ECO220Y1Y, Test #4, Prof. Murdock

April 1, 2022, 9:10 – 11:00 am

U of T E-MAIL: ___________________________________@MAIL.UTORONTO.CA

SURNAME (LAST NAME):																			
GIVEN NAME (FIRST NAME):																			
UTORID: (e.g. LIHAO118)																			

Instructions:

- You have 110 minutes. Keep these test papers and the *Supplement* closed and face up on your desk until the start of the test is announced. You must stay for a minimum of 60 minutes.
- You may use a **non-programmable calculator**.
- There are 4 questions (all with multiple parts) with varying point values worth a total of 95 points.
- This test includes these 8 pages plus the *Supplement*. The *Supplement* contains the aid sheets, statistical tables (Standard Normal, Student t, and F), readings, figures, tables, and other materials for some test questions. For each question referencing the *Supplement*, carefully review *all* materials. ***The Supplement will NOT be collected:*** write your answers on these test papers. When we announce the end of the test, hand these test papers to us (you keep the *Supplement*).
- Write your answers clearly, completely, and concisely in the designated space provided immediately after each question. An answer guide ends each question to let you know what is expected. For example, a quantitative analysis, a fully labelled graph, and/or sentences. Any answer guide asking for a quantitative analysis *always* automatically means that you must show your work and make your reasoning clear.
    - Anything requested by the question and/or the answer guide is required. Similarly, limit yourself to the answer guide. For example, if the answer guide does not request sentences, provide only what is requested (e.g. quantitative analysis).
    - Marking TAs are instructed to accept all reasonable rounding.
- ***Your entire answer must fit in the designated space provided immediately after each question.*** No extra space/pages are possible. You *cannot* use blank space for other questions, nor can you write answers on the *Supplement*. ***Write in PENCIL and use an ERASER as needed*** so that you can fit your final answer (including work and reasoning) in the appropriate space. Questions give more blank space than is needed for an answer (with typical handwriting) worth full marks. ***Follow the answer guides and avoid excessively long answers.***

(1) See ***Supplement for Question (1)***: *Learning by Doing: Evidence from an Automobile Assembly Plant.*

(a) [8 pts] Compute and then *interpret* the 95% confidence interval estimate for the row ln_cum_prod in the Stata regression output. Answer with a quantitative analysis & 1 – 2 sentences.

(b) [3 pts] Compute the standard deviation of the variable named ln_ave_def_pc. Answer with a quantitative analysis.

(2) See ***Supplement for Question (2)***: *Does LEED Certification Save Energy? Evidence from Federal Buildings.*

(a) [8 pts] If there were 19 buildings in the sample, is the correlation between **OS** and **ES** statistically significant? What can we say about the P-value? Answer with hypotheses in formal notation, a quantitative analysis & 1 – 2 sentences.

(b) [6 pts] For a certain number of buildings the P-value is 0.0129 when testing if the correlation between **MS** and **SS** is statistically significant. What can we conclude about the P-value for that same test but between **IS** and **CR**? Why? (Focus on the relevant *concept*, and *not* computational mechanics.) Answer with 1 – 2 sentences.

(3) See ***Supplement for Question (3)***: *Do Well Managed Firms Make Better Forecasts?*

(a) [6 pts] In **Table 1**, the first two rows of results are for management score and employment. For each of these two variables, which best describes the shape of the distribution: "extremely negatively skewed," "somewhat negatively skewed," "somewhat positively skewed," or "extremely positively skewed"? *Explain*. Answer with 2 – 3 sentences.

(b) [2 pts] Given **Table 1**, what percent of firms in these data are *not* family owned? Answer with a quantitative analysis.

(c) [6 pts] Given **Table 1**, for the row "Foreign owned," *why* does the standard deviation not give any further information? *Explain* focusing on the relevant *concepts* and *not* the mechanics of computation. Next, compute that standard deviation. Answer with 2 sentences & a quantitative analysis.

(d) [7 pts] Given **<u>Figure 3</u>**, what is the *approximate* point estimate of the difference between MES replies and professional forecasters for expected GDP growth from **<u>1% to 3%</u>**? *Fully interpret* that point estimate. Assess if it is large or small in this context and support your claim with relevant information from Figure 3. <u>Answer with 2 – 3 sentences.</u>

(e) [7 pts] Continuing with Part **(d)**, how to assess if the difference (between MES replies and professional forecasters for expected GDP growth from 1% to 3%) is statistically significant? Include the hypotheses. What further information is needed to do the test? Also, offer intelligent speculation about the likely outcome of the test supporting your claim with relevant information about this research in *The Supplement*. <u>Answer with 2 sentences & hypotheses in formal notation.</u>

(f) [6 pts] Given **Table 3**, what is the *full interpretation* of 0.063 in Column (1)? Complete the sentence below & add 1 more sentence to discuss the magnitude (size).

Using the 2016 management survey and annual business survey, which include observations for a sample of 7,756 firms (both manufacturing and non-manufacturing firms) with at least 10 employees in the UK,

(g) [5 pts] Given **Table 3**, what is the *interpretation* of -0.015 in Column (4)? (Presume your reader has already read the complete response to the previous part and do not repeat things already explained.) Answer with 1 precise sentence.

(h) [3 pts] Given **Figure 6**, how would the estimated values of the OLS intercept, slope, and R^2 change if the observation with an absolute GDP forecast error near 4.5 were dropped? Fill in the blanks with "increase" or "decrease."

The intercept would ________________, the slope would ________________, and the R^2 would ________________.

(i) [6 pts] The researchers interpret -0.358 in Column (1) of **<u>Table 5</u>** in the <u>excerpt</u> above it. Why don't they do a one-unit change in the x variable? Next, is that point estimate *economically* significant? Explain. <u>Answer with 2 – 3 sentences.</u>

(j) [6 pts] For **<u>Table 5</u>**, what does 0.060 in the last row of Column (6) measure? Is it large or small in this context? What does its size mean? <u>Answer with 2 sentences.</u>

(k) [6 pts] For **<u>Table 5</u>**, compute the P-value that justifies "**" next to the estimate of -0.171 in the first row of results in Column (6). <u>Answer with hypotheses in formal notation & a quantitative analysis.</u>

(4) Consider Rossi (2022) "The Relative Efficiency of Skilled Labor across Countries: Measurement and Interpretation."

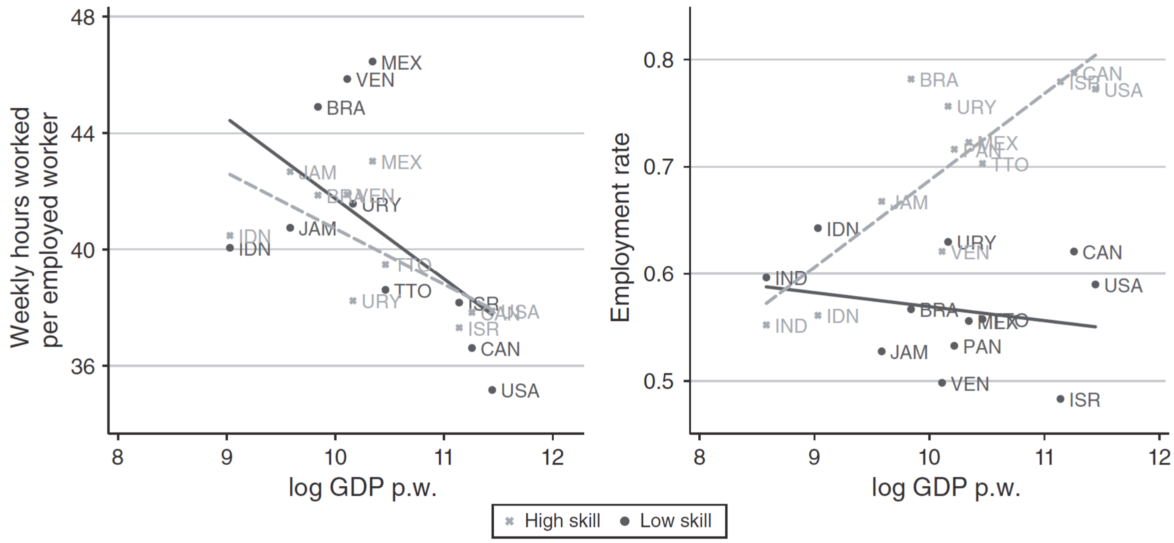


Figure 2. Hours Worked and Employment Rate by Skill Level

Notes: The figure plots the skill-specific average weekly hours per employed worker (left panel) and employment rate (right panel) against log GDP per worker for the countries in the sample. The left panel does not include India and Panama, as no data on hours worked are available for these countries. The solid (dashed) line represents the best linear fit for the low-skill (high-skill) group.

(a) [4 pts] What is the key message of the left panel of Figure 2? Answer with 1 sentence.

(b) [6 pts] For the right panel of Figure 2, which of the point estimates ($b_0,\ b_1,\ b_2,\ b_3$) would be negative in the OLS regression: $\widehat{emp\_rate}_{it} = b_0 + b_1 \ln\left(GDP_{pw}\right)_{it} + b_2 dum\_low\_skill_{it} + b_3 dum\_low\_skill_{it} \times \ln\left(GDP_{pw}\right)_{it}$. *Explain*. Answer with 2 – 3 sentences.

This *Supplement* contains the aid sheets, statistical tables (Standard Normal, Student t, and F), readings, figures, tables, and other materials for some test questions. Review all relevant materials for each question.

Sample mean: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ **Sample variance:** $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$ **Sample s.d.:** $s = \sqrt{s^2}$

Sample coefficient of variation: $CV = \frac{s}{\bar{X}}$ **Sample covariance:** $s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(n-1)}$

Sample interquartile range: $IQR = Q3 - Q1$ **Sample coefficient of correlation:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

Addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ **Conditional probability:** $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

Complement rules: $P(A^C) = P(A') = 1 - P(A)$ $P(A^C|B) = P(A'|B) = 1 - P(A|B)$

Multiplication rule: $P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$

Expected value: $E[X] = \mu = \sum_{all\ x} xp(x)$ **Variance:** $V[X] = E[(X-\mu)^2] = \sigma^2 = \sum_{all\ x}(x-\mu)^2 p(x)$

Covariance: $COV[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = \sum_{all\ x}\sum_{all\ y}(x-\mu_X)(y-\mu_Y)p(x,y)$

<u>Laws of expected value</u>:

$E[c] = c$

$E[X+c] = E[X] + c$

$E[cX] = cE[X]$

$E[a+bX+cY] = a + bE[X] + cE[Y]$

<u>Laws of variance</u>:

$V[c] = 0$

$V[X+c] = V[X]$

$V[cX] = c^2V[X]$

$V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc * COV[X,Y]$

$V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc * SD(X) * SD(Y) * \rho$

where $\rho = CORRELATION[X,Y]$

<u>Laws of covariance</u>:

$COV[X,c] = 0$

$COV[a+bX, c+dY] = bd * COV[X,Y]$

Combinatorial formula: $C_x^n = \frac{n!}{x!(n-x)!}$ **Binomial probability:** $p(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$ for $x = 0,1,2,\ldots,n$

If X is Binomial $(X \sim B(n,p))$ **then** $E[X] = np$ **and** $V[X] = np(1-p)$

If X is Uniform $(X \sim U[a,b])$ **then** $f(x) = \frac{1}{b-a}$ **and** $E[X] = \frac{a+b}{2}$ **and** $V[X] = \frac{(b-a)^2}{12}$

<u>Sampling distribution of $\bar{X}$</u>:

$\mu_{\bar{X}} = E[\bar{X}] = \mu$

$\sigma_{\bar{X}}^2 = V[\bar{X}] = \frac{\sigma^2}{n}$

$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

<u>Sampling distribution of $\hat{P}$</u>:

$\mu_{\hat{P}} = E[\hat{P}] = p$

$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$

$\sigma_{\hat{P}} = SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}}$

<u>Sampling distribution of $(\hat{P}_2 - \hat{P}_1)$</u>:

$\mu_{\hat{P}_2-\hat{P}_1} = E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1$

$\sigma_{\hat{P}_2-\hat{P}_1}^2 = V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$

$\sigma_{\hat{P}_2-\hat{P}_1} = SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}}$

<u>Sampling distribution of $(\bar{X}_1 - \bar{X}_2)$, independent samples</u>:

$\mu_{\bar{X}_1-\bar{X}_2} = E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$

$\sigma_{\bar{X}_1-\bar{X}_2}^2 = V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$\sigma_{\bar{X}_1-\bar{X}_2} = SD[\bar{X}_1 - \bar{X}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

<u>Sampling distribution of $(\bar{X}_d)$, paired $(d = X_1 - X_2)$</u>:

$\mu_{\bar{X}_d} = E[\bar{X}_d] = \mu_1 - \mu_2$

$\sigma_{\bar{X}_d}^2 = V[\bar{X}_d] = \frac{\sigma_d^2}{n} = \frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}$

$\sigma_{\bar{X}_d} = SD[\bar{X}_d] = \frac{\sigma_d}{\sqrt{n}} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}}$

Inference about a population proportion:

***z* test statistic:** $z = \frac{\hat{P}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ **CI estimator:** $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

Inference about comparing two population proportions:

***z* test statistic under Null hypothesis of no difference:** $z = \frac{\hat{P}_2-\hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1}+\frac{\bar{P}(1-\bar{P})}{n_2}}}$ **Pooled proportion:** $\bar{P} = \frac{X_1+X_2}{n_1+n_2}$

CI estimator: $(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$

Inference about the population mean:

***t* test statistic:** $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ **CI estimator:** $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ **Degrees of freedom:** $\nu = n-1$

Inference about a comparing two population means, independent samples, unequal variances:

***t* test statistic:** $t = \frac{(\bar{X}_1-\bar{X}_2)-\Delta_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ **CI estimator:** $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

Degrees of freedom: $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$

Inference about a comparing two population means, independent samples, assuming equal variances:

***t* test statistic:** $t = \frac{(\bar{X}_1-\bar{X}_2)-\Delta_0}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}}$ **CI estimator:** $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$ **Degrees of freedom:** $\nu = n_1 + n_2 - 2$

Pooled variance: $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Inference about a comparing two population means, paired data: (n is number of pairs and $d = X_1 - X_2$)

***t* test statistic:** $t = \frac{\bar{d}-\Delta_0}{s_d/\sqrt{n}}$ **CI estimator:** $\bar{X}_d \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$ **Degrees of freedom:** $\nu = n-1$

SIMPLE REGRESSION:

Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ **OLS line:** $\hat{y}_i = b_0 + b_1 x_i$ $b_1 = \frac{s_{xy}}{s_x^2} = r\frac{s_y}{s_x}$ $b_0 = \bar{Y} - b_1\bar{X}$

Coefficient of determination: $R^2 = (r)^2$ **Residuals:** $e_i = y_i - \hat{y}_i$

Standard deviation of residuals: $s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(e_i-0)^2}{n-2}}$ **Standard error of slope:** $s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$

Inference about the population slope:

***t* test statistic:** $t = \frac{b_1 - \beta_{10}}{s.e.(b_1)}$ **CI estimator:** $b_1 \pm t_{\alpha/2} s.e.(b_1)$ **Degrees of freedom:** $\nu = n - 2$

Standard error of slope: $s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$

Prediction interval for y at given value of x (x_g):

$$\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}} \quad \textbf{or} \quad \hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{\left(s.e.(b_1)\right)^2 \left(x_g - \bar{X}\right)^2 + \frac{s_e^2}{n} + s_e^2}$$

Degrees of freedom: $\nu = n - 2$

Confidence interval for predicted mean at given value of x (x_g):

$$\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}} \quad \textbf{or} \quad \hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{\left(s.e.(b_1)\right)^2 \left(x_g - \bar{X}\right)^2 + \frac{s_e^2}{n}}$$

Degrees of freedom: $\nu = n - 2$

SIMPLE & MULTIPLE REGRESSION:

Model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$

$$SST = \sum_{i=1}^{n} (y_i - \bar{Y})^2 = SSR + SSE \quad SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{Y})^2 \quad SSE = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$s_y^2 = \frac{SST}{n-1} \quad MSE = \frac{SSE}{n-k-1} \quad Root\ MSE = \sqrt{\frac{SSE}{n-k-1}} \quad MSR = \frac{SSR}{k}$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad Adj.R^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = \left(R^2 - \frac{k}{n-1}\right)\left(\frac{n-1}{n-k-1}\right)$$

Residuals: $e_i = y_i - \hat{y}_i$ **Standard deviation of residuals:** $s_e = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - 0)^2}{n-k-1}}$

Inference about the overall statistical significance of the regression model:

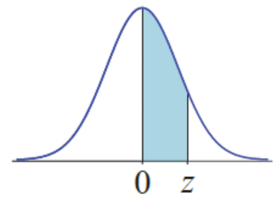
$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{(SST-SSE)/k}{SSE/(n-k-1)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

Numerator degrees of freedom: $\nu_1 = k$ **Denominator degrees of freedom:** $\nu_2 = n - k - 1$

Inference about the population slope for explanatory variable j:

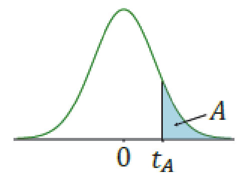
***t* test statistic:** $t = \frac{b_j - \beta_{j0}}{s_{b_j}}$ **CI estimator:** $b_j \pm t_{\alpha/2} s_{b_j}$ **Degrees of freedom:** $\nu = n - k - 1$

Standard error of slope: $s.e.(b_j) = s_{b_j}$ (for multiple regression, must be obtained from technology)



The Standard Normal Distribution:

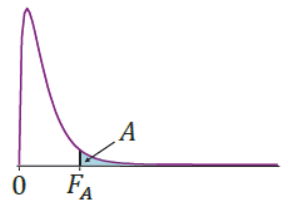
	Second decimal place in z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999



Critical Values of Student t Distribution:

ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$	ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$
1	3.078	6.314	12.71	31.82	63.66	318.3	636.6	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
2	1.886	2.920	4.303	6.965	9.925	22.33	31.60	39	1.304	1.685	2.023	2.426	2.708	3.313	3.558
3	1.638	2.353	3.182	4.541	5.841	10.21	12.92	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	41	1.303	1.683	2.020	2.421	2.701	3.301	3.544
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	43	1.302	1.681	2.017	2.416	2.695	3.291	3.532
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	45	1.301	1.679	2.014	2.412	2.690	3.281	3.520
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	47	1.300	1.678	2.012	2.408	2.685	3.273	3.510
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	49	1.299	1.677	2.010	2.405	2.680	3.265	3.500
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	51	1.298	1.675	2.008	2.402	2.676	3.258	3.492
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	52	1.298	1.675	2.007	2.400	2.674	3.255	3.488
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	53	1.298	1.674	2.006	2.399	2.672	3.251	3.484
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	54	1.297	1.674	2.005	2.397	2.670	3.248	3.480
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	75	1.293	1.665	1.992	2.377	2.643	3.202	3.425
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	140	1.288	1.656	1.977	2.353	2.611	3.149	3.361
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	160	1.287	1.654	1.975	2.350	2.607	3.142	3.352
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	180	1.286	1.653	1.973	2.347	2.603	3.136	3.345
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340
31	1.309	1.696	2.040	2.453	2.744	3.375	3.633	250	1.285	1.651	1.969	2.341	2.596	3.123	3.330
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	300	1.284	1.650	1.968	2.339	2.592	3.118	3.323
33	1.308	1.692	2.035	2.445	2.733	3.356	3.611	400	1.284	1.649	1.966	2.336	2.588	3.111	3.315
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	500	1.283	1.648	1.965	2.334	2.586	3.107	3.310
35	1.306	1.690	2.030	2.438	2.724	3.340	3.591	750	1.283	1.647	1.963	2.331	2.582	3.101	3.304
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	1000	1.282	1.646	1.962	2.330	2.581	3.098	3.300
37	1.305	1.687	2.026	2.431	2.715	3.326	3.574	∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Degrees of freedom: ν



The F Distribution:

ν_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	30	∞
ν_2	**Critical Values of F Distribution for $A = 0.10$:**															
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.24	3.21	3.17	3.10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.24	2.20	2.16	2.06
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	1.97	1.92	1.87	1.76
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.84	1.79	1.74	1.61
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.72	1.67	1.61	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.66	1.61	1.54	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.60	1.54	1.48	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.63	1.60	1.55	1.48	1.41	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55	1.49	1.42	1.34	1.00
ν_2	**Critical Values of F Distribution for $A = 0.05$:**															
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.50	4.36
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.70	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.25	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.04	1.84
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.84	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.74	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.84	1.75	1.65	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.75	1.66	1.55	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.67	1.57	1.46	1.00
ν_2	**Critical Values of F Distribution for $A = 0.01$:**															
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.72	9.55	9.38	9.02
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.25	3.91
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.21	2.87
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.78	2.42
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.39	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.20	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.35	2.20	2.03	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.19	2.03	1.86	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.04	1.88	1.70	1.00
ν_2	**Critical Values of F Distribution for $A = 0.001$:**															
5	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	26.6	26.4	25.9	25.4	24.9	23.8
10	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.59	8.45	8.13	7.80	7.47	6.76
15	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.94	5.81	5.54	5.25	4.95	4.31
20	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.94	4.82	4.56	4.29	4.00	3.38
30	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.11	4.00	3.75	3.49	3.22	2.59
40	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.40	3.14	2.87	2.23
60	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.42	3.32	3.08	2.83	2.55	1.89
120	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.12	3.02	2.78	2.53	2.26	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.84	2.74	2.51	2.27	1.99	1.00

Numerator degrees of freedom: ν_1; Denominator degrees of freedom: ν_2

Supplement for Question (1): Recall Levitt et al. (2013) "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." The Stata output below uses the weekly data. The variable named ln_ave_def_pc is the natural logarithm of the average defects per car produced that week. The variable named ln_cum_prod is the natural logarithm of cumulative production in the previous weeks. The variable named time_trend is the time trend in weeks. Like the authors, the regression below is limited to weeks where at least 100 vehicles are produced.

Note: In the Stata regression output below some numbers have been intentionally erased.

```
. regress ln_ave_def_pc ln_cum_prod time_trend if wk_prod>=100;

      Source |       SS           df       MS      Number of obs   =        47
-------------+----------------------------------   F(2, 44)        =    680.50
       Model |  12.6369307                         Prob > F        =    0.0000
    Residual |  .408538738                         R-squared       =    0.9687
-------------+----------------------------------   Adj R-squared   =    0.9673
       Total |  13.0454694                         Root MSE        =

------------------------------------------------------------------------------
ln_ave_def_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ln_cum_prod |  -.3353905   .0159661   -21.01   0.000
   time_trend |   .0070437   .0021041     3.35   0.002
        _cons |   5.960341   .1185408    50.28   0.000
------------------------------------------------------------------------------
```

Supplement for Question (2): Recall Clay et al. (2021) "Does LEED Certification Save Energy? Evidence from Federal Buildings." The number of buildings (i.e. the sample size) is missing from the table and its description.

Table A.4: Correlation Matrix of Attributes in the LEED Program

	CR	ES	SS	IS	MS	WS	OS
CR	1.000						
ES	0.124	1.000					
SS	0.008	0.148	1.000				
IS	-0.168	-0.444	0.055	1.000			
MS	-0.175	-0.472	0.159	0.671	1.000		
WS	-0.078	0.186	0.072	0.018	0.028	1.000	
OS	-0.253	-0.428	-0.019	0.049	0.493	-0.021	1.000

Notes: This is the correlation matrix for the attributes in the LEED program. CR stands for certified ratio: it is the fraction of a building square footage that is LEED-certified. ES, SS, IS, MS, WS, and OS are the component scores evaluated by the LEED program: energy and atmosphere (ES), sustainable sites (SS), indoor environmental quality (IS), materials and resources (MS), water efficiency (WS), and innovation in operation (OS), respectively.

Supplement for Question (3): Consider Bloom et al. (2021) "Do Well Managed Firms Make Better Forecasts?" See the excerpts, Table 1, Figure 3, Table 3, Figure 6, and Table 5. Well managed firms have structured management practices.

EXCERPT, ABSTRACT: We link a new UK management survey covering 8,000 firms to panel data on productivity in manufacturing and services. There is a large variation in management practices, which are highly correlated with productivity, profitability, and size. Uniquely, the survey collects firms' micro forecasts of their own sales and also macro forecasts of GDP. We find that better managed firms make more accurate micro and macro forecasts, even after controlling for their size, age, industry and many other factors.

EXCERPTS, pp. 3-8: The Management and Expectations Survey (MES) is the largest ever survey on management capabilities in the UK covering both manufacturing and non-manufacturing firms, with its survey design adapted from the World Management Survey. The MES survey reference year is 2016, but it also collects firm-level expectations of turnover, expenditure, investment, and employment growth for 2017 and 2018. The MES was sent to 25,006 firms with ten or more employees. It has the same frame as the Annual Business Survey (ABS) for 2016, allowing us to match to data on value added, employment, output, and investment. The MES asks how many key performance indicators are used and how frequently employees are evaluated against these. It also asks whether targets are set, and if so, how easy or difficult it is to achieve targets. It also asks about incentives asking how much each employee's performance and ability are reflected in performance bonuses or promotion. Each question has a list of options. Respondents chose the option closest to the practices within their firm. For each question, each option is scored on a scale of 0 to 1, where 0 is the least and 1 the most structured management practice. An overall management score is a simple average of a firm's score on all individual questions: a firm scoring 1 overall would have the most structured response to all 12 questions.

EXCERPTS, pp. 9-10: A set of descriptive statistics are in Table 1. Some firms have missing values for a few control variables (e.g. share of non-managers with a college degree). Panel A shows that firms have an average management score of 0.59 with a standard deviation of 0.2. Panel B of Table 1 looks at expectations. Firms were pessimistic about macro growth 2017-18, with an average prediction of 0.1%, even though the actual growth was 1.4%. The forecast error is the (absolute) difference between the firm's estimates and the actual growth.

Note: One number has been intentionally erased from Table 1.

Table 1: Descriptive statistics

Variables	Observations	Mean	Median	Standard Deviation
Panel A: Management				
Management score	7,756	0.586	0.627	0.196
Employment in 2016 (number of employees)	7,756	282.606		1070.225
Age of firm (years since date of incorporation)	7,756	16.695	21	7.418
Family owned but not run	7,717	0.112	0	0.315
Family owned and run	7,717	0.425	0	0.494
Share of managers with a college degree	7,496	0.397	0.350	0.347
Share of non-managers with a college degree	7,109	0.236	0.100	0.269
Foreign owned	7,756	0.132	0	0.338
Panel B: Macro forecasts (shown as percentages)				
Expected GDP growth 2017-18	7,756	0.096	0.000	1.047
Absolute GDP forecast error 2018	7,756	1.410	1.398	0.918

Notes: Panels A and B describe the cleaned sample for the analysis.

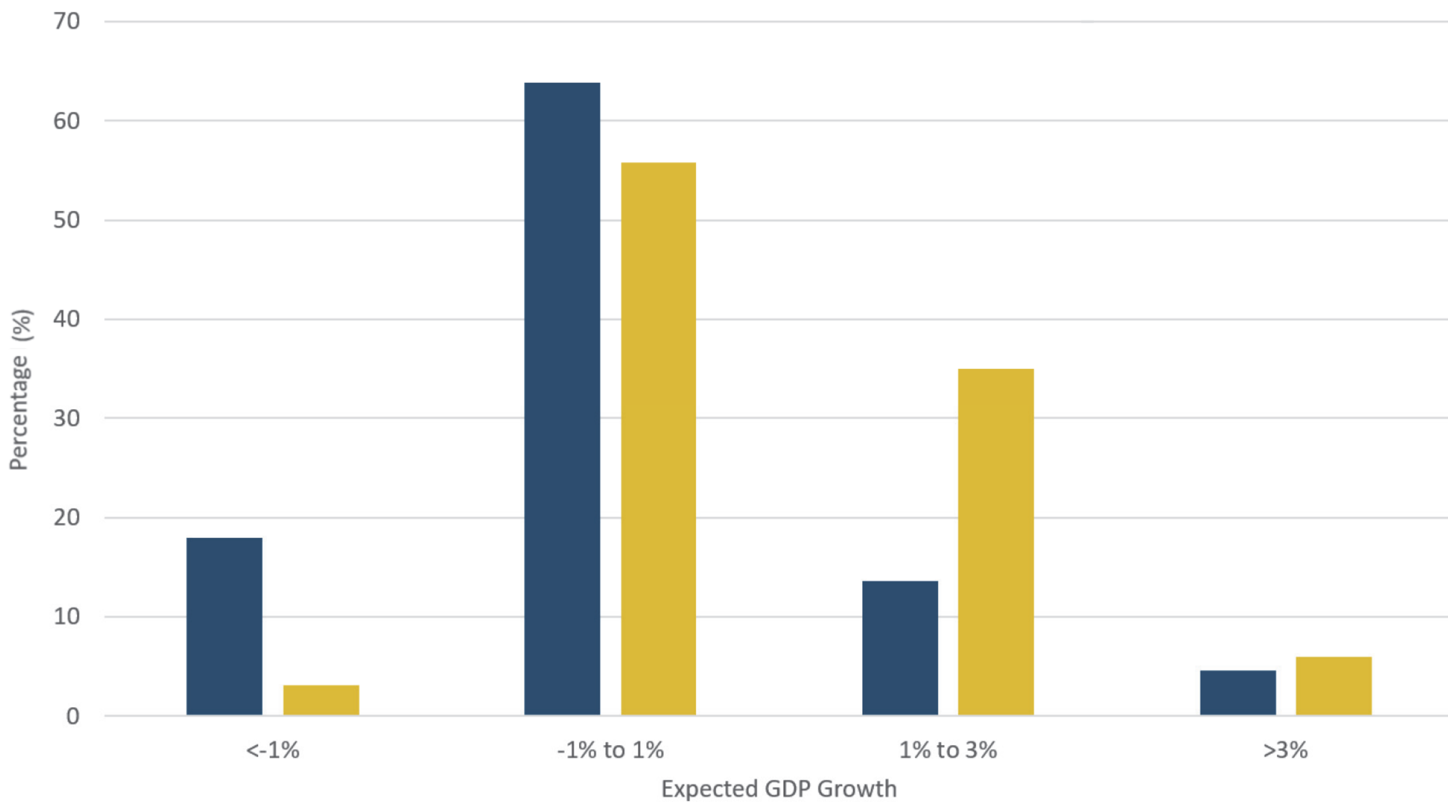
Supplement for Question (3), continues the next page >>>>>

Supplement for Question (3), continued:

To elaborate on the information about "Expected GDP growth 2017-18" in Table 1, the researchers include Figure 3 and discuss it.

EXCERPT, p. 10: We can compare firm estimates to those of professional forecasters in the Bank of England's Survey. Figure 3 gives the distributions showing that businesses were somewhat more pessimistic than professional forecasters.

Figure 3: Businesses Forecasts Compared to Professional Forecasters in the Bank of England Survey of External Forecasters



Notes: The dark colored (blue) bars are the MES responses of the businesses to the macro growth question. The responses of professional forecasters surveyed by the Bank of England are shown with light colored (yellow) bars.

Supplement for Question (3), continues the next page >>>>>

Supplement for Question (3), continued:

Table 3. "Drivers" of Management Scores

Dependent Variable: Management Score				
	Column (1)	Column (2)	Column (3)	Column (4)
Log employment	0.063*** (0.0014)	0.061*** (0.0017)	0.055*** (0.0018)	0.057*** (0.0018)
Family owned but not run			-0.009 (0.0065)	-0.004 (0.0065)
Family owned and run			-0.025*** (0.0050)	-0.015*** (0.0050)
Foreign owned			0.053*** (0.0054)	0.046*** (0.0054)
Log age				-0.016*** (0.0031)
Share of managers with a college degree				0.061*** (0.0079)
Share of non-managers with a college degree				0.058*** (0.0103)
Industry Dummies	No	Yes	Yes	Yes
Location Dummies	No	Yes	Yes	Yes
Other Controls	No	Yes	Yes	Yes
Observations	7,756	7,756	7,756	7,756
R^2	0.212	0.307	0.319	0.341

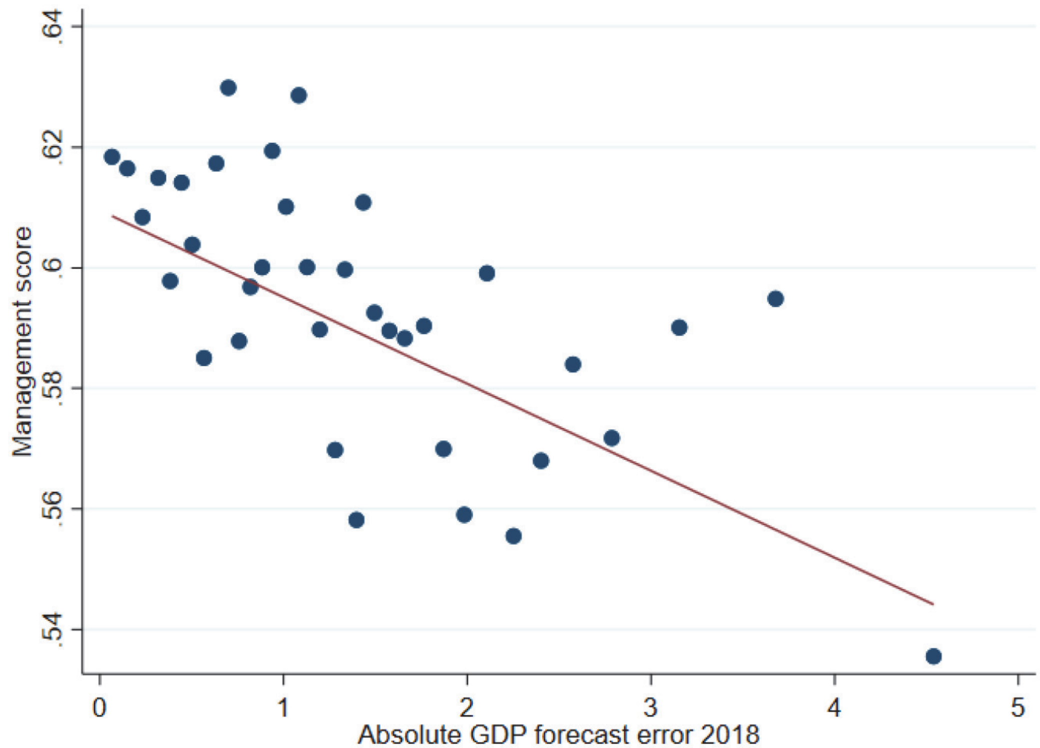
Notes: Estimation by OLS with robust standard errors in parentheses. Management score is the unweighted average of the score for each of the 12 questions, with scores on a scale of 0 to 1 for each question, where 0 was the least and 1 the most structured management practice. "Foreign Owned" is a dummy for whether the firm is an affiliate of a non-UK firm. "Family owned and run" is a firm owned by a family and run by a family member; "Family owned but not run" is a dummy for a firm which is family owned but whose CEO is a non-family member; a firm which is not owned by a family is the omitted base. Industry dummies are created using the first two digits of the four-digit Standard Industrial Classification (SIC) code for each firm. Location dummies are for nine regions in the UK. "Other Controls" includes dummies for the month when the survey was returned, time spent on the survey, multi-site dummy and reporting accuracy indicator. * p < 0.10, ** p < 0.05, *** p < 0.01.

Supplement for Question (3), continues the next page >>>>>

Supplement for Question (3), continued:

EXCERPT, p. 16: Figure 6 shows the relationship between GDP forecast errors and firm management. The horizontal axis has absolute GDP forecast error grouped into 40 equal-sized bins [which means each bin has the same number of firms]. The vertical axis shows the mean values of management scores in each bin. There is a clear negative relationship indicating that better managed firms make lower GDP forecast errors.

Figure 6: GDP Forecast Errors and Management



Notes: Shows the relationship between absolute GDP forecast errors and management score. Horizontal axis shows the value of forecast errors in absolute value. Based on the absolute GDP forecast error, observations are grouped into 40 equal-sized bins. The vertical axis is the mean value of the management score for firms in the same bin.

Supplement for Question (3), continues the next page >>>>>

Supplement for Question (3), continued:

EXCERPT, p. 16: In Table 5, we go beyond the bivariate correlation in Figure 6 and control for many other factors. Column (1) reports the result of regressing a measure of forecast errors on the management score, confirming the statistical significance of the relationship in Figure 6. Column (2) adds in industry dummies, location dummies and the standard other controls. Firms with more structured management practice still make significantly smaller forecast errors. In Column (1), the coefficient of -0.358 implies that an increase in management scores from the 10th to 90th percentile (0.509) is associated with a fall in the absolute value GDP growth forecast errors of 0.18 percentage points, or 13% of the mean of the dependent variable (1.411 as shown in the base of the column).

Note: Columns (3) and (4) are deliberately excluded in the excerpt of Table 5 below.

Table 5. GDP Forecast Errors and Management Score

Dependent Variable: Absolute GDP Forecast Error (percentage)				
	Column (1)	Column (2)	Column (5)	Column (6)
Management score	-0.358*** (0.0609)	-0.293*** (0.0673)		-0.171** (0.0738)
Log employment				-0.054*** (0.0115)
Family owned but not run			0.066* (0.0391)	0.047 (0.0395)
Family owned and run			0.073*** (0.0269)	0.031 (0.0295)
Foreign owned				0.035 (0.0365)
Log age				0.018 (0.0175)
Share of managers with a college degree				-0.030 (0.0474)
Share of non-managers with a college degree				0.087 (0.0641)
Industry Dummies	No	Yes	Yes	Yes
Location Dummies	No	Yes	Yes	Yes
Other Controls	No	Yes	Yes	Yes
Mean of dependent variable	1.411	1.411	1.411	1.411
Observations	7,134	7,134	7,134	7,134
R^2	0.005	0.055	0.054	0.060

Notes: Estimation by OLS with robust standard errors in parentheses. The dependent variable is the absolute value of the difference between expected (in MES 2016) and actual real GDP growth rate 2017-18. "Foreign Owned" is a dummy for whether the firm is an affiliate of a non-UK firm. "Family owned and run" is a firm owned by a family and run by a family member; "Family owned but not run" is a dummy for a firm which is family owned but whose CEO is a non-family member; a firm which is not owned by a family is the omitted base. Industry dummies are created using the first two digits of the four-digit Standard Industrial Classification (SIC) code for each firm. Location dummies are for nine regions in the UK. "Other Controls" includes dummies for the month when the survey was returned, time spent on the survey, multi-site dummy and reporting accuracy indicator. * p < 0.10, ** p < 0.05, *** p < 0.01.