)	(0.18)	(0.173)	(0.325)	(0.326)	(0.357)
$PV_{it}^{CE/K}$	—	0.012	—	—	0.096	—
	—	(0.006)	—	—	(0.015)	—
$PV_{it}^{FW/K}$	—	—	0.014	—	—	0.052
	—	—	(0.009)	—	—	(0.011)
R^2	0.56	0.553	0.553	0.278	0.303	0.317
P	0.063	0.016	0.014	0.00	0.00	0.00
N_{obs}	3260	3260	3260	5140	5140	5140

Adjusted standard errors in parentheses (see appendix).

rables industries. In addition, these industries may have different time-series properties that would be better captured by their own set of time dummies. If so, the forecasting equations obtained from the VAR may perform better. This exercise provides very similar results to those obtained in Table 5. For the durables subsample, fundamental Q provides considerable explanatory power for investment and reasonable estimates of adjustment costs. Nonetheless, present values of financial state variables are still important determinants of investment, in both economic and statistical terms.

The second exercise we perform is to check for robustness by excluding the smallest firms in the sample. Because size and bond rating are correlated, it is useful to know if the results based on sample splits are purely a size effect, generated by the smallest firms in the sample. For a variety of reasons, such firms may respond differently to both fundamental Q and financial Q. To examine this issue, we reconsidered the bond-rating splits after dropping all firms with total assets less than $100 million in real terms (1992 dollars). As one would expect, the financial effect is somewhat weaker when one drops the small firms. We nonetheless still find substantial differences in response between non-bond-rated and bond-rated firms, with bond-rated firms showing no response to financial Q, and non-bond-rated firms showing an economically and statistically significant response. This finding implies that while size is important, it is nonetheless the case that some medium-size firms do not have perfect access to debt and equity markets, and these account for an important component of the overall degree of excess sensitivity measured in the data.

The final exercise we consider is to include lagged investment in the empirical specification. The convex adjustment cost structure developed in this paper suggests that only the fundamentals and financial variables should explain investment and that lagged investment should not matter. Empirically, however, lagged investment may matter for three reasons. First, the adjustment cost structure could be richer than what we have modeled. If this is the case, the model may exhibit more inertia than one would expect absent such misspecification. Second, it is possible that investment itself helps forecast the future fundamentals and/or financial factors, in which case our present value constructs would be measured with an error that is correlated with lagged I/K. Finally, it is possible that the model is well specified but that shocks exhibit serial correlation, in which case the presence of lagged investment would reveal such correlation.

Our empirical results uniformly reject the hypothesis that lagged investment does not matter for current investment, even after controlling for both fundamentals and financial factors. The coefficient on lagged investment is on the order of 0.1–0.2 and highly significant. While this result suggests model misspecification, it is also the case that including lagged investment has little effect on the estimated parameter values and does not reduce the importance of financial factors in the investment equation.[31]

To further investigate the role of lagged investment, we estimated a VAR forecasting system that included investment as one of the system variables. With this specification, lagged investment is then explicitly included in our construction of present-value forecasts. Without imposing model consistency between the forecasting system and the empirical specification of the structural investment equation, we re-estimated the investment equation, allowing lagged investment to enter freely on the right-hand side. Other than raising the coefficient on lagged investment somewhat, this exercise produced little change in any of the coefficient estimates, implying that investment's ability to forecast future MPK and financial variables does not explain the presence of lagged investment. In future work, it would be useful to further investigate the role of alternative explanations like richer adjustment costs and serially correlated shocks. Finally, it is worth noting that lagged investment is a significant explanatory variable in more standard Q regressions as well. Thus, this form of model misspecification does not result from our particularly empirical specification, but is instead endemic to a wide variety of empirical specifications of investment equations.[32]

5.4 THE EMPIRICAL CONTRIBUTION OF FINANCIAL FACTORS

Having estimated the basic model, and having considered a variety of robustness issues, we now use the structural coefficients to gauge the likely empirical contribution of the financial factors to investment. We do this by shocking the VAR used to construct the forecast and tracing out the time path for both fundamental Q and financial Q. To obtain the time path of investment, we feed these two Q-values into the investment

31. To be more precise, omitting lagged investment from both the regressors and the instruments produced very similar coefficients to regressions that include lagged investment as both a regressor and instrument. The main difference in results is that the coefficients on both fundamental and financial terms rise somewhat to offset the inertia introduced by lagged investment in the empirical specification. As a result, the dynamic responses of the models look very similar, whether or not one includes lagged I/K on the right-hand side.

32. While this is especially true of panel data, it also tends to be true for aggregate data as well (see Abel and Blanchard (1986) for example). Kiyotaki and West (1996) provide a notable counterexample with their empirical model of investment using aggregate postwar Japanese data. In particular, they attribute all of the explanatory power of lagged investment to its ability to predict future fundamentals.

Table 8 DYNAMIC INVESTMENT RESPONSE TO FUNDAMENTAL VERSUS FINANCIAL Q

	Response					
Variable	$T=0$	$T=1$	$T=2$	$T=3$	$T=4$	$T=5$
MPK_{it}	0.046	0.032	0.019	0.011	0.006	0.004
$(\text{CE/K})_{it}$	0.042	0.029	0.018	0.011	0.007	0.004
$\hat{\alpha}_1\text{PV}_{it}^{\text{MPK}}$	0.033	0.019	0.011	0.007	0.004	0.002
$\hat{\alpha}_1\text{PV}_{it}^{\text{MPK}}+\hat{\alpha}_1\hat{\gamma}\phi\text{PV}_{it}^{\text{CE/K}}$	0.041	0.024	0.014	0.008	0.005	0.003
Excess response (%)	0.245	0.261	0.275	0.288	0.299	0.308

specification using the parameter values reported in column 2 of Table 5.[33] To gauge the contribution of financial factors, we do this exercise both with and without financial Q. The results are reported in Table 8.

A one-standard deviation shock raises MPK1 by 0.046 units. The ratio of cash and equivalents to capital increases by slightly more than 0.04 units as balance sheets are strengthened in the wake of the expansionary shock. The response of fundamental Q implies an increase in the investment rate next period of 0.033. Adding in the contribution of financial Q raises the overall response of investment to 0.041. Thus in the first period, financial Q adds an additional 25% to the baseline response obtained from shutting down movements in the present value of cash and equivalents. In the next few periods, we also obtain 25–30% magnification. Thus, the financial effect has a substantial contribution to the overall investment response of the average firm.

Using a combination of the coefficient estimates and the impulse response to financial Q, we can compute the implied response to the one-period return that generated the additional movement in investment relative to the baseline model. With $\beta = 0.8$, $\gamma = 0.2$, and an estimate of α around unity, our estimate of ϕ is approximately 0.2.[34] In the initial period, financial Q rises by approximately 0.04. This implies that the one-period return rose by 80 basis points in the first period, before slowly returning to steady state. Given the fact that the average postwar spread between the prime rate and T-bill is 2%, and bank loans are often quoted at 1–2% above or below prime depending on credit quality, this strikes us as a moderate response for the premium on external funds.

33. The model with lagged investment produces similar results.
34. This estimate is very much in line with our ballpark figures of 0.1–0.3 discussed above.

6. Conclusions

In this paper, we argue that by combining careful measurement of MPK with structural VAR methods, it is possible to improve on existing methods for identifying the financing role of cash flow and other financial variables in reduced-form investment equations. We examine two strategies for imposing structure on VARs to identify the effect of financial factors on investment.

In our first strategy, we use a recursive ordering to structure the contemporaneous relationship among shocks in the VAR. This allows us to identify shocks to cash flow that are orthogonal to current MPK. Such shocks elicit a sustained response from investment over a three-year period. Such shocks also predict a fall in future MPK, suggesting that cash flow matters above and beyond its ability to predict investment fundamentals. Because the future response of MPK to an orthogonal cash-flow shock is negative, the investment response likely understates the effect that cash flow has on investment via lower financing costs.

In our second strategy, we estimate a linearized version of a structural model of investment that embeds financial frictions. This structural framework shows that investment depends not only on the present value of the future marginal profitability of capital (MPK), which we call "fundamental Q," but also on the present value of future shadow values of internal funds, which we call "financial Q." In contrast to previous work on financing constraints using firm-level panel data, we first explicitly relate these shadow values to observable financial state variables and then use VARs to construct the present-value terms corresponding to both fundamental Q and financial Q.

Our empirical results using the structural model show that for a wide variety of specification choices, investment is responsive to both fundamental Q and financial Q. We argue that the values of the estimated structural parameters are reasonable on *a priori* grounds, and that the estimated effect of financial factors on investment is quantitatively significant. For the average firm in our sample, financial factors amplify the overall investment response to an expansionary shock by 25%, relative to a baseline model where such effects are shut down. Consistent with the theory underlying financial-market imperfections, small firms and firms without bond ratings show the strangest response to financial factors, while bond-rated firms show little if any response. Because bond-rated firms account for 50% of aggregate manufacturing investment, our results suggest that the overall amplification of manufacturing investment is somewhat less than 25%.

Appendix

This appendix briefly describes the econometrics used in the paper. First, it describes our approach to estimating panel-data VARs, and second, it describes the adjusted standard-error calculations required by our two-step procedure for estimating the structural parameters of the investment equation. Our approach to estimating panel-data VARs follow Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Keane and Runkle (1992), and Arellano and Bover (1995), among others, which consider the treatment of fixed firm effects in the presence of predetermined but not strictly exogenous explanatory variables. Our treatment of the generated regressor problem introduced by our two-step estimation technique follows Newey (1984).

A.1 ESTIMATING VECTOR AUTOREGRESSIONS USING PANEL DATA

Let $y_{it} = \{y^1_{it}, \ldots, y^M_{it}\}'$ be an $M \times 1$ vector of variables observed in panel data, where i indexes cross-section observations and t indexes time-series observations. Then the mth equation of a P-lag VAR can be written as

$$y^m_{it} = x'_{it}b^m + \alpha^m_i + \gamma^m_t + u^m_{it},$$

where $x_{it} = \{y'_{it-1}, \ldots, y'_{it-p}\}'$ is an $MP \times 1$ vector of lagged endogenous variables (the same for each equation of the VAR), b^m is an $MP \times 1$ vector of slope coefficients, α^m_i is a fixed firm effect, γ^m_t is an aggregate shock (time dummy), and u^m_{it} is an idiosyncratic shock satisfying

$$E(u^m_{it}|f^m_i, \gamma^m_t, x_{it}, x_{it-1}, x_{it-2}, \ldots) = 0.$$

This conditional moment implies $E(x'_{it}u^m_{it+s}) = 0$ for all $s \geq 0$. If the model did not include fixed firm and year effects, we could use OLS to obtain estimates of b^m for all m. However, the presence of unobserved fixed effects (which, by virtue of the lagged dependent variable, are correlated with x_{it}) requires panel-data techniques for obtaining consistent estimates of b^m. To deal with fixed year effects is trivial; we can either estimate dummy variables or, more simply, transform the above model again to deviations from year-specific means. In the exposition below, we assume that y^m_{it} and x_{it} have already been transformed to remove year effects.

To remove the fixed effects α^m_i, we transform the model to deviations from forward means. Let $\bar{y}^m_{it}$ and $\bar{x}_{it}$ denote the means constructed from the future values of y_{it}^m and x_{it} available in the data, and let $\tilde{y}_{it}^m$ and $\tilde{x}_{it}$ denote the data transformation given by

$$\tilde{y}_{it}^m = w_{it}(y_{it}^m - \bar{y}_{it}^m),$$
$$\tilde{x}_{it} = w_{it}(x_{it} - \bar{x}_{it}),$$

where $w_{it} = \sqrt{(T_i - t)/(T_i - t - 1)}$, and T_i denotes the last year of data available (among the nonmissing observations) for observation i. Note that in the last year of the data for observation i, the transformation is unavailable (there are no future values for the construction of $\bar{y}_{it}^m$ *and* $\bar{x}_{it}$), so this observation is set to missing. This transformation sets α_i^m to zero, so the transformed model is

$$\tilde{y}_{it}^m = \tilde{x}_{it}'b^m + \tilde{u}_{it}^m.$$

If the original error term u_{it}^m is homoscedastic, this transformation preserves homoscedasticity, and does not induce serial correlation. This transformation preserves instruments, because all current and lagged values of x_{it} remain uncorrelated with the tranformed error term: $E(x_{it-s}\tilde{u}_{it}^m) = 0$ for all $s \geq 0$. These moment conditions suggest the use of an efficient GMM estimator for b^m. In theory, many instruments (all lags of x_{it}) are potentially available in the sample. In practice, to avoid finite-sample problems, we use only current values of x_{it}. That is, we assume $z_{it} = x_{it}$. Combining moment conditions for all equations, our GMM estimator is based on $E(\tilde{u}_{it} \otimes z_{it}) = 0$.

This model can be expressed in matrix notation as follows. Let $\tilde{y}^m = \{\tilde{y}_{11}^m, \tilde{y}_{12}^m, \ldots \tilde{y}_{NT_N}^m\}'$ denote the stacked vector of observations on $\tilde{y}_{it}^m$ for the mth equation, stacking only observations for which $\tilde{y}_{it}^m$, $\tilde{x}_{it}$, or $\tilde{z}_{it}$ are not missing for any m. Similarly, let Z, $\tilde{X}$, and $\tilde{u}^m$ be the stacked observations on z_{it}', x_{it}', and u_{it}^m, respectively. Then the model for the observations in the data (expressed in differences from forward means) can be written

$$\tilde{y}^m = \tilde{X}b^m + \tilde{u}^m.$$

To write the expression for the GMM estimates of the $M^2P \times 1$ vector of slope coefficients $b = \{b^{1\prime}, \ldots, b^{M\prime}\}'$, stack the moments from all M equations to form the $ML \times 1$ vector of moment conditions $E(\tilde{u}_{it} \otimes z_{it}) = 0$, where $\tilde{u}_{it} = \{\tilde{u}_{it}^1, \ldots, \tilde{u}_{it}^M\}'$. Let y be the $MN^* \times 1$ vector $y = \{\tilde{y}^{1\prime}, \ldots, \tilde{y}^{M\prime}\}'$ formed by stacking the vectors of observations on the M equations, and let $X = I_M \otimes \tilde{X}$, $Z = I_M \otimes \tilde{Z}$, and $W = (Z'Z)^{-1}$. It turns out, then, that the vector of slope coefficients b is

$$\hat{b}_{\text{GMM}} = (X'ZWZ'X)^{-1}X'ZWZ'y,$$

where W is a positive semidefinite weighting matrix. The efficient GMM estimator is obtained by choosing $W = \hat{V}_1^{-1}$, where $\hat{V}_1$ is a consistent estimate of the asymptotic covariance of the sample moments, $(1/N^*) \Sigma_{i=1}^{N} \Sigma_{t=1}^{T_i} (\tilde{u}_{it} \otimes z_{it})$. A convenient estimator of V is

$$\hat{V}_1 = \frac{1}{N^*}\sum_{i=1}^{N}\sum_{t=1}^{T_i}(\hat{u}_{it} \otimes z_{it})(\hat{u}_{it} \otimes z_{it})',$$

where $N^* = \Sigma_{i=1}^{N} T_i$ is the total number of observations in the (unbalanced) panel, T_i denotes the number of nonmissing time-series observations available for firm i, and $\hat{u}_{it}$ is the residual estimate of the transformed error term $\tilde{u}_{it}$, constructed using a consistent preliminary estimate of b (two-stage least squares). Note that it is not necessary to include autocovariance terms in the expression for $\hat{V}_1$, since, by assumption, $E(\tilde{u}_{it}\tilde{u}'_{it-s}|z_{it}, \ldots, z_{it-s}) = 0$ for all $s > 0$.

Finally, a robust estimate of the asymptotic covariance of b_{GMM} is given by

$$\text{Est Var}(\hat{b}_{\text{GMM}}) = (\tilde{X}'ZWZ'\tilde{X})^{-1}\tilde{X}'Z\check{V}_1Z'\tilde{X}(\tilde{X}'ZWZ'\tilde{X})^{-1},$$

where $\check{V}_1$ is an estimated like $\hat{V}_1$ using estimates of the transformed residuals derived from the GMM estimate $\hat{b}_{\text{GMM}}$.

A.2 ESTIMATING THE INVESTMENT EQUATION USING GENERATED REGRESSORS BASED ON THE VAR ESTIMATES

Given the GMM estimates $\hat{b}_{\text{GMM}}$ of a VAR system that includes a measure of the marginal profitability of capital, MPK_{it}, and a financial state variable FIN_{it}, we can use a second-stage regression to obtain structural estimates of the parameters for the cost of adjustment and shadow discount rate functions. Consistent estimates of the slope coefficients are easily obtained using GMM. However, the use of generated regressors implies that the usual standard error estimates are inconsistent. This subsection reviews the details of the second-state estimator and provides standard-error estimates that are consistent in the presence of generated regressors.

The investment model in the paper is

$$(\text{I/K})_{it} = \alpha_0 + \alpha_1 (\text{PV}_{it}^{\text{MPK}}) + \alpha_1\gamma\phi(\text{PV}_{it}^{\text{FIN}}) + f_i + d_t + \omega_{it},$$

where $\text{PV}_{it}^{\text{MPK}}$ and $\text{PV}_{it}^{\text{FIN}}$ are present-value terms that are linear in x_{it} but (highly) nonlinear in b. Let $f(b)$ be an $MP \times 2$ matrix defined so that multiplication of $\tilde{x}_{it}$ by $f(b)'$ produces a 2×1 vector of present-values terms, $f(b)'\tilde{x}_{it} = \{\text{PV}_{it}^{\text{MPK}}, \text{PV}_{it}^{\text{FIN}}\}'$. Using the notation from the main text, the first column of $f(b)$ is given by $c_2[I - \beta A(b)]^{-1}bA(b)^2$, and the second column of $f(b)$ is given by $c_3(1-\beta)^{-1}[I - \beta A(b)]^{-1}bA(b)^2$, where the notation $A(b)$ reflects the fact that the companion matrix A is a function of the VAR parameters b.

Letting $i_{it} = (I/K)_{it}$, $\tilde{q}_{it} = f(b)'\tilde{x}_{it}$, and $a = \{\alpha_1, \alpha_2\gamma\phi\}'$, we can write the above investment model as

$$i_{it} = q_{it}'a + f_i + d_t + \omega_{it}.$$

Using forward-mean differences to remove year and firm effects (as discussed in the previous subsection), we can write the transformed model as

$$\tilde{i}_{it} = \tilde{q}_{it}'a + \tilde{\omega}_{it}.$$

For identification, we assume $E(\omega_{it} | f_i, \gamma_t, x_{it}, x_{it-1}, x_{it-2}, \ldots) = 0$. This implies that the same vector of instruments z_{it} used in the estimation of the VAR is also valid for the estimation of the investment equation. If we let i and Q be the matrices of stacked observations on i_{it} and q_{it}, respectively, then a GMM estimator for a is given by

$$\hat{a}_{\mathrm{GMM1}} = (Q'Z\hat{V}_2^{-1}Z'Q)^{-1}Q'Z\hat{V}_2^{-1}Z'i,$$

where $\hat{V}_2 = 1/N^*\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(\hat{\omega}_{it} \otimes z_{it})(\hat{\omega}_{it} \otimes z_{it})'$, and where $\hat{\omega}_{it}$ is the residual estimate of the transformed error term, $\tilde{\omega}_{it}$, constructed using the first stage estimate of a.

Recall that $\hat{a}_{\mathrm{GMM1}}$ is estimated using generated regressors. Hence, the above expression for $\hat{V}_2$ does not consistently estimate the asymptotic covariance of the second-stage sample moments, because it fails to take account of the implicit variation in $\hat{\omega}_{it}$ induced by $\hat{b}$. A consistent estimator that does take account of this variation is

$$\hat{V}_3 = \frac{1}{N^*}\sum_{i=1}^{N}\sum_{t=1}^{T_i} r_{it}r_{it}',$$

where

$$r_{it} = (\hat{\omega}_{it} \otimes z_{it}) - Z'X\hat{G}P_1(\hat{u}_{it} \otimes z_{it})$$
$$P_1 = (X'Z\hat{V}_2^{-1}Z'X)^{-1}X'Z\hat{V}_2^{-1}$$
$$\hat{G} = \frac{\partial}{\partial b}[f(\hat{b})\hat{a}].$$

Consistent standard-error estimates for $\hat{a}_{\mathrm{GMM1}}$ are given by

$$\text{Est Var}(\hat{a}_{\text{GMM1}}) = (\tilde{X}'Z\hat{V}_2^{-1}Z'\tilde{X})^{-1}\tilde{X}'Z\hat{V}_3Z'\tilde{X}(\tilde{X}'Z\hat{V}_2^{-1}Z'\tilde{X})^{-1}.$$

The availability of $\hat{V}_3$ suggests a second, potentially more efficient GMM estimator for a, namely,

$$\hat{a}_{\text{GMM2}} = (Q'Z\hat{V}_3^{-1}Z'Q)^{-1}Q'Z\hat{V}_3^{-1}Z'i.$$

Consistent standard error estimates for $\hat{a}_{\text{GMM2}}$ are given by

$$\text{Est Var }(\hat{a}_{\text{GMM2}}) = (\tilde{X}'Z\hat{V}_3^{-1}Z'\tilde{X})^{-1}\tilde{X}'Z\hat{V}_4Z'\tilde{X}(\tilde{X}'Z\hat{V}_3^{-1}Z'\tilde{X})^{-1},$$

where $\hat{V}_4$ is estimated using the expression for $\hat{V}_3$, but where $\hat{\omega}_{it}$ and $\hat{a}$ are calculated using $\hat{a}_{\text{GMM2}}$. The estimates reported in the main text are based on this estimator.

A derivation of the above results is available from the authors on request. See also Newey (1984).

REFERENCES

Abel, A. B., and O. Blanchard. (1986). The present value of profits and cyclical movements in investments. *Econometrica* 54:249–273.

———, and J. C. Eberly. (1994). A unified model of investment under uncertainty. *American Economic Review* 84:1369–1384.

———, and ———. (1996). Investment and q with fixed costs: An empirical analysis. Wharton School, University of Pennsylvania. Working Paper.

Arellano, M., and S. Bond. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58:277–297.

———, and O. Bover. (1995). Another look at the instrumental variable estimation of error component models. *Journal of Econometrics* 68:29–51.

Bernanke, B., M. Gertler, and S. Gilchrist. (1998). The financial accelerator in a quantitative business cycle framework. Forthcoming in *Handbook of Macroeconomics*, Michael Woodford and John Taylor (eds.).

Caballero, R. J. (1997). Aggregate investment. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 6264.

———, and E. M. R. A. Engel. (1998). Nonlinear aggregate investment dynamics: Theory and evidence. Cambridge, MA: Working Paper. M.I.T.

———, and J. V. Leahy. (1996). The demise of marginal q. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 5508.

Calomiris, C. W., and C. P. Himmelberg. (1998). Investment banking costs as a measure of the cost of access to external finance. Columbia University. Working Paper.

———, ———, and P. Wachtel. (1995). Commercial paper, corporate finance, and the business cycle: A macroeconomic perspective. *Carnegie-Rochester Conference Series on Public Policy* 42:203–250.

Carpenter, R. E., S. Fazzari, and B. Petersen. (1997). Inventory investment, internal-finance fluctuations, and the business cycle. *Brookings Papers on Economic Activity* 2:75–122.

Chirinko, R. S. (1993). Business fixed investment spending: A critical survey of modelling strategies, empirical results, and policy implications. *Journal of Economic Literature* 31(4):1875–1911.

Cummins, J. G., K. A. Hassett, and S. D. Oliner. (1997). Investment behavior, observable expectations, and internal funds. New York University. Working Paper.

Fazzari, S. M., R. G. Hubbard, and B. C. Petersen. (1988). Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1:141–195.

———, ———, and ———. (1996). Financing constraints and corporate investment: Response to Kaplan and Zingales. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 5462.

Gertler, M. (1992). Financial capacity and output fluctuations in an economy with multiperiod financial relationships. *Review of Economic Studies* 59:455–472.

———, and S. Gilchrist. (1994). Monetary policy, business cycles and the behavior of small manufacturing firms. *Quarterly Journal of Economics* 109:309–340.

Gilchrist, S., and C. P. Himmelberg. (1995). Evidence on the role of cash flow for investment. *Journal of Monetary Economics* 36:541–572.

Gomes, J. (1997). Heterogeneity in macroeconomics: Essays on investment behavior and unemployment dynamics. University of Rochester. Unpublished Ph.D. Dissertation.

Goolsbee, A., and D. B. Gross. (1997). Estimating adjustment costs with data on heterogeneous capital goods. University of Chicago. Working Paper.

Gross, D. B. (1997). The Investment and financing decisions of liquidity constrained firms. University of Chicago. Working Paper.

Himmelberg, C. P. (1990). Essays on the relationship between investment and internal finance. Northwestern University. Unpublished Ph.D. Dissertation.

———, and B. Petersen. (1994). R&D and internal finance: A panel study of small firms in high-tech industries. *Review of Economics and Statistics* 76(1):38–51.

Holtz-Eakin, D., W. Newey, and H. S. Rosen. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56:1371–1395.

Hubbard, R. G. (1998). Capital market imperfections and investment. *Journal of Economic Literature* 36(1):193–225.

———, and A. Kashyap. (1992). Internal net worth and the investment process: An application to U.S. agriculture. *Journal of Political Economy* 100(3):506–534.

———, ———, and T. Whited. (1995). Internal finance and firm investment. *Journal of Money, Credit and Banking* 27(3):683–701.

Jaramillo, F., F. Schianterelli, and A. Weiss. (1996). Capital market imperfections before and after financial liberalizations: An Euler equation approach to panel data for Ecuadorian firms. *Journal of Development Economics* 51:367–386.

Kaplan, S. N., and L. Zingales. (1997). Do financing constraints explain why investment is correlated with cash flow? *Quarterly Journal of Economics* 109(3): 565–592.

Kashyap, A., O. Lamont, and J. Stein. (1994). Credit conditions and the cyclical behavior of inventories. *Quarterly Journal of Economics* 109:565–592.

Keane, M. P., and D. E. Runkle (1992). On the estimation of panel data models

with serial correlation when instruments are not strictly exogenous. *Journal of Business and Economic Statistics* 10:1–9.
Kiyotaki, N., and K. D. West. (1986). Business fixed investment and the recent business cycle in Japan. In *NBER Macroeconomics Annual 1996*. Cambridge, MA: The MIT Press.
Moyen, N. (1997). Dynamic investment decisions with a tax benefit and a default cost of debt. University of British Columbia. Working Paper.
Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters* 14:201–206.
Sharpe, S. (1993). Financial market imperfections, firm leverage, and the cyclicality of employment. *American Economic Review* 84(4), 1060–1074.
Whited, T. (1992). Debt, liquidity constraints, and corporate investment. *Journal of Finance* 47(4):1425–1459.

Comment

DAVID B. GROSS
University of Chicago

Understanding investment has always been one of the central goals of macroeconomics. Unfortunately, the neoclassical model does not do a very good job in describing the empirical investment behavior of individual firms. This nice paper by Gilchrist and Himmelberg is part of the large literature which invokes financial-market imperfections to better explain microeconomic investment decisions. One of the key issues in this area is disentangling changes in economic fundamentals from changes in the firm's financial position. Do balance-sheet variables like cash flow have an independent effect on investment, or are both investment and cash flow simply correlated with the firm's underlying state variable? Improvements in economic fundamentals, such as the marginal product of capital (MPK), directly increase investment through the standard neoclassical channel while at the same time improving the firm's financial position. Hence, even if financial markets are perfect, financial variables will be correlated with investment. Initially, this literature used Tobin's q to control for fundamentals and argued that financial variables are still important in explaining investment. However, since empirical estimates of q are most likely highly mismeasured, variables like cash flow may still be capturing the effects of fundamentals on investment rather than imperfect financial markets.

Gilchrist and Himmelberg attempt to solve this problem by obtaining a better measure of economic fundamentals. Instead of using average q, the authors show that for a certain set of production functions, the sales-to-capital ratio will be proportional to the firm's MPK. They then use their preferred measure of economic fundamentals in two specifications. First, they run panel data VARs in I/K, MPK, and CF/K to argue that orthogonal shocks to cash flow matter even after controlling for the marginal product of capital. Second, they linearize a model of investment with financial-market imperfections where firms' discount rates are parametrized as a reduced-form function of balance-sheet variables. The model generates linear regressions similar to the ones used in the earlier financial-market imperfections literature:

$$\frac{I}{K} = \beta_0 + \beta_1 \text{EPDV(MPK)} + \beta_2 \text{EPDV(FIN)}$$

where EPDV denotes the expected present discounted value and FIN is a financial indicator. In this specification, following Abel and Blanchard (1986), economic fundamentals as measured by marginal q are determined by forecasting expected future values of the MPK. Since the forecasts of MPK and the financial variables are based on lags of investment, MPK, and financial variables, this specification is similar to the previous VAR regressions except that the dynamic structure has been constrained by theory.

The results from both specifications are quite nice. The financial variables are significant for small firms and firms without bond ratings, even controlling for fundamentals. Economic fundamentals also are significant, and unlike results using average q, they are economically important. This is an important achievement which should give more confidence in these types of specifications. However, I will argue that the improved performance of the authors' measure of fundamentals should not be surprising and may not be capturing what we originally expected.

The key issue in this paper is appropriately measuring economic fundamentals. Otherwise, the authors could have continued to use average q with stock-market data like most of the literature. What exactly are economic fundamentals? Under the null hypothesis with no financial-market imperfections, economic fundamentals represent the state variable of the firm's investment problem. Gilchrist and Himmelberg follow the financial-market-imperfection literature by arguing that this theoretical state variable is marginal q or EPDV(MPK). Note, however, that marginal q is the correct state variable only in problems with quadratic (or some other convex) adjustment costs. There is a growing empirical literature which argues that quadratic adjustment costs are a poor assumption for investment at the microeconomic level. Instead, at the level of the firm or plant, investments are intermittent and lumpy, consistent with nonconvex adjustment cost functions such as fixed costs of investment. Caballero and Leahy (1996) show that with nonconvex adjustment costs, marginal q is not the appropriate state variable for the firm's problem. In fact, average q is more correlated than marginal q with investment. If marginal q is not the firm's true state variable, then financial variables could be statistically significant in the data, without any financial-market imperfections.

Even if marginal q as measured using the expected future MPK is the correct state variable, there are several difficulties. The authors measure the MPK as

$$\text{MPK} = \theta\left(\frac{\text{sales}}{K}\right)$$

where the constant of proportionality

$$\theta \equiv \left(1 + \frac{1}{\eta}\right)\alpha_k = \frac{r + \delta}{\text{industry average (sales}/K)}$$

is allowed to vary across industries. These identifying assumptions for θ, while true for Cobb–Douglas production functions with a constant elasticity demand curve, are very strong and likely to introduce measurement error just as in the original criticism leveled at average q. For example, there is no reason to believe that θ is constant over time and within an industry. In addition, to solve for θ for an industry, it is assumed that the average marginal product of capital within the industry is just the user cost, $r+\delta$. However, this is not true in general for most specifications of adjustment costs. It is also not true if there are financial-market imperfections, since constrained firms will act as if they faced a higher discount rate than the market rate for unconstrained firms. While the high discount rate is not a problem under the null hypothesis, the authors use this measure of fundamentals to interpret their linearized model of financial imperfections.

The most problematic issue with the construction of fundamentals in this paper is the interpretation of a quantity-based measure such as sales rather than a forward-looking asset price such as stock returns or interest rates. It is well known that including almost any firm-level quantity variable has strong predictive power in investment regressions, while including variables related to the user cost does not work as well. By adding a quantity-based measure such as sales, are we really capturing the expected marginal return on investment, or are we capturing accelerator effects? This question is similar to an earlier debate in the investment literature in which strong interest-rate elasticities for investment were found by substituting output for the interest rate in the firm's first-order condition for investment. These results were not robust to using interest rates directly.

It is certainly possible that quantity measures, such as sales, have less measurement error than price measures, such as average q or the user cost, for measuring economic fundamentals. The problem with arguing that the marginal product of capital is proportional to sales is one of interpretation. How do you separate the effects of fundamentals from financial explanations when financial-market imperfections are the main story behind accelerator models? Models with financial-market imperfections predict that the average product of capital (APK) or other measures of profits should be important for investment in addition to the marginal product. Even in the authors' model, the APK and the MPK will be highly correlated:

$$\text{APK} = \theta_2 \left(\frac{\text{sales}}{K} \right) - \frac{F}{K}$$

Similarly, it is difficult to separate the partial effects of sales and cash flow in the regression

$$\frac{I}{K} = \beta_0 + \beta_1 \text{EPDV} \left(\frac{\text{sales}}{K} \right) + \beta_2 \text{EPDV} \left(\frac{\text{CF}}{K} \right).$$

Since profits or cash flow is equal to sales minus costs, do we really believe that shocks to sales capture economic fundamentals while shocks to costs are only related to financial effects? The authors argue that there is some variation in cash flow which is not related to fundamentals. However, there is also some variation in cash flow which is related to fundamentals, and in sales which is related to the firm's financial condition. As a result, I think of both sales and cash flow as noisy estimates of both fundamentals and financial effects.

Even though it may be difficult to separate fundamentals from financial effects through direct measurement, there is still a lot of potential in this approach. In particular, I encourage the authors to follow the methodology used in their previous paper, Gilchrist and Himmelberg (1995). Rather than trying to distinguish the MPK from cash flow, the authors focused on the strength of this method—dynamics. Using the same variable for both fundamentals and financial effects, it is possible to test for financial effects by looking at whether investment has excess sensitivity to current variables:

$$\frac{I}{K} = \beta_0 + \beta_1 \text{EPDV} \left(\frac{\text{CF}}{K} \right) + \beta_2 \frac{\text{CF}}{K}.$$

Does current cash flow or, even better, the stock of cash predict investment even after controlling for the EPDV of future profits?

In future work, adding more structure to the alternative model of financial-market imperfections should yield differential predictions to help test the validity of financial explanations for investment. In my past work, for example, I showed (Gross, 1997) that imposing liquidity constraints leads to strong nonlinearities in the predicted effects of financial variables, which are confirmed in the data. It would be even more useful to explicitly model the information or agency problem which generated the financial-market imperfection in the first place. It may be too much to ask for an empirically tractable, theoretically justified model of financial-market imperfections which allows for nonconvex adjustment costs and can deal with the heterogeneity present in real-world data. However, it will be difficult to fully understand the microeconomics of investment without such a model.

REFERENCES

Abel, A. B., and O. Blanchard. (1986). The present value of profits and cyclical movements in investments. *Econometrica* 54:249–273.
Caballero, R. J., and J. V. Leahy. (1996). The demise of marginal *q*. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 5508.
Gilchrist S., and C. P. Himmelberg. (1995). Evidence on the role of cash flow for investment. *Journal of Monetary Economics* 36:541–572.
Gross, D. B. (1997). The investment and financing decisions of liquidity constrained firms. University of Chicago. Mimeo.

Comment

KENNETH D. WEST
University of Wisconsin

1. *Introduction*

In this stimulating paper, Gilchrist and Himmelberg (hereafter, GH) use Compustat data to test and quantify the effect of financial factors on investment. I like the paper's innovative willingness to estimate and test a structural model with financial frictions. In my view, this is a welcome advance on a common practice of estimating a model without such frictions and testing whether the model fails in ways consistent with financial frictions. By specifying precisely and completely how financial frictions affect investment, GH can quantify the effects of those frictions,

I thank the National Science Foundation and the Graduate School of the University of Wisconsin for financial support, and John Jones for helpful comments and discussions.

and get a sense for whether still richer models are required to adequately explain investment.

I divide my comments into two parts. First is a summary of the literature on borrowing constraints and investment. Then follows a discussion of the structural model (Sections 3–5 in GH). I conclude that the present paper leaves some important questions unanswered. While I am receptive to GH's interpretation, in the end I do not find it compelling.

2. *Borrowing Constraints and Investment*

In typical parametrizations of the models of Hall and Jorgenson (1967) and Tobin (1969), investment is determined by the ratio of output to the cost of capital or by stock-market-based measures of Tobin's Q, perhaps along with lags of these variables or of investment itself. For want of a better term, I will refer to such variables as the *traditional* determinants of investment. At least since Meyer and Kuh (1957), however, it has been recognized that cash flow is a good predictor of aggregate investment. Clark (1979) concluded that cash flow and output together predict better than do the neoclassical determinants, a finding broadly consistent with the somewhat weaker role for cash flow found in the more recent study by Kopcke (1993).

In a pioneering paper, Fazzari, Hubbard, and Petersen (1988) established a complementary set of results with individual firm data. They ran investment regressions in which right-hand-side variables included some traditional determinants of investment as well as cash flow. They found that cash flow had a statistically significant effect on investment for firms that one might expect to be financially constrained, namely, ones with low dividend payouts. As documented in Hubbard's (1998) survey, many subsequent studies have found that measures of liquidity, such as cash flow or debt, are statistically significant predictors of investment for firms displaying low dividends, no bond rating, or some other presumed signals of financial illiquidity. While a literature has risen in reaction, arguing that in one or another dataset the empirical evidence in fact does not suggest that financial factors are important (e.g., Hayashi, 1997; Kaplan and Zingales, 1997; Cummins, Hassett, and Oliner, 1998), it is my judgment that there now exists a strong case that financial variables have information about investment not contained in the traditional measures.

The predictive power of cash flow suggests a difference between internal and external costs of funds. Indeed, it has been shown formally that in the presence of informational frictions, such as imperfect or costly monitoring of a firm's behavior by a lender (Townsend, 1979; Myers and Majluf, 1984), the shadow value of internal funds will be larger than that of external funds, at least some of the time. Cash flow will therefore likely affect investment, even after controlling for traditional determinants.

In my opinion, however, the empirical literature has not made a good case that informational frictions in fact rationalize the empirical importance of financial variables.[1] It is well recognized that misspecification of the parametric form of the traditional model, or inappropriate accounting for unobservable disturbances or for the informational role of financial variables, could cause spurious significance of financial variables, significance that would disappear if the frictionless model were appropriately specified.

Support for the view that financial frictions rather than misspecification are key could come from establishing that the estimates on financial factors fall in a tight range predicted by an underlying theoretical model, that a model with financial frictions does not reject tests of overidentifying restrictions, that no variables beyond those posited by the model have substantial explanatory power for investment, and so on. To my knowledge, little research in this literature attempts to do so. Instead, the bulk of the literature that I am familiar with evaluates (and usually rejects) a null neoclassical model. No doubt this is in large part because quantitatively tractable models of financial frictions are difficult to come by. Nonetheless, this approach yields diminishing returns for people who, like me, are already persuaded that cash flow and other financial variables help predict the investment of firms that display signs of illiquidity.

I therefore am very receptive to Gilchrist and Himmelberg's effort to formulate and evaluate a model with financial frictions.

3. *Gilchrist and Himmelberg's Model*

Sections 3–5 of GH specify, estimate, and test a model of investment in the presence of financial frictions—that is, the null model is one with financial frictions (though a frictionless model is nested within the null model).

Specifically, this part of GH:

1. Nor, even assuming that these frictions in fact do account for the econometric results, does this or related literature establish that the frictions have nontrivial effects on aggregate output, although it has been shown that such frictions potentially have aggregate effects (Carlstrom and Fuerst, 1997; Kiyotaki and Moore, 1997; Jones, 1998). But further discussion of that point will take us pretty far from the GH paper.

1. Constructs a present value called *fundamental Q*—a specific example of what I called a "traditional" measure above—by positing a parametric functional form, linearizing, and using data on output, capital, and other variables, as in Abel and Blanchard (1986). Fundamentals depend only on the marginal profit of capital (MPK) and, in contrast to Abel and Blanchard (1986) and many other papers, not on discount factors (i.e., not on user costs of capital), which are assumed not to vary.
2. Approximates the effects of financial frictions by (i) showing that these will manifest themselves in time-varying discount factors; (ii) linearizing to separate the discount factor from the frictionless fundamentals; (iii) assuming the linearized term is also linear in observable financial data, namely, cash and equivalents, or working capital less long-term debt.
3. Regresses investment on lagged investment, fundamental Q, and the present value of financial frictions (*financial Q*). The finding is that financial frictions are both statistically and economically important.

As a mechanical matter, the regressions are novel essentially in entering the financial variable in a constrained form, as a present value; the common procedure in this literature is simply to add a variable like cash and equivalents to whatever variables are included by virtue of the traditional model under examination. The important novelty is a willingness to maintain, at least tentatively, that the addition of the financial variable results in a complete model for investment. GH acknowledge that the model in fact does miss an important aspect of investment behavior, in that the data want lagged investment on the right-hand side (Section 5.3), and they make rationalizing this extra term a high priority for future research. But they also clearly feel that the present specification documents strong economic and statistical effects from financial frictions.

While I do find GH's results suggestive, I am not as convinced as are GH that they have documented strong effects. In explaining this viewpoint, I will not promote an alternative interpretation of their results, still less provide a different point estimate of the quantitative effects of financial frictions. Rather, I want to raise some questions, which, until answered, suggest caution in interpreting GH's findings. These questions include:

1. In their theoretical development, GH assume that in the absence of financial frictions the discount rate is constant across firms and time (Section 3.1); had they carried through the algebra with time-varying discount rates, there would have been an additional term, involving

the present value of future values of discount rates (see Abel and Blanchard, 1986). The presence of time dummies in the empirical work implicitly allows time variation in discount rates that is common to all firms; the presence of a fixed effect means one can interpret such variation as occurring around a firm-specific mean discount rate. (That is, these dummies allow for time variation in discount rates that is perfectly correlated across firms, with each firm possibly having a different mean discount rate.) But firm-specific variation (i.e., imperfect correlation across firms) is swept into the regression disturbance. And we have a raft of finance studies indicating that discount rates vary across firms. Such variation may affect the estimates on the coefficients on the two present values included by GH. If the variation is largely uncorrelated with the GH present values (perhaps this is the case for most firms presumed to be financially unconstrained in Tables 5 and 6), the GH interpretation is still legitimate. But if the variation is correlated with the included present values (perhaps firms that are financially constrained undertake riskier projects), omitted-variable bias will invalidate the GH procedure. So: how are the estimates and interpretation affected by cross-sectional variation in discount factors stemming from traditional forces (rather than from financial frictions)?

2. When lagged investment is added to the GH equation, the term proves significant, but, the authors tell us, the point estimates on the two present values change little. It is, however, premature to conclude, as the authors seem to, that once the model is expanded to formally allow for lagged investment, the statistical or economic importance of the two present values will change little. This is because in the expanded model the present values will be calculated in a different way, implying that the present model's regressions contain noisy measures of the relevant objects. If one rationalizes the lag with costs of adjustment or serially correlated shocks, the relevant regression will involve a lag of investment and two present values, and the two present values will be different from the ones presently included. (For example, with costs of adjustment, a certain quadratic will be factored, with forward solution of the unstable root to this quadratic affecting the present value calculation.) If, instead (or in addition), one rationalizes the lag with an information role for investment—that is, the firm forecasts the present values using data not available to the econometrician—the regression will still involve only two present values, but once again the two will be different from the ones presently included (see Kiyotaki and West, 1996). So: what happens when

the model is extended to rationalize the predictive power of lagged investment?

3. GH state that their linearization implies that the external finance premium is proportional to their measure of financial liquidity (either cash and equivalents, or working capital less long-term debt), with a factor of proportionality ϕ. They also conclude that the regression estimate of ϕ is pretty much consistent with a back-of-the-envelope calculation. I could not quite follow the details of this calibration, so I will not comment except to state such a check on the reasonableness of the estimate is an important one. An additional check is suggested by the model's implication that one can read interest-rate spreads off the measure of financial liquidity. If one backs out a series of spreads (presumably identified only up to the addition of a constant), what does the series look like, and how does it compare to observed data on lending rates?

I look forward to the answers to questions such as these in the next paper in the Gilchrist Himmelberg's research program.

REFERENCES

Abel, A., and O. J. Blanchard. (1986). The present value of profits and cyclical movement in investment, *Econometrica* 54(2):249–274.

Carlstrom, C. T., and T. S. Fuerst. (1997). Agency costs, net worth and business fluctuations: A computable general equilibrium analysis. *American Economic Review* 82:430–450.

Clark, P. K. (1979). Investment in the 1970s: Theory, performance and prediction. *Brookings Papers on Economic Activity* 1979(1):73–113.

Cummins, J. G., K. A. Hassett, and S. Oliner. (1998). Investment behavior, observable expectations, and internal funds. New York University. Manuscript.

Fazzari, S. R., R. G. Hubbard, and B. Petersen. (1988). Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1, 141–195.

Hall, R. E., and D. Jorgenson. (1967). Tax policy and investment behavior. *American Economic Review* 53, 391–414.

Hayashi, F. (1997). The main bank system and corporate investment: An empirical reassessment. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 6172.

Hubbard, R. G. (1998). Capital market imperfections and investment. *Journal of Economic Literature* XXXVI:193–225.

Jones, J. (1998). Firm finance with collateral constraints: Business cycle implications. University of Wisconsin. Manuscript.

Kaplan, S. N., and L. Zingales. (1997). Do investment–cash flow sensitivities provide useful measures of financing constraints? *Quarterly Journal of Economics* 112:169–215.

Kiyotaki, N., and J. Moore. (1997). Credit cycles. *Journal of Political Economy* 105, 211–248.

———, and K. D. West. (1996). Business fixed investment and the recent business cycle in Japan. In *NBER Macroeconomics Annual, 1996*, B. Bernanke and J. Rotemberg (eds.), Cambridge, MA: The MIT Press, pp. 277–323.

Kopcke, R. A. (1993). The determinants of business investment: Has capital spending been surprisingly low? *New England Economic Review*, 3–31.

Meyer, J. R., and E. Kuh. (1957). *The Investment Decision*, Cambridge, MA: Harvard University Press.

Myers, S. C., and N. Majluf. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13(2):187–221.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking* 1, 15–29.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory* 21(2):265–293.

Discussion

Much of the discussion focused on the adequacy of the baseline model, particularly the use of the sales-to-capital ratio as a measure of the marginal profitability of fixed capital. Janice Eberly pointed out that the proportionality of profitability and sales requires that variable factors, such as labor, be adjustable instantaneously. If there are costs of adjusting labor, then shocks to wages (for example) will affect profitability and cash flow without having a short-run effect on output. On a related issue, Fabio Schiantarelli noted that, if there are costs of adjusting the capital stock, the net marginal profitability of capital depends on the investment-to-capital ratio as well as sales. Jason Cummins asked about the omission of inventories from the analysis. Charles Himmelberg replied to Cummins that, possibly because of the use of annual data, it makes little difference empirically whether inventories are included, that is, either sales or production can be used as the measure of profitability. More generally, Himmelberg defended the baseline model by pointing out that it seems to work well empirically for bond-rated firms, which are less likely to be financially constrained. Deviations from the baseline model are thus reasonably interpreted as arising from financial factors.

A second issue was the possible effects of unmodeled heterogeneity among firms. Cummins suggested that the restriction that the data-generating process for shocks is identical across firms is unrealistic. He noted that financial analysts produce earnings forecasts using firm-specific models. He reported that, in his own work with coauthors, the use of firm-specific earnings expectations reduced the measured effect of cash-flow variables significantly. Himmelberg disputed the result that earnings expectations measures "drive out" cash-flow variables from investment equations, indicating that this was not the case for the specifications used in the present paper. He also argued that earnings expectations might be interpreted as projections of whether the firm can repay its debt, that is, as an indicator of financial condition, rather than as forecasts of long-term profitability. Also on the theme of heterogeneity, Benjamin Friedman proposed that investment–cash-flow relationships might be quite different at the firm level than at the industry level. For example, a firm might be impelled to invest by an increase in competitive pressure, resulting in a negative relationship of current cash flow and future investment for the firm, even though that relationship might well be positive for the industry as a whole.

Econometric issues also received some discussion. Stephen Oliner proposed the use of a larger set of instruments to reduce problems of endogeneity.