Supplementary Material for "In Search of Ideas: Technological Innovation and Executive Pay Inequality"

Carola Frydman* Dimitris Papanikolaou†

Contents

1 Proofs and Derivations **1**

2 Estimation Details **9**

2.1 Methodology 9

2.2 Definition of target statistics 10

2.3 Estimation of the Optimal Weighting Matrix 12

2.4 Derivation of Influence Functions 13

3 Identification **20**

4 Robustness to using market values as measure of size **31**

*Kellogg School of Management and NBER, c-frydman@kellogg.northwestern.edu

†Kellogg School of Management and NBER, d-papanikolaou@kellogg.northwestern.edu

1 Proofs and Derivations

First, we consider some of the equilibrium relations in order to gain intuition for the overall structure of the solution. Define

$$\zeta_{j,t} = u_{j,t}\, e^{\xi_{\tau(j)}}\, k_{j,t}. \tag{1}$$

and

$$Z_t = \int_{J_t} e^{\xi_{s(j)}}\, u_{j,t}\, k_{j,t}\, dj. \tag{2}$$

The labor hiring decision is static. The firm managing project j chooses L_{jt} as the solution to

$$\pi_{jt} = \sup_{L_{jt}} \left[\zeta_{jt}^{\phi}\, \left(e^{x_t} L_{jt}\right)^{1-\phi} - \mathrm{w}_t\, L_{j,t} \right] \tag{3}$$

The firm's choice

$$L_{j,t}^{\star} = \zeta_{j,t} \left(\frac{(1-\phi)\, e^{(1-\phi)x_t}}{\mathrm{w}_t} \right)^{\frac{1}{\phi}}. \tag{4}$$

After clearing the labor market, $\int_{\mathcal{J}_t} L_{j,t} dj = 1$, the equilibrium wage is given by

$$\mathrm{w}_t = (1-\phi)\, e^{(1-\phi)\, x_t}\, Z_t^{\phi}, \tag{5}$$

and the choice of labor allocated to project j is

$$L_{j,t}^{\star} = \zeta_{j,t}\, Z_t^{-1}. \tag{6}$$

Aggregate output of all projects equals

$$Y_t = \int_{J_t} \zeta_{j,t}\, e^{(1-\phi)\, x_t}\, Z_t^{\phi-1}\, dj = e^{(1-\phi)\, x_t}\, Z_t^{\phi}. \tag{7}$$

The project's flow profits are

$$\pi_{j,t} = \sup_{L_{j,t}} \left[\zeta_{jt}^{\phi}\, \left(e^{x_t} L_{j,t}\right)^{1-\phi} - \mathrm{w}_t\, l_{j,t} \right] = \zeta_{jt}\, \phi\, Y_t\, Z_t^{-1} \tag{8}$$

Because firms' investment decisions do not affect its own future investment opportunities, each investment maximizes the net present value of cash flows from the new project. Thus, the optimal investment in a new project j at time t is the solution to

$$\nu_t \equiv \sup_{k_{j,t}} \mathrm{E}_t \left[\int_t^\infty \frac{\Lambda_s}{\Lambda_t} \pi_{j,s} \, ds \right] - k_{j,t}^{1/\alpha} = \sup_{k_{j,t}} \left[P_t \, k_{j,t} e^{\xi_t} - k_{j,t}^{1/\alpha} \right], \tag{9}$$

where P_t is the time-t price of the asset with the cash flow stream $\exp(-\delta(s-t))p_s$:

$$P_t = \mathrm{E}_t \left[\int_t^\infty \frac{\Lambda_s}{\Lambda_t} e^{-\delta(s-t)} \, p_s \, ds \right]. \tag{10}$$

The optimal scale of each new project is then given by

$$k_t^\star = \left(\alpha \, e^{\xi_t} \, P_t \right)^{\frac{\alpha}{1-\alpha}}. \tag{11}$$

Note that the solution does not depend on the identity of the firm, i.e., all firms, faced with an investment decision at time t, choose the same scale for the new projects. The optimal investment scale depends on the current market conditions, specifically, on the current level of the embodied productivity process ξ_t, and the current price level P_t.

Using (11), we find that, the equilibrium value of a new project is equal to

$$\nu_t = \alpha^{\frac{\alpha}{1-\alpha}} (1-\alpha) \, \left(e^{\xi_t} \, P_t \right) \right)^{\frac{1}{1-\alpha}}. \tag{12}$$

Recall the definition of capital,

$$K_t = \int_0^1 \left(\sum_{j \in \mathcal{J}_{f,t}} e^{\xi_{s(j)}} \, k_{j,t} \right) df. \tag{13}$$

We thus find that the aggregate stock of quality-adjusted installed capital in the intermediate good sector evolves according to

$$dK_t = \left(-\delta \, K_t + \lambda \, e^{\xi_t} \, k_t^\star \right) \, dt = \left(-\delta \, K_t + \lambda \, e^{\xi_t} \, \left(\alpha \, e^{\xi_t} \, P_t \right)^{\frac{\alpha}{1-\alpha}} \right) \, dt, \tag{14}$$

where λ is the aggregate rate of arrival of new projects. In deriving (13), we have conjectured that λ is constant, a fact we verify below.

An important aspect of (13) is that the growth rate of the capital stock K_t depends only on its current level, the productivity level ξ_t, and the price process P_t. Furthermore, as we show below, we can clear markets with the price process P_t expressed as a function of the state vector $X_t = (x_t, \xi_t, K_t)$. Thus, X_t follows a Markov process in equilibrium.

We express equilibrium processes for aggregate quantities and prices as functions of X_t. For instance, the fact that investment decisions are independent of u implies that $Z_t = K_t$. Aggregate investment I_t is given by

$$I_t = i(\omega_t) Y_t = \lambda \left(k_t^\star\right)^{1/\alpha} = \lambda \frac{\alpha}{1-\alpha} \nu_t. \tag{15}$$

where the third equality follows from (12). The aggregate consumption process satisfies

$$C_t = \left(1 - i(\omega_t)\right) Y_t = Y_t - I_t = K_t^\phi e^{(1-\phi)x_t} - \lambda \left(k_t^\star\right)^{1/\alpha}. \tag{16}$$

Prices of long-lived financial assets, such as the aggregate stock market, depend on the behavior of the stochastic discount factor. Standard arguments imply that in equilibrium, the SDF is

$$\Lambda_t = e^{-\rho t} C_t^{-1}. \tag{17}$$

Define

$$\omega_t \equiv \xi_t + \alpha \left(1 - \phi\right) x_t - \left(1 - \alpha\, \phi\right) \log K_t, \tag{18}$$

and use the fact that

$$\log Y_t = \left(1 - \phi\right) x_t + \phi \log K_t, \tag{19}$$

then ω_t and $\log Y_t$ are linear functions of the state vector X_t. In Lemma 1 below, aggregate equilibrium quantities as functions of ω_t and $\log Y_t$.

In the formulation of the lemma, we characterize the value function of a household, as well as prices of financial assets, such as P_t in (10), using differential equations. Verification results, such as (Duffie and Lions, 1992, Sec. 4), show that a classical solution to the corresponding differential equation, subject to the suitable growth and integrability constraints, characterizes the value function. Similarly, the Feynman-Kac Theorem (Karatzas and Shreve, 1991, e.g, Theorem 7.6) provides an analogous result for the prices of various financial assets. Because we solve for equilibrium numerically, we cannot show that the classical solutions to our differential equations exist and satisfy the sufficient regularity conditions. With this caveat in mind, in the following lemma we characterize the equilibrium processes using the requisite differential equations.

Lemma 1 (Equilibrium). *Let the functions $i(\omega)$, $v(\omega)$, $g(\omega)$, $h(\omega)$ solve the following system of four ordinary differential equations,*

$$
\begin{aligned}
0 &= \phi(1-i(\omega))^{-1} - \left(\rho + \lambda^{1-\alpha} e^{\omega}\, i(\omega)^{\alpha}\right) v(\omega) + v''(\omega)\,\frac{1}{2}\left(\sigma_\xi^2 + \alpha^2 (1-\phi)^2 \sigma_x^2\right) \\
&\quad + v'(\omega)\Big(\mu_\xi + \alpha\,(1-\phi)\,\mu_x - (1-\alpha\,\phi)\left(\lambda^{1-\alpha} e^{\omega}\, i(\omega)^{\alpha} - \delta\right)\Big) && (20) \\
0 &= \frac{i(\omega)}{1-i(\omega)}\,\frac{1-\alpha}{\lambda\,\alpha} - \rho\, g(\omega) + g''(\omega)\,\frac{1}{2}\left(\sigma_\xi^2 + \alpha^2 (1-\phi)^2 \sigma_x^2\right) \\
&\quad + g'(\omega)\Big(\mu_\xi + \alpha\,(1-\phi)\,\mu_x - (1-\alpha\,\phi)\left(\lambda^{1-\alpha} e^{\omega}\, i(\omega)^{\alpha} - \delta\right)\Big) && (21) \\
0 &= (1-\phi)\,(1-i(\omega))^{-1} - \rho\, h(\omega) + h''(\omega)\,\frac{1}{2}\left(\sigma_\xi^2 + \alpha^2 (1-\phi)^2 \sigma_x^2\right) \\
&\quad + h'(\omega)\Big(\mu_\xi + \alpha\,(1-\phi)\,\mu_x - (1-\alpha\,\phi)\left(\lambda^{1-\alpha} e^{\omega}\, i(\omega)^{\alpha} - \delta\right)\Big) && (22)
\end{aligned}
$$

and the following algebraic equation

$$
\left(\frac{i(\omega_t)}{\lambda}\right)^{1-\alpha} = \alpha\, e^{\omega}\, v(\omega)\, \left(1 - i(\omega)\right). \tag{23}
$$

Then we can construct price processes and individual policies that satisfy the definition of the competitive equilibrium in Section 1.5 in the paper, so that K_t follows

$$
\frac{dK_t}{K_t} = -\delta\, dt + \lambda\, e^{\omega_t} \left(\frac{i(\omega_t)}{\lambda}\right)^{\alpha} dt. \tag{24}
$$

Proof. We start with a conjecture, which we confirm below, that the equilibrium price process P_t satisfies

$$
P_t = K_t^{-1}\, Y_t\, v(\omega_t)\, (1 - i(\omega_t)). \tag{25}
$$

Under this conjecture, the equilibrium aggregate value of assets in place is

$$
V_t \equiv P_t K_t = Y_t\, v(\omega_t)\, (1 - i(\omega_t)), \tag{26}
$$

the value of growth opportunities for the average firm ($\lambda_f = \lambda$) is

$$
G_t \equiv \lambda\, E_t \int_t^{\tau} \frac{\Lambda_s}{\Lambda_t} \nu_s\, ds = \lambda Y_t g(\omega_t)\, (1 - i(\omega_t)), \tag{27}
$$

and the aggregate value of human capital is

$$
H_t = Y_t h(\omega_t)\, (1 - i(\omega)). \tag{28}
$$

We then characterize the equilibrium SDF and the optimal policies of the firms and households, and show that all markets clear and the above conjectures are consistent with the equilibrium processes for cash flows and the SDF.

We denote the time-t net present value of the new projects (the maximum value in (12)) by ν_t. The aggregate investment process, according to (14), is given by

$$I_t = \lambda \frac{\alpha}{1-\alpha} \nu_t. \tag{29}$$

Using (27) and market clearing (14), K_t follows

$$\frac{dK_t}{K_t} = \left(-\delta\, K_t + \lambda\, e^{\xi_t} \left(\alpha\, e^{\xi_t}\, P_t\right)^{\frac{\alpha}{1-\alpha}}\right) dt = -\delta\, dt + \lambda\, e^{\omega_t} \left(\frac{i(\omega_t)}{\lambda}\right)^{\alpha} dt,$$

where we have used (23), and (21) for the last equality. The equilibrium dynamics of the aggregate quality-adjusted capital stock thus agrees with (22).

Optimality of household consumption and portfolio choices implies that the SDF

$$\Lambda_t = e^{-\rho\, t} C_t^{-1} = e^{-\rho\, t} Y_t^{-1} \left(1 - i(\omega_t)\right)^{-1}, \tag{30}$$

where without loss of generality we have set $\Lambda_0 = 1$. Also, it is helpful to define

$$\pi_t \equiv e^{\rho\, t} \Lambda_t = Y_t^{-1} \left(1 - i(\omega_t)\right)^{-1}. \tag{31}$$

Next, we need to verify that, in equilibrium, the aggregate arrival of projects is constant. Using lemma 3 below, we can write the value of an active match between an executive with perceived talent $p_{f,t}$ and the firm by

$$v_t(p_{f,t}) = p_{ft}\, m_t, \quad \text{where } m_t \equiv\ \lambda_D\, E_t \int_t^{\tau} \frac{\Lambda_s}{\Lambda_t} \nu_s\, ds, \tag{32}$$

where τ is the stochastic time the match is dissolved. A firm will terminate its executive if the value of the current match falls below the value of a new match, excluding training costs,

$$p_{ft}\, m_t \leq \bar{p}\, m_t - c\, \bar{p}\, m_t \Rightarrow p_{ft} \leq p^* \equiv (1-c)\, \bar{p}. \tag{33}$$

Since the firing threshold is constant and the stationary distribution of $p_{f,t}$ exists (see Lemma 2). Hence, the equilibrium level of average executive-firm CEO matches,

$$\lambda \equiv \lambda_L + \lambda_D \int_{p^*}^{1} p_{f,t}\, df, \tag{34}$$

is a constant.

We are now in a position to complete the proof by verifying that the conjectured price processes in (23–26) are consistent with the equilibrium SDF above. ■

Lemma 2 (Stationary distribution for p)**.** *The process p has probability density $f(p,t)$ that for $p \in (p^*, 1)$ solves*

$$f_t(t,p) = \beta\left(\Delta(\bar{p}) - f(t,p)\right) + \lambda_D\, p\,(1-p)\, f_p(t,p) + (\lambda_L + p\,\lambda_D)\left(f\left(t, \frac{p\,\lambda_H}{\lambda_L + p\,\lambda_D}\right) - f(t,p)\right)$$

where $\Delta(\bar{p})$ is the Dirac delta function with point mass at $\bar{p}$. At $p = p^*$ the process is reflected to $\bar{p}$.

Assuming it exists, the stationary density solves the ODE

$$0 = \beta\left(\Delta(\bar{p}) - f(p)\right) + \lambda_D\, p\,(1-p)\, f'(p) + (\lambda_L + p\,\lambda_D)\left(f\left(\frac{p\,\lambda_H}{\lambda_L + p\,\lambda_D}\right) - f(p)\right)$$

for $p \in (p^*, 1)$, and once the process hits $p = p^*$ it is reflected to $\bar{p}$ and

$$\lambda = \lambda_L + \lambda_D \int_{p^*}^{1} p\, f(p)\, dp. \tag{35}$$

Proof of Lemma 2. Lemma follows from a straightforward application of the Kolmogorov Forward equation, taking into account the evolution of p under rational expectations – equation (14) in the text – the fact that matches are exogenously dissolved at rate β or when p reaches the firing threshold p^*.

■

Lemma 3 (Value of Firm-Executive Match)**.** *Define the value of a match between a firm and an executive with perceived talent $p_{f,t}$ as*

$$m(p_{ft}, \omega_t, Y_t) \equiv \pi_{f,t}\, E_t \int_t^{\tau} \lambda_D \frac{\Lambda_s}{\Lambda_t} \nu_s\, ds \tag{36}$$

where τ is the earliest of the exogenous or endogenous termination date, whichever comes first. In equilibrium,

$$m(p_{ft}, \omega_t, Y_t) = p_{f,t}\, \lambda_D\, Y_t \tilde{g}(\omega_t)\,(1 - i(\omega_t))\,, \tag{37}$$

where $\tilde{g}(\omega)$ *solves the following ODE,*

$$
\begin{aligned}
0 = &\frac{i(\omega)}{1-i(\omega)}\frac{1-\alpha}{\lambda\,\alpha} - (\rho+\beta)\,\tilde{g}(\omega) + \tilde{g}''(\omega)\,\frac{1}{2}\left(\sigma_\xi^2 + \alpha^2(1-\phi)^2\sigma_x^2\right) \\
&+ \tilde{g}'(\omega)\Big(\mu_\xi + \alpha\,(1-\phi)\,\mu_x - (1-\alpha\,\phi)\left(\lambda^{1-\alpha}e^{\omega}\,i(\omega)^{\alpha} - \delta\right)\Big).
\end{aligned} \tag{38}
$$

The termination threshold satisfies

$$
p^* = (1-c)\,\bar{p} \tag{39}
$$

Proof. Denote $\hat{m}(p,\omega,Y) \equiv \pi m(p,\omega,Y)$. We know that, in the region $p \in (p^*,1)$, that function satisfies the following PDE

$$
\begin{aligned}
0 = &\,p\,\frac{i(\omega)}{1-i(\omega)}\frac{1-\alpha}{\lambda\,\alpha} - (\rho+\beta)\hat{m}(p,\omega,Y) - \hat{m}_p(p,\omega,Y)p(1-p)\lambda_D \\
&+ (\lambda_L + \lambda_D p)\left(\hat{m}\left(\frac{p\,\lambda_H}{\lambda_L + p\lambda_D},\omega,Y\right) - \hat{m}(p,\omega,Y)\right) + \mathcal{D}_{Y,\omega}\hat{m}(p,\omega,Y)
\end{aligned} \tag{40}
$$

where optimal firing decisions imply that at the firing threshold p^* solves

$$
\hat{m}(p^*,\omega,Y) = (1-c)\,\hat{m}(\bar{p},\omega,Y). \tag{41}
$$

Guess that

$$
\hat{m}(p,\omega,Y) = p\,\lambda_D\,Y\,\tilde{g}(\omega). \tag{42}
$$

Given our guess, equation (38) simplifies to the ODE in (36). Given the definition of $\hat{m}$, we have that

$$
m(p_{ft},\omega_t,Y_t) = p_{f,t}\,\lambda_D\,Y_t\tilde{g}(\omega_t)\,(1-i(\omega_t))\,, \tag{43}
$$

The last thing to check is that given our guess, the termination threshold exists and is given by (37). Indeed,

$$
p^*\,\lambda_D\,Y_t\tilde{g}(\omega_t)\,(1-i(\omega_t)) - (1-c)\,\bar{p}\,\lambda_D\,Y_t\tilde{g}(\omega_t)\,(1-i(\omega_t)) = 0. \tag{44}
$$

■

Lemma 4 (Executive Pay)**.** *The equilibrium level of executive pay (in excess of their compensation for working in production) is given by*

$$
w_{f,t} = \eta\,(p_{ft} - p^*)\,\lambda_D\,i(\omega)\frac{1-\alpha}{\lambda\,\alpha}\,Y_t \tag{45}
$$

Proof. Define the surplus created from having a current CEO of quality p_{ft} is

$$\begin{aligned}S_{f,t} &= m(p_{ft}, \omega_t, Y_t) - (1-c)m(\bar{p}, \omega_t, Y_t)\\ &= (p_{f,t} - p^*)\lambda_D \, Y_t \tilde{g}(\omega_t)\,(1 - i(\omega_t))\,. \end{aligned} \tag{46}$$

Executives capture fraction η of the surplus; hence the present value of CEO pay is $\eta\, S_{f,t}$. Suppose that the firm promises the CEO a flow rate w. Let the CEO's continuation value be W_t; the wage flow satisfies

$$\pi_t W_t = E_t \int_t^{\tau} e^{-(\rho+\beta)(s-t)} \pi_s \, w_s ds \tag{47}$$

where τ is the endogenous termination date. Given our assumptions, we need that this promised value satisfies $W_{f,t} = \eta\, S_{f,t}$. The fact that the discounted gains process associated with W_t is a martingale implies that

$$\begin{aligned} 0 &= \pi\, w - (\rho+\beta)\pi W + \eta\,(p-p^*)\,\lambda_D \,\Big[\tilde{g}''(\omega)\,\frac{1}{2}\left(\sigma_\xi^2 + \alpha^2(1-\phi)^2\sigma_x^2\right)\\ &\quad + \tilde{g}'(\omega)\Big(\mu_\xi + \alpha\,(1-\phi)\,\mu_x - (1-\alpha\,\phi)\left(\lambda^{1-\alpha} e^{\omega}\, i(\omega)^{\alpha} - \delta\right)\Big)\Big] \end{aligned} \tag{48}$$

$$0 = \pi\, w - \eta\,(p-p^*)\,\lambda_D \left[\frac{i(\omega)}{1-i(\omega)} \frac{1-\alpha}{\lambda\,\alpha}\right] \tag{49}$$

where in the last equality we have used the fact that $\pi W = \eta \pi S$ and that $\tilde{g}$ solves (36). To obtain (43) in the text, we can use (14) to replace the investment to output ratio $i(\omega_t)$ in (43) with the value of investment opportunities ν_t.

■

2 Estimation Details

In what follows, we provide some more detail on the estimation methodology (section 2.1), the construction of the target statistics in the data and the model (section 2.2), some details on the estimation of the optimal weighting matrix (section 2.3) and a derivation of the influence functions for the non-standard target statistics that we used (section 2.4).

2.1 Methodology

We estimate the parameter vector θ using the simulated minimum distance method (Ingram and Lee, 1991). We denote by X the vector of target statistics in the data and by $\mathcal{X}(\theta)$ the corresponding statistics generated by the model given parameters θ, computed as

$$\mathcal{X}(\theta) = \frac{1}{S}\sum_{i=1}^{S} \hat{X}_i(\theta), \tag{50}$$

where $\hat{X}_i(p)$ is the 21×1 vector of statistics computed in one simulation of the model. We simulate the model at a weekly frequency, and time-aggregate the data to form annual observations. Each simulation has 1,000 firms. For each simulation i, we first simulate 100 years of data as 'burn-in' to remove the dependence on initial values. We then use the remaining part of that sample, which is chosen to match the longest sample over which the target statistics are computed. Each of these statistics is computed using the same part of the sample as its empirical counterpart. In each iteration we simulate $S = 100$ samples, and simulate pseudo-random variables using the same seed in each iteration.

Our estimate of the parameter vector is given by

$$\hat{\theta} = \arg\min_{\theta}(X - \mathcal{X}(\theta))' W (X - \mathcal{X}(\theta)). \tag{51}$$

where W denotes our choice of weighting matrix. We use the optimal weighting matrix,

$$W = \left(\hat{\Sigma} + \frac{1}{S}\Omega(\hat{\theta})\right)^{-1}, \tag{52}$$

where $\hat{\Sigma}$ is an estimate of the variance-covariance matrix of the target statistics X, and

$$\Omega(\hat{\theta}) = \frac{1}{S}\sum_{i=1}^{S}(\hat{X}_i(\hat{\theta}) - \mathcal{X}(\hat{\theta}))(\hat{X}_i(\hat{\theta}) - \mathcal{X}(\hat{\theta}))' \tag{53}$$

is the estimate of the sampling variation of the statistics in X in simulated data.

Solving each iteration of the model is computationally costly, and thus computing the minimum in (49) using standard numerical methods is computationally infeasible. We therefore use the Radial Basis Function (RBF) algorithm in Björkman and Holmström (2000). The Björkman and Holmström (2000) algorithm first fits a response surface to data by evaluating the objective function at several thousand points. Then, it searches for a minimum by balancing between local and global search in an iterative fashion. We use a commercial implementation of the RBF algorithm that is available through the TOMLAB optimization package.

2.2 Definition of target statistics

Mean and volatility of aggregate consumption growth: We compute the mean and standard deviation of log consumption growth $\Delta \log C$. We use the Barro and Ursua (2008) consumption data for the United States, which covers the 1834-2008 period.

Consumption growth, long-run volatility: Using the Barro and Ursua (2008) consumption data, we compute the estimate of long-run risk using the estimator in Dew-Becker (2014).

Investment, mean share of output: Investment is non-residential private domestic investment. Output is gross domestic product. Data range is 1927-2014.

Investment growth, volatility: Investment is non-residential private domestic investment. We deflate by population and the CPI. Data on the CPI are from the BLS. Population is from the Census Bureau. Data range is 1927-2014. We compute the volatility of log investment growth $\Delta \log I$.

Consumption and investment growth, correlation: We compute the correlation between consumption and investment growth, defined above.

Payout ratio, coefficient of variation: The net payout to assets are from Larrain and Yogo (2008) who use flow of funds data. Data range is 1929-2004. We compute the coefficient of variation as the ratio of the standard deviation of the payout ratio divided by its mean. We focus on net payouts because dividends are not well defined in our model – the model has no unique dividend policy. Depending on the parametrization, net payouts can be negative. Therefore, we target the volatility of the ratio of net payouts to book assets. We normalize by the mean payout ratio to remove scale effects.

Risk-free rate moments: The moments of the real risk free rate are computed using data on the nominal rate and expected inflation from the Cleveland Fed, available here: `https://www.clevelandfed.org/our-research/indicators-and-data/inflation-expectations.aspx`. Data range is 1982-2010.

CEO tenure, median: We compute the median CEO tenure using data on Execucomp. Tenure is defined as the difference between *leftofc* and *becameceo*, that is the dates that a given executive joined the firm as CEO and left the firm.

Size elasticity of pay: We compute the size elasticity of pay as the regression coefficient of log firm-level executive pay on log firm size. To obtain a more precise estimate we use the larger panel dataset of Execucomp. We use the ex-ante value of executive compensation (tdc1), which includes the Black-Scholes value of granted options. We compute firm-level compensation by computing the average compensation in a given year of the top-5 executives. Size is defined as the book value of assets (at) from Compustat. Data range is 1992-2014.

Moments of dispersion in log executive pay: We compute the cross-sectional dispersion (standard deviation) in log firm-level executive pay using the data from Frydman and Saks (2010). We use the ex-ante value of executive compensation, which includes the Black-Scholes value of granted options. Firm-level compensation is defined as the average compensation in a given year of the top-3 executives. Data range is 1936-2014. When estimating the moments of dispersion in executive pay in simulated data, we follow the procedure of Frydman and Saks and sample the 50 largest firms every 20 years.

Moments of mean executive pay to workers: In any given year we compute the logarithm of the cross-sectional mean of the average firm-level executive pay from Frydman and Saks (2010) divided by average worker earnings from the BLS. Data range is 1936-2014.

Firm Investment rate, IQR and serial correlation: Firm investment is defined as the change in log gross PPE (ppegt) from Compustat. When computing correlation coefficients, we winsorize the data by year at the 1% level to minimize the effect of outliers. We simulate the model at a weekly frequency, $dt = 1/50$ and time aggregate the data at the annual level. In the model, firm investment is computed as the growth rate in the firm's capital stock at annual frequencies. Data range is 1950-2014.

Firm Innovation, IQR and serial correlation: Firm innovation in year t is defined as the value of all patents granted to the firm in year t, using the estimated values from Kogan, Papanikolaou, Seru, and Stoffman (2016), scaled by the market value of the firm at the end of year t. We simulate the model at a weekly frequency, $dt = 1/50$ and time aggregate the data at the annual level. Firm innovation in the model is defined as the sum of the market values of all projects acquired by the firm in year t scaled by firm size at the end of year t. Since the estimated values in Kogan et al. (2016) contain a proportional scaling factor that adjusts for the fact that these patent grants are partly anticipated, we scale the firm innovation measures in both the data and the model such that they have a mean equal to 1 conditional on non-zero values.

Firm Profitability, IQR and serial correlation: In the data, firm profitability equals the logarithm of the sum of net income (ni) scaled by the book value of assets (at). When computing correlation coefficients, we winsorize the data by year at the 1% level to minimize the effect of outliers. We simulate the model at a weekly frequency, $dt = 1/50$ and time aggregate the data at the annual level. In the model, we compute firm profitability as $\pi_{f,t}/\hat{K}_{f,t}$, where $\pi_{f,t}$ is the accumulated profits of all projects owned by firm f in year t and $\hat{K}_{f,t}$ is capital at the end of year t. Data range is 1950-2014.

Labor Share: We construct the labor share using Flow of Funds data following Sekyu and Rios-Rull (2009)

2.3 Estimation of the Optimal Weighting Matrix

The estimation of the optimal weighting matrix involves the estimation of the sample covariance matrix of the targeted statistics in Table 1 in the paper. Since not all of these statistics are moments, the estimation of the sample covariance matrix is rather involved. We estimate it by stacking the influence functions following Davis, Fisher, and Whited (2014). We adjust the covariance matrix for the fact that moments are estimated using different samples as in Erickson and Whited (2012). That is, if statistic i is estimated over a sample of n_i observations, while statistic j is computed over a sample of n_j observations, we adjust the covariance as

$$cov(L_t^i, L_t^j) = \frac{n^2}{n_i\, n_j} cov(\hat{L}_t^i, \hat{L}_t^j) \tag{54}$$

where n is the number of overlapping observations and $cov(\hat{L}_t^i, \hat{L}_t^j)$ is the covariance in the overlapping sample.

When estimating the covariance for panel moments, we cluster by firm i and year t. When estimating the covariance for time-series moments, we use a HAC estimator (Newey-West) with 3 lags. Last, when computing the covariance between a panel statistic j and a time-series statistic k, we do so by

$$\frac{1}{T}\, S\left(L_t^k \cdot \frac{1}{N} \sum_{i=1}^{N} L_{i,t}^j \right) \tag{55}$$

where S is an estimate of the asymptotic covariance matrix – we use Newey-West with 3 lags. In words, we first collapse the panel influence functions along the cross-sectional dimension by year, and then estimate the covariance.

Since the resulting moment covariance matrix $\hat{C} = \hat{\Sigma} + S^{-1}\Omega(\theta)$ matrix is ill-conditioned (it has negative eigenvalues), we adjust the estimated matrix using the methodology of

Rousseeuw and Molenberghs (1993).[1] As a robustness check, we also performed a ridge adjustment as in Ledoit and Wolf (2004),

$$W = \left(\hat{C} + \lambda I\right)^{-1} . \tag{56}$$

We used the smallest possible value for λ such that $\hat{C} + \lambda I$ has all positive eigenvalues ($\lambda = 10^{-4}$). As an additional robustness check, we also experimented with an alternative ridge adjustment from van Wieringen and Peeters (2016), where

$$W = \left\{ \left[\lambda I + \frac{1}{4} \left(\hat{C} - \lambda T \right)^2 \right]^{1/2} + \frac{1}{2} \left(\hat{C} - \lambda T \right) \right\}^{-1} , \tag{57}$$

where $T = I$ and with values λ ranging from 10^{-4} to 10^{-6}. The resulting parameter estimates were quantitatively very similar in all three cases.

2.4 Derivation of Influence Functions

Most of the statistics we target have influence functions that are straightforward to derive. In what follows, we provide derivations for some of the non-standard cases.[2]

2.4.1 Influence function for the Mean of the Cross-Sectional Dispersion

Here, we derive expressions for the time-series mean of the cross-sectional variance and standard deviation. First, consider the cross-sectional variance. The parameter of interest is

$$\theta = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} Var_t(x_{it}) := E[Var_t(x_{it})] , \tag{58}$$

where $Var_t(x_{it})$ is the cross section (conditional on t) variance at time period t. The sample analogue estimator is

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{N} \sum_{i=1}^{N} x_{it}^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_{it} \right)^2 \right] . \tag{59}$$

We now derive the influence function for the above estimator by providing a heuristic asymptotic linear expansion, which will satisfy some type of central limit theorem. Note that the asymptotic thought experiment we are thinking of here is one where we are averaging

[1] We are grateful to Luke Taylor for this suggestion.

[2] We thank Vishal Kamat for assistance with these derivations.

over (across t) a growing number of estimates of non identically distributed random variables (i.e. cross section variances), and where both N and T go to infinity. Formally providing assumptions on the data generating process that ensure this heuristic result is valid is fairly non trivial.

To this end, first denote by

$$\bar{x}_t = \frac{1}{N}\sum_{i=1}^{N} x_{it} \tag{60}$$

and then note that the term $(\bar{x}_t)^2$ can be rewritten as

$$\begin{aligned}(\bar{x}_t)^2 &= (\bar{x}_t - \mu_t + \mu_t)^2 &(61)\\ &= (\bar{x}_t + \mu_t)(\bar{x}_t - \mu_t) + \mu_t^2 \ , &(62)\end{aligned}$$

where $\mu_t = E_t[x_{it}]$ is the conditional on t expectation. We can then rewrite the estimator in (57) as

$$\begin{aligned}\hat{\theta} &= \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[x_{it}^2 - E_t[x_{it}^2] + E_t[x_{it}^2]\right] - (\bar{x}_t + \mu_t)(\bar{x}_t - \mu_t) - \mu_t^2\right\} &(63)\\ &= \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[x_{it}^2 - E_t[x_{it}^2]\right] - (\bar{x}_t + \mu_t)\frac{1}{N}\sum_{i=1}^{N}[x_{it} - \mu_t]\right\} + \frac{1}{T}\sum_{t=1}^{T}\left[E_t[x_{it}^2] - \mu_t^2\right] \ , &(64)\end{aligned}$$

which can then be rewritten as

$$\begin{aligned}\sqrt{NT}(\hat{\theta} - \theta) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[x_{it}^2 - E_t[x_{it}^2]\right] - (\bar{x}_t + \mu_t)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[x_{it} - \mu_t]\right\} + o_P(1) &(65)\\ &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t)\right] + o_P(1) \ , &(66)\end{aligned}$$

where $\bar{x}_t$ is replaced by μ_t as it will be converge to it for large enough N under appropriate assumptions. The above will intuitively satisfy some central limit theorem as the two interior terms in (63) are cross section estimates using N observations, and will hence converge to some random variables. The estimator will then average over these T random variables and will again converge to some asymptotic distribution. However, understanding the appropriate assumptions required for formalizing this intuitive explanation is non trivial.

The above explanation then lets us conclude that the 'influence function' in this setting can be viewed as

$$L_t(x_{it}) = x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t) \ , \tag{67}$$

and the corresponding estimated influence function as

$$\hat{L}_t(x_{it}) = x_{it}^2 - \frac{1}{N}\sum_{i=1}^{N} x_{it}^2 - (2\bar{x}_t)(x_{it} - \bar{x}_t) \ . \tag{68}$$

Second, the mean of the cross-sectional standard deviation is

$$\theta = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sqrt{Var_t(x_{it})} := E[\sqrt{Var_t(x_{it})}] \ , \tag{69}$$

where $Var_t(x_{it})$ is the cross section (conditional on t) variance at time period t. The sample analogue estimator is

$$\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T} \sqrt{\hat{\sigma}_t} \ . \tag{70}$$

We apply the delta method here across the cross section (i.e. for a given time period t) in what follows. First, we can obtain:

$$\hat{\theta} - \theta = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2}\sigma_t^{-1/2}(\hat{\sigma}_t - \sigma_t) + o_P(1) \ . \tag{71}$$

Note from the derivation of estimator in (56) we can write

$$\sigma_t^{-1/2}(\hat{\sigma}_t - \sigma_t) = \frac{1}{N}\sum_{i=1}^{N} \sigma_t^{-1/2}\left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t)\right] \ , \tag{72}$$

and plugging this into the above we can get

$$\sqrt{NT}(\hat{\theta} - \theta) = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} \sigma_t^{-1/2}\left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t)\right] + o_P(1) \ . \tag{73}$$

Similar to earlier, the 'influence function' can then be viewed as

$$L_t(x_{it}) = \sigma_t^{-1/2}\left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t)\right] \ , \tag{74}$$

and the corresponding estimated influence function as

$$\hat{L}_t(x_{it}) = \hat{\sigma}_t^{-1/2} \left[x_{it}^2 - \frac{1}{N} \sum_{i=1}^{N} x_{it}^2 - (2\bar{x}_t)(x_{it} - \bar{x}_t) \right] . \tag{75}$$

2.4.2 Influence function for the volatility of the cross-sectional dispersion

First, consider

$$\theta = \lim_{T\to\infty} \left\{ \frac{1}{T} \sum_{t=1}^{T} log(Var_t(x_{it}))^2 - \left[\frac{1}{T} \sum_{t=1}^{T} log(Var_t(x_{it})) \right]^2 \right\} := Var[log(Var_t(x_{it}))] , \tag{76}$$

where the sample analogue estimator is

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} log(\hat{\sigma}_t)^2 - \left[\frac{1}{T} \sum_{t=1}^{T} log(\hat{\sigma}_t) \right]^2 , \tag{77}$$

which can be rewritten to obtain

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} log(\hat{\sigma}_t)^2 - 2E[log(\sigma_t)] \left[\frac{1}{T} \sum_{t=1}^{T} (log(\hat{\sigma}_t) - log(\sigma_t)) \right] + \left(\frac{1}{T} \sum_{t=1}^{T} log(\sigma_t) \right)^2 + o_P(1) . \tag{78}$$

As in Estimator 3, we intuitively apply the delta method here across the cross section (i.e. for a given time period t) in what follows. First, we can obtain:

$$\hat{\theta} - \theta = \frac{1}{T} \sum_{t=1}^{T} \frac{2log(\sigma_t)}{\sigma_t} (\hat{\sigma}_t - \sigma_t) - 2E[log(\sigma_t)] \left[\frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sigma_t} (\hat{\sigma}_t - \sigma_t) \right] + o_P(1) . \tag{79}$$

Note from the derivation of Estimator 1 we can write

$$\frac{2log(\sigma_t)}{\sigma_t} (\hat{\sigma}_t - \sigma_t) = \frac{1}{N} \sum_{i=1}^{N} \frac{2log(\sigma_t)}{\sigma_t} \left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t) \right] + o_P(1) \tag{80}$$

$$\frac{1}{\sigma_t} (\hat{\sigma}_t - \sigma_t) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_t} \left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t) \right] + o_P(1) \tag{81}$$

and plugging into the above we can get

$$\sqrt{NT}(\hat{\theta} - \theta) = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[\frac{2log(\sigma_t)}{\sigma_t} - \frac{2E[log(\sigma_t)]}{\sigma_t} \right] \left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t) \right] + o_P(1) . \tag{82}$$

Similar to above, the influence function can then be viewed as

$$L_t(x_{it}) = 2\left[\frac{log(\sigma_t) - E[log(\sigma_t)]}{\sigma_t}\right]\left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t)\right] \ , \tag{83}$$

and the corresponding estimated influence function as

$$\hat{L}_t(x_{it}) = 2\left[\frac{log(\hat{\sigma}_t)}{\hat{\sigma}_t} - \frac{1}{\hat{\sigma}_t T}\sum_{t=1}^{T} log(\hat{\sigma}_t)\right]\left[x_{it}^2 - \frac{1}{N}\sum_{i=1}^{N} x_{it}^2 - (2\bar{x}_t)(x_{it} - \bar{x}_t)\right] \ . \tag{84}$$

Last, to derive the influence function for

$$\theta' = \sqrt{Var[log(\sqrt{Var_t(x_{it})})]} = \sqrt{Var[\frac{1}{2}log(Var_t(x_{it}))]} \tag{85}$$

we use the chain rule:

$$\hat{L'}_t(x_{it}) = \frac{1}{4\sqrt{Var[log(\sqrt{Var_t(x_{it})})]}}\left[\frac{log(\sigma_t) - E[log(\sigma_t)]}{\sigma_t}\right]\left[x_{it}^2 - E_t[x_{it}^2] - (2\mu_t)(x_{it} - \mu_t)\right] \ . \tag{86}$$

2.4.3 Influence function for the mean of the (log of the) cross-sectional mean

The parameter of interest is

$$\theta = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} log[E_t(x_{it})] := E[log[E_t(x_{it})]] \ , \tag{87}$$

where the sample analogue estimator is

$$\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T} log\left[\frac{1}{N}\sum_{i=1}^{N} x_{it}\right] = \frac{1}{T}\sum_{t=1}^{T} log\left[\bar{x}_t\right] \ . \tag{88}$$

Similar to the earlier estimators, we intuitively apply the delta method here across the cross section (i.e. for a given time period t) in what follows. First, we can obtain:

$$\hat{\theta} - \theta = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\mu_t}\left[\frac{1}{N}\sum_{i=1}^{N}(x_{it} - \mu_t)\right] + o_P(1) \ , \tag{89}$$

to then get

$$\sqrt{NT}(\hat{\theta}-\theta) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{\mu_t}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_{it}-\mu_t)\right] + o_P(1) \tag{90}$$

$$= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{1}{\mu_t}(x_{it}-\mu_t) + o_P(1) \ . \tag{91}$$

The 'influence function' can then be viewed as

$$L_t(x_{it}) = \frac{1}{\mu_t}(x_{it}-\mu_t) \ , \tag{92}$$

and the corresponding estimated influence function as

$$\hat{L}_t(x_{it}) = \frac{1}{\bar{x}_t}(x_{it}-\bar{x}_t) \ . \tag{93}$$

2.4.4 Influence function for the variance of the log of the cross-sectional mean

The parameter of interest is

$$\theta = \lim_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}log(\mu_t)^2 - \left(\frac{1}{T}\sum_{t=1}^{T}log(\mu_t)\right)^2\right] := Var[log[E_t(x_{it})]] \ , \tag{94}$$

where the sample analogue estimator is

$$\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T}log(\bar{x}_t)^2 - \left(\frac{1}{T}\sum_{t=1}^{T}log(\bar{x}_t)\right)^2 \ . \tag{95}$$

The estimator first can manipulated to be rewritten as

$$\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T}log(\bar{x}_t)^2 - 2E[log(\mu_t)]\left[\frac{1}{T}\sum_{t=1}^{T}(log(\bar{x}_t)-log(\mu_t))\right] + \left(\frac{1}{T}\sum_{t=1}^{T}log(\mu_t)\right)^2 + o_p(1) \ . \tag{96}$$

Similar to the earlier estimators, we intuitively apply the delta method here across the cross section (i.e. for a given time period t) in what follows. First, we can obtain:

$$\hat{\theta}-\theta = \frac{1}{T}\sum_{t=1}^{T}\frac{2log(\mu_t)}{\mu_t}\frac{1}{N}\sum_{i=1}^{N}(x_{it}-\mu_t) - 2E[log(\mu_t)]\left[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\mu_t}\frac{1}{N}\sum_{i=1}^{N}(x_{it}-\mu_t)\right] + o_p(1) \tag{97}$$

to then get

$$\sqrt{NT}(\hat{\theta}-\theta) = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} 2 \left[\frac{log(\mu_t)}{\mu_t} - \frac{E[log(\mu_t)]}{\mu_t} \right] [x_{it} - \mu_t] + o_P(1) \tag{98}$$

The influence function can then be viewed as

$$L_t(x_{it}) = 2 \left[\frac{log(\mu_t)}{\mu_t} - \frac{E[log(\mu_t)]}{\mu_t} \right] [x_{it} - \mu_t] \ , \tag{99}$$

and the corresponding estimated influence function as

$$\hat{L}_t(x_{it}) = 2 \left[\frac{log(\bar{x}_t)}{\bar{x}_t} - \frac{1}{\bar{x}_t T} \sum_{t=1}^{T} [log(\bar{x}_t)] \right] [x_{it} - \bar{x}_t] \ . \tag{100}$$

3 Identification

Here, we discuss the identification of the model's structural parameters. In Figures A.1-A.3 we plot the elasticity of the model's implied moments $X(\theta)$ to small changes in parameters around the optimum $\hat{\theta}$. Specifically, we report

$$E_{i,j} = \frac{dX^j(\theta)}{d\theta^i}\frac{\theta^i}{X^j(\theta)}, \tag{101}$$

were the numerical gradient is computed using a 5-point stencil around $\hat{\theta}$.

In addition, in Figures A.4-A.6 we plot the Gentzkow and Shapiro (2014) measure of sensitivity of parameters to moments. We report the measure in elasticity form,

$$\hat{\lambda}_{i,j} = \lambda_{i,j}\frac{X^j(\theta)}{\theta^i}, \tag{102}$$

where $\lambda_{i,j}$ is the element of the sensitivity matrix Λ that corresponds to parameter i and moment j. The matrix Λ is computed as

$$\Lambda = -(G'WG)^{-1}G'W \tag{103}$$

where G is the numerical gradient of the sample moments $g(\theta) = X - \mathcal{X}$ and W is the optimal weighting matrix.

In what follows, we summarize the main patterns in these Figures.

- The parameter e is identified by:
  - The volatility of the executive-pay ratio: The larger e the higher the correlation in pay between executives and workers; hence a more volatile executive-worker pay ratio – holding the other moments fixed – is interpreted as evidence for lower e.
  - The variation in the ratio of payout to assets: The higher e is, the higher the level of executive pay relative to the firm's book value; hence more volatile payout to book ratios are interpreted as evidence for higher e.
- The parameter β is identified by:
  - The length of the executive's tenure; the longer the median tenure, the lower the involuntary separation rate must be, holding all other moments constant.
- The parameter α is identified by:

- The volatility of investment growth: the higher α is, the more responsive investment is to technology shocks.
- The difference between the volatility of investment to output ratio and the volatility of the executive-pay ratio. The higher α is, the more responsive investment is to technology shocks; this implies that the economy's fluctuations along the balanced growth part are amplified (ω becomes more volatile). At the same time, higher α implies the equilibrium value of a project is smaller, which implies that a smaller fraction of executive pay comes from growth opportunities. The model interprets the difference between these two moments as evidence for higher α.

- The parameter η is identified by:
  - The volatility of the executive-worker pay ratio: The higher η is, the larger is the fraction of executive that is attributed to growth opportunities. This implies a more volatile executive-pay ratio.
  - The dispersion in pay across executives: higher values of η magnify the dispersion in pay across executives.

- The parameter σ_x is identified by:
  - The correlation between investment and consumption: x affects investment and consumption symmetrically, hence higher correlation between investment and consumption growth is interpreted as evidence for higher σ_x
  - The volatility of the ratio of executive to worker pay: x affects executive and worker pay fairly symmetrically. Holding the other moments constant, a more volatile worker-pay ratio is interpreted as evidence for lower σ_x.
  - The volatility of consumption and investment growth: x affects output on impact, hence more volatile consumption or investment growth suggest higher σ_x.

- The parameter σ_ξ is identifies by:
  - The difference between the dispersion in firm investment versus dispersion in firm innovation; ξ affects the optimal scale of investment much more than the value of innovation across firms. Higher dispersion in investment rates relative to the dispersion in innovation rates implies that ξ is more volatile.
  - The volatility of the ratio of executive to worker pay: ξ affects executive pay on impact but worker pay with delay, hence higher volatility in the executive-pay ratio is interpreted by the model as evidence for higher σ_ξ.

- The volatility of investment and of the investment-to-output ratio; ξ affects investment more so than consumption on impact, hence more volatile investment growth or investment to output ratio is interpreted as evidence for higher σ_ξ.

- The parameter c is identified by:
  - The median length of executive tenure: higher c implies higher costs of dissolving low quality matches hence longer tenure.
  - The mean dispersion in pay across executives: if poor quality matches are maintained for longer, this implies larger dispersion in pay among employed executives.
  - The dispersion in innovation across firms: if poor quality matches are maintained for longer, this implies larger dispersion in innovation outcomes across firms.

- The parameter $\bar{p}$ is identified by:
  - The volatility of the mean executive pay to worker ratio: higher values of $\bar{p}$ imply higher unconditional mean quality for active matches, and therefore higher contribution of growth opportunities to executive pay and a more volatile executive-worker pay ratio.
  - The length of tenure across executives: A higher value of $\bar{p}$ implies that a smaller fraction of new firm-executive matches are of low quality, and thus a lower rate of executive termination.
  - The volatility of the investment-to-output ratio: A higher value of p implies a higher average quality of active matches and thus a higher aggregate rate of acquisition of new projects. This implies a faster mean-reversion rate for ω and hence lower volatility for the investment-to-output ratio $i(\omega)$.

- The parameter λ_D is identified by:
  - The length of the executive tenure: An increase in λ_D implies faster learning hence faster dissolutions of low quality matches.
  - The dispersion in innovation outcomes across firms: higher λ_D implies larger differences in the rate of innovation across firms.
  - The mean dispersion in pay across executives: higher λ_D implies larger differences in pay among executives and hence higher dispersion.

- The parameter λ_L is identified by:

- The elasticity of exec pay to firm size: the more elastic is executive pay to firm size the tighter is the link between firm growth (which depends on $\lambda_{f,t} = \lambda_L + p_{f,t}\lambda_D$) and firm pay). Holding all the other moments constant, a higher elasticity of firm size to pay will imply a lower value for λ_L.
- The volatility of the executive pay to worker ratio: higher λ_L implies higher λ i.e. more projects; for a given amount of investment this implies less capital is allocated in a given project and therefore lower per project NPVs. Since executive pay partly depends on the equilibrium NPV per project, dispersion is lower.

• The parameter δ is identified by:

  - The average investment-to-output ratio: analogously to deterministic models, a higher depreciation rate of capital implies a higher mean investment to output ratio.

• The parameter μ_x is identified by:

  - Primarily by the mean of consumption growth, though other moments seem to have an effect.

• The parameter μ_ξ is identified by:

  - Average consumption growth: a higher μ_ξ implies the economy grows at a faster rate
  - The moments of the investment to output ratio; higher values of μ_ξ imply that the mean of the stationary distribution of ω is higher, which implies higher investment to output.

• The parameter v_u is identified by:

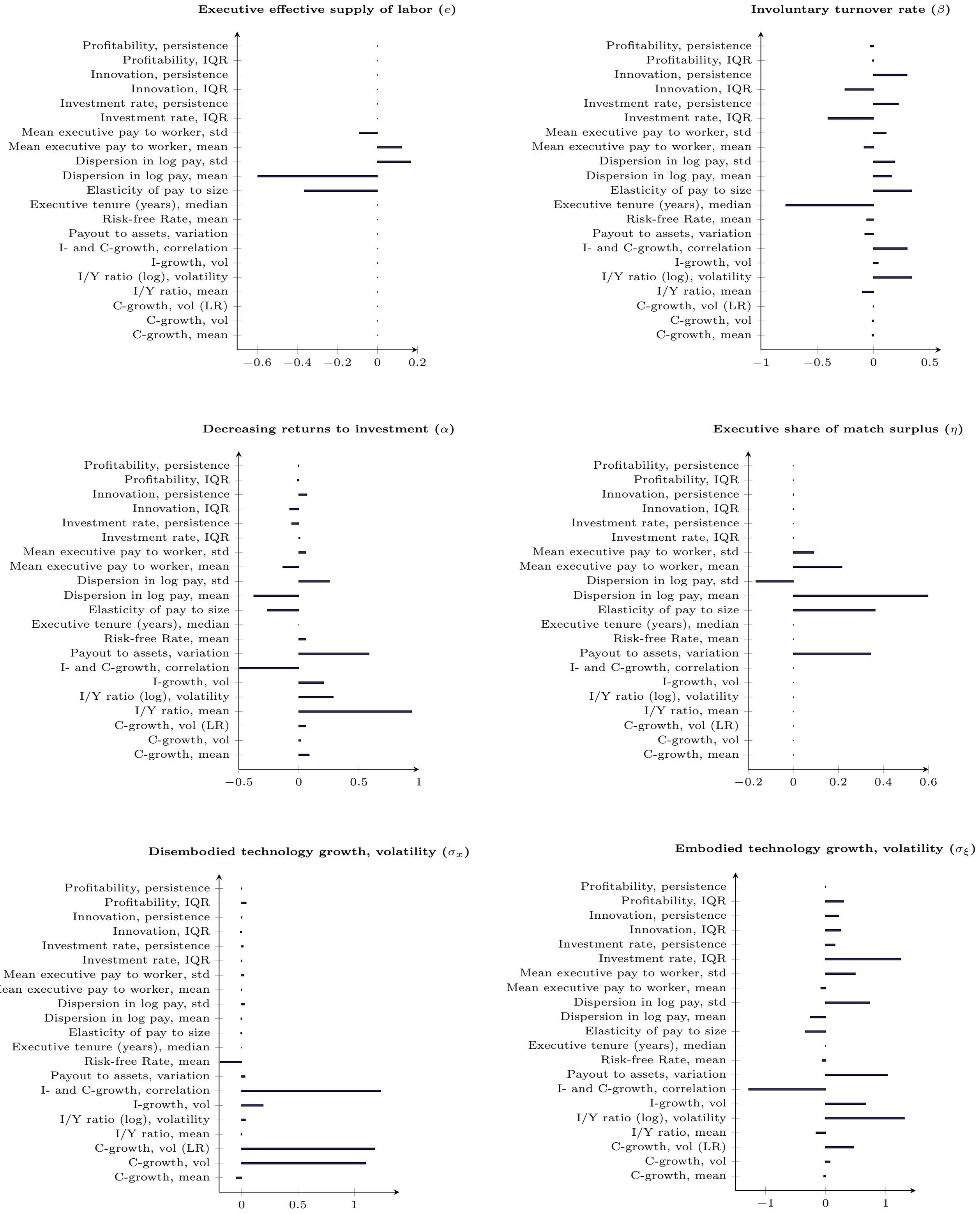
  - the dispersion in profitability across firms: higher v_u implies larger dispersion in u across firms and hence higher dispersion in profitability.

• The parameter ρ is identified by:

  - The difference between the risk-free rate and the mean rate of consumption growth; to a first-order approximation, the risk free rate equals the rate of time preference plus the conditional mean of consumption growth. Hence, a larger difference between the average risk-free rate and the mean consumption growth implies higher values for ρ.

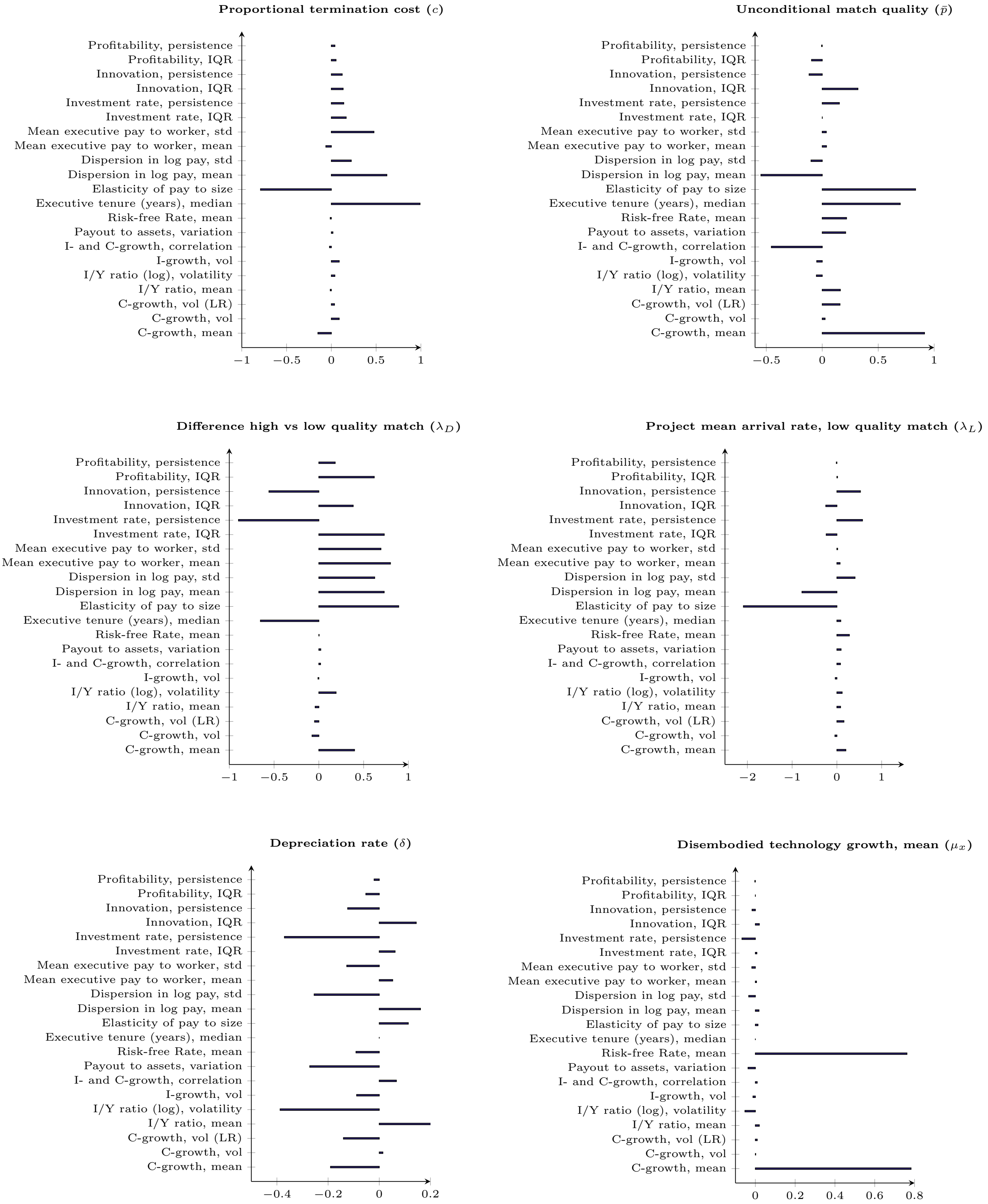
- The parameter θ_u is identified by:
  - the persistence in profitability across firms: holding v_u constant, a higher θ_u implies a faster rate of mean reversion in u, and hence less persistent profitability at the firm level.
- The parameter h is identified by:
  - The mean executive to worker pay ratio; higher h implies a higher level of compensation for production workers, hence a lower executive-worker pay ratio.

Figure A.1: Sensitivity of moments to parameters



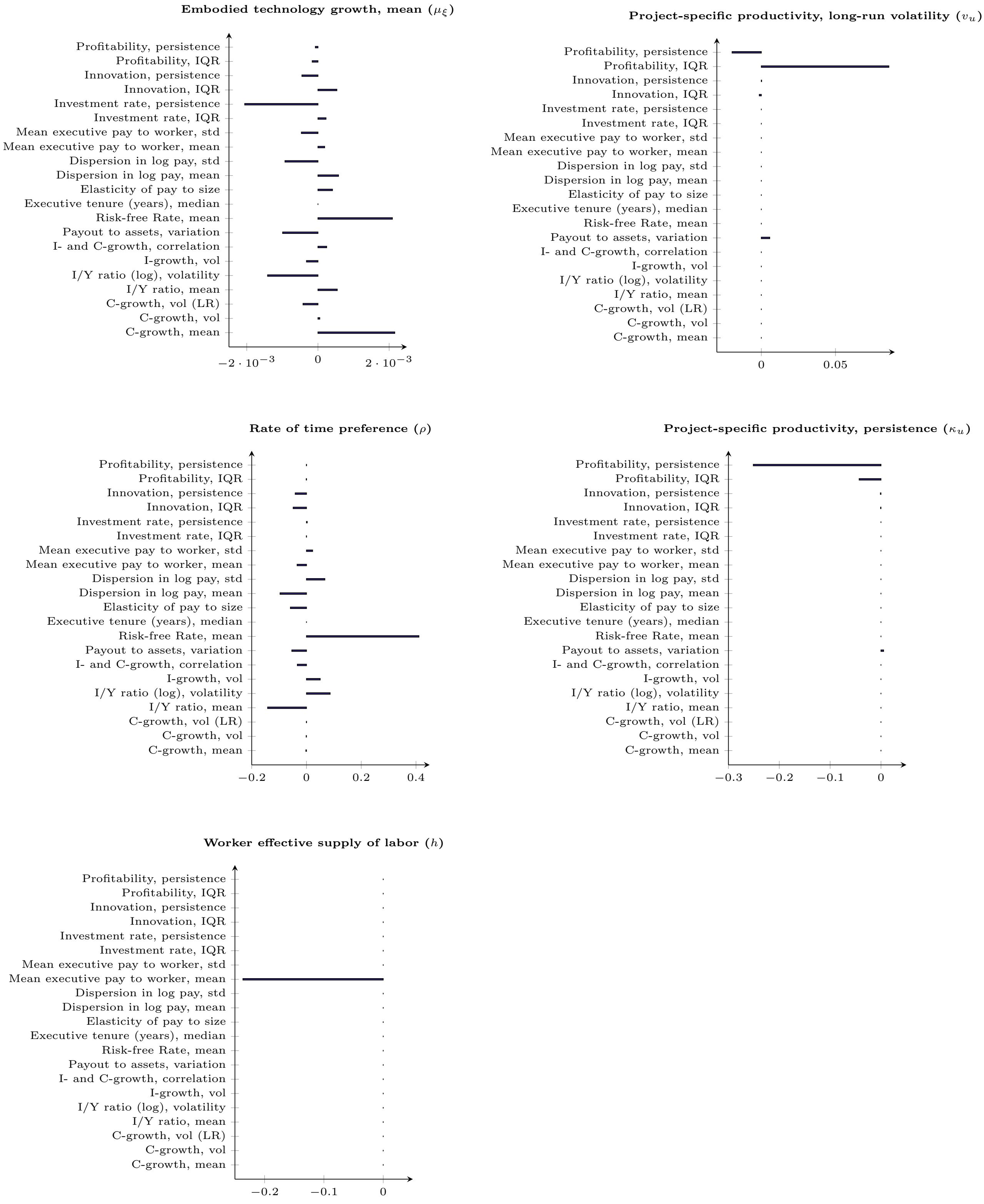
We report the elasticity of moment estimates $\mathcal{X}(\theta)$ to parameters θ. Specifically, we report $\frac{dX^j(\theta)}{d\theta^i}\frac{\theta^i}{X^j}$ where $\frac{dX^j(\theta)}{d\theta^i}$ is the numerical derivative – computed using a 5-point stencil – of moment j to parameter i.

Figure A.2: Sensitivity of moments to parameters



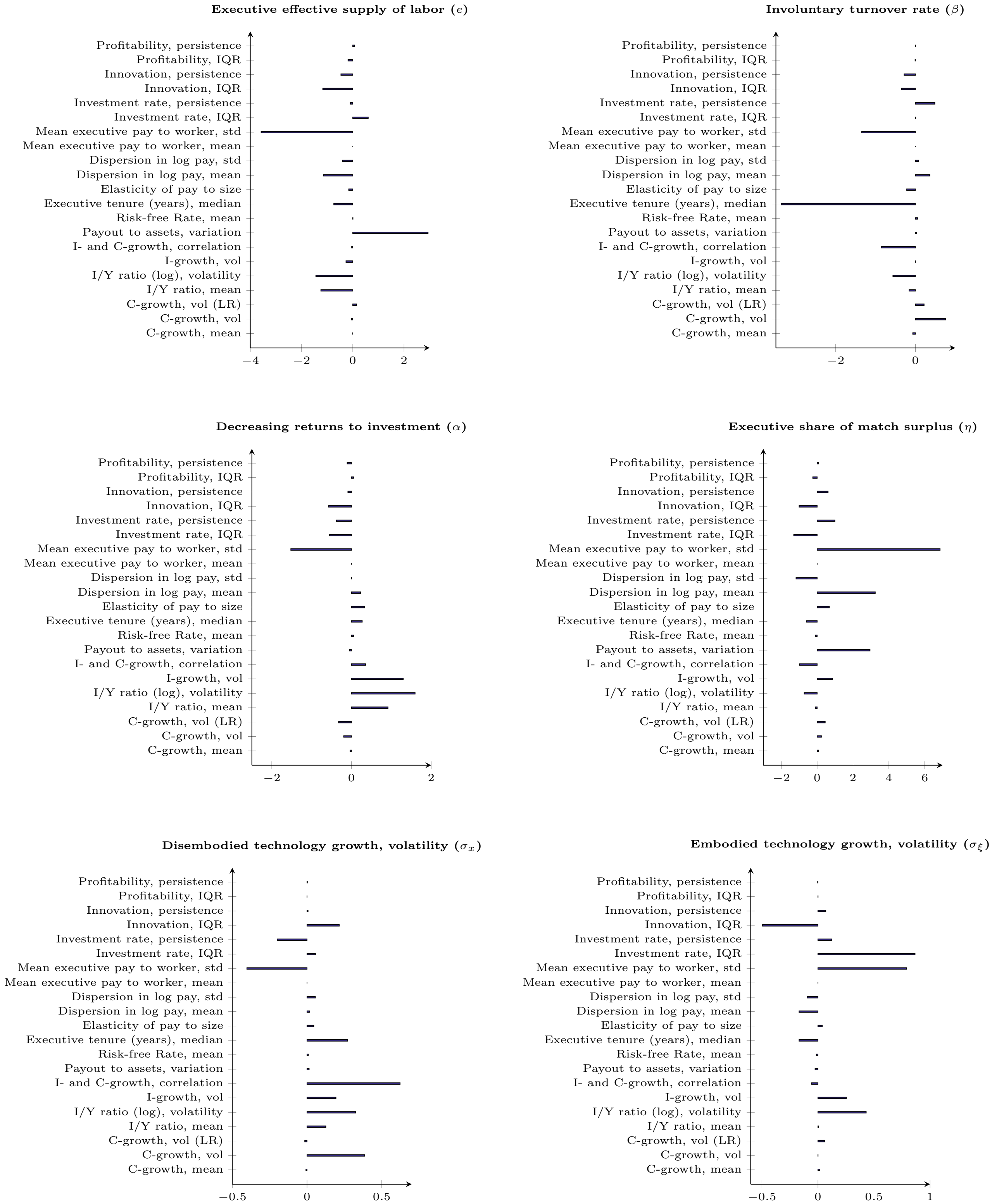
We report the elasticity of moment estimates $\mathcal{X}(\theta)$ to parameters θ. Specifically, we report $\frac{dX^j(\theta)}{d\theta^i}\frac{\theta^i}{X^j}$ where $\frac{dX^j(\theta)}{d\theta^i}$ is the numerical derivative – computed using a 5-point stencil – of moment j to parameter i.

Figure A.3: Sensitivity of moments to parameters



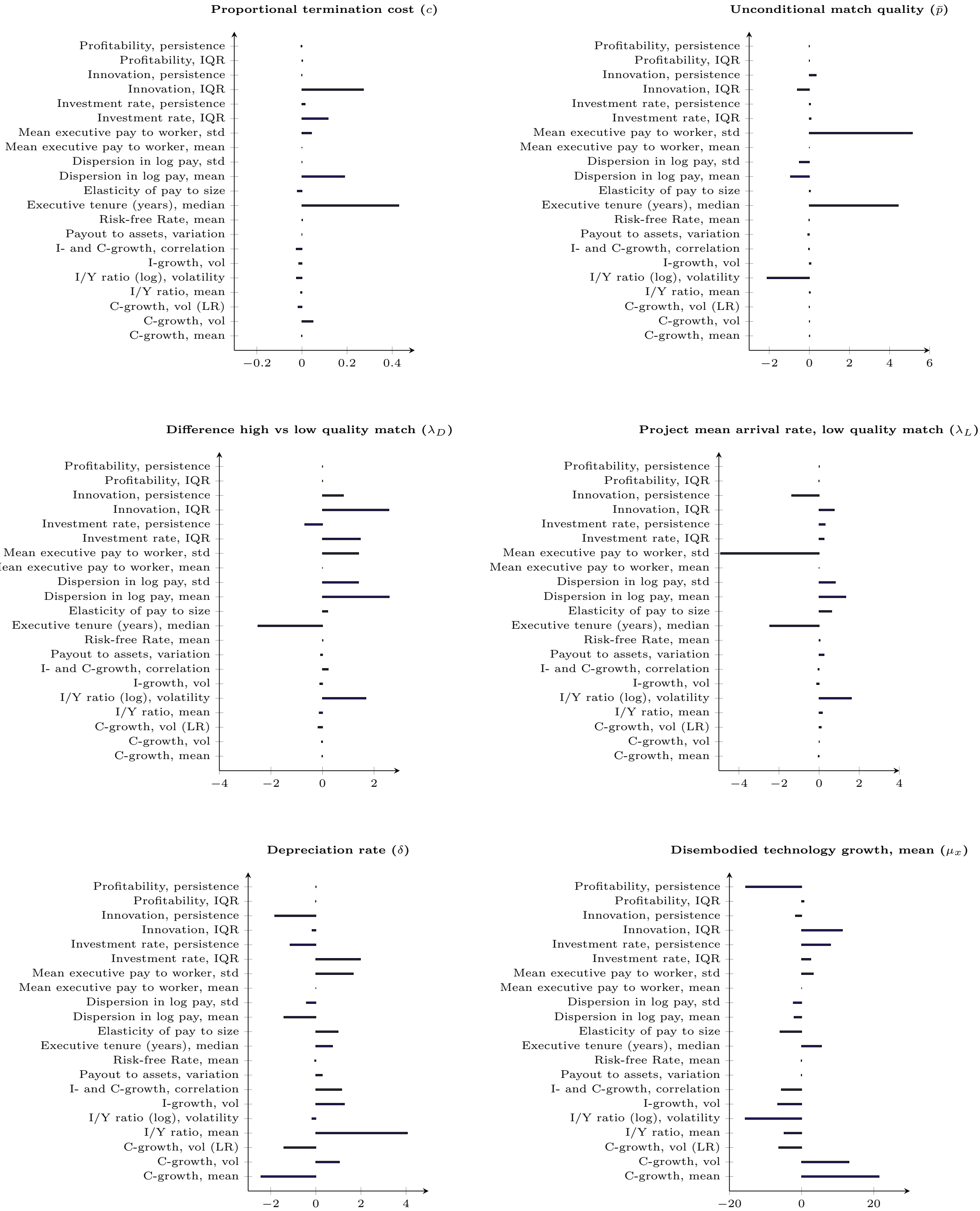
We report the elasticity of moment estimates $\mathcal{X}(\theta)$ to parameters θ. Specifically, we report $\frac{dX^j(\theta)}{d\theta^i}\frac{\theta^i}{X^j}$ where $\frac{dX^j(\theta)}{d\theta^i}$ is the numerical derivative – computed using a 5-point stencil – of moment j to parameter i.

Figure A.4: Sensitivity of parameters to moments (GS)



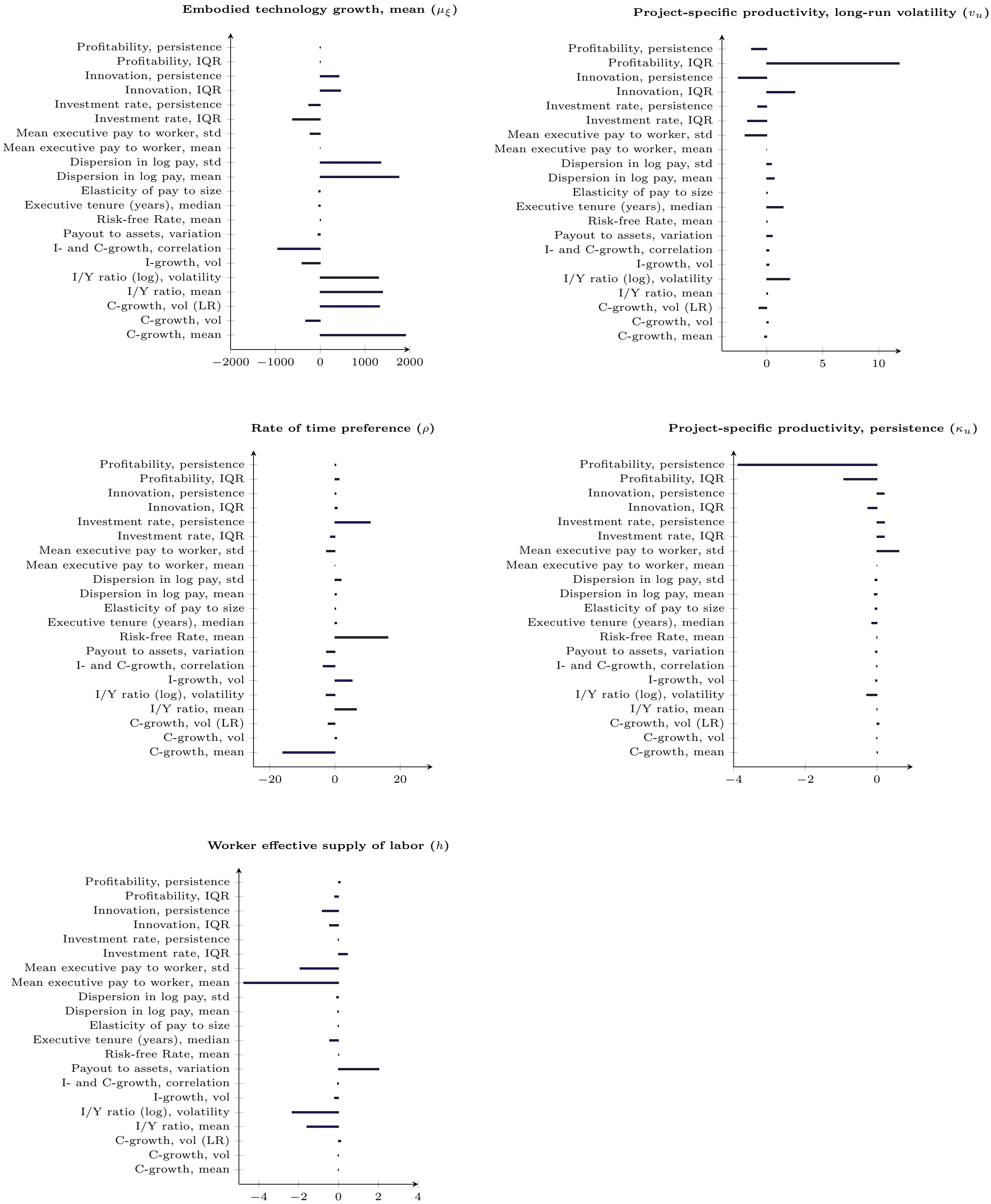
We report the Gentzkow and Shapiro (2014) sensitivity measure of the estimated parameters to moments. We report the measure in elasticity form, $\lambda_{i,j} \frac{X^j}{\theta^i}$, where $\lambda_{i,j}$ is the element of the sensitivity matrix Λ that corresponds to parameter i and moment j.

Figure A.5: Sensitivity of parameters to moments (GS)



We report the Gentzkow and Shapiro (2014) sensitivity measure of the estimated parameters to moments. We report the measure in elasticity form, $\lambda_{i,j}\frac{X^j}{\theta^i}$, where $\lambda_{i,j}$ is the element of the sensitivity matrix Λ that corresponds to parameter i and moment j.

Figure A.6: Sensitivity of parameters to moments (GS)



We report the Gentzkow and Shapiro (2014) sensitivity measure of the estimated parameters to moments. We report the measure in elasticity form, $\lambda_{i,j}\frac{X^j}{\theta^i}$, where $\lambda_{i,j}$ is the element of the sensitivity matrix Λ that corresponds to parameter i and moment j.

4 Robustness to using market values as measure of size

Here, we conduct robustness checks of our empirical results (Tables 4 and 5 in the paper) with respect to two alternative measures of firm size: the market value of the firm, defined as the market value of equity plus the book value of debt (Tables A.1 and A.3) and firm sales (Tables A.2 and A.4). In addition, we perform the same analysis in simulated data from the model (Tables A.5 and A.6).

Table A.1 is the analogue of Table 4 in the paper, in which we control for the market value of the firm (market value of equity plus book value of debt) instead of book assets. Table A.3 does the same for the other measures of growth opportunities (Table 5 in the paper). Our results are mostly robust to this alternative measure of size. We see that in three out of the four cases, the coefficients on the measures of growth opportunities are statistically different from zero. The exception is Tobin's Q. This is hardly surprising, since market values are essentially the numerator in Q. Tables A.2 and A.4 in this document show that when we use firm sales (as a proxy for firm output), the results are quantitatively similar to our baseline results.

Examining more closely the point estimates in Table A.1 and Table A.3, we note that when using market values to control for size, the point estimates on our measures of growth opportunities are smaller (in the Execucomp sample). This pattern is not surprising, but it is a direct prediction of our model. Specifically, the market value of the firm in our model includes not only the market value of its existing assets, but also the present value of its future growth opportunities. That is, market values already incorporate information on future investment opportunities and thus reduce the explanatory power of the other measures.

To illustrate this point more concretely, Tables A.5 and A.6 perform the same analysis in simulated data from the model. Comparing the results when controlling for market values (Table A.5) to either our baseline results (presented in Table 6 in the paper) or results obtained from controlling for firm output (Table A.6), we see that utilizing market values as a proxy for firm size also results in smaller point estimates on growth opportunities in simulated data. These patterns imply that, in the context of the model, the theoretically correct notion of size is a variable that is related to the firm's current scale of operations, rather than market values, since the latter are contaminated by the value of growth opportunities.

Table A.1: Executive Pay and Firm Innovation – Market Value as Size

$\log X_{f,t}$	(1)	(2)	(3)	(4)	(5)
	A. Long sample (1936-2014)				
$\hat{\nu}_{f,t}/B_{f,t}$	0.142***	0.089***	0.059**	0.060**	0.055**
	(0.025)	(0.024)	(0.023)	(0.023)	(0.023)
$\log V_{f,t}$		0.099***	0.093***	0.093***	0.059***
		(0.032)	(0.031)	(0.031)	(0.021)
$\log(1 + ROA_{f,t})$			1.973***	1.832***	1.936***
			(0.527)	(0.540)	(0.322)
$\log(1 + R_{f,t})$				0.107**	0.108***
				(0.044)	(0.035)
Observations	5298	5298	5298	5298	5298
R^2	0.906	0.915	0.918	0.918	0.942
	B. Large panel (1992-2010)				
$\hat{\nu}_{f,t}/B_{f,t}$	0.237***	0.041***	0.038***	0.038***	0.042***
	(0.051)	(0.009)	(0.008)	(0.008)	(0.007)
$\log V_{f,t}$		0.426***	0.432***	0.432***	0.411***
		(0.007)	(0.008)	(0.008)	(0.014)
$\log(1 + ROA_{f,t})$			-0.284***	-0.286***	0.139***
			(0.058)	(0.058)	(0.041)
$\log(1 + R_{f,t})$				0.003	-0.018
				(0.013)	(0.012)
Observations	33551	33551	33551	33551	33551
R^2	0.261	0.660	0.662	0.662	0.797
Year FE	Y	Y	Y	Y	Y
Industry (SIC2) FE	Y	Y	Y	Y	-
Firm FE	-	-	-	-	Y

Table reports estimates of equation (39) in the text. The variable definitions are: $X_{f,t}$ is the firm level of executive compensation, defined as the average compensation (including the ex-ante value of options) in a given year of the firm's top-3 executives (long sample, i.e. FS) or the firm's top 5 executives (large panel, i.e. Execucomp); ν is the firm-level innovation measure from Kogan et al. (2016) (scaled by the book value of assets); $V_{f,t}$ is firm size, defined as the sum of the market value of equity plus the book value of long-term debt (Panel A) or as book assets (at) plus market value of equity (prcc_f times csho) minus common equity (ceq) and deferred taxes (txdb) (Panel B); $ROA_{f,t}$ is firm profitability defined as the ratio of net income scaled by property plant and equipment; $R_{f,t}$ is the firm's stock return in year t. Depending on the specification, we include industry (SIC2) or firm fixed effects. Standard errors are clustered by firm and year and are included in parentheses.

Table A.2: Executive Pay and Firm Innovation – Firm Sales as Size

$\log X_{f,t}$	(1)	(2)	(3)	(4)	(5)
	A. Long sample (1936-2014)				
$\hat{\nu}_{f,t}/B_{f,t}$	0.142***	0.088***	0.053***	0.054***	0.050**
	(0.025)	(0.020)	(0.019)	(0.019)	(0.022)
$\log Y_{f,t}$		0.309***	0.306***	0.307***	0.301***
		(0.021)	(0.021)	(0.021)	(0.031)
$\log(1 + ROA_{f,t})$			2.144***	1.957***	1.434***
			(0.382)	(0.387)	(0.292)
$\log(1 + R_{f,t})$				0.142***	0.144***
				(0.033)	(0.032)
Observations	5230	5230	5230	5230	5230
R^2	0.906	0.935	0.938	0.938	0.951
	B. Large panel (1992-2014)				
$\hat{\nu}_{f,t}/B_{f,t}$	0.237***	0.130***	0.129***	0.126***	0.076***
	(0.051)	(0.020)	(0.020)	(0.019)	(0.010)
$\log Y_{f,t}$		0.378***	0.379***	0.382***	0.345***
		(0.010)	(0.010)	(0.010)	(0.013)
$\log(1 + ROA_{f,t})$			0.010	-0.070	0.361***
			(0.066)	(0.076)	(0.059)
$\log(1 + R_{f,t})$				0.101***	0.058**
				(0.028)	(0.021)
Observations	33551	33551	33551	33551	33551
R^2	0.261	0.585	0.586	0.589	0.780
Year FE	Y	Y	Y	Y	Y
Industry (SIC2) FE	Y	Y	Y	Y	-
Firm FE	-	-	-	-	Y

Table reports estimates of equation (39) in the text. The variable definitions are: $X_{f,t}$ is the firm level of executive compensation, defined as the average compensation (including the ex-ante value of options) in a given year of the firm's top-3 executives (long sample, i.e. FS) or the firm's top 5 executives (large panel, i.e. Execucomp); ν is the firm-level innovation measure from Kogan et al. (2016) (scaled by the book value of assets); $Y_{f,t}$ is firm size, defined as firm sales; $ROA_{f,t}$ is firm profitability defined as the ratio of net income scaled by property plant and equipment; $R_{f,t}$ is the firm's stock return in year t. Depending on the specification, we include industry (SIC2) or firm fixed effects. Standard errors are clustered by firm and year and are included in parentheses.

Table A.3: Executive pay and other measures of growth opportunities – market value as measure of size

$\log X_{f,t}$	(1)	(2)	(3)	(4)	(5)
	A. Tobin's Q				
$\log Q_{f,t}$	0.165***	0.064***	0.073***	0.076***	0.041**
	(0.025)	(0.013)	(0.013)	(0.014)	(0.019)
R^2	0.223	0.658	0.661	0.661	0.796
	B. Firm Investment				
$i_{f,t}$	0.051***	0.063***	0.066***	0.066***	0.039***
	(0.017)	(0.010)	(0.009)	(0.009)	(0.009)
R^2	0.207	0.662	0.665	0.665	0.797
	C. Index of Growth Opportunities (PCA)				
$G_{f,t}$	0.201***	0.083***	0.088***	0.089***	0.073***
	(0.025)	(0.012)	(0.011)	(0.011)	(0.014)
R^2	0.271	0.675	0.678	0.678	0.811
Observations	33551	33551	33551	33551	33551
Industry FE	Y	Y	Y	Y	-
Year FE	Y	Y	Y	Y	Y
Firm FE	-	-	-	-	Y
Size (market value)	-	Y	Y	Y	Y
ROA	-	-	Y	Y	Y
Stock Return	-	-	-	Y	Y

This table reports estimates of equation (39) in the text using three alternative measures of growth opportunities: (A) Tobin's Q, defined as book assets (at) plus market value of equity (prcc_f times csho) minus common equity (ceq) and deferred taxes (txdb) divided by PPE (ppegt); (B) the firm investment rate, defined as the ratio of capital expenditures (capx) scaled by plant and equipment (ppegt); and (C) the first principal component across $\log Q$, investment rate and $\hat{\nu}/B$. Our definition of the size of the firm is the market value of the firm. See the notes to Table 4 in the paper for additional variable definitions. Depending on the specification, we include industry (SIC2) or firm fixed effects. Standard errors are clustered by firm and year. ***,** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table A.4: Executive pay and other measures of growth opportunities – firm sales as measure of size

$\log X_{f,t}$	(1)	(2)	(3)	(4)	(5)
			A. Tobin's Q		
$\log Q_{f,t}$	0.165***	0.272***	0.286***	0.281***	0.305***
	(0.025)	(0.020)	(0.022)	(0.022)	(0.026)
R^2	0.249	0.621	0.626	0.626	0.802
			B. Firm Investment		
$i_{f,t}$	0.051***	0.148***	0.150***	0.149***	0.092***
	(0.017)	(0.013)	(0.013)	(0.014)	(0.012)
R^2	0.207	0.589	0.591	0.594	0.782
			C. Index of Growth Opportunities (PCA)		
$G_{f,t}$	0.201***	0.259***	0.266***	0.261***	0.199***
	(0.025)	(0.011)	(0.011)	(0.012)	(0.012)
R^2	0.243	0.616	0.619	0.620	0.789
Observations	33619	33619	33619	33619	33619
Industry FE	Y	Y	Y	Y	-
Year FE	Y	Y	Y	Y	Y
Firm FE	-	-	-	-	Y
Size (sales)	-	Y	Y	Y	Y
ROA	-	-	Y	Y	Y
Stock Return	-	-	-	Y	Y

This table reports estimates of equation (39) in the text using three alternative measures of growth opportunities: (A) Tobin's Q, defined as book assets (at) plus market value of equity (prcc_f times csho) minus common equity (ceq) and deferred taxes (txdb) divided by PPE (ppegt); (B) the firm investment rate, defined as the ratio of capital expenditures (capx) scaled by plant and equipment (ppegt); and (C) the first principal component across $\log Q$, investment rate and $\hat{\nu}/B$. Our definition of the size of the firm is the value of sales. See the notes to Table 4 in the paper for additional variable definitions. Depending on the specification, we include industry (SIC2) or firm fixed effects. Standard errors are clustered by firm and year. ***,** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table A.5: Model: Executive pay and other measures of growth opportunities – market value as measure of size

$\log X_{f,t}$	(1)	(2)	(3)	(4)	(5)
	A. Firm Innovation (KPSS)				
$\hat{\nu}_{f,t}/B_{f,t}$	0.249	0.149	0.146	0.196	0.074
	(0.056)	(0.040)	(0.042)	(0.037)	(0.027)
R^2	0.119	0.650	0.652	0.661	0.879
	B. Tobin's Q				
$\log Q_{f,t}$	-0.188	0.625	0.655	0.682	0.734
	(0.218)	(0.156)	(0.161)	(0.168)	(0.150)
R^2	0.119	0.831	0.848	0.857	0.931
	C. Firm Investment				
$i_{f,t}$	0.207	0.066	0.064	0.091	0.061
	(0.021)	(0.021)	(0.022)	(0.025)	(0.024)
R^2	0.110	0.493	0.493	0.498	0.805
	D. Index of Growth Opportunities (PCA)				
$G_{f,t}$	0.255	0.085	0.082	0.117	0.094
	(0.040)	(0.020)	(0.021)	(0.020)	(0.024)
R^2	0.110	0.524	0.525	0.535	0.811
Observations	30000	30000	30000	30000	30000
Year FE	Y	Y	Y	Y	Y
Firm FE	-	-	-	-	Y
Size	-	Y	Y	Y	Y
ROA	-	-	Y	Y	Y
Stock Return	-	-	-	Y	Y

Table reports estimates of equation (39) in simulated data from the model using four alternative measures of growth opportunities: (A) the value of new projects (constructed using equation (33) in the text) scaled by the book value of capital; (B) Tobin's Q, defined as the ratio of the market value of the firm scaled by book value of capital; (C) firm investment rate, defined as the change in log capital and (D) the first principal component across (A), (B) and (C). We computed model standard errors as the standard deviation of parameter estimates across $S = 100$ simulations. See the main text and the Online Appendix for more details on the model estimation and simulation methodology.

Table A.6: Model: Executive pay and other measures of growth opportunities – firm output as measure of size

$\log X_{f,t}$	(1)	(2)	(3)	(4)	(5)
	A. Firm Innovation (KPSS)				
$\hat{\nu}_{f,t}/B_{f,t}$	0.2497	0.223	0.201	0.241	0.129
	(0.056)	(0.033)	(0.036)	(0.036)	(0.023)
R^2	0.119	0.371	0.465	0.471	0.792
	B. Tobin's Q				
$\log Q_{f,t}$	-0.188	0.282	0.999	1.021	1.020
	(0.218)	(0.123)	(0.183)	(0.188)	(0.115)
R^2	0.117	0.360	0.717	0.721	0.879
	C. Firm Investment				
$i_{f,t}$	0.207	0.199	0.183	0.215	0.116
	(0.021)	(0.035)	(0.041)	(0.044)	(0.021)
R^2	0.115	0.361	0.458	0.462	0.790
	D. Index of Growth Opportunities (PCA)				
$G_{f,t}$	0.255	0.237	0.241	0.292	0.153
	(0.040)	(0.044)	(0.058)	(0.060)	(0.025)
R^2	0.110	0.378	0.487	0.497	0.796
Observations	30000	30000	30000	30000	30000
Year FE	Y	Y	Y	Y	Y
Firm FE	-	-	-	-	Y
Size	-	Y	Y	Y	Y
ROA	-	-	Y	Y	Y
Stock Return	-	-	-	Y	Y

Table reports estimates of equation (39) in simulated data from the model using four alternative measures of growth opportunities: (A) the value of new projects (constructed using equation (33) in the text) scaled by the book value of capital; (B) Tobin's Q, defined as the ratio of the market value of the firm scaled by book value of capital; (C) firm investment rate, defined as the change in log capital and (D) the first principal component across (A), (B) and (C). We computed model standard errors as the standard deviation of parameter estimates across $S = 100$ simulations. See the main text and the Online Appendix for more details on the model estimation and simulation methodology.

References

Barro, Robert J. and Jose F. Ursua. "Macroeconomic Crises since 1870." *Brookings Papers on Economic Activity*, 39(2008), 255–350.

Björkman, Mattias and Kenneth Holmström. "Global Optimization of Costly Nonconvex Functions Using Radial Basis Functions." *Optimization and Engineering*, 1(2000), 373–397.

Davis, Morris A., Jonas D. M. Fisher, and Toni M. Whited. "Macroeconomic Implications of Agglomeration." *Econometrica*, 82(2014), 731–764.

Dew-Becker, Ian. "How risky is consumption in the long-run? Benchmark estimates from a novel unbiased and efficient estimator." Working paper, Northwestern University (2014).

Duffie, Darrel and Pierre-Louis Lions. "PDE solutions of stochastic differential utility." *Journal of Mathematical Economics*, 21(1992), 577–606.

Erickson, Timothy and Toni M. Whited. "Treating Measurement Error in Tobin's q." *Review of Financial Studies*, 25(2012), 1286–1329.

Frydman, Carola and Raven E. Saks. "Executive Compensation: A New View from a Long-Term Perspective, 19362005." *Review of Financial Studies*, 23(2010), 2099–2138.

Gentzkow, Matthew and Jesse M. Shapiro. "Measuring the Sensitivity of Parameter Estimates to Sample Statistics." Working Paper 20673, National Bureau of Economic Research (2014).

Ingram, Beth Fisher and Bong-Soo Lee. "Simulation estimation of time-series models." *Journal of Econometrics*, 47(1991), 197–205.

Karatzas, Ioannis and Steven Eugene Shreve. *Brownian motion and stochastic calculus*. Graduate texts in mathematics. Springer-Verlag, New York, Berlin, Heidelberg (1991). Autres tirages corrigs : 1996, 1997, 1999, 2000, 2005.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. "Technological Innovation, Resource Allocation, and Growth." *Quarterly Journal of Economics*, forthcoming(2016).

Larrain, Borja and Motohiro Yogo. "Does firm value move too much to be justified by subsequent changes in cash flow?" *Journal of Financial Economics*, 87(2008), 200 – 226.

Ledoit, Olivier and Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, 88(2004), 365 – 411.

Rousseeuw, Peter J. and Geert Molenberghs. "Transformation of non positive semidefinite correlation matrices." *Communications in Statistics - Theory and Methods*, 22(1993), 965–984.

Sekyu, Choi and Jos-Vctor Rios-Rull. "Understanding the Dynamics of the Labor Share: the Role of non-Competitive Factor Prices." *Annales d'Economie et de Statistique*, (2009), 251–277.

van Wieringen, Wessel N. and Carel F.W. Peeters. "Ridge estimation of inverse covariance matrices from high-dimensional data." *Computational Statistics & Data Analysis*, 103(2016), 284 – 303.