Simulation-Based Estimation of a Nonlinear, Latent Factor AggregateProduction Function

Lee Ohanian, Giovanni L. Violante,
Per Krusell, and José-Víctor Ríos-Rull*

January 16, 1997

Abstract

In this paper, we analyze in detail econometric issues associated with the specification and estimation of an aggregate, nonlinear production function with latent variables. The production function has been used to assess the extent to which different substitution possibilities between capital, unskilled labor, and skilled labor can account for the recent increase in wage inequality. We use Monte Carlo methods to evaluate the performance in our environment of three different simulation-based estimation procedures, Stochastic Integration, Extended Kalman Filter with Indirect Inference correction and Simulated Pseudo ML, with a focus on how reliable these techniques are in small samples and when the latent variables have trends. We find that when the unobservable states are modeled as trend stationary processes the estimators peform much better than when the states are specified as I(1) series. For the trend stationary specification, the simulated MLE was judged to be the best method on the ground of precision and computational efficiency. The results of the estimation suggest that capital-skill complementarity, coupled with the sharp recent drop in relative prices of equipment, triggered the observed rise in wage inequality.

*Krusell, University of Rochester, Ohanian, University of Minnesota and University of Pennsylvania, R´ıos-Rull, Federal Reserve Bank of Minneapolis and University of Pennsylvania, and Violante, University of Pennsylvania.

1 Introduction

The purpose of this paper is to explore in detail a number of econometric issues associated with the specification and estimation of a non-linear, latent variable aggregate production function developed in Krusell, Ohanian, Ríos-Rull, and Violante (1995) (hereafter KORV). In particular, we discuss how different simulation-based methods can be used to address a number of difficult problems associated with our particular model, and evaluate the relative performance of these methods. Since some of the difficult issues with the specification and estimation of our model have not been analyzed in much detail using simulation-based methods, the findings reported here may be of use to other researchers working in similar environments.

In our earlier paper, we developed an aggregate production function that differs substantially from the standard production function used in macroeconomic analysis. The development of our alternative model was motivated by a key fact of the U.S. economy. In the last 30 years, a substantial difference has emerged in the growth rates of wages for workers with different educational levels. Juhn, Murphy, and Pierce (1993) report that the median wage earner among college graduates experienced a 15 percent increase in inflation-adjusted wages between 1964 and 1988, while the median wage earner among high school graduates experienced a 5 percent decline in real wages over the same period. The widening gap in the relative wage of skilled to unskilled labor is commonly referred to as the "wage premium" or "skill premium".

This observation has a number of important economic and social implications. First, not only do these data represent a striking change from the past, but the trend in wage disparity seems to be accelerating, suggesting the possibility of even *greater* economic inequality in the future. While individuals with higher educational attainment historically have received higher wages, the wage premium to education was relatively constant over time, and in fact narrowed considerably in the 1940s. Increased wage inequality is a relatively new phenomenon, and has increased substantailly in the last 15 years. Moreover, these observations are at variance with standard economic theory, which predicts that factor price differentials in the long run will tend to narrow as producers substitute away from relatively expensive factors into cheaper factors. These observations also have a number of key implications for government policy. For example, they play a central role in current discussions of redistribution, welfare and social insurance reform, since the extent to which aid recipients can be moved succesfully into employment depends in an important way on the wages they can earn.

Given these implications, a number of economists have tried to account empirically for the main factors responsible for the growing wage gap between skilled and unskilled workers. Despite substantial research on this topic, no consensus explanation for these observations has been established. Perhaps the most comprehensive work in this area is by Bound and Johnson (1992) who empirically evaluate four popular explanations for these changes: (1) The loss of high wage manufacturing jobs associated with the U.S. trade deficit, (2) The decline in the ability of labor unions to negotiate high wages, (3) Changes in technology, and (4) A slowdown in the growth rate of the college-educated population. Using regression methods, they find that these potential explanations account for only a fraction of the observed changes. Instead, most of the growth in the wage premium is explained by a regression-residual category, which they summarize as "*... a combination of skill-biased technical change and changes in unmeasured labor quality*".

Our theory can be used to interpret structurally the empirical findings of Bound and Johnson, and others. The main premise of our theory is based on the economic interaction between the behavior of the wage premium and two other key observations. First, there has been a rapid decline in the relative price of capital equipment over the past 25 years (Gordon (1990)). Associated with this falling price of capital equipment has been rapid growth in the stock of business capital equipment. Second, there is considerable empirical evidence (Hamermesh (1993) lists several references) that suggests *unskilled* labor and physical capital are relatively good substitutes, while *skilled* labor and physical capital are relatively complementary. Our theory proposes the following explanation for the change in the wage premium: the rise in the relative wages of skilled labor reflects the sharp increase in the stock of physical capital equipment given that they are relative complements in production. Moreover, the secular decline in the price of equipment puts downward pressure on the competitive price of unskilled labor, since they are relatively good substitutes.

To implement this theory empirically, we developed an aggregate production function with four inputs. This production technology differs considerably from the standard model in macroeconomics in that it explicitly distinguishes between skilled and unskilled labor, and between capital structures and capital equipment. We estimate the parameters of the production function, and study quantitatively the extent to which differences in complementarities between the two types of labor and capital equipment can account for increase in wage inequality. This strategy allows us to interpret and understand the main empirical findings of Bound and Johnson within the context of a fully articulated model. In particular, it allows us to separate the effects on the wage premium of the two main factors cited by Bound and Johnson: (1) technical improvement in capital equipment, and the different ways in which skilled and unskilled labor interact with equipment, and (2) changes in unmeasured quality of the two types of labor.

While the focus of our initial paper was on the development of the theory, data construction, and economic interpretation of the empirical findings, the main theme of this paper is a detailed analysis and presentation of the specification and estimation of the parameters of the four-input nonlinear production function.

There are a number of challenging and interesting econometric issues associated with estimating the parameters of this production function. The three main issues are:

- unmeasured labor quality of both skilled and unskilled workers enters the production function *nonlinearly* as a *latent* variable,
- we recognize that labor efficiency has changed over time, so we model the unobservable state as a *nonstationary* stochastic process,
- the sample size we will work with is relatively small.

In particular, we will consider two different stochastic specifications for labor efficiency (which could be interpreted as human capital). The first specification is a trend stationary latent state and the second is a first-difference stationary latent state with drift. Across these two specifications, we contrast three different procedures for estimating the parameters of interest. The first method is based on the exact likelihood of the model which is obtained by integrating out the unobservable state. The second method is based on a simple, but misspecified auxiliary model corrected by Indirect Inference, and the third method is based on a Simulated Pseudo-ML estimator. We provide Monte Carlo evidence on the performance of these estimators in nonstationary environments and in small samples and discuss their computational efficiency.

By providing a detailed analysis of the issues associated with parameter estimation in our specific model, we hope that our findings will be useful to other researchers analyzing problems that share some of the characteristics of our environment.

The remainder of the paper is organized as follows. In section 2, we describe the aggregate production function, and summarize the data used in the analysis. In section 3 we present the different specifications of the model, different according to the way we deal with the unmeasurable labor quality. In section 4, we outline the possible estimation methods when labor efficiency is assumed to be stochastic and compare their performance in a Monte Carlo analysis. Section 5 presents the results of the estimation, and the economic interpretation. Section 6 presents a summary and our conclusions.

2 The Model and Data

2.1 The Production Function

To develop the production function, we consider two categories of labor input: skilled workers, and unskilled (or less skilled) workers. The two categories are different from the point of view of production. The skill level is assumed to be exogenous, i.e. we do not model the individual's choice of skill level. We disaggregate the capital stock into two components, structures and equipment, because the most striking changes in the relative price of capital have occurred in the equipment category.

We choose to represent the production technology with a general nested CES functional form. This function is relatively simple and parsimonious, and allows for different substitution possibilities across factors. The production function is:

$$y_t = A_t k_{st}^{\alpha}[\mu u_t^{\sigma} + (1-\mu)(\lambda k_{et}^{\rho} + (1-\lambda)s_t^{\rho})^{\frac{\sigma}{\rho}}]^{\frac{1-\alpha}{\sigma}}$$

This technology has several distinguishing features. The technology has (stochastic) neutral technical progress A_t and is a CRS Cobb-Douglas function in the two inputs (i) capital structures, k_s, and (ii) the composite term $[\mu u_t^{\sigma} + (1-\mu)(\lambda k_{et}^{\rho} + (1-\lambda)s_t^{\rho})^{\frac{\sigma}{\rho}}]^{\frac{1}{\sigma}}$. This second term is a CES aggregate over unskilled labor, u, with share parameter μ, and a *second* CES aggregate (with share parameter $1-\mu$) over capital equipment, k_e, and skilled labor, s. (The relevant share parameters within this second CES term are λ and $1-\lambda$, respectively). The parameter α measures the structures share of income. The parameters σ and ρ are the key elements from the standpoint of our theory, since they govern the amount of substitutability between (i) skilled labor and the CES aggregate of capital equipment and unskilled labor, and (ii) unskilled labor and capital equipment, respectively. To satisfy regularity conditions, both σ and ρ are restricted to lie in $(-\infty, 1)$. Higher values of σ or ρ than zero correpond to more substitutability than Cobb-Douglas for the corresponding aggregate. The elasticity of substitution between equipment (or skilled labor) and unskilled labor is $\frac{1}{1-\sigma}$, and the elasticity of substitution between equipment and skilled labor is $\frac{1}{1-\rho}$. For simplicity, we assume that neutral technical progress A_t is an i.i.d. random variable, with mean normalized to 1.

The labor inputs are measured in efficiency units: each input type is a product of the raw number of labor hours and an hourly efficiency index. In particular, $s \equiv H_s n_s$ and $u \equiv H_u n_u$, where n_i is the number of hours and H_i the quality per hour of type i. It is clear that H_i can be given two distinct interpretations: it could be human capital, accumulated by the individual, or it could represent a skill-specific technology level, which may have come about by research and development activity. In the absence of direct measures of these two, they cannot be distinguished. Given our *i.i.d.* restriction on neutral technological change, we implicitly assume that all the technical progress in the economy is factor specific.

The wage premium, π, is defined as the ratio of skilled wages to unskilled wages. Assuming that factor prices are equal to marginal products, we have:

$$\pi_t = \frac{(1-\mu)(1-\lambda)}{\mu}\left[\lambda(\frac{k_{et}}{s_t})^{\rho} + (1-\lambda)\right]^{\frac{\sigma-\rho}{\rho}} \left(\frac{n_{ut}}{n_{st}}\right)^{1-\sigma} \left(\frac{H_{st}}{H_{ut}}\right)^{\sigma}.$$

The main premise of our theory regarding the relative complementarity of skilled labor and capital equipment, and the relative substitutability between unskilled labor and equipment corresponds to the case in which $\sigma > \rho$. In this case, as argued in KORV, the wage premium necessarily rises in response to an increase in the amount of equipment, *ceteris paribus*. Thus, the extent to which our theory featuring capital-skilled labor complementarity is useful to interpret the behavior of the wage premium can be evaluated by estimating the parameters of the production function, specifically the substitutability parameters σ and ρ.

2.2 The Data

2.2.1 Data on Labor Input

This section presents a summary of the data construction procedures[1]. The sources of our data on labor input and wages are **CPS Annual Demographic March Files** for the years 1964-1993. We have restricted our attention to all persons between 16 and 70 years old, excluding self-employed. The procedure for constructing the series for skilled and unskilled labor input and wage consists of two steps. In the *first step* we generate a partition of the sample into more than 200 demographic groups constructed on the basis of age, race, gender and educational level. For each one of these groups, we have a measure of average labor input and labor earnings. Hourly wage is simply the ratio between yearly labor income and yearly measure of labor input in terms of hours worked. The *second step* consists of sorting these groups into two categories: skilled workers and unskilled workers, and compute measures of total annual labor input for skilled workers n_{st} and of their hourly wage w_{st}, as well as total annual labor input for the unskilled n_{ut} and their hourly wage w_{ut}. Since there are no generally accepted definitions of skilled and unskilled (or less skilled) labor input, in KORV we explored several different methods to perform this aggregation, some based on education and others on the distribution of labor income across demographic groups. In this paper we will focus on one particular education-based criterion that defines unskilled workers as those individuals with less than 12 years of school (high school graduate). In order to aggregate

[1]A detailed data appendix from our other paper is available upon request.

group specific measures into the broad class of skilled and unskilled defined by the above criterion, we assume that the groups within a class are perfect substitutes (they simply add). Moreover, for the aggregation we use as weights the group wages of a given base year (1980, for consistency with the construction of the capital series) in very much the same way as real GDP is computed in National Income and Product Accounts (**NIPA**)[2]. Panel (a) and (b) of Figure 1 report the labor input ratio and the wage premium implied by our definition of skills.

2.2.2 Data on Capital

We have two types of capital in our model, Producers' Durable Equipment (PDE) and nonresidential structures. A key aspect of the data is that the relative price of capital equipment has declined secularly. Since in the data the behavior of the prices of structures and personal consumption has behaved similarly, we treat them identically, deflating by the price index for consumer nondurables and services.

The measurement of capital equipment is more difficult, since we are interested in *efficiency units* of equipment entering the production function. To obtain efficiency units, we need to compute the relative price of equipment. Because of the enormous improvement in the quality of equipment goods occurred in the last 10-15 years, especially for information technologies, quality-adjusted price indexes are needed for a meaningful and reliable measure of this relative price. The method we have used to aggregate all the quality-adjusted prices for the different categories of PDE and obtain a unique corrected price index for capital equipment is simply the Implicit Price Deflator (IPD).

For the period 1963-1983 we can rely on Robert Gordon's quality adjusted prices for what he defines as the *primary* 16 subcategories of PDE[3]. For the post 1983 sample period, we are not aware of any existing quality-adjusted series of PDE, except for computers and peripherals on which a fairly large literature of quality-adjustment procedures has emerged. Therefore, we have aggregated the 16 categories used by Gordon through the IPD with base year 1980 into the 4 main categories classified in the NIPA: Office and Information Processing Equipment (OIP), Industrial Equipment (INDEQ), Transportation Equipment (TRANSP) and Other Equipment (OTHER). The price indexes for the INDEQ, TRANSP and OTHER categories did not appear to experience dramatic changes in the 80's, and we assume that the relationship between quality-adjusted and official price indexes documented by Gordon has

[2]Note that we have CPS data for the period 1964-1993, but since wages and labor input data in the survey refer to one year earlier, our sample spans the period 1963-1992.

[3]See Appendix B.18 in Gordon (1990). His quality adjustment involves a detailed study on more than 100 specific products and the use of hedonic regressions and matching model procedures.

been stable over time. This allows us to estimate this relation by using Gordon's adjusted data as a dependent variable and the official series as regressors in a VAR on the levels for 1963-1983, and then to use the estimates and the exogenous variables, which are available up to 1992, to forecast the adjusted-price series. We specify the relationship in levels to take advantage of potential cointegration among variables[4].

The computation of an accurate quality-adjusted series for OIP is an important issue, since this is the group within PDE that recorded the largest change in price and relative share. We first divided the OIP group into Computer and Peripherals (COMP), and all others. For all the OIP categories other than computers and peripherals, we have used the official price index series. We do this for two reasons. First, we are not aware of any attempt of quality-adjustment which covers the 1984-1992 sample period. Second, it is our view that using the forecasting described above could generate very misleading results, given that dramatic changes have occurred in the relative prices and shares of these goods in the 80's. For instance, the NIPA official price index of communication equipment[5] grew at an increasing rate in the 60's and 70's, but in the 80's this rate decreased steadily. These observations suggest that the relation between official and quality-adjusted indexes for these goods may have changed structurally in the 80's.

Substantial research has been done on computing quality adjusted price indexes for computers and peripherals. We distinguished between PC's, mainframes (a broad category that includes also supercomputers, workstations and midrange) and peripherals, and derived quality-adjusted measures of prices for these 3 groups from the existing literature. Berndt, Griliches and Rappaport (1995) provide an hedonic price index for PC's, Brown and Greenstein (1995) compute a price index for mainframes and Cole et al. (1986) compute one for peripherals. We then constructed the series for quality-adjusted price index for COMP by aggregating these prices through an IPD for the period 1984-1990[6].

We then aggregated the price indexes for the 4 main categories of PDE with the IPD procedure. The relative price of equipment is constructed by dividing the aggregate IPD price series through the price index for consumption. Investment in capital equipment in efficiency units is constructed by deflating the nominal series of Investment in Equipment from NIPA through our quality-adjusted price index for equipment and then it is used together with a

[4]We fit a VAR(1) on q_t^a, with lagged exogenous variables q_{t-1} and $DCOINC_{t-1}$, where q^a is the vector of quality-adjusted prices for INDEQ, TRANSP, and OTHER obtained from Gordon's data, q is the vector of the corresponding official NIPA price indexes and $DCOINC$ is a coincident indicator of the business cycle.

[5]This price index is computed as the ratio between the series in current dollars GIPCE and the series in constant (1987) dollars GIPCEQ.

[6]As we could obtain data on COMP only up to 1990, this will be our endpoint of the sample in the estimation.

given depreciation rate to construct a series for the stock of equipment. The following Table summarizes the average growth rates of the relative price and the capital stock of equipment in efficiency units in the three decades of the sample.

Table1 – CapitalEquipment			
Growth Rate	**1963-1972**	**1973-1982**	**1983-1992**
Relative Price	-.023	-.028	-.139
Efficiency Units	.070	.069	.114

Panel (c) in Figure 1 reports the growth rates of structures and equipment over time and panel (d) depicts the decline of the relative price of equipment, especially sharp in the 80's.

3 Unobserved Labor Efficiency and Model Specifications

To estimate the parameters of the production function, we will use the first-order necessary conditions of a competitive profit maximizing firm. We assume that the firm observes the efficiency factors $\{H_{st}, H_{ut}\}$ of its labor inputs and the relative price of equipment p_t before making its choice on the quantities of the four inputs to hire but does not observe the *i.i.d.* neutral technical change A_t. On the other hand, none of these are observable by the econometrician.

For the labor efficiency terms, we distinguish three cases that provide a benchmark to evaluate the performance of different estimation techniques. Define h_t as $\log(H_t)$. In the first case (denoted as DT) the log of labor efficiency is simply a deterministic trend. In the second case (TS) h_t is a trend stationary process with no autocorrelation. In the third case (FDS) it is assumed to follow a first-difference stationary process with drift. These three specifications imply the following laws of motion for the unobservable states:

$$\begin{aligned}
&\text{(DT):} && h_t = h_0 + \gamma t \\
&\text{(TS):} && h_t = h_0 + \gamma t + \omega_t \\
&\text{(FDS):} && h_t = \gamma + h_{t-1} + \omega_t
\end{aligned}$$

Here ω_t is a (2×1) vector of disturbances distributed as $N(0, \Omega)$, h_0 is a (2×1) vector of constants, and γ is a (2×1) vector which specifies the (constant) growth rates of the two efficiency factors. We treat H_t as a latent variable, but only in the two latter specifications it evolves stochastically.

We exploit three restrictions to estimate the parameters of the production function. The first restriction is that the ratio of the wage bill of skilled to unskilled workers be equal to that observed in the data:

$$\frac{w_{st}N_{st}}{w_{ut}N_{ut}} = wbr_t(H_t, X_t; \phi) \tag{1}$$

The left-hand side of the equation is the ratio of the wage bill in the data, and the right hand side is the wage bill ratio from the model, which has as arguments the marginal products of the two types of labor, the vector of efficiency factors, a vector of exogenous observable variables $X_t = \{K_{et}, K_{st}, N_{ut}, N_{st}\}$, and a vector of structural parameters, ϕ.

The second restriction is that aggregate labor's share of income in the model (lsh_t) be equal to that observed in the data:

$$\frac{w_{st}N_{st} + w_{ut}N_{ut}}{Y_t} = lsh_t(H_t, X_t; \phi) \tag{2}$$

The data for the LHS of this restriction is the ratio of unambiguous labor income -defined as wages, salaries, and benefits- divided by the sum of the numerator and unambiguous capital income, which includes corporate profits, depreciation, rental income, and interest payment, following the methodology illustrated in Cooley (1995).

The third restriction is a no-arbitrage condition that equates the expected gross rates of return of the two types of capital:

$$1 - \delta_s + mpks_t(H_t, X_t; \phi) = (1 - \delta_e)\frac{E_t(p_{t+1})}{p_t} + \frac{1}{p_t} mpke_t(H_t, X_t; \phi) \tag{3}$$

The left-hand side of the equation is the net of depreciation return to capital structures, while the right-hand side of the equation is the expected net of depreciation return to capital equipment. The terms $mpke$ and $mpks$ denote "marginal product of capital equipment" and "marginal product of capital structures", respectively. δ_s and δ_e are the depreciation rates for structures. For the estimation, we will use as the first term on the right-hand side $(1-\delta_e)\frac{p_{t+1}}{p_t} + \varepsilon_t$, where ε_t is an i.i.d. forecast error which is assumed to have the Gaussian structure $\varepsilon_t \sim N(0, \eta_\varepsilon)$.

Note that the A_t term drops out of the first two conditions, and does not appear in the third one as the *i.i.d.* assumption allows us to integrate it out. Therefore the growth in the model economy is driven by the technological progress embedded in new capital equipment and by the growth in the two labor efficiency factors.

Given the relatively small sample and the large number of parameters, we have reduced the dimensionality of the parameter space by adopting values for those parameters that have been estimated or calibrated in prior studies. Following Greenwood, Hercowitz, and Krusell (1994) we calibrated δ_s, δ_e, and α. Furthermore, we used data on the relative price of equipment to estimate the standard deviation of the one-step forecast error η_ε. This was accomplished by fitting to the deviations of $\frac{p_{t+1}}{p_t}$ from a linear trend an $ARMA(1,1)$ model [7]; we then measured the standard deviation of the residuals and we multiplied it by

[7] The results of the estimation were (call $\pi_{t+1} = \frac{p_{t+1}}{p_t}$):

$(1-\delta_e)$. The number obtained -which is quite robust to different specifications for the linear conditional mean- is the calibrated counterpart of η_ε. The results of the calibration for these parameters are in the following table.

Table2 – CalibratedParameters			
$\alpha = .11$	$\delta_s = .05$	$\delta_e = .125$	$\eta_\varepsilon = .03$

Since we calibrate these parameters, it is worth noting that we do not completely take into account parameter uncertainty in the estimation. However the use of this procedure increases considerably degrees of freedom.

4 Estimation and Inference with Stochastic Labor Efficiency

Although attractive for its simplicity, the crucial shortcoming of the deterministic approach (DT) to the estimation of our model is that the error terms in the wage bill and the labor share equation have no clear economic interpretation, so all the sources of uncertainty in the model would remain not be identified, which makes the interpretation of the results very difficult. Sargent (1981), among others, has argued that the practice of adding error terms to equations is not a legitimate procedure. This is not an issue, however, for the stochastic specifications of labor efficiency, since the shocks enter into the exogenous laws of motion of the unobservable labor efficiency factors.

The model with stochastic labor efficiency is a relatively difficult one to estimate in that the latent variables -the (2×1) vector of labor efficiency factors H_t- are embedded within a nonlinear function. Given this complication, the model becomes a nonlinear latent variable model of the following form:

$$\begin{array}{ll} \text{ME:} & Z_t = f(X_t, H_t; \phi) + \varepsilon_t \\ \text{SE:} & \left\{ \begin{array}{l} h_t = h_0 + \gamma t + \omega_t \\ \text{or} \\ h_t = \gamma + h_{t-1} + \omega_t \end{array} \right. \end{array} \tag{4}$$

where ME indicates the 3 measurement equations, and SE stands for state equations. The function f contains the three nonlinear observational equations (1), (2) and (3) obtained as

$$\begin{aligned} \pi_t &= .95 - .008t + .75\pi_{t-1} - .95\varepsilon_{t-1} + \varepsilon_t \\ with\overline{R}^2 &= .57, DW = 2.3, \hat{\sigma}_\varepsilon = .035. \end{aligned}$$

restrictions from the model. Z_t and X_t are respectively the (3×1) and (4×1) vectors of dependent and exogenous variables described in Section 3, h_t is the (2×1) vector of the logs of the latent labor efficiency factors, ε_t and ω_t are respectively (3×1) and (2×1) vectors[8] of i.i.d. multivariate normal disturbance terms with mean zero and covariance η_ε and Ω. Define $\phi \equiv \{\delta_s, \delta_e, \alpha, \mu, \lambda, \sigma, \rho, \eta_\varepsilon\}$ as the (8×1) vector of parameters of the measurement equation, and define $\varphi \equiv \{h_0, \gamma, \Omega\}$ as the (7×1) vector of parameters of the state equations. Again, for notational simplicity call $\theta = \{\phi, \varphi\}$. Since we have only 28 observations, we chose to reduce the dimensionality of the parameter vector to be estimated using a priori information from the calibration of $\delta_s, \delta_e, \alpha$ and η_ε as described above and some additional restrictions. First we assume that the innovations to the log of skilled and unskilled labor efficiency do not covary and have the same standard deviation, hence we can rewrite $\Omega = \eta_h^2 I_2$. Second, we fix the initial level of skilled labor efficiency h_{0s} to a given scaling factor. This latter restriction is needed as our model does not have any implication on the *levels* of variables. It actually turns out that given a value for the initial quality index of skilled labor h_{0s}, then the other parameters λ, μ and h_{0u} can be estimated, but all four parameters together are not identified. The dimensionality of the parameter vector to be estimated, θ, is therefore reduced to eight.

Given the nonlinear state-space representation of our model, as in (4), there are a variety of techniques that in principle could be used to estimate the parameters of the production function. The source of the difficulties in estimating our model originates exactly from the *joint presence* of the nonlinearity and the unobservability of some key endogenous variables. In fact, the nonlinearity of the measurement equations prevents us from using the classical Kalman Filter methods to integrate out the latent variable. The Kalman Filter in this case still yields the *best linear predictor,* but the best predictor need not be linear and more in general, as Meinhold and Singpurwalla (1989) emphasize, may not be close to the linear one. The Kalman Filter estimator is therefore *nonrobust* to nonlinear perturbations of the model, which makes it particularly unsuitable for our case.

Indicating with the superscript T the vector of all sample observations, the joint pdf of our model is $\psi(Z^T, H^T | X^T; \theta)$. Since H_t is latent, we can only observe $\psi(Z^T | X^T; \theta)$, therefore in the structural estimation it is necessary to collapse the first pdf into the second in order to map the model into the data.

There are several simulation-based approaches that can be used to address this problem. A necessary condition for the implementation of all of them is that there be no feedback from the endogenous variables to the exogenous variables (or, the exogenous variables must

[8]Since ε_t enters only the third ME, the first two components of this vector are identically zero.

be *strongly exogenous*). This is an issue in our analysis because capital and labor are kept fixed throughout the simulation, while in equilibrium they will tend to respond to variations in H_t. The only way of dealing with this problem explicitly would be to extend the model to a dynamic general equilibrium setting, in which one could solve for decision rules for capital accumulation and labor supply. In this case, capital and labor could be simulated based on the draws of the error terms. Of course, this would be a much more complicated model without an analytic solution, and would also involve a higher number of parameters in the estimation. The basic setup of our environment suggests that the scope of this problem may not be very large. First, the disturbance terms are i.i.d., so that shocks today to labor efficiency or the relative price of capital equipment are not expected to persist. Second, while shocks may affect investment, which is a flow, the overall effect on the *stock* of capital will be relatively small, since on average business equipment investments are about eight percent of the capital stock in the US. Finally, if the innovation variance of the shocks is fairly small, this will tend to limit the range of values the shocks can take. We have indeed found that the innovation variance is quite small, as it is discussed in more detail in section 5.

The simulation-based methods we consider in this paper are stochastic integration of the state vector (which we contrast with a numerical integration algorithm), estimation of an auxiliary model with Indirect Inference correction, and simulated pseudo-maximum likelihood. In what follows we will outline these three methods, run some Monte Carlo experiment and, in the last part of this section, compare their performance in our environment.

4.1 Numerical and Stochastic Integration

The first way to attack the problem described above is integrating out the latent state vector to obtain the likelihood solely as a function of observable variables and parameters. This requires in principle performing an integral of dimension[9] $(T \times 2)$, precisely $\int \psi(Z^T, H^T | X^T; \theta) dH^T$. When maximizing the criterion function for the estimation, the integral has to be computed each time the hill-climbing algorithm evaluates the function in the parameter space. For this reason, finding an efficient (but sufficiently precise) way of computing multidimensional integrals is the crucial step in this approach to the estimation of nonlinear state-space models.

The two specifications (TS) and (FDS) have very different implications in terms of the best strategy to adopt, given that one wants to follow the approach of integrating out the latent state and construct the exact observable likelihood. The (TS) specification does

[9]More in general, the dimension of the integral is the number of elements in the state vector multiplied by the sample size.

not have autocorrelation in the labor efficiency factor, therefore the integration is greatly simplified as it reduces to a sequence of T two-dimensional integrals. For this latter case we have compared two integration methods: numerical quadrature and stochastic simulation.

Quadrature is a numerical method that approximates the integral on the state vector at time t as follows:

$$\int \psi(Z_t, H_t|X_t;\theta)dH_t \simeq \sum_{j=1}^{N} \psi(Z_t^j|H_t^j, X_t;\theta)\psi_\omega(H_t^j;\theta)w^j, \tag{5}$$

where the N quadrature points $\left\{H_t^j\right\}_{j=1}^N$ and their respective weights $\{w^j\}_{j=1}^N$ are deterministically preassigned in the given quadrature rule. Z_t^j is the vector of values of the dependent variables implied by the $j-th$ node of the latent state vector. Note that in our case both $\psi_\omega(H_t;\theta)$ and $\psi(Z_t^j|H_t^j, X_t;\theta)$ are normal pdf. Obviously, the approximation error in the computation of the integral will be smaller, the larger is the number of nodes N.

The second approach is through stochastic simulation of the integral. We can write:

$$\int \psi(Z_t, H_t|X_t;\theta)dH_t \simeq \frac{1}{S}\sum_{i=1}^{S} \psi(\tilde{Z}_t^i|\tilde{H}_t^i, X_t;\theta) \tag{6}$$

where the equality holds exactly as $S \to \infty$. Here $\tilde{H}_t^i$ denotes the $i-$th draw of the state from $\psi_\omega(H_t;\theta)$ and $\tilde{Z}_t^i$ the implied vector of endogenous variables obtained from $f(\cdot)$. Once we are able to compute the integral in (5) or (6), the exact likelihood function for the model can be constructed and evaluated at each point in the parameter space attained by the numerical optimization algorithm which will deliver, respectively, a quadrature based MLE (QUAD) and a simulated MLE (SMLE). Note that for the SMLE case, throughout all the maximization of the objective function the *same* set of $(T \times S)$ random numbers must be used to prevent the likelihood from becoming a random function.

It is important to indicate that the methodology of integrating out the state requires to specify some additional error terms to the measurement equations for the maximized likelihood not to be degenerate. As we discussed earlier, we are not confortable with the specification of additive arbitrary error terms to the first two equations. An alternative would be to introduce measurement error explicitly in one of the series, but this would generate a (generally) non-gaussian model with an unknown distribution that, together with the nonlinearity of the measurement equation, becomes fairly untractable. In the Monte Carlo experiment that follows we have simply attached additively a gaussian error term to each equation, so that the likelihood -once the state is integrated over- is a correctly specified gaussian density. Given the shortcoming of the specification of this model, we present the Monte Carlo results for heuristic purposes, and will not use this procedure when formally estimating the model parameters from U.S. data.

We have compared across 200 Monte Carlo trials the Gauss-Legendre quadrature method with $N = 256$ quadrature points in the two-dimensional space of the latent state vector against stochastic simulation based on $S = 10, 50, 100$. We only estimate the key parameters from the perspective of our theory $\sigma, \rho, \gamma_u, \gamma_s$ and we set to the the remaining 4 parameters to the values used in the DGP. The Appendix to the paper contains a detailed description of all the Monte Carlo experiments we have performed.

Table 3 - QUAD MLE (TS) CASE			
PERCENTAGE BIAS			
	$N = 256$		
	Mean	Median	S.D.
σ	-1.62	-1.72	1.31
ρ	1.01	1.01	1.98
γ_u	-6.43	-6.51	9.68
γ_s	.23	-.13	8.35

Comparing results from Tables 3 and 4 which summarize the empirical distribution of the percentage bias in Figure 2, it emerges that all coefficients except γ_u show very little mean and median bias. It is interesting to note how the smoothed empirical distribution in the quadrature case and in the SMLE (especially with $S = 100$) show a strikingly similar shape for all four parameters. In general, the quadrature method dominates the stochastic simulation in terms of mean bias. It is quite remarkable however, that computing the stochasting integration with only 10 draws delivers such a small estimation bias. Note that there is little improvement in going from $S = 10$ to $S = 100$ in terms of mean bias, while a non negligeable reduction in the standard deviation (S.D) of the bias is obtained. Even in terms of S.D., the quadrature method is superior. For S above 100 only very marginal improvement was noticed in the stochastic integration based maximum likelihood estimator, at the expense of a large increase in computational time. It appears that only by associating some variance reduction device to the Monte Carlo integration, such as importance sampling[10], the latter procedure could achieve the same precision as quadrature with a reasonable number of draws.

[10]See Geweke (1989) and Danielsson and Richard (1993).

Table4 – SMLE – TSCASE									
PERCENTAGE BIAS									
	$S=10$			$S=50$			$S=100$		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
σ	-2.30	-2.32	1.62	-2.32	-2.4	1.46	-2.31	-2.4	1.41
ρ	1.37	1.39	2.13	1.34	1.28	2.03	1.36	1.39	1.99
γ_u	-11.17	-10.08	12.22	-11.47	-12.19	10.98	-11.20	-11.56	10.70
γ_s	1.38	.94	9.73	1.19	.65	8.81	1.29	.95	8.76

In the FDS case the dependence in the process prevents from breaking down the $(T \times 2)$ integral into a sequence of T double integrals. It is clear that performing such a multidimensional integration requires a huge number of simulations (or quadrature points) that would tremendously slow down the underlying optimization algorithm, so alternative methods are required. An alternative technique that provides an exact solution and in principle is implementable in this particular case is the Gibbs Sampler associated with the Simulated EM (SEM) algorithm. In the EM method (see Ruud (1991) for a survey) the function to be maximized with respect to θ^{j+1}, for a given θ^j computed at the $j-th$ step of the hill-climbing algorithm is:

$$Q(\theta^{j+1};\theta^j) = \int \log \psi(Z^T|H^T, X^T; \theta^{j+1}) \psi_\omega(H^T|Z^T;\theta^j) dH^T.$$

The advantage of this method compared to a standard likelihood maximization is that the objective function, when maximized, strictly increases towards the maximum and can be very fast in reaching the neighborhood of the optimum. This is because the state is integrated out with respect to a more informative density (compared to the SMLE) which is conditional also to Z^T.

In the simulated version of the EM, the integral of the $Q(\cdot)$ function cannot be computed analytically and it is therefore simulated. For the simulation one needs to draw paths of the latent state vector from $\Psi_\omega(H^T|Z^T;\theta^j)$. These draws are more complicated as this density is a priori unknown and therefore require the implementation of Markov Chain Monte Carlo techniques. Carlin, Polson and Stoffer (1992) (hereafter CPS) describe a method which is suitable for our case based on a single move Gibbs sampler and an acceptance-rejection step.

This approach which couples the Gibbs sampler and the SEM, however, has three major drawbacks. First, the process for h_t in the (FDS) specification does not have stationary distribution, hence the Gibbs sampler would never achieve convergence. Even if one is willing to transform the (FDS) process into a very persistent stationary process in levels, the algorithm will be extremely slow to reach convergence as the CPS method is based on a single-move sampler. Second, the acceptance/rejection step embedded in the CPS procedure furtherly slows down the algorithm, especially when the parameter values are far from the true ones. Finally, it is well known that the EM algorithm is very inefficient once in the neighborhood of the maximum and to our knowledge there is no proof of convergence for its simulated counterpart[11]. Our initial experience with this method proved to be extremely slow and inefficient.

4.2 Auxiliary Model and Indirect Inference

A second method that can be used in our environment is a two-step procedure. In the first step an auxiliary model which provides an approximation to the more complex structural model is estimated. This allows one to sidestep the problem of multidimensional integration. In the second step, one can recover consistent estimates of the parameters by a standard application of Indirect Inference methods. The auxiliary model we chose to work with is the Extended Kalman Filter (EKF) (see Anderson and Moore (1979)). A first-order Taylor series approximation of the true model is made around the predicted value of the latent factors at each point in time and the standard Kalman recursion is run on the linearized model.

First, define $h_{t|t-1}$ as the one step ahead linear predictor of the (log of the) latent state vector, changing according to the specification (TS) or (FDS) that we choose. Denote also $Z_{t|t-1} = \frac{\partial f(X_t,h_t;\theta)}{\partial h_t'}\Big|_{h_t=h_{t|t-1}}$ and $f_{t|t-1} = f(X_t, h_{t|t-1};\theta)$. It follows that $Z_t = f_{t|t-1} + Z_{t|t-1}(h_t - h_{t|t-1}) + \tilde{\varepsilon}_t$. The main source of approximation is that the error term $\tilde{\varepsilon}_t$ has no longer necessarily zero mean, in general is correlated with h_t and is not normal, while in the EKF it will be treated as a gaussian white noise

[12]. The linearization above justifies the EKF recursion which runs as follows:

$$
\begin{aligned}
&P_{t|t-1} = P_{t-1|t-1} + \Omega,\\
&k_t = P_{t|t-1}Z_{t|t-1}'(Z_{t|t-1}P_{t|t-1}Z_{t|t-1}' + \Sigma)^{-1},\\
&P_{t|t} = P_{t|t-1} - k_t(Z_{t|t-1}P_{t|t-1}Z_{t|t-1}' + \Sigma)k_t',\\
&h_{t|t} = h_{t|t-1} + k_t(Z_t - f_{t|t-1}).
\end{aligned}
$$

where Σ is the covariance matrix of the errors to the measurement equations (nonzero only in the no-arbitrage condition) and $P_{t|t-1}$(respectively $P_{t|t}$) is the MSE associated with the predicted (updated) linear projection of h_t.

[11]Chang and Ledolter (1995) show that it is only possible to prove that after a certain number of iterations the SEM with high probability will get close to the maximizer of the likelihood of the observed data. Experience with the SEM seems to suggest that at some point the algorithm starts fluctuating randomly around a constant value but does not converge.

[12]No approximation is needed on the state equations, since they are linear.

In terms of initialization of the filter, $h_{1|0}$ was started through a randomization, i.e. by drawing from $N(h_0, P_{0|0})$, where $P_{0|0}$ was set to $\eta_\omega^2 I_2$. The innovation form of the likelihood function for the time t observation was constructed as

$$\ell(Z_t; X_t, \theta) = -\frac{1}{2}\ln\left|\Lambda_{t|t-1}\right| - \frac{1}{2}\left(Z_t - f_{t|t-1}\right)' \Lambda_{t|t-1}^{-1}\left(Z_t - f_{t|t-1}\right),$$

where $\Lambda_{t|t-1} = Z_{t|t-1}P_{t|t-1}Z_{t|t-1} + \Sigma$ and it was maximized over θ.

We expect that this approximation will generate biased estimates, as the auxiliary linearized model is misspecified except for a neighborhood of the conditional mean of the states. The bias we will obtain is in principle a combination of the small-sample bias, and the misspecification bias, but the two components cannot be distinguished. In general, for a given sample size, the error is higher the larger is the difference between the first and second moments of the structural model and their respective approximations $f_{t|t-1}$and $\Lambda_{t|t-1}$.

To evaluate the magnitude of this bias, we performed a Monte Carlo experiment on both the (TS) and the (FDS) specifications. Table 5 collects the main results about mean, median percentage bias and its empirical S.D. and Figures 3 and 4 depict the smoothed empirical distributions of the percentage bias of the parameter estimates. Here we distinguished two cases: the first with low S.D. and the second with high S.D. (larger by a factor of 2.5) in the unobservable states.

Table 5 - Extended Kalman Filter

	PERCENTAGE BIAS								
	(TS) CASE						(FDS) CASE		
	Low Variance			High Variance			Low Variance		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
σ	.10	.12	1.08	.34	.20	2.52	-.19	-.30	.89
ρ	-.26	-.50	1.94	-.59	-.60	3.65	-.06	.31	4.74
γ_u	-.07	-.35	7.42	1.80	1.47	18.61	28.13	-1.29	114.31
γ_s	-.65	-1.01	8.44	1.31	2.30	17.30	-3.59	14.33	135.36

Unexpectedly, in the (TS) specification for all the parameter estimates the mean bias is always below 2%. The empirical distribution of the bias is somewhat less concentrated for the growth rate parameters, indicating the danger of a potential high bias. The mean bias is larger for the specification with high variance, but still extremely small (both estimates of σ and ρ exhibit less than .6% bias on average), although the S.D. doubles as it is evident from Figure 3. This result is very surprising given that the model is quite nonlinear and the EKF it is just a linear approximation around the predicted value of the state. The reason for this result might reflect the fact that in the presence of the deterministic trend which allows forecasting the state very accurately, so the EKF is indeed a good approximation.

In the (FDS) case there is a marked difference between the mean bias of the substitution parameters, in line with that of the (TS) model, and the bias of the drift parameters which is more than 30 times higher on average. The direction of the median bias for all four parameters is the same as in the (TS) case. For the drift terms, a huge S.D. of the bias was recorded which indicates that estimates with very large distortions are probable. Also, as clear from Figure 4, the empirical densities, especially for the drift parameters, show large skewness.

Although the bias we found in the (TS) specification and for the σ and ρ parameters in the (FDS) model is extremely small, the EKF is, in principle, inconsistent. As a result, it seems appropriate to implement an Indirect Inference correction. Denote by $\hat{\theta}_T$ the EKF inconsistent estimator lying in the parameter space Θ and denote the "true" parameter vector β_0 lying in B. We can define a bias-function $b : B \to \Theta$ which maps "true values" β_0 into biased estimates $\hat{\theta}_T$. This function is obtained by simulating the endogenous variables $\left\{Z^{Ti}(\beta)\right\}_{i=1}^{M}$ from the model for any given parameterization β in B and re-estimating a vector $\hat{\theta}_{MT}(\beta)$, where M is the number of simulations used for the estimation. The premise of the Indirect Inference correction is exactly to use these simulations in order to correct the initial EKF estimate $\hat{\theta}_T$. The Indirect Inference estimator of β_0 is simply:

$$\hat{\beta}_{MT}(\Lambda) = \arg\min_{\beta \in B} \left(\hat{\theta}_{MT}(\beta) - \hat{\theta}_T\right)' \Lambda \left(\hat{\theta}_{MT}(\beta) - \hat{\theta}_T\right)$$

where Λ is a symmetric negative-definite matrix, defining a metric. The Indirect Inference algorithm therefore searches for that value of β which minimizes the distance, taken in the $\Lambda-$metric, between the estimate of the auxiliary model when data are generated under that $\beta-$ parametrization of the model and the original estimate $\hat{\theta}_T$. Gouriéroux, Monfort and Renault (1993) and Gouriéroux and Monfort (1995) develop the asymptotic theory of the Indirect Inference estimator and derive the form for the optimal weighting matrix Λ. Gouriéroux, Renault and Touzi (1995) investigate its small sample properties through Edgeworth expansions. Since the asymptotic theory and the small-sample characterization exist only for stationary cases, we believe it is useful to perform some Monte Carlo experiments to assess the properties of the Indirect Inference correction in the type of nonstationary environment we consider.

Figure 5 shows the results of a series of 50 Monte Carlo trials were the model in the (TS) specification was first estimated by EKF and then by Indirect Inference with EKF as auxiliary model. The Indirect Inference estimator was based on $M = 15$ and $S = 50$. The weighting matrix was set to the identity matrix, but as the numbers of parameters to be estimated is the same in the auxiliary and structural model, this is the optimal choice. Perhaps not surprisingly, since the bias of the EKF is so small, the Indirect Inference correction is often indistinguishable from its first step counterpart and for some cases it even tends to worsen the initial estimate, but this may be due to the low number of trials that were performed.

Table 6 - EKF and INDIRECT INFERENCE						
PERCENTAGE BIAS						
	Mean		Median		S.D.	
	EKF	INDINF	EKF	INDINF	EKF	INDINF
σ	-6.84	-6.04	-6.58	-6.02	1.49	2.08
ρ	-3.19	-3.73	-4.75	-4.92	6.35	5.94
γ_u	-19.70	-.16	-12.25	1.25	41.29	21.57
γ_s	-1.37	-10.51	-2.66	-9.66	11.22	12.59

In the (FDS) case (see Table 6 above), although the mean bias of the EKF is larger than in the (TS) case, the Indirect Inference is not always effective. Only for the drift term of unskilled labor it provides a non-negligeable adjustment of the estimate towards the true value. These results are based on 25 Monte Carlo trials, but we think that even once the simulation uncertainty is substantially reduced, the pattern of our findings would not change dramatically.

4.3 Simulated Pseudo-MLE

An alternative method of bypassing the multidimensional integration problem is the use of simulation techniques jointly with a Pseudo-Maximum Likelihood Estimation (PMLE). The theory of PMLE was developed by White (1982) and Gouriéroux, Monfort and Trognon (1984). The extension to the simulated case is due to Laroque and Salanié (1989, 1993, 1994). The main premise of the PMLE method is that one can approximate the unknown and untractable distribution on which the "observable" likelihood is based with a simpler, but possibly misspecified, objective function constructed only from some empirical moments of the dependent variables. When these moments cannot be obtained analytically, under fairly general conditions they can be simulated. The simulated version of the PMLE or Simulated PMLE (SPMLE) reduces the computational burden from one $(T \times 2)-$dimensional integral to T different $2-$dimensional integrals, since the moments of the dependent variables are simulated at each period t to construct the likelihood. In particular, given our distributional assumptions on the error terms, we generate S realizations of the dependent variables each indexed by i, by following these 2 steps:

$$\text{step 1:} \quad \begin{cases} h_t^i = h_0 + \gamma t + \omega_t^i \\ \text{or} \\ h_t^i = \gamma + h_{t-1}^i + \omega_t^i \end{cases}$$

$$\text{step 2:} \quad Z_t^i = f(X_t, H_t^i, \varepsilon_t^i; \phi)$$

In step 1 a realization of ω_t is drawn from its distribution and used to construct a time t value for h_t. In step 2 this realization of h_t, together with a draw of ε_t allows to generate a realization of Z_t. As mentioned for the SML estimator, the set of $(T \times S)$ draws must be kept fixed during the optimization of the objective function. We chose the SPMLE2, hence based on the first and second simulated moments of the unknown distribution, respectively:

$$m_S(X_t; \theta) = \tfrac{1}{S} \textstyle\sum_{i=1}^S f(X_t, H_t^i(\varphi), \varepsilon_t^i; \phi),$$

$$V_S(X_t; \theta) = \tfrac{1}{S-1} \textstyle\sum_{i=1}^S \left(Z_t^i - f(X_t^i, h_t^i(\varphi), \varepsilon_t^i; \phi)\right)\left(Z_t^i - f(X_t^i, h_t^i(\varphi), \varepsilon_t^i; \phi)\right)'.$$

The simulated pseudo likelihood as a function of $\theta \equiv \{\phi, \varphi\}$ becomes

$$\begin{aligned} \ell_S(Z^T; X^T, \theta) = \; & \tfrac{1}{2T} \sum_{t=1}^{T} \left\{ [Z_t - m_S(X_t;\theta)]' \, V_S(X_t;\theta)^{-1} \, [Z_t - m_S(X_t;\theta)] \right. \\ & \left. + \log\det\left(V_S(X_t;\theta)\right) \right\}, \end{aligned} \tag{7}$$

and the SPML2 estimator $\hat{\theta}_{ST}$ is defined as the maximizer of the function in (7).

The SPMLE has in principle 3 sources of approximation to the correctly specified model. The first one comes from the potential misspecification in the objective function. Instead of the untractable exact likelihood, we use a simpler loss function. The second source of approximation originates from the fact that the true moments are replaced by the simulated ones, inducing an error due to the finiteness of the number of simulations S. In addition to finite number of simulations bias and misspecification bias, there is a third source of potential bias, which is finite sample bias.

Laroque and Salanié (1994) proved that in a stationary environment, as $S \to \infty$, $T \to \infty$ and $\frac{\sqrt{T}}{S} \to o$, the SPLME has the same asympotic properties of the PMLE, hence the SPMLE estimator converges to the pseudo-true value and it is asymptotically nomal. Gouriéroux, Monfort and Trognon (1984) proved that if the pseudo-distribution belongs to the quadratic exponential family, as in our specification, then there is no misspecification bias and the PMLE is consistent and asymptotically normal. This gain in the robustness of the estimator comes at the cost of a loss in efficiency compared to the MLE. As far as the simulation bias is concerned, Laroque and Salanié (1994) prove that this bias has the order of $\frac{1}{S}$ and according to their Monte-Carlo experiments, for $S = 20$ the bias is already quite small.

The theory underlying the use of SPMLE is asymptotic and for stationary environments, but in our analysis, the regressors and the unobservable variables have trends and the sample size is small. Therefore, to evaluate the performance of SPMLE in our setting, we conducted a Monte Carlo analysis. In the first specification the latent variables were trend stationary (TS). In the second specification they were random walks with drift (FDS).

The results of the Monte Carlo experiment are presented in Figures 6-7 and Tables 7-8. These figures show the empirical distributions of the percentage bias of the key coefficients for $S = 10$, $S = 20$ and $S = 50$ where S is the number of simulations used to generate the empirical moments in the simulated pseudo-likelihood in (7).

For the (TS) case, Figure 6 and Table 7 show that in general there is very little mean and median bias in the estimation even for $S = 10$, which is a very encouraging result. For $S = 50$ the mean bias is essentially null. The bias of the substitution parameters σ and ρ is never larger than 6% and averages respectively .06% and -.16%. Although still roughly symmetric in a neighborhood of zero, the distribution of the bias in the estimation of the trend growth rate coefficients for the labor efficiency factors is more spread with substantial mass for values between 10% and 20%.

Table 7 - SPMLE - (TS) CASE

PERCENTAGE BIAS									
	$S = 10$			$S = 20$			$S = 50$		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
σ	.20	.09	1.05	.11	.09	.93	.06	.03	.92
ρ	-.33	-.20	1.54	-.22	-.34	1.26	-.16	-.17	1.1
γ_u	.85	1.21	7.45	.27	.09	6.61	-.16	-.38	7.19
γ_s	-1.09	-.70	7.64	-1.01	-1.1	7.18	-.73	-.65	5.75

The performance of SPMLE2 is more disappointing when the latent states are integrated, however. This evidence is summarized in Table 8. Although the mean bias for the substitution parameters is still very small, it increases enormously on the estimates of the drift terms. Another surprising feature of these estimates is that as S increases, the mean bias does not seem to be reduced, rather it seems to be larger. The same surprising pattern emerges for the S.D. of the drift parameters. As shown in Figure 7, the empirical distribution of the estimators are remarkably skewed which involves the danger of a severe bias in the estimates. Recall that the same type of skewness (although less pronounced) was found in the EKF estimator, hence skewness seems to be a feature of the small sample distribution of the estimators in the (FDS) case.

Table 8 - SPMLE - (FDS) CASE									
PERCENTAGE BIAS									
	$S = 10$			$S = 20$			$S = 50$		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
σ	-.41	-.65	2.90	-.61	-.705	2.48	.89	-1.01	2.23
ρ	.17	.54	3.58	.52	.786	3.17	.77	1.14	2.79
γ_u	77.6	27.87	168.04	102.72	68.35	199.54	190.49	153.58	223.22
γ_s	-214.35	-144.87	251.40	-224.02	-125.05	285.85	-295.48	188.47	291.02

One could in principle improve on the SPMLE2 by running an Indirect Inference estimation with SPMLE2 as a first step. This procedure, given the consistency of the first-step estimator would provide a correction for the potential small-sample bias, as pointed out in Gouriéroux and Renault and Touzi (1995). Given the negligeable bias of the SPMLE2 in the (TS) specification, and the unsatisfactory performance of Indirect Inference in the (FDS) case obtained for the EKF, we will not pursue this correction in either case.

4.4 A Comparison of the Methods

In order to make a meaningful comparison of the perfomance of the different methods we outlined in this section, it is necessary to assess also their computational efficiency. We have standardized the speed by setting to 1 the time needed for the fastest method, the SMLE with $S = 10$, which takes about .15 minutes for one complete estimation.

Table 9 - Numerical Efficiency - (TS) case						
QUAD	SMLE		EKF	IND. INF.	SPMLE	
$N = 256$	$S = 10$	$S = 100$		$S = 50$, $M = 15$	$S = 10$	$S = 50$
22	1	9.5	1.5	290	7	38

The numbers above suggest that for the simulation-based estimators evaluated in our paper, time increases approximately linearly with the number of simulations. As our results did not suggest the existence of huge deficiencies when S is low compared to a high S, an efficient strategy in estimation seems to be starting to explore the parameter space with a small S, and only when sufficiently good initial conditions are found increase S so to reduce the simulation uncertainty.

Note also that the EKF method is obviously very fast, as for every iteration of the hill-climbing algoritm it has just to compute the derivative of the measurement equations at the value of the conditional mean for the state and then run the Kalman Filter recursion to construct the likelihood. Unfortunately, since it is inconsistent it should always be coupled with Indirect Inference, which is on the contrary extremely slow. This reduces substantially the speed advantage of EKF.

Taking into account both bias and computational efficiency, our Monte Carlo analysis suggests the following observations:

- In the (FDS) specification, the EKF performs better than SPMLE, since for the latter we found a huge bias in the drift term of the labor efficiency factors and higher variability of estimates. Moreover the SPMLE surprisingly shows higher bias when S is larger.

- The performance of the estimators in terms of mean and median bias seems largely superior in the (TS) case compared to the (FDS) model. This is especially true for the growth rate parameters, where the high mean biases we recorded potentially put a heavy burden on working with this specification. For all parameters, the different methods -when operated in the (FDS) environment- generate estimates with much higher variability.

- All methods in the (TS) specification show very little mean and median bias. The S.D. of the estimates is in general larger for the growth rate parameters γ_u and γ_s.

- The quadrature based MLE is superior to the simulated MLE when no variance reduction technique is associated to the latter.

- In the (TS) case both EKF and SPMLE show less bias and less variability in the estimates than the numerical and stochastic integration methods. This holds true particularly for estimates of γ_u. Moreover, the integration procedures depends crucially on the specification of additive error terms to the measurement equations which, as we argued already, are of difficult interpretation.

- Indirect Inference does not make any improvement upon the inconsistent first-step EKF estimator in the (TS) case, most likely because of the very little bias of the EKF, but even in the (FDS) case, where the bias of the first step estimator is larger, Indirect Inference is not effective.

- In the (TS) case, SPMLE is comparable to EKF, in that the small sample bias is practically absent, but it is superior as it has 2 advantages over the EKF. First, by increasing S the bias and the S.D. (the latter already smaller than in the EKF) of the SPMLE can be further reduced. Second, although SPMLE is slower for S large enough to give reliable results, it is a consistent estimator and does not need any correction by Indirect inference which, in our experience, dramatically slows down the whole estimation.

Based on these findings, we conjecture that the negligible small-sample bias in the (TS) case may reflect an analog to "super-consistency" of the estimators of coefficients associated with regressors which have a deterministic trend. When we simulate the labor efficiency vector to generate the moments needed in the pseudo-likelihood and perform the estimation, we introduce a deterministic trend nonlinearly in the measurement equation. As all the parameters of interest interact in some way with the latent states, the speed of convergence of the estimates may benefit from the trend.

Second, it is possible that when latent states are modeled as integrated processes, the population objective function based on simulated estimators (as SPMLE and Indirect Inference) become a random function, generating inconsistent estimates. This would explain the very large variance that we found for some of the parameter estimates and the surprising finding that the bias does not decline and sometimes increases with higher number of simulations.[13] The higher reliability of the EKF estimator in the (FDS) specification is probably due to that its objective function is based on the conditional first and second moments of the state which, even in the I(1) specification exist and they are finite. On the other hand, the simulated pseudo likelihood function is based on moments of the endogenous variables constructed from simulated paths of the latent vector, hence they are affected by the unconditional moments of the latter, which are not finite.

To conclude, in the evidence we have accumulated the specification with trend-stationary labor efficiency seems much more reliable and, for this specification, estimation by SPMLE2 is accurate and relatively fast. In section 5 we will present the results of the SPMLE2 estimation of our model for the (TS) case.

5 Results of the Estimation

Based on our analysis described earlier, we have estimated the model described by the (TS) state equations and the measurement equations in (1), (2) and (3) by SPMLE2. Recall that there are eight parameters to estimate: σ defines the elasticity of substitution between unskilled labor and equipment $\frac{1}{1-\sigma}$, ρ defines that of skilled labor with equipment $\frac{1}{1-\rho}$, γ_u and γ_s are the growth rates of labor efficiency, λ and μ are the weights in the CES nestings

[13] This point was also made in an example by Laroque and Salanié (1993).

of the production function, h_{0u} is the initial value for unskilled labor efficiency and η_ω^2 is the variance of both efficiency factors. In what follows we will discuss only the estimates of the two elasticity parameters, the two growth rate of labor efficiency and their variance, which are all the parameters of economic interest. The model has been estimated from 1963 to 1990, using the time series on labor input, wages, capital and relative prices described in the section on data construction.

An important choice in the analysis is the number of simulations, S. The estimation algorithm was run with $S = 500$. We chose this number, since at this point changes in the parameter vector due to simulation uncertainty were negligeable.

The main difficulty encountered in the estimation procedure was to pin down the variance of the latent state. In fact, when η_ω was estimated together with the other parameters, the estimation algorithm would always tend to increase quickly the estimated variance generating an implausibly high variability in the implied simulated path of labor efficiency. Moreover, due to this high variance in the latent states, the growth rate coefficients were extremely difficult to estimate precisely. Therefore, we splitted the estimation in two steps. We fixed a very fine grid of values over a realistic[14] range for η_ω and for each of these points, we estimated the rest of the parameter vector. The maximum of the pseudo likelihood function across these grid points delivered the final estimates.

When dealing with such a large parameter space and with simulation-based estimation techniques there is uncertainty on whether the maximum found be local or global. To assess the global nature of the optimum, we used Veall's (1990) test based on the asymptotic confidence interval for the first order statistics computed by De Haan (1981). Call $\hat{\ell}$ the value of the objective function at the optimum, ℓ^1 and ℓ^2 the highest two values of a vector of M random evaluations of the likelihood in the parameter space. We reject the null hypothesis that the optimum is not a global maximum at $(1-p)$ level of confidence if $\ell^p > \hat{\ell}$, where $\ell^p = \ell^1 + (\ell^1 - \ell^2) / \left(p^{\frac{-2}{k}} - 1\right)$ and k is the dimension of the parameter space. We performed $M = 60,000$ random evaluations of the likelihood in our 8−dimensional parameter space.[15] We found that we could reject the null hypothesys only at a level of $p = .185$, therefore at all conventional confidence levels, the hypothesis is not rejected.

In addition, we used the information collected through these random evaluations in a subsequent analysis by identifying points far from our estimate where the likelihood was

[14]We simulated the model with different values of the variance of the efficiency factors and set a range for which the model generated data showed approximately the same variance as in our sample.

[15]It is not clear what should be the optimal ratio M/k. Veall finds that in cases with 3 parameters 10 random draws would deliver satisfactory results in terms of test power, but other examples with only 1 parameter needed at least 200 draws.

relatively high and re-started the maximization routine at those new initial guesses, but only lower local maxima were found. Moreover these alternative extrema predict almost perfect substitution between either of the two labor inputs and equipment, which is difficult to accept from the viewpoint of economic theory. This further evidence confirms the global nature of our estimate.

Table 10 shows the SPMLE2 estimates and standard deviations and also reports the Indirect Inference correction for small-sample bias for the parameters which are economically most interesting $(\sigma, \rho, \gamma_u, \gamma_s)$. Standard deviations of the SPMLE2 estimator were computed on the basis of the asymptotic covariance matrix $\frac{1}{T}\left(J_0^{-1} I_0 J_0^{-1}\right)$ where:

$$\begin{aligned} I_0 = &\quad E_0 \left\{ \frac{\partial \ell(Z_t;X_t,\theta_0)}{\partial \theta} \frac{\partial \ell(Z_t;X_t,\theta_0)}{\partial \theta'} \right\}, \\ J_0 = &\quad E_0 \left\{ -\frac{\partial^2 \ell(Z_t;X_t,\theta_0)}{\partial \theta \partial \theta'} \right\}. \end{aligned}$$

E_0 denotes the expectation with respect to Z_t and X_t taken under θ_0. The sample counterparts of I_0 and J_0 are :

$$\begin{aligned} \hat{I}_{ST} &= \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{\partial \ell(Z_t;X_t,\hat{\theta}_{ST})}{\partial \theta} \frac{\partial \ell(Z_t;X_t,\hat{\theta}_{ST})}{\partial \theta'} \right\} \text{ and} \\ \hat{J}_{ST} &= \frac{1}{T} \sum_{t=1}^{T} \left\{ -\frac{\partial^2 \ell(Z_t;X_t,\hat{\theta}_{ST})}{\partial \theta \partial \theta'} \right\}. \end{aligned} \tag{8}$$

In the last row of Table 8 the implied net average rate of return on capital r in the economy is reported. This latter number was computed as the sample mean of the average, for each year, between the model prediction of rates of return on structures and on equipment, weighted by their respective share of total capital in efficiency units. We found r to be about 7.2%, which is very close to the 6.9% calibrated in Cooley (1995).

Table10 – EstimationResults		
Parameter	SPML	I I
σ	.486	.487
	(.048)	-
ρ	-1.172	-1.179
	(.110)	-
γ_u	-.041	-.040
	(.016)	-
γ_s	.033	.032
	(.003)	-
η_ω	.086	-
	(.013)	-
r	.072	-

It is striking that the Indirect Inference estimates[16] are extremely close to the initial SPLE2 estimates, which provided evidence for our conjecture based on the Monte Carlo experiments in section 4.3. These results are consistent with our theory, in that they indicate much stronger complementarity between skilled labor and equipment that between unskilled labor and equipment. The elasticity of substitution in the latter two inputs is more than 4 times higher. These estimates are also broadly in line with the existing literature that is surveyed in Hamermesh (1993).

The growth rate of unskilled labor efficiency is estimated to be negative but its statistical significance is low, while the efficiency of skilled labor grew at an average rate of 3.2%, according to the estimates. We need to note that as the unobservable states and some of the exogenous variables have trends, the convergence rate of the SPMLE estimator might be higher than $\sqrt{T}$ at least for some of the parameters. The evidence from the Monte Carlo experiment points in this direction, as the empirical standard deviations were found to be smaller than the asymptotic standard deviations reported in Table 10.

In any case a negative value for γ_u should not be surprising. Recall that in our classification, unskilled individuals are those who did not receive a high-school diploma. In the last 25 years the average level of education has increased dramatically and therefore it is not unreasonable to presume that as a matter of composition, the quality of workers without high-school degree has worsened over time.

[16]In the Indirect Inference estimation we set $S = 200$ and $M = 10$. The matrix Λ is simply the identity matrix.

Figure 8 reports the fit of the 3 equations used in the estimation and the implied wage premium vs. the data. The rate of return differentials are quite small for the whole sample, although their variability tends to increase slightly over time. The labor share in the model has the same long-term movements as those in the data, decreasing in the 1970's and increasing after 1984. The model does not capture the increase in the first 5 years of the sample, but looking at the scale, the discrepancy between model and data is very small. Perhaps the most striking result is that the predicted wage premium (which is not estimated as a dependent variable, but generated from the estimated parameter values) matches the data extremely well, tracking very closely all three swings in the data: the increase of the 1960's, the decline in the 1970's and the sharp acceleration in the 1980's.

6 Summary and Conclusions

In this paper, we explored in detail a number of issues associated with the estimation of a latent variable, nonlinear aggregate production function, and used the estimated parameters to evaluate the extent to which different substitution possibilities between skilled labor and capital, and unskilled labor and capital, in conjunction with recent declines in the relative price of equipment, could account for variations in the wage gap between skilled and unskilled workers over the last 30 years. Our model involved a number of complications from the perspective of identification and estimation, in particular, a small sample, trending regressors, and nonstationary latent variables that entered the function nonlinearly. We examined three types of estimation procedures for this problem: (i) numerical and stochastic integration, (ii) the extended Kalman Filter with Indirect Inference correction, and (iii) simulated pseudo-maximum likelihood.

Our Monte Carlo evidence clearly shows that the simulation based estimation procedures that we have compared perform much better when latent variables are trend stationary rather than integrated. Under the (FDS) specification, SPMLE was found to have very large bias for the drift parameters and the bias did not tend to reduce with the number of simulations. Indirect Inference, even when in the (FDS) case the first step estimator showed substantial bias, did not prove useful in adjusting the estimates towards their true value. The Extended Kalman Filter on the contrary showed good properties even when the unobservable variables were modeled as first difference stationary processes. In the trend stationary specification, all methods showed very little bias, but the SPMLE based on the first two moments was judged to be superior to numerical and stochastic integration and to the EKF. Finally, the estimated (by SPMLE2) parameters of the production function are consistent with our theoretical conjecture on the role of capital-skill complementarity coupled with the extensive growth in equipment in the rising wage premium.

Given that not much work has been done in evaluating the performance of these types of simulation-based estimation procedures in non-standard environments, these results are of particular interest in that they indicate that these methods can be useful in estimating models for certain types of nonstationarity and in small samples but also they point at the potential dangers in using the same methods with other types of nonstationarities. To determine whether our findings generalize to other types of models and environments, additional work along these lines will prove fruitful in providing researchers with new tools to attack difficult econometric problems.

A Appendix

This appendix provides a complete description of the design of our Monte Carlo experiments.

The DGP for the (TS) and (FDS) specifications are summarized in the following Table:

Parameters for DGP in Monte Carlo Experiments									
Model	σ	ρ	γ_u	γ_s	h_{0u}	h_{0s}	μ	λ	η_ω
TS	.50	-1.15	-.04	.03	2.1	6.0	.77	.012	.086
FDS	.50	-.40	-.04	.03	2.1	6.0	.60	.06	.086

The other parameters were fixed at their calibrated values (see Table 1 in the text). In the high variance case of the EKF, η_ω was set to .22. In the QUAD and SMLE procedures, a measurement error was added to the wage bill and to the labor share equation with same standard error as the disturbance in the no-arbitrage equation, i.e. $\eta_\varepsilon = .03$.

For every MC trial, $T = 30$ observations of the endogenous variables were generated. Only four parameters were estimated, $\sigma, \rho, \gamma_u, \gamma_s$ while all the other parameters were fixed at their value used in the DGP. The number of replications for each experiment was set to 200. The 200 sample paths of data generated from the (TS) or the (FDS) model were the same across all estimation methods, to minimize simulation uncertainty. The experiments involving Indirect Inference were based on 50 trials in the (TS) case and on 25 in the (FDS) case.

The starting values for the estimates of the parameter vector were drawn randomly at each MC trial in the admissible region of the parameter space. From the point of view of economic theory, both σ and ρ are constrained to be less than 1, but γ_u and γ_s are in principle unconstrained, although very high positive or negative values are economically implausible. Moreover, high negative values for these growth rates can generate a negative labor share or wage bill ratio, which are clearly meaningless. For this reason, we constrained the absolute value of the growth rates of labor efficiency to be less than 25 percent per year in the initial draws.

Call θ_0 the parameter set of the DGP and $\hat{\theta}$ the estimates with any of the methods above, then the percentage bias b is computed as $b = \frac{\hat{\theta}-\theta_0}{|\theta_0|} \times 100$.

For the kernel estimates of the empirical distributions of the percentage bias in parameter estimates, we used a Gaussian kernel with smoothing parameter fixed according to Silverman's "rule of thumb" (see Silverman (1986) pp. 45).

The codes for QUAD MLE, SMLE, EKF and SPMLE were written in Gauss 3.1 with extensive use of the OPTMUM library and were run on a Dell XPS P100, 16Mb RAM. The convergence criteria was always set to 10e-4. The code for Indirect Inference was written in Fortran (and made use of the IMSL) optimally compiled with IBM AIX XL Fortran Compiler/6000 and run on a IBM SPARC 10.

References

Anderson, B.D.O., and Moore J.B., *Optimal Filtering*, Englewood-Cliffs, N.J., Prentice-Hall, 1979.

Berndt, E., Griliches, Z., and Rappaport, N., "Econometric Estimates of Price Indexes for Personal Computers in the 1990's," *Journal of Econometrics*, 68, 1995, 243-268.

Bound J., and Johnson, G., "Change in the Structure of Wages in the 80's: An Evaluation of Alternative Explanations," *American Economic Review*, 78(2), 1992, 124-128.

Brown, K., and Greenstein, S., "How Much Better is Bigger, Faster & Cheaper? Buyer Benefits from Innovation in Mainframe Computers in the 1980's," NBER working paper N. 5138, May 1995.

Carlin, B., Polson N., and Stoffer, D., "A Monte-Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493-500, 1992.

Chan, K.S., and Ledolter, J., "Monte Carlo Estimates for Time Series Involving Counts," *Journal of the American Statistical Association*, 90, 242-251, 1995.

Cole, R., Chen, Y., Barquin-Stolleman, J., Dulberger, E., Helvacian, N., and Hodge, J., "Quality-adjusted Price Indexes for Computer Processors and Selected Peripherals Equipment," *Survey of Current Business*, January 1986.

Cooley, T., editor, *Frontiers of Business Cycle Research*, Princeton, N.J., Princeton University Press, 1995.

Danielsson J., and Richard, J. F., "Accelerated Gaussian Importance Sampling," *Journal of Applied Econometrics*, 8, S153-S173, 1993.

De Haan, L., "Estimation of the Minimum of a Function Using Order Statistics," *Journal of the American Statistical Association*, 76, 467-469, 1981.

Geweke, J., "Bayesian Inference in Econometrics Using Monte-Carlo Integration," *Econometrica*, 57, 1317-1340, 1989.

Gordon, R. *The Measurement of Durable Goods Prices*, Chicago, IL, University of Chicago Press, 1990.

Gouriéroux, C., and Monfort, A., "Simulation Based Econometric Methods," manuscript, 1994.

Gouriéroux, C., Monfort, A., and Renault, E., "Indirect Inference," *Journal of Applied Econometrics*, 8, 85-118, 1993.

Gouriéroux, C., and Monfort, A., Trognon, A., "Pseudo-Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681-700, 1984.

Gouriéroux, C., Renault, E., and Touzi, N., "Calibration by Simulation for Small Sample Bias Correction," manuscript 1995.

Greenwood, J., Hercowitz, Z., and Krusell, P., "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, forthcoming, 1995.

Hamermesh, D.S, *Labor Demand*, Princeton, N.J., Princeton University Press, 1993.

Juhn, C., Murphy, K.M., and Pierce, B., "Wage Inequality and the Rise in the Returns to Skill," *Journal of Political Economy*, 101, 1993, 410-442.

Kitagawa, G., "Non-Gaussian State-Space Modeling of Nonstationary Time-Series," (with discussion), *Journal of the American Statistical Association*, 82, 1032-1063, 1987.

Krusell P., Ohanian L., Rios-Rull J., and Violante, G. "Capital-Skilled Labor Complementarity and Inequality", manuscript, University of Minnesota, 1995.

Laroque, G., and Salanié, B.,"Estimation of Multi-Market Fix-Price Models: An Application of Pseudo Maximum Likelihood Methods," *Econometrica*, 57, 831-860, 1989.

Laroque, G., and Salanié, B.,"Simulation Based Estimation of Models with Lagged Dependent Variables," *Journal of Applied Econometrics*, 8, 119-133, 1993.

Laroque, G., and Salanié, B.,"Estimating the Canonical Disequilibrium Model: Asymptotic Theory and Finite Sample Properties," *Journal of Econometrics*, 62, 165-210, 1994.

Meinhold, R.J., and Singpurwalla, N.D., "Robustification of Kalman Filter Models," *Journal of the American Statistical Association*, 84, 479-486, 1989.

Ruud, P., "Extensions of Estimation Methods Using the EM Algorithm," *Journal of Econometrics* 49, 305-341.

Kim, S., and Shephard, N., "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," manuscript, 1994.

Sargent, T. "Interpreting Economic Time Series," *Journal of Political Economy*, 1981.

Silverman, B.W., *Density Estimation*, London UK, Chapman & Hall, 1986.

Veall, M. R., "Testing for a Global Maximum in an Econometric Context," *Econometrica*, 58, 1459-1465, 1990.

White, H., "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-28, 1982.