Online Appendix for

Are Technology Improvements Contractionary?

Susanto Basu, John Fernald, and Miles Kimball

July 2005

This document contains the appendices from Basu, Fernald, and Kimball (2004, NBER Working Paper). The intention is to post this document on the web site for the *American Economic Review*. The sections are as follows:

- Section I discusses data and instruments, and is a fleshed-out version of the appendix attached to the revised paper.
- Section II presents new econometric results on the generated-regressor problem.
- Section III assesses classical measurement error.
- Section IV discusses small sample biases of instrumental variables.

I. Data and Instruments

We use industry data even though the theory probably applies most naturally to firms. Unfortunately, no firm-level data sets span the economy. Narrowing the focus to a subset of the economy—e.g., using the Longitudinal Research Database—would require sacrificing a macroeconomic perspective, as well as panel length and data quality.

Jorgenson dataset. We use updated data described in Jorgenson, Gollop, and Fraumeni (1987).[1] (Barbara Fraumeni, Mun Ho, and Kevin Stiroh were major contributors to various vintages of the data.)

We merged the main dataset, which runs from 1958 to 1996, with an earlier vintage of the datset that runs 1948-1989. We used growth rates from 1949 to 1958 from the older dataset and growth rates 1959 to 1996 from the newer dataset. Growth rates for the post-1959 overlap period generally line up closely, particularly in the early years, so there are not major inconsistencies between the two data series around the merge point. In addition, qualitative results are robust to using the two datasets separately.

We generally construct indices and aggregates as Tornquist indices, with log-changes weighted by average nominal shares in periods t and $t-1$. However, to construct industry input aggregates, we use factor shares averaged over the entire sample period. We use average factor shares because we are concerned that observed factor payments may not be allocative period-by-period, e.g., because of implicit contracts. This leads us to take an explicit first-order approximation to the industry production function. Results do not appear at all sensitive to this choice, however: Results appear virtually identical using time-varying shares.

We assume that industries earn zero economic profits, so that factor shares sum to one. In U.S. data, pure profits generally appear small (see, e.g., Rotemberg and Woodford 1995). In previous work with older versions of the Jorgenson dataset, we estimated payments to capital as in Hall (1990); estimated profits were generally small, and results were virtually indistinguishable from those that assumed zero profits.

Hours-per-worker. Where available, we used BLS data on hours/worker for production workers. Where necessary, particularly in early years of the sample, we used supplemental employment and hours data provided by Dale Jorgenson and Kevin Stiroh to construct a long time series for each industry. We then detrended hours-per-worker using Christiano-Fitzgerald's (2003) band pass filter, isolating frequency components between 2 and 8 years. By detrending, our utilization series has zero mean and no trend. We then took the first-difference in this detrended series as our measure of hours-per-worker growth dh. (Detrending log hours/worker with an HP filter or a simple first-difference filter makes little difference to results. In addition, using Jorgenson's hours/worker data yields very similar results, although the resulting technology series is a bit more volatile.)

National accounts data. All series were downloaded from the Bureau of Economic Analysis, via Haver

[1] Downloaded from http://post.economics.harvard.edu/faculty/ jorgenson/data/35klem.html (Oct 2002).

Analytics database, on April 7, 2004. (We also downloaded the Federal Reserve Board's broad real and nominal exchange rate indices from this source.)

Instruments

Monetary Shocks. We use quarterly VAR monetary innovations, following Christiano, Eichenbaum, and Evans (1999), Burnside (1996), and others. Following Burnside (1996), we measure monetary policy as innovations to the 3-month Treasury bill rate, since the fed funds market did not exist until the mid-1950s (from 1954:1 through 2003:1, the quarterly average 3-month T-bill rate has a correlation with the fed funds rate of over 0.99). More specifically, we measure monetary shocks as the innovations to the 3-month T-bill rate from a VAR with GDP, the GDP deflator, an index of commodity prices, the 3-month T-bill rate, and M1. (We thank Charles Evans for providing RATS code that estimated the VAR and innovations).[2]

We sum the quarterly series for the preceding year to obtain an annual series. In principle, we could use the four quarterly shocks separately as instruments, but the first-stage F-statistic falls sharply.

Government Spending. We use the average quarterly growth rate of real government defense spending from the preceding year, i.e., from the fourth quarter of *t-2* to the fourth quarter of *t-1*, as the instrument for annual input growth from year *t-1* to year *t*.[3]

Petroleum prices. Following Mork (1989), we base our oil instrument on the "composite" refiner acquisition price (RAP) for crude oil, a series produced by the Department of Energy. The composite price is refiners' average purchase price of crude oil, i.e., the appropriate weighted average of the domestic and foreign prices per barrel. Conceptually, the major difference between RAP and the PPI for crude petroleum arises from the Nixon price controls imposed in the second half of 1971; controls were not completely removed until the early 1980s and bind particularly sharply in early 1974.[4]

RAP is available monthly from January 1974 on. However, an annual average series is available from the late 1960s on. We follow Mork (1989) in linking the PPI and the annual composite RAP to create an estimated quarterly refiner acquisition price.[5] We assume that before 1974, the refiner price moves one-for-one with the PPI, since the annual growth rate in the composite refiner price moves quite closely with the annual growth in the crude petroleum PPI. In particular, domestic purchases accounted for about 80 percent of refiner purchases and price controls were a minor factor: In 1973, for example, the average RAP for domestic crude oil was $4.17 a barrel while the average RAP for imported oil was $4.08 a barrel. (In 1974, by contrast, the domestic price rose to $7.18/barrel, while the imported price rose to $12.52.)[6]

We normalize the estimated pre-1974 oil prices so that the average monthly price in 1973 matches the average price in the reported annual composite RAP. Since the PPI is an index, we make a levels-adjustment to get a monthly oil price for the pre-1974 period. The annual composite RAP in 1973 averaged $4.15, so we normalized our derived monthly series to have an average value of $4.15 in 1973. [7]

---

[2] GDP and the GDP deflator are from the BEA via Haver (downloaded May 22, 2003). The 3-month T-bill rate is the rate quoted on the secondary market, from Federal Reserve Board publication H.15 via Haver. M1 is from the Philadelphia Fed real-time dataset. We spliced data from 2003Q1 (from 1959 onwards) to data from 1973Q4 (which covers the pre-1959 period). Charles Evans provided us with the PCOM data used in Christiano, Eichenbaum, and Evans (1999). We extended his PCOM variable back one year, to 1947, by splicing his series with Conference Board data on raw materials spot prices *SMP* (Haver mnemonic U0M023, downloaded Aug 15, 2003). Following Evans, we filter *SMP* as follows: $PCOM(t) = 1.451 \cdot PCOM(t-1) - 0.586 \cdot PCOM(t-2) + 0.134 \cdot \Delta \ln(SMP(t))$.

[3] Downloaded from the Bureau of Economic Analysis via Haver Analytics December 12, 2002.

[4] Mark French and Rob Vigfusson independently pointed out to us the problems with the PPI.

[5] Vigfusson (2002) uses the IMF world spot price of oil. This series, which is not available for our full sample period, moves reasonably closely with the PPI until the first quarter of 1974—when the log change in the IMF world price is 1.37 while the log change in the PPI is 0.32. (IMF Data from International Financial Statistics July 2002 CD-ROM). These changes bracket the price change for U.S. purchasers of oil: In annual data, the 1974 log change in the refiner acquisition price is 0.78; the log-change in the IMF world price is 1.27, compared with 0.52 for the PPI. In sum, we view the composite refiner acquisition price as a better measure of the relative price shock that hit the U.S. economy.

[6] From Haver Analytics, we downloaded the PPI for crude petroleum (mnemonic P0561) and the Composite Refiner Acquisition Price for Crude Oil (PZRAC). We downloaded annual RAP data from the Department of Energy at http://www.eia.doe.gov/emeu/aer/txt/ptb0519.html. (All downloads were December 11, 2002)

[7] Results appear little affected by alternative ways of linking data over the price-control period 1971 to 1974. For

Hamilton (1996) recommends focusing on oil price *increases* above the peak level over the preceding 12 months. First, Hamilton and others find a nonlinearity: oil price increases are more contractionary than oil price declines are expansionary. Second, he argues that oil price increases have a larger effect if they follow stable prices than if they simply reverse an earlier decline. Thus, we measure the quarterly oil price 'shock' as the difference between the log of the quarterly real oil price and the maximum oil price in the preceding four quarters. (In all cases, we measure the quarterly oil price using the last month of the quarter.) For annual data, we take as our instrument the sum of the quarterly shocks in the preceding calendar year.

II. The Purified Technology Series as a Generated Regressor[8]

Our application focuses on hypothesis testing. Arguing that some variable s_{t+j} (say current or forwarded employment or investment) covaries negatively with the true technology series $\zeta_t(\Gamma)$, is equivalent to arguing that the true value of θ in the regression

$$s_{t+j} = \alpha + \theta\zeta_t(\Gamma) + \nu_t$$

is negative. The problem is that $\hat{\theta}$ is estimated from the OLS regression

$$s_{t+j} = \alpha + \theta\zeta_t(\hat{\Gamma}) + \nu_t,$$

where $\hat{\Gamma}$ is the estimated value of the "first-step" parameters. Testing the null hypothesis that θ is equal to zero would require no generated regressor correction for the asymptotic hypothesis test.

But arguing that the covariance is negative requires rejecting not only the hypothesis that θ=0, but also rejecting any positive value of θ. Because the test statistic is not monotonic in the true value of θ, rejecting any positive value of θ requires one additional condition beyond rejecting the hypothesis that θ=0. As shown below, the *additional* condition does not depend on any characteristics of s_{t+j}, and so can be interpreted as an overall "quality control" condition on the generated regressor $\zeta_t(\hat{\Gamma})$. As long as this overall "quality control" condition on the generated regressor is satisfied, the uncorrected test statistic is valid for asymptotic tests of the hypothesis that $\theta \geq 0$. The remainder of the appendix demonstrates this claim and spells out the "quality control" condition on the generated regressor.

Our estimation involves a two-step procedure. In the first step, after stacking the industries on top of each other, we can use an instrumental variable row vector q_i to estimate the parameter vector Γ in

$$dy_t = \xi_t \Psi \Gamma + \varepsilon_t,$$

where t is time, dy_t is a vector of changes in industry log gross output, ξ_t is a matrix whose "diagonal" elements are the row vectors $[1, \chi_{(t>73)}, dx_i, dh_i]$ for industry i, Ψ embodies the cross-equation restrictions, and ε_t is a vector of the demeaned industry-level technology changes, which is the first-step error term. For an appropriately defined r_t (see Jeffrey M. Wooldridge 2002, page 140), the estimator satisfies

$$\sqrt{T}(\hat{\Gamma} - \Gamma) = T^{-1/2}\sum_{t=1}^{T} r_t + o_p(1).$$

We assume that r_t is serially uncorrelated. (Since autocorrelation-robust standard errors in the first step are quite similar to uncorrected standard errors, and the estimated aggregate technology shocks themselves are serially uncorrelated, deviations from this assumption are unlikely to be substantial enough to seriously alter the bottom line below.) We can write the usual estimator of the variance-covariance matrix of $\hat{\Gamma}$ as

$$\hat{\Phi} = T^{-1}\sum_{t=1}^{T} \hat{r}_t \hat{r}_t'.$$

Denote the generated demeaned aggregate technology change by $\hat{\zeta}_t = \zeta_t(\hat{\Gamma})$. $\zeta_t(\hat{\Gamma})$ is a linear

---

example, we tried a further levels adjustment to exactly match the average price in our constructed series to the actual average 1972 composite price. In addition, we tried deflating by the GDP deflator and also using actual log-change in oil prices rather than using oil price increases only. All of these made little or no perceptible difference to results.

[8] We thank Jeff Wooldridge and Serena Ng for helping us with this appendix. All errors remain our own.

combination of the estimated first-step errors $\hat{\varepsilon}_i$, and so is a function of the data and the estimated value of Γ. By construction, $\zeta_t(\hat{\Gamma})$ has a mean of zero. As noted in the main text, we cannot reject the hypothesis that the (demeaned) aggregate technology change is white noise. Imposing the assumption that the true (demeaned) technology change $\zeta_t(\Gamma)$ is white noise is helpful in delineating the issues that arise because the estimated technology change $\zeta_t(\hat{\Gamma})$ is a generated regressor. If technology changes are white noise, then the covariance of a current technology change with a given variable conditional on other leads and lags of technology is the same as the unconditional covariance of the current technology change with that variable. The unconditional covariance can be consistently estimated by univariate OLS. The simplicity of univariate OLS greatly clarifies the generated regressor problem in this context.

As above, let s_{t+j} be any variable that is of interest because it might be affected by the technology change at time t. For example, s_{t+j} could be the current level of aggregate inputs, or a lead of the aggregate input level. In the second step we estimate

$$s_{t+j} = \alpha + \theta\zeta_t(\hat{\Gamma}) + v_t$$

by OLS. Because our focus is on hypothesis testing, we want to know the variance of the estimate $\hat{\theta}$ conditional on a range of values of the true θ. Following Wooldridge (2002, pp 139-141),

$$\sqrt{T}(\hat{\theta} - \theta) \sim Normal(0, V),$$

where $V = \text{plim}(\hat{A}\theta^2 - 2\hat{B}\theta + \hat{C}),$ with

$$\hat{A} = \hat{D}^{-2}T^{-1}\sum_{t=1}^{T}\{\hat{H}\hat{r}_t\}^2 = \hat{D}^{-1}\hat{H}\hat{\Phi}\hat{H}'\hat{D}^{-1},$$

$$\hat{B} = \hat{D}^{-2}T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t\hat{v}_t\hat{H}\hat{r}_t,$$

$$\hat{C} = \hat{D}^{-2}T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t^{\,2}\hat{v}_t^{\,2},$$

$$\hat{D} = T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t^{\,2}$$

and

$$\hat{H} = T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t[\nabla_\Gamma\zeta_t(\hat{\Gamma})],$$

where $\nabla_\Gamma\zeta_t(\hat{\Gamma})$ is the gradient of $\zeta_t(\Gamma)$ with respect to Γ, evaluated at $\hat{\Gamma}$ and expressed as a row vector. The Cauchy-Schwartz inequality implies that $\hat{A}\hat{C} - \hat{B}^2 \geq 0$. Note that $\hat{A}$ is independent of the particular variable being represented by s_{t+j}, since $\hat{v}_t$ does not appear in its formula. All the extra information one needs to know from the first-step in order to calculate $\hat{A}$ is $\hat{\Phi}$, the estimate of the variance-covariance matrix of the first-step parameter vector, together with the gradient vector $\nabla_\Gamma\zeta_t(\hat{\Gamma})$.

We are interested in showing that θ is negative. (For the most important cases, this is the natural direction. Cases in which we want to show that a covariance of technology shocks with a variable is positive can be handled by defining s_{t+j} as the negative of the variable of interest.) Showing that θ is negative can be formalized as a rejection of any hypothesis that has a nonnegative value for θ. That is, for all $\theta \geq 0$, if κ is the designated critical ratio, we need

$$f(\theta) = \frac{\sqrt{T}(\theta - \hat{\theta})}{\sqrt{\hat{A}\theta^2 - 2\hat{B}\theta + \hat{C}}} > \kappa.$$

If the test statistic $f(\theta)$ is monotonically increasing in θ, showing that $f(0) > \kappa$ is enough to guarantee that $f(\theta) > \kappa$ for all $\theta > 0$ as well. However, $f(\theta)$ is not, in general, monotonically increasing in θ. Instead, we demonstrate the following Lemma.

Lemma: *if* $\hat{\theta} < 0$, *then*

$$\min_{\theta \in \Re^+} f(\theta) \geq \min(f(0), f(+\infty)).$$

As a consequence, showing that f(0)>κ and that

$$f(+\infty) = \lim_{\theta \to +\infty} f(\theta) = \sqrt{\tfrac{T}{\hat{A}}} > \kappa$$

are together sufficient to guarantee that f(θ)> κ for all θ≥0.

Remarks: As noted above, $\hat{A}$ depends only on the details of the first-step estimation and not on the identity of s_{t+j}, so the condition $\sqrt{\tfrac{T}{\hat{A}}} > \kappa$ can be seen as an overall "quality control" condition for the generated regressor. (It is often true in applications that no generated regressor correction is needed for rejecting a zero value of a parameter. The complication here is that we need to reject *θ>0* as well, which also requires $\sqrt{\tfrac{T}{\hat{A}}} > \kappa$.)

For our measure of aggregate technology change, we calculated $\hat{A}$, which equals 0.0061. Thus, $\sqrt{\tfrac{T}{\hat{A}}}$ equals $\sqrt{48/0.0061}$ = 89 —far in excess of what's needed to pass this "quality control" condition at any reasonable level of significance.

Proof: The easiest way to demonstrate that $\min_{\theta \in \Re^+} f(\theta) \geq \min(f(0), f(+\infty))$ as promised is to show that *f(θ)* is either (a) monotonically increasing on $\Re^+$, (b) monotonically decreasing on $\Re^+$, or (c) first increasing, then decreasing on $\Re^+$. The derivative of the test statistic is

$$f'(\theta) = T^{1/2}\{\hat{A}\theta^2 - 2\hat{B}\theta + \hat{C}\}^{-3/2}[(\hat{A}\hat{\theta} - \hat{B})\theta + (\hat{C} - \hat{B}\hat{\theta})].$$

Thus, the sign of $f'(\theta)$ is the same as the sign of the linear function $(\hat{A}\hat{\theta} - \hat{B})\theta + (\hat{C} - \hat{B}\hat{\theta})$. If $\hat{C} - \hat{B}\hat{\theta} \geq 0$, then $f'(0) \geq 0$, and *f(θ)* must be either (a) monotonically increasing on $\Re^+$, or (c) first increasing, then decreasing on $\Re^+$, depending on the sign of $\hat{A}\hat{\theta} - \hat{B}$. If $\hat{C} - \hat{B}\hat{\theta} \leq 0$, then the Cauchy-Schwartz inequality $\hat{A}\hat{C} - \hat{B}^2 \geq 0$ implies that

$$\hat{B}^2 \leq \hat{A}\hat{C} \leq \hat{A}\hat{B}\hat{\theta}.$$

Since $\hat{\theta} < 0$, $\hat{B} < 0$ and dividing both sides by $\hat{B}$ indicates that

$$\hat{B} \geq \hat{A}\hat{\theta},$$

so that $\hat{A}\hat{\theta} - \hat{B} \leq 0$. Therefore, if $\hat{C} - \hat{B}\hat{\theta} \leq 0$, then $\hat{A}\hat{\theta} - \hat{B} \leq 0$ as well, and

$$f'(\theta) = T^{1/2}\{\hat{A}\theta^2 - 2\hat{B}\theta + \hat{C}\}^{-3/2}[(\hat{A}\hat{\theta} - \hat{B})\theta + (\hat{C} - \hat{B}\hat{\theta})] \leq 0$$

for all θ>0, implying in turn that *f(θ)* is (b) monotonically decreasing on $\Re^+$.

III. Classical Measurement Error in Inputs

Classical measurement error in inputs could, in principle, lead to counter-cyclical measurement error in our technology residuals. However, a simple model suggests that such measurement error cannot explain our results. First, for plausible parameterizations of the importance of measurement error, the "true" correlation remains negative. Second, the observed covariance between measured output and technology, which is zero or negative, bounds the covariance between true technology and true inputs, again suggesting a negative "true" correlation.

In our empirical work, we take the entire regression residual as "technology," implicitly assuming that our utilization proxies control fully for all variations in utilization. If they do not, but merely provide unbiased estimates of utilization, then the residual includes non-technological “noise” that is completely analogous to classical measurement error. Our model here abstracts from variations in utilization and does not explicitly consider aggregation across industries; neither changes the basic message.

Suppose the true economic model is given by

$$dy^* = \gamma\, dx^* + dz^* , \tag{A.1}$$

where the starred variables are unobserved, true values. Both output and inputs are measured with error:

$$dy = dy^* + \eta \tag{A.2}$$

$$dx = dx^* + \varepsilon , \tag{A.3}$$

where η and ε are iid, mean-zero variables with variances σ_η^2 and σ_ε^2, respectively. Note that the estimated variances of *dy* and *dx* always exceed their true values: $\sigma_{dx}^2 = \sigma_{dx^*}^2 + \sigma_\varepsilon^2$ and $\sigma_{dy}^2 = \sigma_{dy^*}^2 + \sigma_\eta^2$.

Now suppose we estimate (A.1) by instrumental variables. If the instruments are uncorrelated with the measurement error, then the estimate of γ is consistent. Hence, in the limit, the only source of error in our estimate of technology change is the measurement error in *dy* and *dx*:

$$dz = dz^* + \eta - \gamma\varepsilon . \tag{A.4}$$

Abstracting from estimation error in γ, equation (A.4) implies that $\sigma_{dz}^2 = \sigma_{dz*}^2 + \sigma_\eta^2 + \gamma^2\sigma_\varepsilon^2$. Note that for given observed variance of measured technology, as measurement error becomes larger, the variance of true technology shocks *dz**must fall. Using equation (A.4), the covariances of estimated technology change with output and input growth are:

$$\mathrm{cov}(dz, dy) = \mathrm{cov}\left(dz^*, dy^*\right) + \sigma_\eta^2 \tag{A.5}$$

$$\mathrm{cov}(dz, dx) = \mathrm{cov}\left(dz^*, dx^*\right) - \gamma\sigma_\varepsilon^2 . \tag{A.6}$$

Measurement error hence biases up both the estimated covariance between output and technology, and the estimated standard deviation of technology. If the true correlation between output growth and technology change is positive, then the estimated correlation may be biased either towards or away from zero, but cannot turn negative. However, suppose the true correlation between output growth and technology change is negative. Then the estimated correlation is unambiguously towards zero. Thus, our point estimates a negative correlation between output growth and technology change cannot be attributed to measurement error.

However, if the true covariance $\mathrm{cov}(dz^*, dx^*)$ is positive, then the estimated correlation is biased down. If the true input covariance is negative, then the estimated correlation might be biased up or down. To assess the the input-mismeasurement bias, we rewrite (A.6) in terms of correlations: Some algebra yields:

$$Corr(dz^*, dx^*) = \left[Corr(dz, dx) + \gamma \frac{\sigma_\varepsilon^2}{\sigma_{dz}\sigma_{dx}} \right] \left[\frac{\sigma_{dz}\sigma_{dx}}{\sigma_{dz*}\sigma_{dx*}} \right]$$

By specifying returns to scale and variances, we can calibrate this equation to observed correlations and variances. Suppose returns to scale are constant and that output is measured without error (output measurement error strengthens our case by reducing the variance of true technology), then the *maximum* σ_ε is 1.41 percent, given that this is the standard deviation of measured technology (since $\sigma_{dz*}^2 = \sigma_{dz}^2 - \sigma_\eta^2 - \gamma^2\sigma_\varepsilon^2 \geq 0$). In this case, there is no variation in true technology and the true correlation of inputs and technology is undefined. If instead we assume σ_ε is 1 percent—still a high number—then σ_{dz*} is also 1 percent. If we define true inputs as the sum of observed utilization plus measured utilization, then observed σ_{dx} is 3.3 percent per year; the “true” correlation between technology and inputs is –0.37, compared with the observed correlation with inputs of –0.50. Even if σ_ε is 1.35 percent, the true correlation remains at 0.15.

Finally, we are mostly interested in the signs of the correlations rather than their sizes. We can use the upward-biased output covariance to bound the input-covariance from above. Equation (4.1) implies that

$$\mathrm{cov}\left(dz^*, dy^*\right) \geq \mathrm{cov}\left(dz^*, dx^*\right), \tag{A.7}$$

(since the variance of dz^* is positive and $\gamma \geq 1$). But we see from equation (4.5) that

$$\mathrm{cov}(dz, dy) \geq \mathrm{cov}\left(dz^*, dy^*\right).$$

Our estimated covariance of output and technology appears to be either approximately zero or even negative. Thus, we conclude that the true covariance of technology and inputs must also be zero or smaller. Thus, our surprising results about the effects of technology improvements survive considerations of measurement error.

Since we cannot observe measurement error directly, we cannot say how much it affects our results. However, since the bias works against our finding that technology improvements reduce output, it seems likely that technology improvements are in fact contractionary. Furthermore, unlike the simple model used here, our technology change series takes a weighted average of technology shocks across sectors. If measurement error is relatively independent across industries, averaging should attenuate any biases.

IV. Small Sample Properties of Instrumental Variables

Could our results arise from a weak instruments problem? For example, the average F statistic from the first-stage regression of industry inputs *dx* on the instruments is 5.4—high enough to be statistically significant. But Staiger and Stock (1997) suggest that instrumental variables estimators sometimes have poor small sample properties when the first-stage F statistic is less than about 10.

Nevertheless, the small sample properties of instrumental variables do not appear to drive our results. First, Staiger and Stock note that LIML has better small sample properties than TSLS. LIML gives results that are qualitatively similar, though with much higher variance, than our preferred results. Second, when we throw out the industries for which the instruments are *particularly* bad (with first-stage F-statistics for dx_i of less than 2), the correlation of technology with hours remains significantly negative.

Third, and more substantively, we pooled industries within groups in order to raise the significance level of the first stage regression; we still find a robust negative correlation of technology and hours. To implement the pooled approach, we stacked industries within groups (durables, non-durables, and non-manufacturing) and then estimated a single regression for each group (using the regression equation from section I). We thus end up with a separate estimate of γ and β for each group. (In all cases, we removed industry fixed effects by demeaning all variables). The instruments generally appear highly relevant for the stacked regressions, with F statistics for *dx* that range from 15 to 40; the F statistic for *dh* ranges from 8 to 28.[9] After estimating the pooled regressions, we unstack the residuals into industry residuals, and aggregate as before. The resulting technology series has a correlation of 0.9 with our preferred technology series from Tables 2 and 3. The contemporaneous correlation between technology and hours is a statistically significant -0.39. (It is not surprising that the correlation is a bit less negative than before, given that we lose some of the "reallocation" effects that come from allowing for differences in γ 's.)

Finally, we simulated 1000 draws of random, irrelevant instruments and ran our system, deriving 1000 artificial technology series. We then assessed the actual small sample distribution of coefficients and t-statistic from an OLS regression of actual hours growth on estimated technology (contemporaneous only) under the null that the instruments are, in fact, irrelevant. As expected, coefficients are biased towards the OLS estimates—which yield a small positive coefficient, not the negative coefficient we find. In 123/1000 cases, the t-statistic at least as negative as -2; and in 54/1000 cases (5.4 percent), the t-statistic was as negative as (the OLS coefficient) from our main results of -3.6 (this differs slightly from Table 3, since it's a bivariate regression and it does not calculate Newey-West-corrected standard errors). These frequencies are considerably higher than one would expect from a normal distribution, but it nevertheless suggests it is very unlikely that random instruments explain our results.

Since the pooled specification largely replicates our overall results, we examined what the first-stage F statistics with generated instruments look like. With the pooled data, the median (for all three groups) of the first-stage F statistics with random instruments was 0.9 for *dx* and 0.8 for *dh.* Indeed, for *none* of the 1000 replications were *any* of the first-stage F statistics for any of the six variables—*dx* and *dh* for durable

---

[9] *dh* in non-manufacturing equal to 8.2 is lower than we would like. But pooled results are virtually unaffected by doing LIML rather than TSLS (since LIML is more robust to small sample issues, though often more variable); and indeed, results are robust to focusing on manufacturing alone.

manufacturing, non-durable manufacturing, and non-manufacturing—as large as we actually found in the actual data. Hence, it is exceedingly unlikely that weak instruments alone could explain both our large negative regression coefficient and our relatively high first stage F statistic in the pooled specification.

In addition, we looked more closely at the underlying cases where we found a large negative t-statistic. These generally appear to be cases where one or more of the point estimates of returns to scale are extremely large (e.g., 3 or more). In particular,

- Quantitatively, the negative correlation disproportionately represented the effect of a single industry,[10] in contrast to our results;
- The variance of the derived technology shocks tended to be much larger than for our actual purified technology series. (The median ratio of the variances was 2.8, and in only 1 of the cases was the variance smaller with the random instruments.).

In sum, although weak instruments are a concern, they cannot explain our results.

[10] Since $dz = \sum_i w_i (dz_i / (1 - s_{Mi}))$, the arithmetic contribution *CONT*$_i$ of each sector to the aggregate correlation (so that $Corr(dz, dx^V) = \sum Cont_i$) is $Cov(w_i dz_i / (1 - s_{Mi}), dx^V) / (stdev(dz) \bullet stdev(dx^V))$. In our reported results, 23 of the 29 industries contribute negatively, with the largest (negative) arithmetic contribution being -0.11 (construction). Of the 123 simulated cases with a negative t-statistic of -2 or larger in magnitude, only 1 simulation had a single contribution as small in magnitude as -0.11. For the cases with a t-statistic at least as large in magnitude as -3, *none* had a single-industry contribution as small in magnitude as -0.11.

References

Burnside, Craig (1996). "What do Production Function Regressions Tell Us about Increasing Returns to Scale and Externalities?" *Journal of Monetary Economics 37* (April): 177-201.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (1999). "Monetary Policy Shocks: What Have We Learned and to What End?" *Handbook of macroeconomics*. Volume 1A, pp. 65-148, Handbooks in Economics, vol. 15. New York: Elsevier Science, North-Holland,

Christiano, Lawrence J. and Terry J. Fitzgerald (2003). "The Band Pass Filter." *International Economic Review*, May 2003, v. 44, iss. 2, pp. 435-65.

Hall, Robert E. (1990). "Invariance Properties of Solow's Productivity Residual." In Peter Diamond (ed.) *Growth,Productivity, Unemployment* (Cambridge, MA: MIT Press).

Hamilton, James D. (1996). "This Is What Happened to the Oil Price-Macroeconomy Relationship." *Journal of Monetary Economics*, v. 38, iss. 2 (October), pp. 215-20

Jorgenson, Dale W.; Gollop, Frank and Fraumeni, Barbara (1987). *Productivity and U.S. Economic Growth.* Cambridge: Harvard University Press.

Mork, Knut Anton (1989). "Oil and the Macroeconomy When Prices Go Up and Down: An Extension of Hamilton's Results." *Journal of Political Economy,* v. 97, iss. 3 (June), pp. 740-44

Rotemberg, Julio J. and Michael Woodford. (1995). "Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets." In Thomas F. Cooley (ed.) *Frontiers of Business Cycle Research* (Princeton, NJ: Princeton University Press).

Staiger, Douglas and James H. Stock. (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, v. 65, iss. 3 (May), pp. 557-86

Vigfusson, Robert (2002). "Why Does Technology Fall after a Technology Shock?" Manuscript.

Wooldridge, Jeffrey M. (2002) *Econometric Analysis of Cross-Section and Panel Data*. The MIT Press, Cambridge Massachusetts.