Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 05-012

"Elasticity of Substitution Between Capital and Labor and its applications to growth and Development"

by

Samuel de Abreu Pessoa, Silvia Matos Pessoa, and Rafael Rob

Elasticity of Substitution Between Capital and Labor and its applications to growth and Development

Samuel de Abreu Pessoa* Silvia Matos Pessoa† Rafael Rob‡

March 4, 2005

Abstract

This paper estimates the elasticity of substitution of an aggregate production function. The estimating equation is derived from the steady state of a neoclassical growth model. The data comes from the PWT in which different countries face different relative prices of the investment good and exhibit different investment-output ratios. Then, taking advantage of this variation we estimate the long-run elasticity of substitution. Using various estimation techniques, we find that the elasticity of substitution is 0.7, which is lower than the elasticity, 1, that is traditionally used in macro-development exercises. We show that this lower elasticity reinforces the power of the neoclassical model to explain income differences across countries as coming from differential distortions.

JEL Classification numbers: D24, D33, E25, O11, O47, O49.

Key words: Demand for Investment, Dynamic Panel Data, Elasticity of Substitution.

*Graduate School of Economics (EPGE), Fundação Getulio Vargas, Praia de Botafogo 190, 1125, Rio de Janeiro, RJ, 22253-900, Brazil. Fax number: (+) 55-21-2553-8821. Email address: pessoa@fgv.br.

†Department of Economics, University of Pennsylvania, Email address: smatos@econ.ssc.upenn.edu.

‡Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA. Fax number: (+) 1-215-573-2057. Email address: rrob@econ.sas.upenn.edu. NSF support under grant number 01-36922 is greatfully acknowledged.

1 Introduction

This paper estimates the elasticity of substitution of an aggregate production function. To this end we use the Summers-Heston (2002) Penn World Table (PWT). The PWT contains international data on investment prices and investment-output ratios, which varies considerably over time and across countries. Then, using this variation we estimate the elasticity of the investment-output ratio with respect to the investment price. Assuming that the data is generated at the steady state of a neoclassical growth model and that the aggregate production function exhibits a constant elasticity of substitution, our estimate is also interpreted as the elasticity of substitution of this aggregate production function.

The main motivation for this exercise is that it helps assess whether differential distortions explain the huge per capita income differences that exist across countries of the world. A common approach to this question is to view different countries as having different distortions to the capital accumulation decision, reflected in different prices of investment goods. Then, prices of investment goods affect investment-output ratios (and thereby capital-worker ratios) and the latter affect, via the mechanics of the neoclassical growth model, per-capita incomes. Following up on this approach several papers (see Hall and Jones (1999)) find that this causality link is not *quantitatively* significant, i.e., that a lot of income variation remains unexplained after the role of investment prices is accounted for. Other papers (see Barro et al. (1995)) find that this causality link is significant, but only if one assumes a non-traditional and unusually high value for the capital share of income (2/3). Our estimation and calibration results offer a simple resolution to this dilemma. While previous papers assume a Cobb-Douglas production function for which $\sigma = 1$ (from this point onwards σ denotes the elasticity of substitution of the aggregate production function), our estimation results point towards $\sigma = 0.7$. Moreover, calibrating and simulating the model, we show, under $\sigma = 0.7$, that a sizeable fraction of per capita incomes is accounted for as coming from differential prices of the investment good. Thereby, our results highlight the role of the aggregate production function in explaining income gaps.

An important precursor to our work is the paper by Restuccia and Urittia (2001), where the hypothesis that the aggregate production function is Cobb-Douglas ($\sigma = 1$) is accepted. The Restuccia and Urittia (2001) estimation procedure is predicated, however, on all countries having the same total factor productivity (TFP). On the other hand, many researchers (including Klenow and Rodriguez-Clare (1997), Hall and Jones (1999) and Romer (2001)) argue that TFP varies considerably across countries, and is correlated with investment and with per-capita incomes. We take this possibility into account, allowing different countries to have different TFP's and other country specific effects, and allowing correlations to exist between TFP and per-capita incomes. Once these possibilities are accounted for, we estimate σ to be 0.7 and reject the hypothesis that it is 1. To corroborate this empirical finding we theoretically compute the bias that would occur if one were to ignore country specific effects. We find that the estimator of σ is then biased upwards, which explains why we obtain a lower estimate.

In somewhat greater detail we execute the following econometric exercises. The first exercise is to take annual panel data and derive a static estimate of σ, taking into account country specific effects. This yields an estimate of 0.5. We suspect that the true value of σ is higher than 0.5 because the time intervals between observations are short (annual), while the relationship we estimate is a long run relationship. In addition and related to this, the error terms in the annual panel data set are serially correlated. We address this problem by taking long run averages of the variables. We constructed two panels that average the variables from the annual panel data set over 6 and 7 years. Using this averaged data we obtain new estimates for σ, using the within group two stage least square procedure. The numbers we get are $\sigma = 0.650$ for the 6 year averages, and $\sigma = 0.674$ for the 7 year averages. We also find, for both panels, that serial correlation is not a problem once the data is averaged. A third finding is that a Wald test rejects the Cobb-Douglas hypothesis $\sigma = 1$ at the 3% and the 10% significance levels, respectively.

To confirm these results we do a third exercise using methods developed in Arellano and Bond (1991). We use the original, annual panel and apply dynamic panel estimation techniques to it. These techniques allow one to distinguish between the short and the long-run elasticities of substitution and to include all relevant variables. In addition, these techniques allow one to accommodate shocks to the regressors that are manifested in future periods. Using these techniques, we obtain 0.69 for the long run σ for both the within group and the extended GMM procedures, and we reject the Cobb-Douglas hypothesis $\sigma = 1$ at the 10% significance level.

All in all, our conclusion is that the evidence points towards a σ that is around 0.7. This conclusion is further supported by the work of Collins and Williams (1999). These authors consider a data set comprising of OECD countries over the period 1870-1950. Then, performing cross country regressions (which do not control for country specific effects), they obtain an estimate of $\sigma = 0.7$. We have done the analogue of the Collins and Williams exercise with our data set, i.e., we confined attention to OECD countries, and estimated σ to be close to 0.7 as well. This result agrees - naturally - with the results we get when we use a larger and, therefore, less homogenous set of countries, but when we control for country specific effects.

Traditionally the elasticity of substitution is estimated using industry (micro) data. Early examples include Arrow *et al.* (1961), using cross section data and Lucas (1969), using time series data. A recent study in the same tradition, employing static panel estimation techniques and using U.S. cross-industry data is Chirinko (2002). As reported in that study σ is somewhere between 0 and 1 and, most likely, between 0.5 and 1. The estimates we obtain here are well within this range, which is the expected result (given that industry studies are based on micro data and that our estimates are based on macro data).

As stated earlier, our interest in estimating σ stems from the fact that it determines the quantitative effect that investment distortions have on per-capita incomes. To make this point we calibrate parameters of the model - other than σ - to U.S. data and then simulate the model for several values of σ. These simulations show that the impact of distortions under $\sigma = 0.7$ is significantly stronger (in a sense to be made precise below) than under $\sigma = 1.0$. This improves the explanatory power of the neoclassical model to explain income gaps as coming from differential distortions, and suggests that policies that reduce distortions in poor (highly distorted) countries are more effective than they would appear under $\sigma = 1.0$. In addition, we perform a development decomposition exercise à la Hall and Jones (1999), and show that the correlation between per-capita income and TFP is smaller under $\sigma = 0.7$ than under $\sigma = 1.0$. Finally, as an application of our estimation results, we assess what portion of the distortions that our model formulation is based on is reflected in the PWT.

The calibration approach we employ in these exercises is not completely standard and, as such, may be of independent interest. Most notably, rather than "commit" to an aggregate production function in advance (usually the Cobb-Douglas), which is the usual procedure in calibration exercises, we use a production function that we estimate from empirical (and relevant to the problem at hand) data. This approach is necessitated by the fact that σ is a curvature parameter of a production function, so data on a single country does not offer enough variation to pin it down. As a consequence what we offer here is a hybrid methodology relying on estimation and calibration, which, as we suggest later, may prove fruitful in other contexts.

The rest of the paper is organized as follows. Section 2 presents a theoretical model and derives the equation that is to be estimated. Section 3 describes the econometric procedures used for estimating this equation. The numerical results of our estimation are then reported and discussed in Section 4. Section 5 calculates the bias in the regression that would occur if one were to ignore the country specific effects. Section 6 conducts quantitative exercises, showing how our estimation results are applied to the question of income gaps. Section 7 concludes.

2 Model Specification

2.1 Theoretical Model

Although the theoretical model we present now is routine in some respects, we include it because it contains features that specifically tailor it to the data we have, to the literature that our paper relates to, and to the set of applications we conduct later. Readers interested mainly in the estimation part of the paper can skip over to equation (17), which is the estimating equation, and then proceed to Sections 3 and 4.

We consider a two sector neoclassical growth model. Time is continuous and the horizon is infinite.

Sector 1 produces a consumption good, using labor and capital. The per-capita output y_1 of this sector is

$$y_1 = Al_1 f(k_1), \tag{1}$$

where l_1 is the fraction of the labor force employed in sector 1, k_1 is the capital-labor ratio in sector 1, A is total factor productivity, and f is the production function specified in (3).

Sector 2 produces an investment good, using labor and capital. The per-capita output y_2 in that sector is

$$y_2 = ABl_2 f(k_2), \tag{2}$$

where B is an investment sector productivity parameter. The function f is specified as

$$f(k_i) = \left(1 - \alpha + \alpha k_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

i.e., f exhibits a constant elasticity of substitution between capital and labor that we denote by σ. α is a parameter relating to income shares.

The economy is populated by a continuum of identical, infinitely-lived individuals that act as consumers, workers and owners of capital. The supply of labor of each individual is inelastic at 1 unit per unit time, and there is no disutility from working. The measure of individuals is 1. Individuals take prices as given and make intertemporal consumption/savings decisions, where savings are effected by buying capital goods and renting them out to firms.

There is a continuum of profit maximizing, price-taking firms that buy inputs (labor and capital services) from individuals and sell output back to individuals.

The lifetime utility of a representative individual is

$$\int_0^\infty e^{-\rho t} \frac{C_t^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} \mathrm{d}t, \tag{4}$$

where C_t is the date t consumption flow. ρ is the subjective, instantaneous rate of time preference and γ is the intertemporal elasticity of substitution.

Consider some fixed point in time, say t. Then at that point a representative individual receives the flow wage of w_t and the flow rental rate of q_t per unit of capital good that she rents out to firms. Let p_t be the price of the investment good. All prices are denominated in terms of the consumption good, which is the numeraire commodity. Then, a representative individual faces the following budget constraints

$$C_t + p_t I_t = q_t K_t + w_t + x_t, \tag{5}$$

where K_t is the individual's capital stock, I_t is the individual's addition to this capital stock, and x_t is a lump-sum transfer (specified below).

The capital accumulation equation is

$$\dot{K}_t = I_t - \delta K_t, \tag{6}$$

where δ is the physical depreciation rate. The individual's initial endowment of capital is exogenously specified and denoted by K_0.

Substituting (6) into (5) we get

$$p_t \dot{K}_t - (q_t - \delta p_t) K_t = w_t + x_t - C_t. \tag{7}$$

Let us introduce the interest rate

$$r_t \equiv \frac{q_t - \delta p_t}{p_t} = \frac{q_t}{p_t} - \delta. \tag{8}$$

Then, maximizing (4) subject to the budget constraints (7), one gets the Euler equation

$$\frac{\dot{C}_t}{C_t} = \gamma (r_t - \rho). \tag{9}$$

We are going to focus on a steady state, where $\dot{C}_t = 0$, and where investment is solely used to replace depreciating capital.

To this economy we add distortions that come either from government policy or from "institutional" considerations. The effect of these distortions is to drive a wedge between the equilibrium prices that would have prevailed in their absence and the prices that individuals and firms actually face. To determine the prices that individuals and firms face, we first find the prices that would have prevailed in the absence of distortions. Then, we tack on distortions to these prices.

As stated above, we focus on a steady state. Then, equilibrium prices are time invariant. Since any input combination that produces one unit of the consumption good produces B units of the investment good, the relative price of capital is

$$p = \frac{1}{B}.$$

Given this, the rental rate of capital is $q = Af'(k)$, the interest rate is $r = ABf'(k) - \delta$, and the wage rate is $w = A[f(k) - kf'(k)]$.

Next we consider distortions. The government imposes a tax on the investment good at the rate of τ_{I} or, alternatively, imposes a tariff on the importation of investment goods in case the economy is open.[1] In addition, the government imposes a tax on capital income at the rate of τ_{K}.[2] We assume that tax proceeds are returned to individuals in the form of lump sum transfers, and appear as x_t in the individual budget constraints, (5).

As a consequence of these distortions, individuals pay

$$p = \frac{T_{\mathrm{I}}}{B} \equiv \frac{1+\tau_I}{B}$$

for the investment good, and receive

$$q = \frac{Af'(k)}{T_{\mathrm{K}}} \equiv (1-\tau_{\mathrm{K}})Af'(k)$$

as net rental rate on capital.

[1] Considering an open economy requires slight notational modifications. However, the equation to be estimated in the end is the same.

[2] Instead of interpreting τ_i's as taxes, one may interpret them as "distortions" that stem from cultural, historical and sociological features of real life economies. For example, one may interpret τ_{K} as the fraction of earnings that organized crime extorts from owners of capital. Or τ_{K} maybe money that owners of capital must pay to corrupted government officials to be able to run their businesses (which, in effect, means that capital income is taxed).

Combining (8) and (9) we have that

$$r = \frac{q}{p} - \delta = \frac{Af'(k)}{T_{\mathrm{K}}}\frac{1}{p} - \delta = \rho,$$

which implies

$$f'(k) = p\frac{T_{\mathrm{K}}}{A}(\rho + \delta) = \frac{T_{\mathrm{I}}}{B}\frac{T_{\mathrm{K}}}{A}(\rho + \delta). \tag{11}$$

Now, since f exhibits a constant elasticity of substitution, (3) implies that

$$\frac{k}{f(k)} = \left(\frac{f'(k)}{\alpha}\right)^{-\sigma}. \tag{12}$$

Substituting (11) into (12), we get

$$\frac{k}{f(k)} = \left[\frac{T_{\mathrm{I}}T_{\mathrm{K}}}{BA}\frac{\rho + \delta}{\alpha}\right]^{-\sigma}. \tag{13}$$

Next we compute national income statistics at the steady state. The per capita GDP of the economy y is defined as

$$y = y_1 + \frac{y_2}{B}.$$

Using (1) and (2) and substituting the equilibrium condition for the labor market, $l_1 + l_2 = 1$, we see that $y = Af(k)$, where k is the stock of capital per-capita. Then, using the fact that the steady state investment, δk, is equal to sector 2's output, $ABl_2 f(k)$, the economy's resource constraint is

$$y = Af(k) = c + \mathrm{inv} = c + \frac{\delta k}{B}, \tag{14}$$

where 'inv' is the per capita flow of investment goods (we reserve the letter i for the investment-output *ratio*).

2.2 Taking the Model to Data

This completes the derivations of the theoretical relationships that hold for a single economy. Let's consider now a cross section of economies, indexed by j. Each economy is characterized by its own TFP parameter A_j, its own investment sector productivity parameter B_j, and its own distortions $T_{\mathrm{I},j}$ and $T_{\mathrm{K},j}$. To make consumption, investment and GDP comparable across countries we evaluate them in terms of international prices[3] and, without loss of

[3] The issue here is that the investment-consumption price ratios are not equal across countries. We adopt the procedure developed by Restuccia and Urittia (2001) to address this issue.

generality, we let the international price of investment be one.[4] Then, if i_j is the steady state investment-output ratio in country j, (14) tells us that

$$i_j \equiv \frac{\mathrm{inv}_j}{y_j} = \frac{\delta}{A_j}\frac{k_j}{f(k_j)}. \tag{15}$$

Substituting (13) into (15), the long run investment-output ratio is

$$i_j = \frac{\delta}{A_j}\left[p_j \frac{T_{\mathrm{K},j}}{A_j}\frac{\rho+\delta}{\alpha}\right]^{-\sigma}. \tag{16}$$

Taking logarithms on both sides of equation (16), we get a log-linear relationship between the relative price of capital and the investment-output ratio

$$\ln i_j = \ln FE_j - \sigma \ln p_j, \tag{17}$$

where

$$\ln FE_j \equiv \ln\left[\delta\left(\frac{\alpha}{\rho+\delta}\frac{1}{T_{\mathrm{K},j}}\right)^{\sigma}\right] - (1-\sigma)\ln A_j. \tag{18}$$

FE_j is referred to as the jth economy fixed effect.

In Sections 3 and 4 we estimate the long run relationship (17). As stated in the introduction, several studies indicate that total factor productivity is correlated with investment and output; it has been argued, for example, that high productivity economies happen to be less distorted, meaning that A is correlated with T_{I} or with B. Another possibility is that T_{K} is correlated with T_{I} or with B (for example, distortions to capital creation may be related to distortions to capital remuneration). If such correlations exist, then the fixed effect is correlated with prices and if this correlation is ignored, the estimation results are going to be biased. An important feature of our estimation procedure is that we account for these correlations.

Scope of the estimation results. Before we proceed to the estimation, we note that our results apply beyond the particular model we presented above. In particular:

(1) The model is adaptable to the case in which there is population growth at the rate n and disembodied, labor-augmenting technological progress at the rate g. In that case equation (18) is replaced by

$$\ln FE_j \equiv \ln\left[\delta_{\mathrm{EF}}\left(\frac{\alpha}{\rho+\delta+\frac{g}{\gamma}}\frac{1}{T_{\mathrm{K},j}}\right)^{\sigma}\right] - (1-\sigma)\ln A_j,$$

[4] We do this by re-defining the unit of measurement for the investment good.

where

$$\delta_{\mathrm{EF}} \equiv \delta + g + n.$$

Equation (17) remains intact. Then the estimation procedure is identical to the one we present here.

(2) The model is also adaptable to the case in which there are other distortions (apart from the investment and the capital income distortions). Most notably, one can adapt the model to the case where labor is supplied elastically and is taxed, and/or where consumption is taxed. The estimating equation is again similar, and details are found in Appendix D.

(3) Our estimation results are also applicable to an environment in which technological progress is embodied and firms periodically upgrade their capital stocks. As we show elsewhere (see Pessoa and Rob (2003)), the relationship between the price of capital p and the investment-output ratio is, to a large degree of approximation, the same as (18). Therefore, one is able to translate the estimate we obtain here for σ to an estimate of the parameters of an individual firm production function in the model with embodied technological progress.

(4) Finally our model and estimation results may be interpreted from a differential productivity rather than a distortion point of view. Jones (1994) advanced the hypothesis that income gaps among countries are due to distortions and used investment prices from the PWT to empirically assess this hypothesis. On the other hand, Parente and Prescott (2000) and more recently Hsieh and Klenow (2003) advance the alternative hypothesis that investment price differences are due to differential productivities of the investment good sector. Our formulation encompasses both views. T_{I} and T_{K} reflect distortions while B reflects differential productivity. Correspondingly, our estimation results and the quantitative exercises we perform can be interpreted from either point of view.

3 Empirical Implementation

Our ultimate goal is to estimate the long run price elasticity of investment, i.e., the parameter σ in equation (17). Assuming that all countries share a common value for σ, the investment demand is the same for all countries apart from a country specific effect (or an "intercept"), which comes either from differences in the TFP terms A_j or from the policy/institutional variable, $T_{\mathrm{K},j}$, or both. These country specific effects are subsumed in the fixed effect term (18). To account for these effects as well as to distinguish between short-run and long-run price elasticities, we employ dynamic panel data techniques.

Dynamic panel data techniques are advantageous in our context for three reasons. First, the regression analysis relies on data that exhibit greater variability as compared to pure time series or pure cross section data. Second, panel data techniques allow us to identify country specific effects, which would have been impossible if we were to use pure cross section techniques. Third, using dynamic panel data techniques, we are able to distinguish between the short and the long-run price elasticities of demand for investment.

For completeness and to verify the plausibility of our approach, we work with two econometric specifications. In the first **static** specification, the lagged dependent and the lagged independent variables are not included on the RHS of the regression equation. In the second **dynamic** specification these lagged variables *are* included. The next two subsections describe these specifications and the econometric exercises that we perform on them.

3.1 Static Panel

Based on the theory above, see equation (17), we consider the following static specification of the demand for investment

$$\begin{aligned} \ln i_{jt} &= \ln FE_j + \beta_0 \ln p_{jt} + \varepsilon_{jt}, \\ \varepsilon_{jt} &\backsim \text{iid}(0, \sigma_\varepsilon^2), \\ j &= 1, 2, ..., N, \\ t &= 1, 2, ..., T, \end{aligned} \tag{19}$$

where $\ln FE_j$ is an unobserved time invariant country specific effect, ε_{it} is an error term, subscript j is a country index and subscript t is a time index. N is the number of countries in our sample and T is the number of time periods. Depending on the exercise (see below), the time period is either one year or an average over either 6 or 7 years. β_0 is the same as σ in Section 2.

If we use the original, annual data set, four issues need to be addressed. First, we need to determine whether to use estimation techniques that consider the country specific effect as a fixed-effect (FE) or as a random-effect (RE). Second, we need to account for the possibility that error terms are heteroskesdastic, i.e., that they have different variances for different countries. Third, we need to account for the possibility that the explanatory variable $\ln p_{jt}$ is correlated with the error term ε_{jt} (the so called endogeneity issue). Fourth, we need to test and correct for the possibility that error terms are serially correlated. We describe now how each of these issues is dealt with.

- FE versus RE. The FE model is estimated by the Within Group estimator (WG). To do that we first average equation (19) over time to get the cross section equation

$$\ln \bar{i}_j = \ln FE_j + \beta_0 \ln \overline{p}_j + \overline{\varepsilon}_j, \tag{20}$$

where $\ln \bar{i}_j = \frac{1}{T}\sum_{t=1}^{T} \ln i_{jt}$, $\ln \overline{p}_j = \frac{1}{T}\sum_{t=1}^{T} \ln p_{jt}$, and $\overline{\varepsilon}_j = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{jt}$. Second we subtract equation (20) from (19) for each t, which gives a transformed equation. Third we run an OLS regression on the transformed equation. The RE model, on the other hand, is estimated by the GLS random effects estimator. This procedure is more involved so we refer the interested reader to Baltagi (1995), Chapter 2, where it is fully described.

Comparing the two procedures, the GLS random effect estimator is more efficient, but it yields consistent estimates only if the country specific effects are not correlated with the regressors. On the other hand, the FE estimator is consistent regardless of the correlation between the country specific effects and regressors. To find out which estimator is more appropriate we apply a Hausman test to assess how large is the difference between the estimated parameters according to these two procedures. If the difference is large, then we conclude that there is correlation between the country specific effects and the regressors, and we adopt the FE estimator.

- Heteroskedasticity. In order to deal with heteroskedasticity, we report the consistent standard error of the WG estimator. The advantage of this standard error, which has been derived in Arellano (1987), is that it is robust to heteroskedasticity.[5]

- Correlation between the explanatory variable and the error terms. We relax the commonly held assumption that $\ln p$ is strictly exogenous. Then, $\ln p$ may be correlated with ε for some leads and lags. To account for that possibility, we let the lagged value of the regressor be an instrument. Then, to assess whether $\ln p_{it-1}$ and ε_{jt} are correlated, we apply the Sargan test of over identifying restrictions.

- Serial correlation of error terms. Given that we use a static specification and that we work with annual data, it is possible that the error terms are serially correlated. Most notably, this may occur because relevant variables are omitted. To check for

[5] The formula for the consistent standard error of the WG estimator of $\widehat{\beta}_0$ is

$$\mathrm{var}(\widehat{\beta}_0) = (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}(\sum_{j=1}^{N} \widetilde{\mathbf{X}}_j'\widehat{\boldsymbol{\varepsilon}}_j\widehat{\boldsymbol{\varepsilon}}_j'\widetilde{\mathbf{X}}_j)(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1},$$

where $\widetilde{\mathbf{X}}_j = \ln \mathbf{p}_j - \ln \overline{\mathbf{p}}_j$, $\widehat{\boldsymbol{\varepsilon}}_j = (\ln \mathbf{i}_j - \ln \bar{\mathbf{i}}_j) - \widehat{\beta}_0(\ln \mathbf{p}_j - \ln \overline{\mathbf{p}}_j)$ and all bold face variables are $T \times 1$ vectors.

As Arellano (1987) shows, this standard error formula is valid under the presence of any heteroskedasticity or serial correlation in the error terms - as long as T is small relative to N.

that, we apply a first order serial correlation test. If this test indicates the presence of serial correlation, one can remedy this by explicitly allowing for serial correlation. One commonly used remedy is to assume that error terms are AR(1)

$$\varepsilon_{jt} = \rho\varepsilon_{jt-1} + \nu_{jt}, \qquad (21)$$
$$\nu_{jt} \backsim \text{iid}(0, \sigma_\nu^2).$$

Then under this AR(1) assumption the estimation procedure is carried out as follows. First, the AR(1) coefficient, ρ, is estimated using the residuals from WG estimation. After estimating ρ, the data is transformed and the AR(1) component is removed. Finally, the WG estimator is applied to the transformed data.

A potential problem with this remedy is that it is very limited. It assumes a particular form of serial correlation, namely AR(1), and it does not deal directly with the source of the serial correlation, namely the omission of relevant variables. We show these limitations below by comparing the regression equation that an AR(1) transformation produces with the regression equation that one gets with a more general formulation, i.e., a formulation in which no variables are omitted.

All this describes the issues that arise if one uses the original annual data set. One way to get around these issues is to create a new, low frequency data set. This is done by averaging the original data over (say) six or seven year disjoint time blocks. Then, one estimates the static equation (19), where each time period is one of these blocks. The downside of this estimation strategy is that it reduces the number of data points available as inputs into the regression analysis and, thus, reduces the efficiency of estimators.[6] The upside is that the estimates one gets *are* long run estimates, which is what we are interested in, and that they are unbiased. Chirinko *et al.* (2002) provide a detailed description of this 'averaging' estimation strategy. We pursue this strategy in our context and report estimation results for it.

An alternative estimation strategy is to keep using the original annual data set but enrich the econometric specification to cope with the above four issues. Recall that our goal is to estimate the long run price elasticity of investment demand. The problem we run into is that the data presents us with short run fluctuations of the price of investment goods and with consequent short run adjustments to them. If we use the static specification (19),

[6] Another commonly used approach in the macroeconometrics literature is to 'smooth' the data. That approach however distorts the available information and, as such, it has been widely criticized.

these short run adjustments introduce correlations between contemporaneous investment, the lagged values of investment, and the lagged values of prices. To cope directly with these correlations we introduce a dynamic formulation into which these lagged values are integrated, and then estimate the dynamic formulation. We describe this approach in the next subsection.

3.2 Dynamic Panel

The dynamic econometric specification is

$$\begin{aligned} \ln i_{jt} &= \ln FE_j^{\mathrm{D}} + \beta_1 \ln i_{jt-1} + \beta_2 \ln p_{jt} + \beta_3 \ln p_{jt-1} + \epsilon_{jt}, \\ \epsilon_{jt} &\backsim \mathrm{iid}(0, \sigma_\epsilon^2), \end{aligned} \tag{22}$$

where $\ln FE_j^{\mathrm{D}}$ are unobserved time invariant country specific effects, superscript 'D' stands for dynamic and ϵ_{jt} are the error terms.[7]

The parameter β_2 in equation (22) is interpreted as the *short run* price elasticity of investment demand. The corresponding long run price elasticity is derived from (22) by setting $\epsilon_{jt} = 0$, $\ln i_{jt} = \ln i_{jt-1}$ and solving the resulting relationship between $\ln i$ and $\ln p$. Then the long run price elasticity is[8]

$$\beta_{\mathrm{LR}} = \mathrm{LR}(\beta_1, \beta_2, \beta_3) \equiv \frac{\beta_2 + \beta_3}{1 - \beta_1}. \tag{23}$$

Econometrically (22) is estimated via OLS and WG techniques, as in section 3.1, and also

[7]The dynamic specification (22) is related to the static specification (19) with AR(1) error terms as follows. Substituting (21) into (19), the AR(1) regression equation is written as

$$\ln i_{jt} = (1-\rho) \ln FE_j + \rho \ln i_{jt-1} + \beta_0 \ln p_{jt} - \beta_0 \rho \ln p_{jt-1} + \nu_{jt},$$

where

$$\nu_{jt} \backsim \mathrm{iid}(0, \sigma_\nu^2).$$

Then, if we set $\beta_1 = \rho$, $\beta_2 = \beta_0$, $\beta_3 = -\beta_0\rho$ and $\ln FE_j^{\mathrm{D}} = (1-\rho)\ln FE_j$, this regression equation is the same as (22). In general, however, (22) contains three parameters whereas (19) with AR(1) error terms contains only two. Therefore, (19) with AR(1) is a (very) special case of (22).

[8]A limitation of the static specification (19) with AR(1) error terms is now revealed: It does not allow one to distinguish between the short and long run price elasticities of investment demand. By the equation in the foregoing footnote and by (23), both elasticities are equal to β_0. Indeed

$$\beta_{\mathrm{LR}}^{AR(1)} = \frac{\beta_2 + \beta_3}{1 - \beta_1} = \frac{\beta_0 - \beta_0 \rho}{1 - \rho} = \beta_0.$$

via the Generalized Method of Moments (GMM) technique. For a detailed description of GMM techniques see Chamberlain (1984), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998a).[9]

Applying GMM techniques to the problem at hand, we report estimation results following two approaches. In the first approach, which is based on Arellano and Bond (1991), country specific effects are eliminated by taking first differences of the regression equation.[10] Applying this to (22), we get

$$\begin{aligned} \ln i_{jt} - \ln i_{jt-1} = \beta_1(\ln i_{jt-1} - \ln i_{jt-2}) + \beta_2(\ln p_{jt} - \ln p_{jt-1}) \\ + \beta_3(\ln p_{jt-1} - \ln p_{jt-2}) + \epsilon_{jt} - \epsilon_{jt-1}. \end{aligned} \tag{24}$$

Assuming that the ϵ_{jt}'s are serially uncorrelated (i.e., that $E(\epsilon_{jt}\epsilon_{js}) = 0$ for $t \neq s$), $\ln i_{jt-s}$ are valid instruments in these first differenced equations if $s \geq 2$. Then using these instruments we get the following $T - 3$ moment restrictions

$$E(\ln i_{jt-s}(\epsilon_{jt} - \epsilon_{jt-1})) = 0 \quad \text{for } s \geq 2 \text{ and } t = 3, ..., T. \tag{25}$$

Assuming furthermore that $\ln p$ is weakly exogenous,[11] we get additional moment restrictions

$$E(\ln p_{jt-s}(\epsilon_{jt} - \epsilon_{jt-1})) = 0 \quad \text{for } s \geq 2 \text{ and } t = 3, ..., T. \tag{26}$$

Arellano and Bond (1991) developed a consistent estimator, which is referred to as GMM-DIF, for this first difference approach. This estimator works well when the instruments are highly correlated with the regressors. Blundell and Bond (1998a) show, via Monte-Carlo simulations, that if β_1 is close to 1 (and in our case it is), then the lagged values of variables are weak instruments for the corresponding differenced variables, causing the asymptotic and the small sample performance of the GMM-DIF estimator to be poor.[12] Blundell and Bond

[9] Details concerning how these GMM techniques are applied to the problem at hand are found in Appendix A.

[10] Subtracting the average as we do with the WG estimator of the static panel is not going to work here. This is because the transformed lagged dependent variable and the transformed error terms are correlated, and this correlation does not vanish as the number of data points increases to infinity. This is shown in Nickell (1981).

[11] The assumption of weak exogeneity of $\ln p_{jt}$ is that $E(\epsilon_{js} \ln p_{jt}) = 0$ for $s > t$.

[12] Although the autocorrelation of the $\ln i_{jt}$ series is sufficiently below 1 that we can reject the unit root hypothesis (see below), $\ln i_{jt}$ are still positively and highly correlated, i.e., β_1 is positive and 'large.' Because of that, the instrumental variables for $\ln i_{j,t-2}$, $\ln i_{j,t-3}$,..., $\ln i_{j,1}$ are weak instruments, i.e., they are not strongly correlated with the regressors, and this poses problems for applying the GMM-DIF estimator.

also show that the GMM-DIF estimator of β_1 exhibits a downward asymptotic bias and large standard errors in small samples.[13] Furthermore, recent empirical work (see Blundell and Bond (1998b), Loyaza, Schmidt-Hebbel and Serven (2000) and Bond, Hoeffer and Temple (2001)) shows that the estimate of β_1 under GMM-DIF is close to the estimate of β_1 under WG estimation, which, as we discuss later, is biased downwards. This empirical work also points out that GMM-DIF estimators are inefficient, i.e., the standard errors of the estimates are large.

To overcome these biases and imprecisions we pursue a second, 'system' approach, referred to as GMM-SYS (or extended GMM) estimation. This approach combines, in a system, regressions in differences with regressions in levels, as in Arellano and Bover (1995). The work of Blundell and Bond (1998a) shows - theoretically and via Monte Carlo simulations - that the level restrictions under GMM-SYS are informative in cases where the first differenced instruments are not (even if β_1 is large). In addition the empirical work mentioned above shows that standard errors under GMM-SYS are smaller than under GMM-DIF.

This GMM-SYS estimator works as follows. The instruments for the regression in differences are the lagged values of the corresponding level variables as before. Symmetrically, the instruments for the regression in levels are the lagged differences of the corresponding variables. These are suitable instruments under the additional condition that there is no correlation between the differences of the right hand side variables and the country specific effects, which is written as[14]

[13] Blundell and Bond (1998a) evaluate the performance of the GMM-DIF estimator via Monte-Carlo simulations. In particular, they consider the pure AR(1) case

$$y_{it} = \eta_i + \alpha y_{it-1} + v_{it}.$$

Then, they illustrate their results with a dynamic labor demand equation, which includes wage and capital stock as explanatory variables

$$n_{it} = \eta_i + \alpha n_{it-1} + \beta_0 w_{it} + \beta_1 w_{it-1} + \gamma_0 k_{it} + \gamma_1 k_{it-1} + v_{it}.$$

[14] This assumption doesn't require that there is no correlation between the levels of $\ln p_{jt}$ and $\ln FE_j^{\mathrm{D}}$. Instead, this assumption follows from the stationarity property

$$E(\ln i_{jt+m} \ln FE_j^{\mathrm{D}}) = E(\ln i_{jt+n} \ln FE_j^{\mathrm{D}}) \text{ for any } m \text{ and } n,$$
$$E(\ln p_{jt+m} \ln FE_j^{\mathrm{D}}) = E(\ln p_{jt+n} \ln FE_j^{\mathrm{D}}) \text{ for any } m \text{ and } n.$$

$$E((\ln i_{jt-1} - \ln i_{jt-2}) \ln FE_j^{\mathrm{D}}) = 0$$
$$E((\ln p_{jt-1} - \ln p_{jt-2}) \ln FE_j^{\mathrm{D}}) = 0.$$

Then, adding to this the standard condition that $E((\ln i_{jt-1} - \ln i_{jt-2})\epsilon_{jt}) = 0$ and $E((\ln p_{jt-1} - \ln p_{jt-2})\epsilon_{jt}) = 0$, we get the following additional moment restrictions[15]

$$E((\ln i_{jt-1} - \ln i_{jt-2})(\ln FE_j^{\mathrm{D}} + \epsilon_{jt})) = 0 \quad \text{for } t = 3, ..., T, \tag{27}$$
$$E((\ln p_{jt-1} - \ln p_{jt-2})(\ln FE_j^{\mathrm{D}} + \epsilon_{jt})) = 0 \quad \text{for } \ t = 3, ..., T. \tag{28}$$

Another advantage of the system GMM over the first-difference GMM estimator is that it allows us to study not only the time series relationship (between price and demand for investment) but also their cross section relationship.[16] In any event, we report estimation results for both GMM-DIF and GMM-SYS.

To assess the empirical results of GMM-DIF and GMM-SYS, we apply two specification tests proposed by Arellano and Bond (1991). The first specification test is the Sargan test of over identifying restrictions, which tests for the overall validity of the instruments. The second test examines the hypothesis that the ϵ_{jt}'s are not second order serially correlated.[17]

We also test the validity of the additional instruments in the level equations. The set of instruments used for the equations in GMM-DIF is a subset of that used in GMM-SYS, so a test of these extra instruments is naturally defined. We apply a "difference" Sargan test by comparing the Sargan statistic for the GMM-SYS estimator and the Sargan statistic for the corresponding GMM-DIF estimator.

Measurement Error. So far we assumed that variables are measured without errors. Measurement errors are not unlikely for our data set, so we now indicate how our procedures are extended to cope with them. Suppose that $\ln i_{jt}$ and $\ln p_{jt}$ are not directly observed and that, instead, we observe

[15] Arellano and Bover (1995) show that further lagged differences would result in redundant moment restrictions if all available moment restrictions in first differences are exploited.

[16] The GMM-DIF estimator eliminates the unobserved fixed effects, while regression in levels does not.

[17] By construction, it is likely that $E((\epsilon_{jt} - \epsilon_{jt-1})(\epsilon_{jt-1} - \epsilon_{jt-2})) \neq 0$. Therefore, even if the original error terms are not serially correlated, the differenced error terms *are*, which means that the hypothesis that they are not serially correlated would likely be rejected.

$$\ln \widetilde{i}_{jt} = \ln i_{jt} + m^i_{jt}, \qquad (29)$$
$$\ln \widetilde{p}_{jt} = \ln p_{jt} + m^p_{jt},$$

where m^i_{jt} and m^p_{jt} are measurement errors that are uncorrelated with all of $\ln i_{jt}$ and $\ln p_{jt}$, and that are uncorrelated over time. Then, if one substitutes $\ln i_{jt}$ and $\ln p_{jt}$ from equation (29) into equation (25), one gets

$$\widetilde{\epsilon}_{jt} - \widetilde{\epsilon}_{jt-1} = \epsilon_{jt} - \epsilon_{jt-1} + m^i_{jt} - m^i_{jt-1} - \beta_1(m^i_{jt-1} - m^i_{jt-2})$$
$$- \beta_2(m^p_{jt} - m^p_{jt-1}) - \beta_3(m^p_{jt-1} - m^p_{jt-2}).$$

By the condition that the measurement errors are uncorrelated over time,[18] we obtain the following moment restrictions

$$E(\ln \widetilde{i}_{jt-s}(\widetilde{\epsilon}_{jt} - \widetilde{\epsilon}_{jt-1})) = 0 \quad \text{for } s \geq 3 \text{ and } t = 4, ..., T,$$
$$E(\ln \widetilde{p}_{jt-s}(\widetilde{\epsilon}_{jt} - \widetilde{\epsilon}_{jt-1})) = 0 \quad \text{for } s \geq 3 \text{ and } t = 4, ..., T,$$
$$E((\ln \widetilde{i}_{jt-2} - \ln \widetilde{i}_{jt-3})(\ln FE^{\mathrm{D}}_j + \widetilde{\epsilon}_{jt})) = 0 \quad \text{for } t = 4, ..., T,$$
$$E((\ln \widetilde{p}_{jt-2} - \ln \widetilde{p}_{jt-3})(\ln FE^{\mathrm{D}}_j + \widetilde{\epsilon}_{jt})) = 0 \quad \text{for } t = 4, ..., T.$$

Once we have these moment restrictions we apply GMM estimation, following the same steps as before. Specification tests for the validity of the instruments are analogous too.

4 Data and Results

The data we use comes from the Penn World Table, PWT 6.0 (Heston *et al.* 2002). To balance the data, we extracted a sub-sample of 113 countries, observed over 37 years, from 1960 to 1996. Table 1 at the end of the paper lists all the countries in our sample. The relative price of investment that we use is the ratio of the 1996 international price level of investment, PWT variable **pi**, and the 1996 international price level of consumption, PWT variable **pc**. The investment-output ratio is the investment share of real GDP per capita

[18] Alternatively, we could assume that measurement errors follow a moving average process of order 1. In that case we would use instruments that are lagged one more period than what would be necessary if measurement errors were serially uncorrelated.

evaluated at 1996 international prices, PWT variable **ki**.

We also constructed two 'average' panel data sets, derived from the above raw data. In the first panel we averaged the data over six disjoint time blocks with six years in each block: 60-65, 66-71, 71-77, 78-83, 83-89, and 90-95. Each block t is considered one time period and we have six time periods altogether, $T = 6$. In the second panel we averaged the data over five disjoint blocks with seven years in each block: 60-66, 67-73, 74-80, 81-88 and 89-96. Then we have five time periods altogether, $T = 5$.

As a first step we checked whether the $\ln i_{jt}$ and the $\ln p_{jt}$ series are stationary. To do that, while accounting for possible trends, we ran the regressions

$$\begin{aligned} \ln i_{jt} &= \delta_0 + \delta_1 t + \rho_1 \ln i_{jt-1} + \nu_{jt} \\ \ln p_{jt} &= \delta_2 + \delta_3 t + \rho_2 \ln p_{jt-1} + \mu_{jt}, \end{aligned}$$

using the STATA module *xtdftest*.[19] Based on these regressions we test for stationarity, using the Fisher version of the Dickey-Fuller test under the assumption of no cross country correlation among the errors. We have chosen the Fisher test because, as shown in Madalla and Kim (1998), it is more robust than other tests to violations of the no correlation assumption. Applying this test we find that non-stationarity is rejected, i.e., we reject the hypotheses $\rho_1 = 1$ and $\rho_2 = 1$, the p value being 0.00. Therefore our series reflect stationary fluctuations around (perhaps) a deterministic trend.[20] This allows us to proceed with the statistical procedures below.

Having done that, we present estimation results for the price elasticity of investment demand. To shorten the language we discuss the absolute values of the price elasticities in the text, which are positive, even though the corresponding numbers reported in the tables are negative. The overall conclusion that emerges from our analysis is that the estimates of the long run price elasticity are, for the most part, between 0.5 and 1. They tend to equal 1 when country specific effects are ignored and this is true regardless of whether we use static or dynamic panel techniques, and whether we control for the endogeneity of the regressor (price) or not. At the other end of the spectrum, the estimates tend to be close to 0.5 when country specific effects *are* taken into account but when serial correlation or, more generally, dynamic linkages are ignored. When both dynamic linkages and country

[19] We thank Luca Nunziata for kindly providing us with this module.

[20] This result may seem to contradict some literature reporting that the investment series exhibits a unit root. Note, however, that our estimates are based on a panel data and not on time series of a single economy, which is what said literature is based on in. Furthermore, we consider not investments but the investment-output ratios.

specific effects are controlled for, then, depending on the particular procedure we use, the estimates fall somewhere between 0.5 and 1, and in the majority of cases are close to 0.7. The order of presentation of these estimates follows the order of presentation of the econometric specifications in Section 3.

4.1 Static Panel

4.1.1 Annual panel data

Table 2 reports estimation results for the static specification (19), using our raw annual data.[21] In column [1] we report the results of an OLS regression and in column [2] the results of a 2SLS regression. Both regressions do not control for country specific effects. The first regression ignores price endogeneity as well, while the second regression does not. As can be seen, the estimated price elasticity in columns [1] and [2] is around 1. Whether we control or do not control for price endogeneity, the Wald test does not reject the hypothesis that the price elasticity is 1. This result agrees with the results of Restuccia and Urittia (2001) who, likewise, do not control for country specific effects.

These results change dramatically when country specific effects are controlled for. This can be seen in columns [3]-[7], which report regression results when fixed effects are (potentially) different across countries. The reported estimates in these columns are all well below 1, and actually close to 0.5. In particular, the WG regression [3] yields price elasticity of 0.522 and, correcting for price endogeneity in column [4], we get a slightly higher estimate, 0.558. The Sargan tests of over identifying restrictions for the 2SLS regressions, columns [2] and [4], do not indicate a problem with the validity of instrumental variables.

In column [6] we check for first-order serial correlation, AR(1), of the error terms - continuing to control for country specific effects (i.e., running a WG regression). We find strong and positive serial correlation. The estimated AR(1) coefficient ρ is high, 0.725, and the Bhargava *et al.* (1982) Durbin Watson test rejects $\rho = 0$. The estimate of the price elasticity in this column, 0.385, appears excessively low. Recall however that when error terms are AR(1) correlated, the short and the long-run price elasticities are constrained to be equal (see footnote 9). Since the short-run elasticity is smaller than the long-run elasticity we interpret this estimate for β_0 as an average between the short and the long-run elasticities. A more satisfactory approach obviously is to explicitly distinguish between the short-run and the long-run elasticities in the econometric specification, which we do below.

[21] All results in Tables 2, 3, 4 and 5 are computed using Stata 7.0. The test of first-order serial correlation is taken from the DPD98 software for GAUSS developed by Arellano and Bond (1998).

Column [7] of Table 2 reports regression results when country specific effects are treated as random effects, i.e., when equation (19) is estimated via GLS with random effects. The estimate we get then, 0.566, is sufficiently different from the WG estimate we get under a fixed effect treatment, 0.522. Because of that the Hausman test rejects the hypothesis of no correlation between the fixed effects and the regressors. Consequently, we consider country specific effects as fixed effects from this point onwards.

The net result from all this is that working with the static specification and with annual data is inappropriate. Error terms are serially correlated, when we naively correct for them via AR(1) we get excessively low estimates of the price elasticity, and short run and long run elasticities are not distinguished. This suggest we should consider either transformed data or an alternative specification. We first present results for transformed (i.e., averaged) data. Then we present results for the dynamic specifications.

4.1.2 Average panel data

Table 3 presents the results for the 6 and 7 year average panels. The first thing to note here, see columns [1], [2], [5], and [6], is that, when the fixed effect is constrained to be equal across countries, the price elasticity is still around 1. Therefore averaging the data may (and as we shall see, does) remedy for serial correlation, but it is no panacea for ignoring country specific effects. The second thing to note is that the estimates in the remaining columns are larger than the corresponding estimates in Table 2, but are still significantly lower than 1. And the third thing to note is that accounting for price endogeneity here makes a bigger difference than in Table 2, i.e., it increases the estimates by a bigger margin. In the end, when we control both for country specific effects and for price endogeneity, we get 0.650 for the six year average (column [4]) and 0.674 for the seven year average (column [8]).

Another thing we did was to check whether the addition of a time variable makes a difference. To do that we re-ran the previous regressions with time dummies. The results are shown in Table 4. As this table shows, if we do not control for the fixed effect or for price endogeneity (columns [1] and [5]), the estimated elasticity is still 1 and the dummies are significant. On the other hand, when we control for price endogeneity but not for country specific effects, only one time dummy is significant (columns [2] and [6]). The WG estimates (columns [3] and [7]) without controlling for price endogeneity deliver values for β_0 very close to columns [3] and [7] of Table 3 and, likewise, columns [4] and [8] are similar in the two tables. Furthermore, the WG estimates that control for price endogeneity (columns [4] and [8]) indicate that price dummies are insignificant. Finally the Wald Test rejects $\beta_0 = 1$ in columns [3], [4], [7] and [8]. All in all, the addition of time dummies makes little difference, especially when controlling for cross country heterogeneity and price endogeneity.

In summary, if we had to pick one estimate to report from this averaged panel exercise it would be the one for the six year average (column [4]), 0.66, with a robust standard error of 0.16; the corresponding estimate for the 7 year average has a higher robust standard error so we consider it inferior. The upside of this estimation strategy is that we get estimates of the long run elasticity and that serial correlation tests come back negative. Moreover, time dummies are significant only when we do not control for price endogeneity. The downside is that all standard errors are higher when we work with the averaged data than with the annual data. In particular, the WG-2SLS robust standard errors are doubled, compare columns [4] in Tables 2 and 3. This comes from the fact that we have less data points to work with when the data is averaged. Also, this approach does not make a distinction between the short run and long run price elasticity of demand. The approach we turn to next makes this distinction.

4.2 Dynamic Panel

We implemented the dynamic panel specification, (22), employing OLS, WG and GMM estimators. Before we comment on the numerical results we obtained, we discuss what estimates we report, how we obtained these estimates and how one should go about interpreting them.

The first issue to be discussed is that the usual GMM procedure that uses all lagged values as instruments becomes computationally infeasible when T gets large. This is shown in full detail in Arellano and Bond (1998). Furthermore, Monte-Carlo experiments (see Judson and Owen (1996)) indicate that increasing the number of instruments used creates a trade off. On the one hand, it increases the efficiency but, on the other hand, it increases the bias of the estimated β_1.[22] To deal with this issue, we used a "restricted GMM" procedure in which the number of lagged values used as instruments was at most two.

The second issue is that we had to decide whether to report numbers from the one step or the two step GMM (we describe these procedures in Appendix A). The one step GMM is predicated on the error terms ϵ_{jt} being independent and homoskedastic - both cross sectionally and over time. But then standard errors and test statistics are not robust to heteroskedasticity. The two step GMM remedies this problem by constructing a consistent estimate of the variance-covariance matrix of the moment conditions (based on first step residuals) and then re-running the estimator.[23] The problem with the two step GMM estimator however is

[22] An empirical cross country study that lends further support to this result is Loyaza, Schmidt-Hebbel and Serven (2000).

[23] If the error terms are spherical (homoskedastic), the one step and the two step GMM estimators are

that the standard errors it produces are biased downward in small samples.[24] This problem is pointed out in Blundell and Bond (1998a). The same authors also show - via Monte-Carlo simulations - that the precision of the one step GMM is not much lower than the precision of the two step GMM. Following up on these findings, we report the following estimates. For the point estimates of β's we report the estimates from one step GMM; for standard errors we report the estimates from one step GMM - corrected by the variance-covariance matrix computed from the first step residuals; and for specification tests and checking for second-order serial correlation we report the estimates from two step GMM. This last choice is guided by the fact that the Sargan test, based on the two step GMM, is the only one that is heteroskedasticity consistent. Also, the asymptotic power of the second-order serial correlation test increases in the efficiency of the GMM estimator,[25] and the two step GMM is more efficient.

A third issue is whether to include lagged price $\ln p_{jt-1}$ on the right hand side of the regression equation. As far as the generality of econometric procedure, $\ln p_{jt-1}$ should be included.[26] As far as economic theory, $\ln p_{jt-1}$ should be excluded. This is because a price shock in period $t-1$ affects investment in period $t-1$, $\ln i_{jt-1}$, and $\ln i_{jt-1}$ affects $\ln i_{jt}$. Once this chain of effects is accounted for, there is no further, independent effect of $\ln p_{jt-1}$ on $\ln i_{jt}$. Nonetheless, and for completeness sake, we report estimation results both when $\ln p_{jt-1}$ is included and excluded.

Let us now discuss now how to interpret the estimates, i.e., which of the various estimates we report (OLS, WG, GMM) in Table 5 is more reasonable. As Nickell (1981) and, more recently, Blundell and Bond (1998a) show, the transformation underlying WG estimation (see Section 3) biases the estimated coefficient β_1 downwards.[27] Furthermore, well-known results - in simpler settings - show that, when variables are omitted, the estimate of β_1 is biased upwards under OLS regression; Appendix C extends these results to our setting. As far as GMM estimation, it is known that if T is small relative to N, then GMM estimators are consistent, whereas WG estimators are not. In our case however T is not so small relative to N ($T = 37$, $N = 113$), and theoretical results comparing GMM and WG in this case are

asymptotically equivalent for GMM-DIF. Otherwise, the two step GMM is more efficient.

[24]Windmeijer (2000) created a procedure to correct the standard errors of the two step GMM estimator and embedded it into the DPD98 program for Gauss. He has kindly provided us with this procedure. We applied it to our problem and the standard errors we got were similar to those we got by correcting for heteroskedasticity.

[25]See Arellano and Bond (1991).

[26]In order to pin down the correct specification one should start with a broad specification then let the statistical results dictate which variable(s) to keep.

[27]They show this however for the "pure" AR(1) case without exogenous regressors.

just starting to emerge. The first such result is found in Alvarez and Arellano (2002). They consider the case where T/N tends to a positive constant and show that WG and GMM estimators exhibit negative asymptotic biases.[28] However, they also report several Monte-Carlo simulations where $T \leq N$, and where the bias of the GMM estimator is always smaller than the bias of the WG estimator. Therefore even if N and T are of (approximately) the same order of magnitude, it seems that GMM estimation is less biased.

Now we are ready to discuss the numerical results for the dynamic panel, as shown in Table 5.[29] Odd numbered columns report estimates when $\ln p_{jt-1}$ is included on the RHS of the regression equation, and even numbered columns report estimates when $\ln p_{jt-1}$ is excluded. As can be seen, the coefficient of $\ln p_{jt-1}$ is not significant in columns [5] and [7]. The first four columns of Table 5 report OLS and WG estimates of the parameters β_1, β_2 and β_3 together with estimates of the robust standard errors. As discussed above, the OLS estimates of β_1 are biased upwards while the WG estimates are biased downwards.[30] Computing the long run price elasticity β_{LR} from OLS estimation, we find that we cannot reject the hypothesis that it equals 1.

Columns [5] to [8] report the results of GMM estimation. In all GMM regressions we take the conservative approach of allowing for measurement errors that are uncorrelated across time. The validity of the lagged level variables $t-3$ and $t-4$ as instruments in the GMM-DIF equation [5] is not rejected by the Sargan tests. Likewise the $t-3$ lagged level variables combined with $t-2$ lagged first differenced variables as instruments in GMM-SYS in [7] is not rejected by the Sargan tests. Similar statements apply to regressions [6] and [8] where $\ln p_{jt-1}$ is not included as an explanatory variable.[31] We have tested for second order serial correlation and rejected that possibility.

As stated earlier, the WG estimates of β_1 are known to be biased downwards. Columns [5] and [6] show that GMM-DIF estimates are smaller yet. So this suggests that the instruments used in the GMM-DIF estimator are indeed weak.

Interpreting the overall message of Table 5, we would say that estimates under GMM-

[28] However, Alvarez and Arellano (2002) show this result for a first-order autoregressive model AR(1) without explanatory variables, with homoskedasticity and only the one step GMM estimator is considered.

[29] All results in Tables 5, 6, and 7 are computed using the DPD98 software for GAUSS. See Arellano and Bond (1998).

[30] This is because the OLS estimator ignores not only the unobserved country specific effects but also the endogeneity of the explanatory variables. WG estimator deals with the first problem, but still ignores the second one.

[31] In this case, we use the lagged level $t-2$ as instruments in the first-differenced equation (24). Also we use $t-2$ as instruments in the first-differenced equations, combined with lagged first-differenced variables dated $t-1$ as instruments in the level equations in (22) for $\ln p$.

SYS, columns [6] and [8], seem the most reasonable. The estimated coefficients of $\ln i_{jt-1}$ are higher than the WG estimates, which are known to be biased downwards, and lower than the OLS estimates, which are known to be biased upwards. Furthermore, the estimated coefficient of $\ln i_{jt-1}$ under GMM-DIF is lower than under WG, so the GMM-DIF procedure seems to go in the wrong direction. If we compare standard errors, there is a gain in precision from exploiting the additional moment restrictions. And, finally, the difference Sargan statistic that tests the additional moment restrictions confirms their validity. Comparing columns [7] and [8] suggests that $\ln p_{jt-1}$ can be omitted. Therefore, if we consider regression [8] as the most reasonable, the coefficient on the lagged dependent variable is 0.744, the short run price elasticity is 0.177, and the two together imply a long run price elasticity of investment demand of 0.691 (0.174). We tested the hypothesis that the long run price elasticity is 1, and rejected it at the 10% significance level.[32]

Although the estimates reported under [8] seem the most reasonable, it is worth noting that the point estimate for β_{LR} from WG estimation, regression [4], is very close to the point estimate from the GMM-SYS estimation, column [8]. Although the WG estimation results are biased, Nickell (1981) shows that this bias is of order $\frac{1}{T}$. Therefore, since T is fairly large in our data set, this bias is quantitatively small. Note also that WG estimation rejects $\beta_{\text{LR}} = 1$ at the lower, 5%, significance level.

In Table 6 we report OLS and WG estimates for the dynamic specification with time dummies added to the RHS of the regression equation.[33] We obtained very similar results to those in Table 5 (columns [2] and [4] respectively). In particular, the WG estimation indicate long run price elasticity of investment demand of 0.707 (0.093).

For completeness we tried a more general lag structure of the dynamic specification, which includes a second lag of the price and investment variables. The last two columns of Table 7 report GMM-SYS estimations of this generalized equation. It turns out that both lagged variables $\ln i_{jt-2}$ and $\ln p_{jt-2}$ are not significant.

What we can say as an overall summary from this analysis is that putting lagged investment on the RHS of the regression equation shows a positive and significant coefficient β_1

[32] The standard error of β_{LR} is obtained by using the Delta method. See appendix B.

[33] As before, it was infeasible to apply GMM estimators when time dummies are included. This is because the total number of instruments would then be excessively large relative to the cross section dimension. This implies that the two step GMM estimator, cannot be computed because the matrix $\mathbf{W}_2 = (\frac{1}{N}\sum_{j=1}^{N} \mathbf{Z}_j^{D\prime}\widehat{\boldsymbol{\epsilon}}_j^{*}\widehat{\boldsymbol{\epsilon}}_j^{*\prime}\mathbf{Z}_j^{D})^{-1}$ is not invertible. See appendix A and, for a full treatment of invertibility issues, Arellano and Bond (1998).

and eliminates the need to add an arbitrary serially dependent error term.[34] In addition it allows us to distinguish between the short-run and the long-run price elasticities and, as it turns out, this distinction is quantitatively significant; the long-run price elasticity of investment demand is more than three times bigger than the short-run elasticity.[35] And finally if country specific effects are not controlled for, we continue to get a long-run estimate of 1 even with dynamic panel data techniques.

5 The Bias of OLS Estimation

A repeatedly appearing result in Section 4 is that, when country-specific effects are ignored, the Cobb-Douglas hypothesis $\sigma = 1$ is accepted. In this section we investigate what gives rise to this result. We do this by calculating the bias that comes from not considering country specific effects, and adding this bias to the estimated value of σ when these effects *are* considered. As it turns out, the sum of the two is indeed 1.

To begin with, let's define

$$\begin{aligned} \mathbf{i}' &\equiv \left(\mathbf{i}'_1, ..., \mathbf{i}'_j, ..., \mathbf{i}'_N\right) \text{ where } \mathbf{i}'_j \equiv (i_{j1}, ..., i_{jt}, ..., i_{jT}) \quad \text{and} \\ \mathbf{p}' &\equiv \left(\mathbf{p}'_1, ..., \mathbf{p}'_j, ..., \mathbf{p}'_N\right) \text{ where } \mathbf{p}'_j \equiv (p_{j1}, ..., p_{jt}, ..., p_{jT}) \,. \end{aligned}$$

The variance-covariance matrix of the PWT data is

$$\mathbf{M} = \begin{bmatrix} \text{var}(\ln \mathbf{i}) & \text{cov}(\ln \mathbf{i}, \ln \mathbf{p}) \\ \text{cov}(\ln \mathbf{i}, \ln \mathbf{p}) & \text{var}(\ln \mathbf{p}) \end{bmatrix} = \begin{bmatrix} 0.605 & -0.307 \\ -0.307 & 0.306 \end{bmatrix}.$$

And the OLS estimate of the static panel satisfies

$$\widehat{\beta}_0^{\text{OLS}} = -1.00 = \frac{\text{cov}(\ln \mathbf{i}, \ln \mathbf{p})}{\text{var}(\ln \mathbf{p})} = \frac{\text{cov}((\ln \mathbf{FE} + \beta_0 \ln \mathbf{p}), \ln \mathbf{p})}{\text{var}(\ln \mathbf{p})} = \frac{\text{cov}(\ln \mathbf{FE}, \ln \mathbf{p})}{\text{var}(\ln \mathbf{p})} + \beta_0.$$

This implies that OLS estimation will bias upwards the estimated value of β_0 whenever $\text{cov}(\ln \mathbf{FE}, \ln \mathbf{p}) < 0$, which, as the next paragraph shows, is the case.

An analogous - although more involved - proof applies to the dynamic panel. In Appendix

[34] Note that the estimated value of β_1, 0.744, is quite close to the estimated ρ that we obtained with the static AR(1) specifications, 0.725.

[35] We also conducted a wide array of sensitivity analyses to verify the robustness of our results. First, we consider two alternative sub samples, broken up according to 'early' and 'late' periods. The first sub sample has observations from 1960 to 1978 and the second from 1979 to 1996. Moreover, we conducted the estimations with and without Sub Saharan countries. Overall, the GMM-SYS estimates are pretty robust across these alternative data sets and the long run price elasticity are between 0.72 and 0.78.

C, using the fact that $\text{var}(\ln \mathbf{p}) \approx -\text{cov}(\ln \mathbf{i}, \ln \mathbf{p}) \approx \frac{1}{2}\text{var}(\ln \mathbf{i})$, that $\widehat{\beta}_3 = \widehat{\beta}_{3,\text{Bias}} = 0$, and assuming that all economies are on a balanced growth path in the first period, we show that

$$\frac{\partial \widehat{\beta}_{\text{LR}}^{\text{OLS}}}{\partial \text{cov}\left(\ln \widehat{\mathbf{FE}}^{\text{D}}, \ln \mathbf{p}\right)} > 0, \text{ where } \widehat{\beta}_{\text{LR}}^{\text{OLS}} = \frac{\widehat{\beta}_2 + \widehat{\beta}_{2,\text{Bias}}}{1 - \left(\widehat{\beta}_1 + \widehat{\beta}_{1,\text{Bias}}\right)}. \tag{30}$$

Thus OLS estimation biases upwards the estimated value of β_{LR} for the dynamic panel as well. Furthermore, using the estimated values of $\text{var}\left(\ln \mathbf{FE}^{\text{D}}\right)$ and $\text{cov}\left(\ln \mathbf{FE}^{\text{D}}, \ln \mathbf{p}\right)$, we calculate $\widehat{\beta}_{\text{LR}}^{\text{OLS}}$ directly, obtaining 1.04.[36] This helps explain why ignoring country specific effects biases the estimate of β upwards and leads to the erroneous conclusion that the aggregate production function is Cobb-Douglas.

To further substantiate this result and relate it to previous literature, we have done the following exercise. We restricted our time averaged data set to the more or less homogeneous set of 15 OECD economies. Table 8 displays estimation results for this sub-panel when country specific effects are ignored. As shown in that table, the price elasticity estimates we get for β_0 are between 0.54 (for 6 year averaging) and 0.76 (for 7 year averaging).[37] These results are what we had expected. When attention is restricted to a small set of similar countries, country specific effects are approximately the same. Then the estimates we get should be close to the ones we get when we consider a large set of dissimilar countries, but when country specific effects are controlled for. This result is also in conformity with results reported by Collins and Williams (1999), using a similar approach, i.e., restricting attention to OECD economies.

To illustrate what country specific effects add to the statistical quality of results, we present the scatter plots of

$$\left(\Lambda \ln i \equiv \frac{\ln i_{jt} - \widehat{\beta}_1 \ln i_{jt-1} - \ln \widehat{FE^{\text{D}}}_j}{1 - \widehat{\beta}_1}, \ln p_j\right)_{t=1,\ldots,36\ N=1,\ldots,113}$$

Figure 1 shows this scatter plot for OLS estimation and Figure 2 shows it for GMM-SYS estimation. These Figures show that the scatter plot is tighter around the regression line for

[36] This estimate is obtained by computing $\text{var}(\ln \widehat{FE}^{\text{D}})$ $\text{cov}\left(\ln \widehat{\mathbf{FE}}^{\text{D}}, \ln \mathbf{p}\right)$ (which fall out of the estimation) and from them get $\widehat{\beta}_{\text{LR}}^{\text{OLS}}$ directly. Note that this direct estimate is not far off the estimate we report in Table 5, column 2.

[37] The estimates we get for β_0 are, not surprisingly, of poor statistical quality. This is indicated by the high robust standard errors. The reason for this is that we lose a lot of observations because the data is both time-averaged *and* because we delete many economies.

GMM-SYS, giving us a better fit of the data when country specific effects are included.

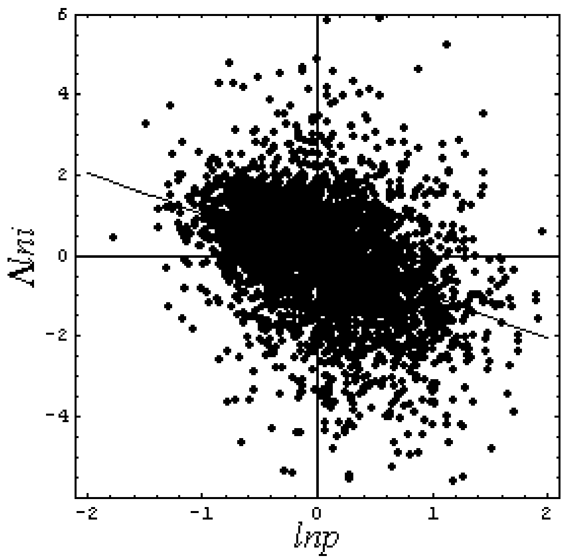

Figure 1: Scatter plot under OLS

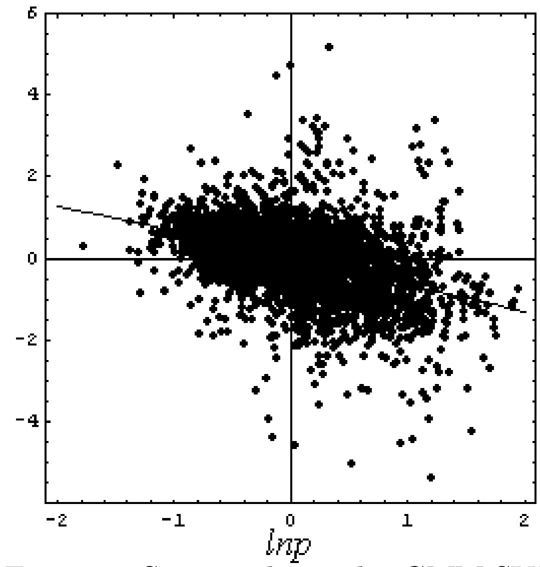

Figure 2: Scatter plot under GMM-SYS

6 Quantitative Exercises

In this section we perform several quantitative exercises, showing what bearing our estimation results have on several important issues of economic development and, in particular, on the question whether investment distortions explain income gaps across countries. In Subsection 6.1 we make the point that if we compare our estimated σ, $\sigma = 0.7$, to the traditionally used σ, $\sigma = 1$, then our σ magnifies the effect of distortions and thereby improves the neoclassical model's capability to explain income differences. In Subsection 6.2 we perform a development decomposition exercise à la Hall and Jones (1999). This exercise demonstrates that $\sigma = 0.7$ reduces the correlation between TFP and per capita incomes and again magnifies the role of distortions. In Subsection 6.3 we assess how much of the distortions that our model formulation is based on are captured by the investment good price in the Summers Heston data set.

6.1 How income jointly varies with P and σ

In this subsection we show the quantitative effects of distortions. We do this by (i) showing the extent to which incomes vary as distortions vary over a "reasonable" range. And (ii) by computing the elasticity of per-capita incomes with respect to distortions. To highlight the role of σ, we do both exercises for several values of σ, showing that a small σ accentuates the impact of distortions.

Our approach is to calibrate the model to U.S. data, and then simulate the model by letting the distortion parameter vary over a certain range and by letting σ assume several values (while holding other parameters constant).

To do this let's go back to Section 2, which shows that steady state per-capita income depends on the distortion parameters T_{I} and T_{K}, on the productivity parameters A and B, and on other parameters of the model. To make this dependence explicit we solve (11), assuming an interior solution. We get

$$k = \left\{ \frac{\alpha}{1-\alpha} \left[\left(\frac{P}{P(\sigma)} \right)^{\sigma-1} - 1 \right] \right\}^{-\frac{\sigma}{\sigma-1}}, \tag{31}$$

where

$$P(\sigma) \equiv A \frac{\alpha^{\frac{\sigma}{\sigma-1}}}{\rho+\delta} \text{ and } P \equiv \frac{T_{\mathrm{I}} T_{\mathrm{K}}}{B}. \tag{32}$$

This solution is interior, i.e., $k > 0$ if and only if $\sigma < 1$ and $P < P(\sigma)$, or $\sigma > 1$ and $P > P(\sigma)$.[38]

Substituting (31) into (3) and recalling that $y = Af(k)$, we get

$$y(P, \sigma) = A \left[\frac{1-\alpha}{1-\alpha_{\mathrm{K}}(P)} \right]^{\frac{\sigma}{\sigma-1}}, \tag{33}$$

where

$$\alpha_{\mathrm{K}}(P) \equiv k \frac{f'(k)}{f(k)} = \left[\frac{P}{P(\sigma)} \right]^{1-\sigma} \tag{34}$$

is the capital share of income.

Log-differentiating (33), the elasticity of income with respect to distortions is written as

$$\eta(P, \sigma) \equiv -\frac{P}{y} \frac{\mathrm{d}y}{\mathrm{d}P} = \sigma \frac{\alpha_{\mathrm{K}}(P)}{1-\alpha_{\mathrm{K}}(P)}. \tag{35}$$

(33) and (35) are the objects we are interested in, exhibiting the dependence of per-capita income y, and its elasticity η, on distortions P and on the aggregate production function σ. As they stand, however, (33) and (35) depend not only on P and σ but also on α and A.

[38] If $\sigma < 1$ and $P \geq P(\sigma)$ capital demand drops to zero (the economy is poverty trapped), whereas, if $\sigma > 1$ and $P \leq P(\sigma)$, capital demand is unbounded.

To isolate the role of P and σ, we assign values to α and A by calibrating the model to US data. To do this, consider the following three equations

$$
\begin{aligned}
y &= A\left[1-\alpha+\alpha k^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
\alpha_{\mathrm{K}} &= \frac{\alpha}{(1-\alpha)\,k^{-\frac{\sigma-1}{\sigma}}+\alpha} \\
\kappa \equiv \frac{k}{y} &= \frac{k}{A\left(1-\alpha+\alpha k^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}.
\end{aligned} \tag{36}
$$

y, α_{K} and κ in these equations have the status of observed variables (from US national income statistics), while α and A have the status of unobserved parameters. σ has the status of a "free parameter" (i.e., we solve for α and A as functions of σ). We normalize $y = 1$ and solve system (36) for α and A in terms of κ, α_{K} and σ, which gives

$$
A = \frac{\left[\alpha_{\mathrm{K}}+(1-\alpha_{\mathrm{K}})\,\kappa^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}}{\kappa} \tag{37}
$$

and

$$
\alpha = \frac{\alpha_{\mathrm{K}}}{\alpha_{\mathrm{K}}+(1-\alpha_{\mathrm{K}})\,\kappa^{\frac{\sigma-1}{\sigma}}}. \tag{38}
$$

Also, using (34) and substituting (37) and (38) into (32), we get

$$
P_{\mathrm{C}}(\sigma) = \frac{\alpha_{\mathrm{K}}^{\frac{\sigma}{\sigma-1}}}{\rho+\delta}\frac{1}{\kappa} \text{ and } P_{\mathrm{C}} = \frac{\alpha_{\mathrm{K}}}{\rho+\delta}\frac{1}{\kappa}, \tag{39}
$$

where C stands for 'calibrated.'

Having solved for α and A we simulate the model, i.e., we ask what US per-capita income would have been for hypothetical values of the distortion parameter, P. We let $P = P_{\mathrm{C}}\mathrm{P}$, where P is a hypothetical distortion parameter for the US economy. We substitute (39) into (34), which gives

$$
\alpha_{\mathrm{K}}(\mathrm{P}) = \alpha_{\mathrm{K}}\mathrm{P}^{1-\sigma}. \tag{40}
$$

Then we substitute (40) into (33) and (35), and get

$$
y\,(\mathrm{P},\sigma) = \left(\frac{1-\alpha_{\mathrm{K}}}{1-\alpha_{\mathrm{K}}\mathrm{P}^{1-\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{41}
$$

and

$$\eta\left(\mathrm{P},\sigma\right)=\sigma\frac{\alpha_{\mathrm{K}}\mathrm{P}^{1-\sigma}}{1-\alpha_{\mathrm{K}}\mathrm{P}^{1-\sigma}}. \tag{42}$$

Equations (41) and (42) is what we call the simulated model. They are graphically illustrated in Figures 3 and 4 for a range of P values. The figures exhibit the dependence of per-capita income and its elasticity on distortions for three distinct values of σ: $\sigma = 0.25$, 1, and 4.

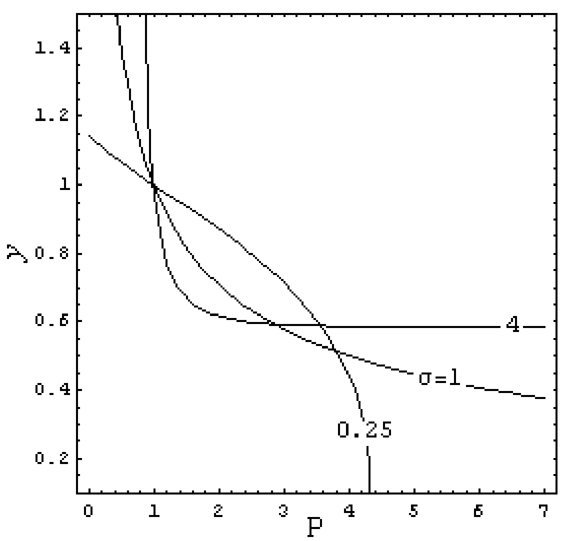

Figure 3: Output

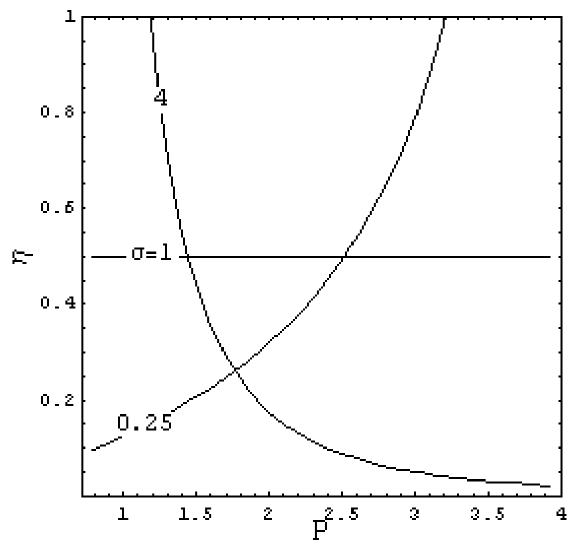

Figure 4: Elasticity

These figures show the role of σ in explaining per-capita income differences in the simulated model. Consider Figure 3 first. Then, letting P vary over the domain $[1, \mathrm{P}_{\max}]$, where $\mathrm{P}_{\max}$ is the largest P for which the equilibrium is interior, we see that income varies over a larger range the smaller σ is (or, in symbols, that $l(\sigma) \equiv y(1,\sigma) - y(\mathrm{P}_{\max},\sigma)$ is decreasing in σ). In this sense a smaller σ magnifies the impact of distortions. The reason we consider the domain $[1, \mathrm{P}_{\max}]$ is that most economies have a calibrated value of P above 1, and that $\mathrm{P}_{\max}$ is the largest P at which income is still positive.

Consider now Figure 4, which shows the elasticity of per-capita income with respect to distortions. As Figure 4 shows (and unlike Figure 3) σ has an ambiguous effect on this elasticity. If P is small, then a large σ makes η bigger,[39] whereas if P is large, then a *small* σ makes η bigger. The analytical counterpart to this is the following Proposition

Proposition 1 *There exists a* $\overline{\mathrm{P}}$ *so that*

$$\frac{\partial\eta}{\partial\sigma} \lesseqgtr 0 \text{ if and only if } \mathrm{P} \gtreqless \overline{\mathrm{P}}.$$

[39]This finding is consistent with Mankiw's (1995) work.

Proof. *Differentiating (42) we get*

$$\frac{\partial \eta}{\partial \sigma} = \eta\left(\mathrm{P}\right)^2 \frac{1 - \alpha_K \mathrm{P}^{1-\sigma} - \ln \mathrm{P}^{\sigma}}{\alpha_K \mathrm{P}^{1-\sigma}}.$$

Therefore the sign of $\frac{\partial \eta}{\partial \sigma}$ *depends on the sign of the numerator. Let us study then the numerator, which is a continuous function of* P

$$H\left(\mathrm{P}\right) \equiv 1 - \alpha_K \mathrm{P}^{1-\sigma} - \ln \mathrm{P}^{\sigma}. \tag{43}$$

We prove first that H *is decreasing in P whenever the solution is interior. Indeed*

$$\frac{dH\left(\mathrm{P}\right)}{d\mathrm{P}} \equiv -\frac{\sigma}{\mathrm{P}^{\sigma}}\left[\frac{1-\sigma}{\sigma}\alpha_K + \mathrm{P}^{\sigma-1}\right].$$

And this is negative when $\sigma < 1$ *and* $\mathrm{P} < \alpha_K^{\frac{1}{\sigma-1}}$ *or when* $\sigma > 1$ *and* $\mathrm{P} > \alpha_K^{\frac{1}{\sigma-1}}$ *(which, it can be shown, is equivalent to the condition for interior maximum).*

Second if $\sigma = 1$, $H\left(\mathrm{P}\right) \equiv 1 - \alpha_K - \ln \mathrm{P}$, *so we can explicitly solve* $H\left(\mathrm{P}\right) = 0$ *and get* $\overline{\mathrm{P}} = e^{1-\alpha_K}$. *If* $\sigma < 1$, *we have* $\lim_{\mathrm{P}\searrow 0} H\left(\mathrm{P}\right) = \infty$ *and* $H(\alpha_K^{\frac{1}{\sigma-1}}) = \frac{\sigma}{1-\sigma}\ln \alpha_K < 0$. *Thus there must be a* $\overline{\mathrm{P}} \in [0, \alpha_K^{\frac{1}{\sigma-1}})$ *so that* $H\left(\overline{\mathrm{P}}\right) = 0$. *If* $\sigma > 1$, $H(\alpha_K^{\frac{1}{\sigma-1}}) = -\frac{\sigma}{\sigma-1}\ln \alpha_K > 0$ *and* $lim_{\mathrm{P}\nearrow\infty} H\left(\mathrm{P}\right) = -\infty$. *So again there must be a* $\overline{\mathrm{P}}$ *so that* $H\left(\overline{\mathrm{P}}\right) = 0$. *Since* H *is decreasing this* $\overline{\mathrm{P}}$ *is unique.* ■

Given this Proposition we know there must be a $\widehat{\mathrm{P}}$ so that $\eta(\widehat{\mathrm{P}}, 0.7) = \eta(\widehat{\mathrm{P}}, 1.0)$, and after some computations we find that $\widehat{\mathrm{P}} = 2.01$. Therefore, if $\mathrm{P} > 2.01$ distortions under $\sigma = 0.7$ have a greater impact on per-capita incomes than under $\sigma = 1$. Using the calculations in Subsection 6.3, we find that roughly 40% of the (poorest) economies in the PWT have distortions in this range. Thus, policies that reduce distortions in such economies will have a greater impact under a CES production function with $\sigma = 0.7$ than under a CD production function with $\sigma = 1.0$.

6.2 TFP when $\sigma = 0.7$

In this section we calculate the total factor productivity implied by our model and how it correlates with per-capita incomes. We do this in our model with $\sigma = 0.7$, and compare the results to those calculated by Hall and Jones (1999), which use the Cobb-Douglas specification, $\sigma = 1$. Hall and Jones (1999) use data for per-worker capital and per-worker output controlled for education for 127 economies and measured in efficiency units. The data is for

1988 and the per-worker output excludes the output of the mineral sector. We use the same data for the exercise of this section.

The analogue of the Hall-Jones exercise in our framework works as follows. We substitute (37) and (38) into the production function and get

$$y_j = \mathrm{A}_j \left[1 - \alpha_{\mathrm{K}} + \alpha_{\mathrm{K}} \left(\frac{k_j}{\kappa} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \qquad (44)$$

where $\mathrm{A}_j \equiv \frac{A_j}{A}$ is the TFP of the jth economy, A is the US calibrated value from Subsection 6.1, $\alpha_{\mathrm{K}} = \frac{1}{3}$ and $\kappa = 3$ (observed US values).

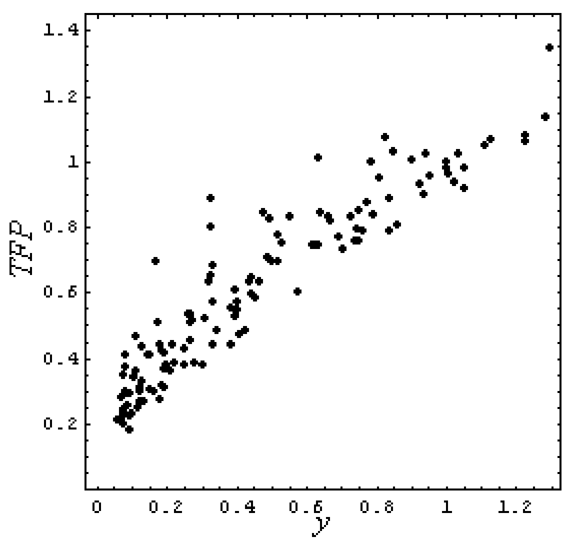

Figure 5: TFP with $\sigma = 1.0$

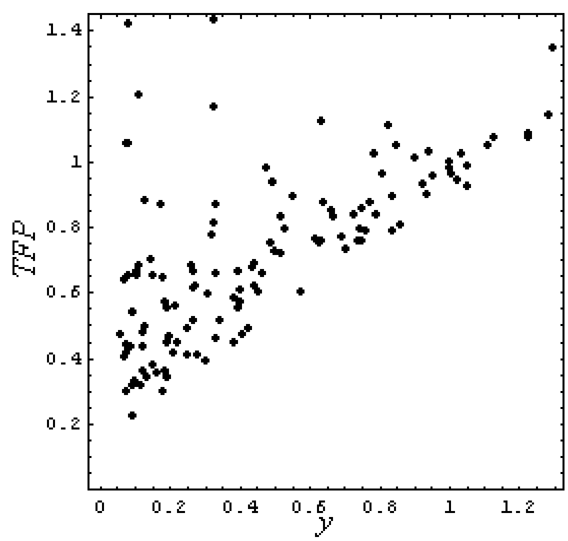

Figure 6: TFP with $\sigma = 0.7$

Then we take the values of y_j and k_j as reported in Hall and Jones (1999). Plugging those into equation (44), we compute the implied A_j for each economy. Then we plot those implied A_j against the corresponding GDP's y_j. The plot we get along with the plot that Hall and Jones get are shown in Figures 5 and 6.

Inspecting these figures and doing some calculations two features are revealed. First, the correlation between the implied A and y is reduced: $corr(y, A)$ is now (under $\sigma = 0.7$) 0.49, whereas before (under $\sigma = 1$) it was 0.86. Second, the average implied A increases from 0.61 to 0.73.

6.3 How much of the distortions are captured by S-H data?

The model formulation in Section 2 accommodates both observed ($T_{\mathrm{I},j}$) and unobserved ($T_{\mathrm{K},j}$) distortions. This raises the question what portion of the overall distortions are reflected by the price of capital in the Summers-Heston data set. As it turns out, our estimation results can be used to address that question. To do this, note that equation (18) implies

$$\frac{FE_j}{FE_{\mathrm{US}}} = \left(\frac{T_{\mathrm{K,US}}}{T_{\mathrm{K},j}}\right)^{\sigma} \left(\frac{A_{\mathrm{US}}}{A_j}\right)^{1-\sigma},$$

which, after some manipulations, gives

$$\frac{T_{\mathrm{K},j}}{T_{\mathrm{K,US}}} = \left(\frac{FE_{\mathrm{US}}}{FE_j}\right)^{\frac{1}{\sigma}} \mathrm{A}_j^{\frac{1-\sigma}{\sigma}}. \tag{45}$$

We plug the implied values of A_j (as we computed them in Subsection 6.2) along with the estimated values of the fixed effects FE_j (using the dynamic panel data approach, Subsection 4.2) into (45).[40] Then we compute the implied value of $T_{\mathrm{K},j}$. We find that $T_{\mathrm{K},j} \in [0.4, 12]$, that average$(T_{\mathrm{K},j}) = 1.5$ and that

$$\begin{bmatrix} \mathrm{var}\,(\ln \mathbf{A}) & \mathrm{cov}\,(\ln \mathbf{A}, \ln \mathbf{T}_{\mathrm{K}}) & \mathrm{cov}\,(\ln \mathbf{A}, \ln \overline{\mathbf{p}}) \\ \mathrm{cov}\,(\ln \mathbf{A}, \ln \mathbf{T}_{\mathrm{K}}) & \mathrm{var}\,(\ln \mathbf{T}_{\mathrm{K}}) & \mathrm{cov}\,(\ln \mathbf{T}_{\mathrm{K}}, \ln \overline{\mathbf{p}}) \\ \mathrm{cov}\,(\ln \mathbf{A}, \ln \overline{\mathbf{p}}) & \mathrm{cov}\,(\ln \mathbf{T}_{\mathrm{K}}, \ln \overline{\mathbf{p}}) & \mathrm{var}\,(\ln \overline{\mathbf{p}}) \end{bmatrix} = \begin{bmatrix} 0.16 & -0.08 & -0.05 \\ -0.08 & 0.43 & 0.08 \\ -0.05 & 0.08 & 0.24 \end{bmatrix},$$

where $\overline{\mathbf{p}}$ is the vector of cross time price averages. Looking at this table we see that the cross-country variance of investment prices (in the PWT data), var$(\ln \overline{\mathbf{p}})$ is 0.24, which represents roughly 36% of the total cross country variability of incentives to the investment decision, var$(\ln \mathbf{T}_{\mathrm{K}})$ +var$(\ln \overline{\mathbf{p}}) = 0.67$.

For completeness we have done an analogous exercise using the estimated FE from the static 6 years averaged panel. The results are in the same ball park: $T_{\mathrm{K},j} \in [0.37, 14]$, average$(T_{\mathrm{K},j}) = 1.7$, var$(\ln \mathbf{T}_{\mathrm{K}}) = 0.50$, cov$(\ln \mathbf{A}, \ln \mathbf{T}_{\mathrm{K}}) = -0.09$, cov$(\ln \mathbf{T}_{\mathrm{K}}, \ln \overline{\mathbf{p}}) = 0.10$, and

$$\frac{\mathrm{var}\,(\ln \overline{\mathbf{p}})}{\mathrm{var}\,(\ln \mathbf{T}_{\mathrm{K}}) + \mathrm{var}\,(\ln \overline{\mathbf{p}})} = \frac{0.24}{0.24 + 0.50} = 0.33.$$

Interestingly, when we do the same exercise under a Cobb-Douglas specification, $\sigma = 1.0$, we get the significantly larger portion 66%.

[40] We used the SYS-GMM estimates of $\ln FE_j^{\mathrm{D}}$, and set $\ln FE_j = \frac{\ln FE_j^{\mathrm{D}}}{1-\beta_1}$.

7 Conclusion

This paper presents several econometric exercises aimed at estimating the elasticity of substitution of an aggregate production function. Our results indicate that this elasticity hovers around 0.7 and that it is decidedly less than 1. Once we obtain a value for σ we use it to address the question whether the neoclassical model accounts for income gaps as coming from differential distortions, and to do related macro-development exercises.

Let us close by suggesting that the methodology we advance in this paper may be of interest in other contexts. Our methodology is such that, with the exception of σ, we determine the values of model parameters by calibrating the model to the U.S. economy, but we determine a value for σ by *estimating* an econometric model, using international data. This methodology is necessitated by the fact that one cannot assign a value to σ using data on a single economy. σ is a curvature parameter of the production function, so we need to observe some variation along the production function in order to infer its curvature, and for that purpose international data is important (since different countries are at different points along the production function, which is due to the fact that they have different distortions). What is not so standard about this methodology is that it *combines* calibration with estimation.

This same methodology is applicable for other purposes. Let us mention three. The first is still within the macro-development context. Following up on Lewis (1954) work, two-sector models of development have recently been popularized; see for example Hansen and Prescott (2002). One sector is the traditional or the agricultural sector, whereas the other is the modern or the industrial sector. In such context one studies how economies transit from traditional to industrial production or, conversely, how poorly managed (highly distorted) economies revert back to traditional production. Just as above, a key determinant of this process is the elasticity of substitution and, hence, one can quantitatively evaluate the process under an empirically estimated σ. Compared to what we have done here, in this two-sector framework, there may no longer be such thing as a poverty trap (which happens in our one-sector model). Compared with the previously studied two-sector model, the calibration and simulation results are expected to be different when $\sigma = 0.7$ (as opposed to $\sigma = 1$).

A second application is to the micro-policy question of quantifying the effects of investment tax credits. The investment equation we derive here, (17), shows that investment tax credits have a greater stimulative effect on investment the greater is σ. Then, as in Chirinko (2002), one can study how government policy along with properties of the production function affect investments. A third application is to the real business cycle literature. A key ingredient in this literature is the aggregate production function. Therefore, whatever quantitative exercises are done in this literature can be re-done by first estimating a production function from relevant data and then using it in the quantitative exercises.

References

[1] **Alvarez, Javier and Manuel Arellano 2002.** "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators." Forthcoming in *Econometrica (*http://www.cemfi.es/~arellano/#WPapers).

[2] **Arellano, Manuel 1987.** "Computing Robust Standard Error for Within Groups." *Oxford Bulletin of Economics and Statistics*, 49: 431-434.

[3] **Arellano, Manuel, and Stephen Bond 1991.** "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58: 277-297.

[4] **Arellano, Manuel, and Stephen Bond 1998.** "Dynamic Panel Data Estimation using DPD98 for Gauss: A Guide for Users" Institute for Fiscal Studies, London (http://www.ifs.org.uk/staff/steve_b.shtml).

[5] **Arellano, Manuel, and Olympia Bover 1995.** "Another Look at the Instrumental Variables Estimation of Error Components Models." *Journal of Econometrics*, 68: 29-51.

[6] **Arrow, K. J., H. B. Chenery, B. S. Minhas and R. M. Solow 1961.** "Capital-Labor Substitution and Economic Efficiency." *Review of Economic and Statistics*, 43(3): 225-250.

[7] **Baltagi, Badi 1995**. *Econometric Analysis of Panel Data.* New York: John Wiley & Sons.

[8] **Barro, R. J., N. G. Mankiw, and X. Sala-I-Martin 1995.** "Capital Mobility in Neoclassical Models of Growth." *American Economic Review,* 85(1): 103-115.

[9] **Bhargava, A., L. Franzini, and W. Narendranathan 1982**. "Serial Correlation and the fixed effects model." *The Review of Economics Studies*, 49: 533-549.

[10] **Blundell, Richard, and Stephen Bond 1998a.** "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87: 115-14

[11] **Blundell, Richard, and Stephen Bond 1998b.** "GMM Estimation with Persistent Panel Data: an Application to Production Functions." The Institute for Fiscal Policies, Working Paper Series No W99/4 (http://www.ifs.org.uk/working papers/wp994.pdf).

[12] **Bond, Stephen, Anke Hoeffler, and Jonathan Temple 2001.** "GMM Estimation of Empirical Growth Models." *Discussion Paper* # 01/21 (http://www.nuff.ox.ac.uk/economics/papers/2001/w21/bht10.pdf).

[13] **Chamberlain, Gary 1984.** "Panel Data," in: *Handbook of Econometrics*, Vol 2, Eds: Z. Griliches and M. D. Intriligator, Elsevier, Amsterdam, 1247-1313.

[14] **Chirinko, Robert S. 2002.** "Corporate Taxation, Capital Formation, and the Substitution Elasticity Between Labor and Capital." Emory University, Department of Economics Working Papers # 02-01 (March) (http://userwww.service.emory.edu/%7Eskrause/wp/chirinko_02_01_cover.htm).

[15] **Chirinko, Robert S., Steve M. Fazzani and Andrew P. Mayer 2002.** "That Elusive Elasticity: A Long Panel Approach to Estimating The Price Sensitivity of Business Capital." Emory University, Department of Economics Working Papers # 02-02 (March) (http://userwww.service.emory.edu/%7Ecozden/chirinko_02_02_cover.html).

[16] **Collins, Williams J. and Jeffrey G. Williams 1999.** "Capital Goods Prices, Global Capital Markets and Accumulation: 1870-1950." NBER Working Paper No. 7145 (May).

[17] **Hall, Robert, and Charles Jones 1999.** "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, 114(1): 83-116.

[18] **Hansen, Gary and Edward Prescott 2002.** "Malthus to Solow." *American Economic Review*, 92(4): 1205-1218.

[19] **Heston, Alan, Robert Summers and Betina Atten 2002.** "Penn-World Table Version 6.1." Center for International Comparisons at the University of Pennsylvania, October (http://pwt.econ.upenn.edu/).

[20] **Holtz-Eakin, Douglas, White Newey, and Harvey Rosen, 1988.** "Estimating Vector Autoregressions with Panel Data." *Econometrica*, 56(6): 1371-1395.

[21] **Hsieh, Chang-Tai and Peter J. Klenow 2003.** "Relative Prices and Relative Prosperity." Working Paper (April) (http://www.klenow.com/).

[22] **Jones, Charles I. 1994.** "Economic Growth and the Relative Price of Capital." *Journal of Monetary Economics*, 34: 359-382.

[23] **Judson, Ruth, and Ann Owen 1996.** "Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists." Federal Reserve Board of Governors.

[24] **Klenow, Peter J. and Andrés Rodríguez-Clare 1997.** "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual*, Ben S. Bernanke and Julio J. Rotemberg (editors), The MIT Press: 73-103.

[25] **Loayza, Norman, Klaus Schmidt-Hebbel, and Luis Serven 2000**. "What Drives Private Saving Across the World?" Forthcoming, *Review of Economics and Statistics.*

[26] **Lucas, R. E., Jr. 1969.** "Labor-Capital Substitution in US Manufacturing." in *The Taxation of Income from Capital*, ed. Arnold C. Harberger and Martin J. Bailey, The Brookings Institution, Washington, D.C.

[27] **Lewis, W. A. 1954.** "Economic Development with Unlimited Supplies of Labor." *Manchester School of Social Science*, 22: 139-191.

[28] **Maddala, G. S. and I. Kim 1998.** "Unit Roots, Cointegration and Structural Change." Cambridge: Cambridge University Press.

[29] **Mankiw, N. Gregory 1995.** "The Growth of Nations." *Brookings Papers on Economic Activity*, 1995 (1): 275-326.

[30] **Nickell, Stephen 1981.** "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49(6): 1417-1426.

[31] **Parente, Stephen and Edward Prescott 2000**. *Barriers to Riches.* The MIT Press: Cambridge, Massachusetts.

[32] **Pessoa, Samuel and Rafael Rob 2003**. "The Implications of Embodiment and Putty-Clay to Economic Development," Mimeo, University of Pennsylvania.

[33] **Romer, David 2001.** *Advanced Macroeconomics*, Second Edition, McGraw-Hill.

[34] **Restuccia, Diego and Carlos Urritia 2001.** "Relative Prices and Investment Rates." *Journal of Monetary Economics*, 47: 93-121.

[35] **Summers, Robert and Alan Heston 1991.** "The Penn-World Table: An Expanded Set of International Comparisons, 1950-1988." *Quarterly Journal of Economics* 106 (May): 327-368.

[36] **Windmeijer, Franck** 2000. "A Finite Sample Correction for the Variance of Linear Two-Step GMM Estimators." The Institute of Fiscal Policies, Working Paper Series n. W00/19.

A Dynamic panel Estimation with GMM

GMM-DIF Estimation. The first two observations used for estimating equation (24) are lost to lags and differencing. At $t = 3$, $\ln i_{j1}$ is a valid instrument for $\ln i_{j2} - \ln i_{j1}$, and $\ln p_{j1}$ is a valid instrument for $\ln p_{j2} - \ln p_{j1}$ and $\ln p_{j3} - \ln p_{j2}$. Similarly, at $t = 4$, $\ln i_{j1}$ and $\ln i_{j2}$ are valid instruments for $\ln i_{j3} - \ln i_{j2}$, and $\ln p_{j1}$ and $\ln p_{j2}$ are valid instruments for $\ln p_{j3} - \ln p_{j2}$ and $\ln p_{j4} - \ln p_{j3}$, respectively. Consequently, the instrument matrix has one row for each time period, giving $T - 2$ rows altogether, and $M = 2 \times \sum_{m=1}^{T-2} m$ columns. The instruments matrix is

$$\mathbf{Z}_j^D = \begin{pmatrix} \mathbf{Z}_j^{D1}, & \mathbf{Z}_j^{D2} \end{pmatrix},$$

where

$$Z_j^{D1} = \begin{pmatrix} \ln i_{j1} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \ln i_{j1} & \ln i_{j2} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \ln i_{j1} & \ln i_{j2} & \cdots & \ln i_{jT-2} \end{pmatrix},$$

$$Z_j^{D2} = \begin{pmatrix} \ln p_{j1} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \ln p_{j1} & \ln p_{j2} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \ln p_{j1} & \ln p_{j2} & \cdots & \ln p_{jT-2} \end{pmatrix}.$$

Let $\mathbf{X}_{jt} = (\ln i_{jt-1}, \ln p_{jt}, \ln p_{jt-1})$ be the 1×3 vector of covariates for j and t and Θ the 3×1 vector of coefficients. Define the first-differenced variables as

$$y_j^* = \begin{pmatrix} \ln i_{j3} - \ln i_{j2} \\ \ln i_{j4} - \ln i_{j3} \\ \vdots \\ \ln i_{jT} - \ln i_{jT-1} \end{pmatrix}, \ \mathbf{X}_j^* = \begin{pmatrix} \mathbf{X}_{j3} - \mathbf{X}_{j2} \\ \mathbf{X}_{j4} - \mathbf{X}_{j3} \\ \vdots \\ \mathbf{X}_{jT} - \mathbf{X}_{jT-1} \end{pmatrix}, \text{ and } \boldsymbol{\epsilon}_j^* = \begin{pmatrix} \epsilon_{j3} - \epsilon_{j2} \\ \epsilon_{j4} - \epsilon_{j3} \\ \vdots \\ \epsilon_{jT} - \epsilon_{jT-1} \end{pmatrix}.$$

The moment restrictions (25) and (26) can be written as $E(\mathbf{Z}_j^{D\prime}\epsilon_j^*) = \mathbf{0}$, where $\mathbf{0}$ is an $M \times 1$ vector of zeros. The GMM estimator based on these moment restrictions minimizes the expected quadratic distance between $\boldsymbol{\epsilon}^{*\prime}\mathbf{Z}^D\mathbf{W}\mathbf{Z}^{D\prime}\boldsymbol{\epsilon}^*$ and the true vector of parameters for the metric $\mathbf{W}$, where $\mathbf{Z}^{D\prime}$ is the $M \times N(T-2)$ matrix $(\mathbf{Z}_1^{D\prime}, \mathbf{Z}_2^{D\prime}, ..\mathbf{Z}_N^{D\prime})$ and $\boldsymbol{\epsilon}^{*\prime}$ is the $N(T-2)$

vector $(\boldsymbol{\epsilon}_1^{*\prime}, \boldsymbol{\epsilon}_2^{*\prime}, ...\boldsymbol{\epsilon}_N^{*\prime})$. This gives the GMM estimator of Θ as

$$\widehat{\Theta} = (\mathbf{X}^{*\prime}\mathbf{Z}^D\mathbf{W}\mathbf{Z}^{D\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Z}^D\mathbf{W}\mathbf{Z}^{D\prime}y^*,$$

where $y^{*\prime}$ is an $N(T-2)$ vector and $\mathbf{X}^*$ is an $N(T-2)\times 3$ matrix.

Arellano and Bond (1991) suggest two choices for the weights $\mathbf{W}$, giving rise to two GMM estimators: one and two step estimators. In the one step estimator it is assumed that the ϵ_{jt} are independent and homoskedastic both across units and over time. Then the optimal choice of $\mathbf{W}$ is given by $\mathbf{W}_1 = (\frac{1}{N}\sum_{j=1}^{N}\mathbf{Z}_j^{D\prime}\mathbf{H}^D\mathbf{Z}_j^D)^{-1}$, where $\mathbf{H}^D$ is the $(T-2)\times(T-2)$ variance-covariance matrix of $E(\boldsymbol{\epsilon}_j^*\boldsymbol{\epsilon}_j^{*\prime})$

$$\mathbf{H}^D = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}.$$

The variance-covariance estimator of the parameter $\widehat{\Theta}$ that is robust to heteroskedasticity is

$$\widehat{VC}(\Theta) = N(\mathbf{X}^{*\prime}\mathbf{Z}^D\mathbf{W}^{-1}\mathbf{Z}^{D\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Z}^D\mathbf{W}^{-1}(\sum_{j=1}^{N}\mathbf{Z}_j^{D\prime}\widehat{\boldsymbol{\epsilon}}_j^*\widehat{\boldsymbol{\epsilon}}_j^{*\prime}\mathbf{Z}_j^D)\mathbf{W}^{-1}\mathbf{Z}^{D\prime}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{Z}^D\mathbf{W}^{-1}\mathbf{Z}^{D\prime}\mathbf{X}^*)^{-1},$$

where $\widehat{\boldsymbol{\epsilon}}_j^*$ are the estimated residuals.

For the two step estimator the previous assumptions about ϵ_{jt} are relaxed. In the first step we obtain the $\widehat{\boldsymbol{\epsilon}}_j^*$ and then we use them to construct a consistent estimate of the variance-covariance matrix of the moment restrictions. In this case, the optimal choice of $\mathbf{W}$ is given by $\mathbf{W}_2 = (\frac{1}{N}\sum_{j=1}^{N}\mathbf{Z}_j^{D\prime}\widehat{\epsilon}_j^*\widehat{\epsilon}_j^{*\prime}\mathbf{Z}_j^D)^{-1}$.

Both GMM estimators are consistent when N is much larger than T, although they may differ in their asymptotic efficiency. Also, in the special case of i.i.d. disturbances, both are asymptotically equivalent.

System GMM. The additional moment conditions (27) and (28) can be expressed as

$$E(\mathbf{Z}^{L'}\boldsymbol{\epsilon}_j) = 0,$$

where

$$\mathbf{Z}_j^L = \begin{pmatrix} y_{j2}^* & 0 & \cdots & 0 & \triangle \ln p_{j2} & 0 & \cdots & 0 \\ 0 & y_{j3}^* & \cdots & 0 & 0 & \triangle \ln p_{j3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{jT-1}^* & 0 & 0 & \cdots & \triangle \ln p_{jT-1} \end{pmatrix},$$

with $\triangle \ln p_{jt} = \ln p_{jt} - \ln p_{jt-1}$. Now, we can construct a GMM estimator which exploits both sets of moment restrictions. The instrument matrix for GMM-SYS is written as

$$\mathbf{Z}_j = \begin{pmatrix} \mathbf{Z}_j^D & 0 \\ 0 & \mathbf{Z}_j^L \end{pmatrix}.$$

The GMM-SYS estimator combines both sets of moment restrictions

$$E(\mathbf{Z}_j' \boldsymbol{\xi}_j) = 0,$$

where

$$\boldsymbol{\xi}_j = \begin{pmatrix} \boldsymbol{\epsilon}_j^* \\ \boldsymbol{\epsilon}_j \end{pmatrix}.$$

Note that the one step GMM estimator is not asymptotically equivalent to the two step estimator - even when disturbances are i.i.d. The natural candidate for a weighting matrix for the one step estimator is $\mathbf{W}_1^{SYS} = (\frac{1}{N}\sum_{j=1}^{N} \mathbf{Z}_j' \mathbf{H} \mathbf{Z}_j)^{-1}$, where $\mathbf{H}$ is

$$\mathbf{H}_j = \begin{pmatrix} \mathbf{H}_j^D & 0 \\ 0 & \mathbf{I}_j \end{pmatrix},$$

which is always asymptotically inefficient relative to the two step estimator, because with level equations included in the system, the optimal weighting matrix depends on unknown parameters.

The construction of the two step GMM-SYS estimator is then analogous to that described under GMM-DIF, except that we use $\mathbf{H}_j = \widehat{\boldsymbol{\xi}}_j \widehat{\boldsymbol{\xi}}_j'$.

Monte Carlo simulations of Blundell and Bond (1998a) show that the finite sample distributions of the one step and two step system GMM estimators are similar.

B Estimated Standard Error of the Long-Run Price Elasticity

In order to compute the estimated standard error of the long run price elasticity of investment demand β_{LR} we apply the Delta Method. Define $\boldsymbol{\beta} \equiv (\beta_1, \beta_2, \beta_3)'$. Then β_{LR}, as a function of $\boldsymbol{\beta}$, is given by equation (23). Applying a first order Taylor series approximation to this function around the true value of $\boldsymbol{\beta}$ we get

$$\text{LR}(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3) \approx \text{LR}(\beta_1, \beta_2, \beta_3) + (\nabla LR)'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{46}$$

where

$$(\nabla LR)' = \left[\begin{array}{ccc} \frac{\beta_2+\beta_3}{(1-\beta_1)^2}, & \frac{1}{1-\beta_1}, & \frac{1}{1-\beta_1} \end{array} \right]$$

is the gradient of $\text{LR}(\beta_1, \beta_2, \beta_3)$. The variance of $\text{LR}(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)$ (which is a nonlinear function of $(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)$) is approximately equal to the variance of the right hand side of (46), which is

$$\text{var}(\text{LR}(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)) = \frac{\partial \text{LR}}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \text{var}(\widehat{\boldsymbol{\beta}}) \frac{\partial \text{LR}}{\partial \boldsymbol{\beta}}|'_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \ ,$$

where $\text{var}(\widehat{\boldsymbol{\beta}})$ is the estimated variance-covariance matrix of $\boldsymbol{\beta}$

$$\text{var}(\widehat{\boldsymbol{\beta}}) = \left[\begin{array}{ccc} \text{var}(\widehat{\beta}_1) & \text{cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \text{cov}(\widehat{\beta}_1, \widehat{\beta}_3) \\ \text{cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \text{var}(\widehat{\beta}_2) & \text{cov}(\widehat{\beta}_2, \widehat{\beta}_3) \\ \text{cov}(\widehat{\beta}_1, \widehat{\beta}_3) & \text{cov}(\widehat{\beta}_2, \widehat{\beta}_3) & \text{var}(\widehat{\beta}_3) \end{array} \right].$$

C The Bias of the OLS Estimation

We assume in this Appendix that the price of investment is an exogenous variable. Then

$$\widehat{\boldsymbol{\beta}}_{\text{Bias}} = \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} \mathbf{X}'\mathbf{X} \right)^{-1} \frac{1}{N} \mathbf{X}'\mathbf{FE}, \tag{47}$$

where

$$\underset{3\times((T-1)N)}{\mathbf{X}'} = \left[\begin{array}{ccccccccccc} 1 & \dots & 1 & & 1 & \dots & 1 & & 1 & \dots & 1 \\ \ln i_{12} & \dots & \ln i_{1T} & \cdots & \ln i_{j2} & \dots & \ln i_{jT} & \cdots & \ln i_{N2} & \dots & \ln i_{NT} \\ \ln p_{12} & \dots & \ln p_{1T} & & \ln p_{j2} & \dots & \ln p_{jT} & & \ln p_{N2} & \dots & \ln p_{NT} \end{array} \right]$$

and

$$\underset{1\times((T-1)N)}{\mathbf{FE}'} = \left[\begin{array}{ccccccccccc} \ln \overline{FE}_1^{\mathrm{D}} & \ldots & \ln \overline{FE}_1^{\mathrm{D}} & \cdots & \ln \overline{FE}_j^{\mathrm{D}} & \ldots & \ln \overline{FE}_j^{\mathrm{D}} & \cdots & \ln \overline{FE}_N^{\mathrm{D}} & \ldots & \ln \overline{FE}_N^{\mathrm{D}} \end{array} \right].$$

$\ln \overline{FE}_j^{\mathrm{D}} \equiv \ln FE_j^{\mathrm{D}} - \frac{1}{N} \sum_{j'=1}^{N} \ln FE_{j'}^{\mathrm{D}}$ is the centred fixed effect for the j-th economy. The first observation is deleted due to the dynamics.

Evaluating $\frac{1}{N}\mathbf{X}'\mathbf{FE}$, we have

$$\frac{1}{N}\mathbf{X}'\mathbf{FE} = \left[\begin{array}{c} \frac{1}{N} \sum_{j=1}^{N} \ln \overline{FE}_j^{\mathrm{D}} \\ \frac{1}{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \ln i_{jt} \ln \overline{FE}_j^{\mathrm{D}} \\ \frac{1}{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \ln p_{jt} \ln \overline{FE}_j^{\mathrm{D}} \end{array} \right].$$

Now we are going to compute each of the three components of $\frac{1}{N}\mathbf{X}'\mathbf{FE}$. The first component is zero by the way $\ln \overline{FE}_j^{\mathrm{D}}$ is defined. It remain then to compute the other two components.

To compute the third component of $\frac{1}{N}\mathbf{X}'\mathbf{FE}$, we assume that the covariance between the price and the fixed effect is time invariant

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{j=1}^{N} \ln p_{jt} \ln \overline{FE}_j^{\mathrm{D}} = \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{j=1}^{N} \ln p_{jt'} \ln \overline{FE}_j^{\mathrm{D}} \text{ for any } t, t' \in \{1, ..., T\}. \tag{48}$$

Then

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \ln p_{jt} \ln \overline{FE}_j^{\mathrm{D}} = \underset{N\to\infty}{\text{plim}} \frac{T-1}{N} \sum_{j=1}^{N} \ln p_j \ln \overline{FE}_j^{\mathrm{D}}.$$

To compute the second component of $\frac{1}{N}\mathbf{X}'\mathbf{FE}$, we apply equation (22) in the text. Then, using $\sigma = -\frac{\beta_2+\beta_3}{1-\beta_1}$ and the fact that β_3 is estimated to be zero, equation (22) is reduced to

$$\ln i_{jt} = \ln FE_j^{\mathrm{D}} + \beta_1 \ln i_{jt-1} - \sigma (1-\beta_1) \ln p_{jt} + \epsilon_{jt}. \tag{49}$$

Now we repeatedly substitute (49) into itself, obtaining

$$\ln i_{jt} = \frac{1-\beta_1^{t-1}}{1-\beta_1} \ln FE_j^{\mathrm{D}} + \beta_1^{t-1} \ln i_{j1} - \sum_{k=0}^{t-2} \beta_1^k \left[\sigma (1-\beta_1) \ln p_{j,t-k} - \epsilon_{j,t-k} \right]. \tag{50}$$

Then assuming that all economies are initially on a balanced growth path we have

$$\ln i_{j1} = \ln FE_j^{\mathrm{D}} - \sigma \ln p_{j1} + \epsilon_{j1}. \tag{51}$$

Plugging (51) into (50) we get

$$\ln i_{jt} = \frac{1-\beta_1^t}{1-\beta_1} \ln FE_j^{\mathrm{D}} - \beta_1^t \sigma \ln p_{j1} - \sigma \left(1-\beta_1\right) \sum_{k=0}^{t-1} \beta_1^k \ln p_{j,t-k} + \sum_{k=0}^{t-1} \beta_1^k \epsilon_{j,t-k}.$$

The above equation allows us to write the second component of $\frac{1}{N}\mathbf{X}'\mathbf{FE}$ as

$$\begin{aligned}\sum_{j=1}^{N}\sum_{t=2}^{T} \ln i_{jt} \ln \overline{FE}_j^{\mathrm{D}} &= \sum_{t=2}^{T} \frac{1-\beta_1^t}{1-\beta_1} \sum_{j=1}^{N} \ln FE_j^{\mathrm{D}} \ln \overline{FE}_j^{\mathrm{D}} - \sigma \sum_{t=2}^{T} \beta_1^t \sum_{j=1}^{N} \ln p_{j1} \ln \overline{FE}_j^{\mathrm{D}} \\ &- \sigma \left(1-\beta_1\right) \sum_{t=2}^{T}\sum_{k=0}^{t-1} \beta_1^k \sum_{j=1}^{N} \ln p_{j,t-k} \ln \overline{FE}_j^{\mathrm{D}} \\ &+ \sum_{t=2}^{T}\sum_{k=0}^{t-1} \beta_1^k \sum_{j=1}^{N} \epsilon_{j,t-k} \ln \overline{FE}_j^{\mathrm{D}}.\end{aligned}$$

To simplify this last expression, we re-use our assumption (48) and, on top of that, assume that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \epsilon_{jt} \ln \overline{FE}_j^{\mathrm{D}} = 0, \text{ for all } t.$$

Then after some calculations we find that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N}\sum_{t=2}^{T} \ln i_{jt} \ln \overline{FE}_j^{\mathrm{D}} = \operatorname*{plim}_{N\to\infty} \frac{D}{N} \sum_{j=1}^{N} \ln FE_j^{\mathrm{D}} \ln \overline{FE}_j^{\mathrm{D}} - \operatorname*{plim}_{N\to\infty} \sigma \frac{T-1}{N} \sum_{j=1}^{N} \ln p_j \ln \overline{FE}_j^{\mathrm{D}},$$

where

$$D \equiv \frac{\left(1-\beta_1\right)\left(T-1\right) - \beta_1^2 \left(1-\beta_1^{T-1}\right)}{\left(1-\beta_1\right)^2}.$$

This completes the computation of the second component. Introducing simplifying notation and stacking up the three components we have

$$\frac{1}{N}\mathbf{X}'\mathbf{FE} = N \begin{bmatrix} 0 \\ D \operatorname{var}\left(\ln \mathbf{FE}^{\mathrm{D}}\right) - \sigma \left(T-1\right) \operatorname{cov}\left(\ln \mathbf{FE}^{\mathrm{D}}, \ln \mathbf{p}\right) \\ \left(T-1\right) \operatorname{cov}\left(\ln \mathbf{FE}^{\mathrm{D}}, \ln \mathbf{p}\right) \end{bmatrix}.$$

Going back to (47) we now find that

$$\widehat{\boldsymbol{\beta}}_{\text{Bias}} = \begin{bmatrix} \widehat{\beta}_{0,\text{Bias}} \\ \widehat{\beta}_{1,\text{Bias}} \\ \widehat{\beta}_{2,\text{Bias}} \end{bmatrix} = \left(\underset{N\to\infty}{\text{plim}} \frac{\det \mathbf{X}'\mathbf{X}}{N^3 (T-1)^2} \right)^{-1} \begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \text{var}(\ln \mathbf{p}) & -\text{cov}(\ln \mathbf{i}, \ln \mathbf{p}) \\ \bullet & -\text{cov}(\ln \mathbf{i}, \ln \mathbf{p}) & \text{var}(\ln \mathbf{i}) \end{bmatrix}$$
$$\times \begin{bmatrix} 0 \\ D\text{var}\left(\ln \mathbf{FE}^{\text{D}}\right) - \sigma (T-1) \text{cov}\left(\ln \mathbf{FE}^{\text{D}}, \ln \mathbf{p}\right) \\ (T-1) \text{cov}\left(\ln \mathbf{FE}^{\text{D}}, \ln \mathbf{p}\right) \end{bmatrix}.$$

Now that we found the biases, we plug them into (30) and derive the following result

$$\frac{\partial \widehat{\beta}_{\text{LR}}^{\text{OLS}}}{\partial \text{cov}(\mathbf{FE}, \mathbf{p})} = \left(\underset{N\to\infty}{\text{plim}} \frac{\det \mathbf{X}'\mathbf{X}}{N^3 (T-1)^3} \right)^{-1} \frac{2 - \sigma + \widehat{\beta}_{\text{LR}}^{\text{OLS}} (1-\sigma)}{1 - \left(\widehat{\beta}_1 + \widehat{\beta}_{1,\text{Bias}}\right)} \text{var}(\ln \mathbf{p}). \tag{52}$$

In deriving this result we use the fact that, according to the data,

$$\text{var}(\ln \mathbf{i}) \approx 2\text{var}(\ln \mathbf{p}) \text{ and } \text{var}(\ln \mathbf{p}) \approx -\text{cov}(\ln \mathbf{i}, \ln \mathbf{p}).$$

Consider now equation (52). The first and third terms are positive. The denominator of the second term is also positive because all our regression results are such that $\widehat{\beta}_1 + \widehat{\beta}_{1,\text{Bias}} < 1$ (and more generally because unit root in the investment process had been ruled out). Therefore if we can show that the numerator of the second term is positive we will have that

$$\frac{\partial \widehat{\beta}_{\text{LR}}^{\text{OLS}}}{\partial \text{cov}\left(\ln \mathbf{FE}^{\text{D}}, \ln \mathbf{p}\right)} > 0.$$

To show that this numerator is positive we note that

$$-\frac{2-\sigma}{1-\sigma} \approx -\frac{2 - \widehat{\beta}_{\text{LR}}}{1 - \widehat{\beta}_{\text{LR}}} = -\frac{1.3}{0.3} \approx -4.$$

where the equality comes from our estimation result $\widehat{\beta}_{\text{LR}} = 0.7$ (see Table 6, column [8]). Therefore if we can show that $0 \geq \widehat{\beta}_{\text{LR}}^{\text{OLS}} > -4$ we would be done.

Let's then explicitly calculate $\widehat{\beta}_{\text{LR}}^{\text{OLS}}$. To do that we use the full information from the data

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.0070 & -0.0023 & -0.0026 \\ -0.0023 & 0.0008 & 0.0008 \\ -0.0026 & 0.0008 & 0.0016 \end{bmatrix},$$

and the following estimated values from our GMM-SYS estimation

$$\sum_{j=1}^{N} \ln p_j \ln \overline{FE}_j^{\mathrm{D}} = -3 \text{ and } \sum_{j=1}^{N} \ln FE_j^{\mathrm{D}} \ln \overline{FE}_j^{\mathrm{D}} = 1.5.$$

Furthermore, the dynamic panel has $T = 36$ and we know from the GMM-SYS estimation that $\widehat{\beta}_1 = 0.744$ and $\widehat{\beta}_2 = -0.177$ (Table 6, column [8]). Using all these values we get $D \approx 128$ and $\sigma (T-1) \approx 25$. Then after plugging this into equation (30) we get

$$\widehat{\beta}_{\mathrm{LR}}^{\mathrm{OLS}} = \frac{\widehat{\beta}_2 + \widehat{\beta}_{2,\mathrm{Bias}}}{1 - \left(\widehat{\beta}_1 + \widehat{\beta}_{1,\mathrm{Bias}}\right)} = -1.04,$$

where

$$\widehat{\beta}_{1,\mathrm{Bias}} = 0.13 \text{ and } \widehat{\beta}_{2,\mathrm{Bias}} = 0.042.$$

D Other Distortions

In this Appendix we endogenize labor supply and consider distortions other than on the investment good price and on capital income. Analysis of this case shows that the estimating equation, which, as before, comes from the steady state, is the same as (17). Therefore, the estimate of the elasticity of substitution that we obtain remains valid in this more general environment.

To analyze this case we let each individual allocate one unit of time (at each point of time) between leisure and work. We denote by L the endogenously chosen amount of work, $0 \leq L \leq 1$.

The production function of sector 1 is now written as

$$y_1 = ALl_1 f(k_1), \tag{53}$$

where l_1 is again the fraction of per capita total labor employed in sector 1, k_1 is the capital-labor ratio in sector 1, A is total factor productivity, and f is the C.E.S. production function specified in the text.

Likewise, the production function of sector 2 is

$$y_2 = ABLl_2 f(k_2), \tag{54}$$

with analogous interpretation of variables.

The period subutility from leisure and consumption is general, $u(C, 1-L)$, so the lifetime utility of a representative individual is specified now as

$$\int_0^\infty e^{-\rho t} u(C, 1-L)\,\mathrm{d}t. \tag{55}$$

Given this the short run consumption-leisure decision satisfies

$$u_1(C, 1-L) = \frac{\mu}{p} \quad \text{and} \quad u_2(C, 1-L) = \frac{w\mu}{p}, \tag{56}$$

where μ is the Lagrange multiplier on the consumer's budget constraint (equation (5) in the text). The new Euler equation is now

$$\frac{\dot{\mu}}{\mu} = r_t - \rho.$$

Equations (5)-(8) remain intact.

Regarding distortions, τ_{I} and τ_{K} remain as is. On top of those we add a consumption tax at the rate τ_{C} and a wage tax at the rate τ_{L}.

As a consequence of these distortions, individuals pay

$$p = \frac{T_{\mathrm{I}}}{B} \equiv \frac{1+\tau_I}{B} \text{ and } T_{\mathrm{C}} \equiv 1 + \tau_{\mathrm{C}}$$

for the investment and consumption goods, respectively. As far as payments for factors of production, individuals receive

$$q = \frac{Af'(k)}{T_{\mathrm{K}}} \equiv (1-\tau_{\mathrm{K}})Af'(k) \text{ and } \frac{w}{T_{\mathrm{L}}} \equiv (1-\tau_{\mathrm{L}})\,w$$

as net rental rate and wages, respectively.

Equations (11)-(14) are the same as before.

When the model is taken to data, we let each country have its own distortions parameters $T_{\mathrm{I},j}$, $T_{\mathrm{K},j}$, $T_{\mathrm{C},j}$ and $T_{\mathrm{L},j}$. Then, doing the same manipulations as before the investment equation reads

$$i_j \equiv \frac{\mathrm{inv}_j}{y_j} = \frac{\delta}{A_j}\frac{K_j}{L_j f(k_j)} = \frac{\delta}{A_j}\frac{k_j}{f(k_j)},$$

and the long run investment-output ratio is

$$i_j = \frac{\delta}{A_j}\left[p_j \frac{T_{\mathrm{K},j}}{A_j}\frac{\rho+\delta}{\alpha}\right]^{-\sigma}. \tag{57}$$

Therefore, when we take logarithms we get the same estimating equation as in the text, (17).