Production Function Estimation with Imperfect Proxies*

Mert Demirer
MIT

PRELIMINARY DRAFT

November 22, 2019

Abstract

The 'proxy variable' approach is often used to estimate production functions. This approach is not robust to measurement error, and it relies on some strong assumptions, including strict monotonicity, scalar productivity, and timing. In this paper, I develop partial identification results that are robust to deviations from these assumptions and measurement errors in inputs. In particular, my model (i) allows for multi-dimensional unobserved heterogeneity, (ii) relaxes strict monotonicity to weak monotonicity, (iii) accommodates . I show that under these assumptions production function parameters are partially identified by an 'imperfect proxy' variable via moment inequalities. Using these moment inequalities, I derive bounds on the parameters and propose an estimator. An empirical application is presented to quantify the informativeness of the identified set.

*MIT Economics. Email: mdemirer@mit.edu. This paper has benefited from feedback and suggestions from Anna Mikusheva, Nikhil Agarwal, Kirill Evdokimov, Jerry Hausman, Whitney Newey, Alex Olssen, Kamil Yilmaz, Nathan Zorzi as well as seminar participants at the MIT Econometrics lunch.

1 Introduction

Production functions are a critical input in many economic studies. These studies typically require estimating a production function using firm-level data. A major challenge in production function estimation is the endogeneity of inputs. Firms observe their productivity before choosing production inputs; however, productivity is unobservable to the researcher. This results in inputs being correlated with productivity, a standard endogeneity problem.

A commonly used method to address the endogeneity problem is the proxy variable approach. Introduced by Olley and Pakes (1996) (henceforth OP), this approach relies on using a variable, which is called a proxy, to control for unobserved productivity. OP use investment as a proxy variable, which is assumed to be a strictly increasing function of productivity conditional on capital. By inverting this unknown function, they essentially recover the productivity shock, and control for it in the estimation. The proxy variable approach has become the workhorse for estimating production functions and has been extended by several papers. Levinsohn and Petrin (2003) (LP) have proposed using materials as a proxy, and Ackerberg et al. (2015) (ACF) have introduced a unified framework of proxy variable approach that deals with some practical concerns.

A limitation of proxy variable approach is that it relies on strong assumptions, such as single-dimensional unobserved heterogeneity and strict monotonicity. These assumptions have important economic implications, as observed by others (Ackerberg et al. (2007), Ackerberg et al. (2015)). First, firms are differentiated only by a single productivity shock, which restricts firm-level heterogeneity. Second, there is no heterogeneity in adjustment costs and investment prices, as investment depends only on productivity. Third, estimation requires restricting competition in the output market. Moreover, the proxy variable approach is not robust to measurement errors in inputs, an important concern, especially for capital.

In this paper, I develop a partial identification approach that is robust to some deviations from proxy variable assumptions and measurement errors in inputs. In particular, my model (i) allows for multi-dimensional unobserved heterogeneity, (ii) relaxes strict monotonicity to weak monotonicity, (iii) accommodates a more general timing assumption, and (iv) is robust to measurement errors in all inputs. With these changes, the standard proxy variable becomes an 'imperfect proxy,' which can be used to derive moment inequalities for identification. Using these moment inequalities, I characterize the identified set for the parameters and propose an estimator.

An 'imperfect proxy' variable contains information about productivity, but it cannot be directly used to control for productivity in estimation. Instead, an imperfect proxy gives a stochastic ordering of productivity distributions, which can be used for identification. To show this result, I first group firms into 'high' investment firms, firms that invest more than a cutoff value, and 'low' investment firms, firms that invest less than the cutoff value. Then, I show that the productivity distribution of 'high' investment firms first-order stochastically dominates the productivity distribution of 'low' investment firms. The main idea for identification is to use this stochastic ordering in the form of moment inequalities to obtain bounds for production function parameters.

I derive moment inequalities and study identification under a wide range of assumptions. The first identification result relies only on the assumption that productivity shocks follow an exogenous Markov process. This is the least restrictive specification, and therefore, gives the widest bounds. The other identification results exploit modeling assumptions fully to derive moment inequalities via the imperfect proxy. These moment inequalities give an identified set for the main specification of the paper. I also show how to tighten the identified set under additional distributional assumptions and shape restrictions. Analyzing identification under a wide range of assumptions makes the role of each assumption in identification transparent. For example, one can start with the most general model to impose as few restrictions as possible. If the estimated set is not informative, then a nested model can be considered to shrink the identified set. Also, comparing the results from a nested model and a general model tests the restrictions imposed by the nested model.

The partial identification approach allows me to have a model with rich heterogeneity. My model includes two productivity shocks, one persistent and the other transitory. The firm can observe both of these shocks, so both can create endogeneity. Moreover, my model includes unobserved variables that affect the firm's choice of investment. Consequently, it allows for heterogeneity in input prices and adjustment costs as well as demand shock in the output market. Finally, the identification approach is robust to measurement errors in all inputs. This robustness is particularly crucial for capital, which is most prone to measurement error.

My method is generic in that it applies to production functions under different specifications. First, one can use my method to partially identify the parameters of both value-added or gross Cobb-Douglas production functions. Second, the model is agnostic about which proxy variable to use, so both investment and materials can serve as an imperfect proxy for estimation. Third, the model can accommodate different timing assumptions about capital. One can assume that capital is chosen one period in advance, as in prior approaches, or that firms choose capital after (partially) observing productivity shocks. Finally, the model is not specific to the Cobb-Douglas production function. A nonlinear production function that is known up to a finite-dimensional parameter vector can be considered.

This paper contributes to the large literature on production function estimation using proxy variables (OP, LP, ACF; Gandhi et al. (2018) (GNR)).[1] OP find the conditions under which investment can be used as a 'proxy' to control for unobserved productivity. Motivated by 'zero' and 'lumpy' investment problem, LP propose using materials as a proxy variable. ACF point out a collinearity issue in these papers and propose an alternative proxy variable approach that avoids

[1] The production function estimation literature goes back to Marschak and Andrews (1944), who first recognized the endogeneity problem. First attempts to address the endogeneity problem have used panel data methods (Mundlak and Hoch (1965), Mundlak (1961)). However, in practice, these methods do not give satisfactory answers, as summarized by Griliches and Mairesse (1995). See also Blundell and Bond (2000).

the collinearity problem. My paper extends these approaches by showing how to make inferences when the standard proxy variable approach assumptions are violated.

A few recent papers study production function estimation with measurement errors in capital (Hu et al. (2011), Collard-Wexler and De Loecker (2016) and Kim et al. (2016)). These papers require either an instrumental variable or another proxy variable to address measurement errors. In contrast, my method does not require an additional variable, but it gives a bound rather than a point estimate.

This paper is related to the literature on monotone instrumental variables (Manski (1997), Manski and Pepper (2000)). This literature assumes that the means of potential outcomes can be ordered conditional on an observed variable, which is called the monotone instrument. In my model, the monotone instrument corresponds to the indicator variable that specifies whether the proxy variable is greater than a cutoff. My approach differs from the monotone instrument variable approach in that the monotone instrument costructed from inside the model.

Notation. I use the notation $F_a(t)$ and $F_a(t \mid b)$ to donate the distribution of variable a and the distribution of a conditional on b, respectively. Similarly, I use $f_a(t)$ and $f_a(t \mid b)$ to denote the probability density function of random variable a and the probability density function of a conditional on b, respectively.

2 Model

In this section, I describe the production function model and assumptions. The model builds on the proxy variable framework introduced by OP, but allows for deviations from some of OP's assumptions. I discuss how my model differs from the proxy variable framework and the implications of the differences for identification.

2.1 Production Function

I consider a value-added Cobb-Douglas production function to demonstrate the main results of the paper.[2] The production function is given by

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \omega_{it} + \epsilon_{it}, \tag{2.1}$$

where y_{it} denotes log-output, k_{it} denotes log-capital, and l_{it} denotes log-labor input.[3] The model includes two unobserved productivity shocks, ω_{it} and ϵ_{it}. ω_{it} represents the persistent component of productivity; it is correlated over time. On the other hand, ϵ_{it} represents the transitory component of productivity; it is independently and identically distributed over time, and it does not provide

[2] The identification strategy applies to other forms of production functions. I show how the model can be extended to other commonly used production functions in Section 5.

[3] Lowercase letters correspond to the logarithm of uppercase variables.

information about future productivity. Firms observe ω_{it} before choosing inputs, whereas ϵ_{it} can be partially or fully observed. Therefore, both productivity shocks can be correlated with inputs and can generate endogeneity.

The data consists of a panel of firms observed over periods $t = 1, \ldots, T$. To simplify the exposition I assume that $T = 3$. Observations are independently and identically distributed across firms, but they can be serially correlated within the firm. The objective is to estimate the production function parameters, (θ_l, θ_k). Since inputs are endogenous, OLS estimation would give biased coefficient estimates.

I assume that capital is a dynamic input, meaning that the firm's current period capital level affects the firm's future production.[4] As a result, capital is a state variable in the firm's dynamic optimization problem. However, unlike the standard proxy variable framework, I do not assume that capital is a predetermined input. That is, capital may be chosen after persistent productivity, ω_{it}, is (partially) observed by the firm.[5] The model is agnostic about labor input, so it can be a dynamic or static input.

2.2 Assumptions

My assumptions follow the structure of the proxy variable approach assumptions but relax them in several ways. This section presents the assumptions and describes how they lead to a less restrictive model than the proxy variable model, in terms of their economic implications. The first assumption defines the firm's information set.

Assumption 2.1 (Information Set)**.** *Let $\mathcal{I}_{it}$ denote the firm i's information set at period t. I assume that past and current persistent productivity shocks are in firm's information set, that is, $\{\omega_{i\tau}\}_{\tau=-\infty}^{t} \in \mathcal{I}_{it}$. The transitory shocks satisfy $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it-1}] = 0$.*

This assumption distinguishes the roles of two productivity shocks. The persistent productivity, ω_{it}, is observed by the firm. The transitory productivity, ϵ_{it}, can be observed, partially observed, or not observed. I also assume that the transitory shock cannot be predicted by the firm in the sense that $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it-1}] = 0$. Note that since this includes mean independence, not full independence, the firm's dynamic decision can still be affected by ϵ_{it}, as the distribution of ϵ_{it} can give information about future production. The next assumption restricts the distribution of persistent productivity shock.

Assumption 2.2 (Markov Property)**.** *Persistent productivity shocks follow an exogenous first-order Markov process*

$$P(\omega_{it+1} \mid \mathcal{I}_{it}) = P(\omega_{it+1} \mid \omega_{it}),$$

[4] This can happen, for example, due to adjustment cost in investment.

[5] It is important to note that my results do not rely on this assumption since the model can accommodate predetermined capital. This requires a minor modification in the estimation procedure, as discussed in Subsection 5.3.

and the distribution is stochastically increasing in ω_{it}.

This assumption is standard in the literature and states that the only information about future productivity (in the firm's information set) is given by current productivity. An implication of this assumption is that the transitory productivity shock ϵ_{it} is not informative about the distribution of future productivity shocks, once we condition on ω_{it}. However, this assumption does not restrict the correlation between two productivity shocks, as we might expect a positive correlation between them.

The second part of the assumption says that the distribution of future persistent productivity is stochastically increasing in ω_{it}. This assumption was first made by OP and indicates that more productive firms at the current period are likely to be more productive next period.[6] This assumption is critical for the moment inequality approach developed in this paper. Even though the literature makes this assumption, this is the first paper to use it for identification.

Assumption 2.3 (Capital Accumulation)**.** *Capital accumulates according to*

$$k_{it} = \delta k_{it-1} + i_{it}.$$

Capital is depreciated at the rate of δ, and firms make investments to accumulate capital.[7] An important feature of this assumption is that investment made at time t is productive immediately. Therefore, it makes capital a dynamic but not necessarily a predetermined input. This assumption relaxes the standard timing assumption, which assumes that capital is lagged by one period, that is, $k_{it} = \delta k_{it-1} + i_{it-1}$. This timing assumption is critical for identification in standard methods.

I am able to relax the timing assumption because my method uses an imperfect proxy instead of the perfect proxy. Also, my goal is partial identification, rather than point identification. I also emphasize that Assumption 2.3 is not required for my identification results, and therefore, my method can accommodate the standard timing assumption. I show this as an extension in Section 5.

Assumption 2.4 (Investment Function)**.** *The firm's investment decision is given by,*

$$i_{it} = f_t(k_{it-1}, \omega_{it}, \xi_{it}),$$

where $\xi_{it} \in \mathbb{R}^L$ is a vector of unobserved random variables that affects firm's investment and it is assumed to be jointly independent of ω_{it} conditional on k_{it-1}.

According to this assumption, investment depends on the two standard state variables, capital, and persistent productivity, as well as other unobserved variables denoted by ξ_{it}. The unobserved vector

[6] We say a distribution is stochastically increasing if $P(\omega_{it+1} \mid \bar{\omega}_{it})$ first-order stochastically dominates $P(\omega_{it+1} \mid \tilde{\omega}_{it})$ if and only of $\bar{\omega}_{it} \geqslant \tilde{\omega}_{it}$.

[7] My framework allows for a more complicated capital accumulation function; however, I do not consider it for simplicity.

ξ_{it} can include variables that affect the firm's dynamic problem, such as heterogeneity in adjustment cost, investment prices, or demand shocks.[8] As noted by ACF, OP do not allow heterogeneity in these variables, as the only unobservable that affects the firm's investment decision is assumed to be ω_{it}. Note that ξ_{it} can also include ϵ_{it}, the transitory productivity. According to Assumption 2.3 the firm chooses k_{it} at time t after observing the persistent productivity shock ω_{it}.

Since ξ_{it} can include demand shocks to produced goods, I do not restrict market power and competitive conduct. This is important because, in prior approaches, identification is possible only under particular competition assumptionsx'z' in the output market. For example, OP, LP, and ACF implicitly assume a perfect competition or monopolistic competition with identical demand curves. GNR consider a perfectly competitive market. This shows the importance of allowing multidimensional heterogeneity in a model of a firm to capture a richer competition structure. This is especially relevant when production function estimates are used for calculating markups (De Loecker et al. (2018)).

Another condition in Assumption 2.4 is that ξ_{it} is independent of persistent productivity conditional on the last period's capital. Therefore, given the firm's capital level, the variables that affect the firm's investment decision are not informative about the persistent productivity shock, ω_{it}. Although this assumption is restrictive, it allows for multi-dimensional heterogeneity in investment function and productivity. Some of the identification results presented later do not require this assumption, so it is still possible to make inferences on the parameters without this assumption. However, this assumption gives additional moment inequalities, which are likely to make the identified set tighter.

Finally, Assumption 2.4 also accommodates measurement error in capital, as one interpretation of ξ_{it} could be measurement error in investment. I discuss this point in Section 5.5, since measurement error in capital is an important concern in production function estimation.

Assumption 2.5 (Imperfect Proxy). *$f_t(k_{it-1}, \omega_{it}, \xi_{it})$ is weakly increasing in ω_{it} conditional on (k_{it-1}, ξ_{it}).*

This assumption relaxes the standard condition that investment is strictly monotonict in productivity. It instead assumes a weak monotonicity. Strict monotonicity is the key restriction in the proxy variable approach, which makes it possible to invert and essentially 'observe' the productivity using the proxy variable. Under my assumption, investment is no longer a 'perfect proxy' because the one-to-one relationship between investment and productivity disappears. However, by weak monotonicity, investment is still informative about productivity, so it becomes an imperfect proxy. My identification approach relies on capturing the information in an imperfect proxy variable via

[8] There is strong evidence for heterogeneity in adjustment cost. For example, Goolsbee and Gross (2000) present empirical evidence on heterogeneity in adjustment cost. Cooper and Haltiwanger (2006) argue that there is substantial heterogeneity in capital associated with heterogeneity in adjustment costs. Hamermesh and Pfann (1996), in a review paper, claim that heterogeneity in adjustment cost is a key source of heterogeneity across firms and should be included in models of firm behavior.

moment inequalities.

Relaxing strict monotonicity has important practical implications. As observed by LP and other papers, investment is often lumpy in the data. Moreover, several firms do not invest in some periods. This suggests that investment is not continuous in productivity. OP drop firms with zero investment to overcome this problem. LP propose using materials instead of investment as a proxy. My approach is robust to both zero and lumpy investments in the data.

Assumption 2.6 (Measurement Error in Labor)**.** *Labor is measured with error*

$$l_{it} = l^*_{it} + \eta^l_{it}$$

where l^*_{it} denotes true labor input and η^l_{it} is measurement error in labor. I assume that measurement error is orthogonal to the information set at $t-1$, $\mathbb{E}[\eta^l_{it} \mid I_{it-1}] = 0$.

This assumption addresses an important concern in production function estimation literature as production inputs are prone to measurement errors. My estimation method can accommodate measurement error in labor as long as it is orthogonal to the firm's information set $t-1$. Note that η^l_{it} is more general than the classical measurement error because it can be correlated with right-hand side variables. Also, this assumption nests the situation where labor does not have measurement error when we set $\eta^l_{it} = 0$.

2.3 Discussion

Overall, Assumptions 2.4 and 2.5 are the key differences of this paper from the standard assumptions, which assume that $i_{it} = f_t(k_{it}, \omega_{it})$ and $f_t(k_{it}, \omega_{it})$ is strongly increasing in ω_{it}. These assumptions limit the dimension of unobserved heterogeneity that impacts firm behavior. I relax these two strong assumptions on the functional form of investment by assuming that (i) investment is weakly increasing in productivity, and (ii) there are other unobservables affecting the investment decision. Under these assumptions, f_t is not invertible, which is the key step in OP to control for unobserved productivity. I deal with controlling for ω_{it} using moment inequalities. This allows me to have multi-dimensional unobserved heterogeneity.

3 Identification

This section derives a set of moment inequalities from the assumptions presented in the previous section. I derive moment inequalities and study identification under a wide range of assumptions. The first identification result relies only on the exogenous Markov assumption, so it is the least restrictive and gives the largest bound. Other identification results make use of other modeling assumptions and tighten the bounds.

3.1 Identification with Markov Assumption

In this section, I show that the Markov property of ω_{it} in Assumption 2.1 provides moment inequalities and set identifies the production function. This result relies on the following proposition.

Proposition 3.1. *Under Assumption 2.2 we have*

$$\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \leqslant \mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big]. \tag{3.1}$$

Proof. See Appendix A.

This proposition states that the difference between productivity shocks across two periods is smaller for the same firm than for two different firms. The key assumption to obtain this result is that conditional distribution of ω_{it} is stochastically increasing in ω_{it-1}. Therefore, firm i's current period productivity, $\omega_{it} + \epsilon_{it}$, is closer to firm i's previous period productivity than firm j's previous period productivity in Euclidean distance. Now, using the Cobb-Douglas functional form, I write the productivity shocks in Proposition 3.2 as

$$\Delta\omega_{it} + \Delta\epsilon_{it} = \Delta y_{it} - \theta_k \Delta k_{it} - \theta_l \Delta l_{it},$$
$$\Delta\omega_{ijt} + \Delta\epsilon_{ijt} = \Delta y_{ijt} - \theta_k \Delta k_{ijt} - \theta_l \Delta l_{ijt},$$

where I use $\Delta z_{it} := z_{it} - z_{it-1}$ and $\Delta z_{ijt} := z_{it} - z_{jt-1}$. Combining this with Proposition 3.1, I construct a moment inequality

$$\mathbb{E}\Big[\Delta y_{it} - \theta_k \Delta k_{it} - \theta_l \Delta l_{it}\Big] \leqslant \mathbb{E}\Big[\Delta y_{ijt} - \theta_k \Delta k_{ijt} - \theta_l \Delta l_{ijt}\big)\Big], \tag{3.2}$$

which consists only of data and parameters, so it can be used for estimating bounds for the parameters. The result uses only two assumptions (i) persistent productivity shock follows an exogenous Markov process, and ii productivity shocks are additively separable in the production function. Therefore, moment inequalities are obtained under very general conditions. First, inputs could be dynamic or static. Second, we do not need to observe a proxy variable to control for productivity shocks. Finally, the variables that affect the firm's dynamic or static decisions are unrestricted. Of course, this flexibility might come with a cost, as the identified set might not be very informative.

In Proposition 3.1, I use the Euclidean distance to derive moment inequalities. However, one can consider other distance measures and obtain different moment inequalities. In that case, different distances would give different identified sets, which can be intersected to obtain tighter bounds.

3.2 Identification with Other Assumptions

This section derives a set of conditional moment inequalities based on all the modeling assumptions. The derivation relies on using investment as an imperfect proxy variable. The first step in constructing moment inequalities is stochastically ordering the productivity distributions that involve firms with ‘high’ and ‘low’ investment.

Proposition 3.2. *Assumptions 2.1-2.6, along with some regulatory conditions, imply that*
(i) *For all z in the support of i_{it},*

$$\frac{f_{\omega_{it}}(t \mid k_{it-1}, i_{it} > z)}{f_{\omega_{it}}(t \mid k_{it-1}, i_{it} < z)}$$

is increasing in t, that is, it satisfies the Monotone Likelihood Ratio Property (MLRP).
(ii) *The distribution function of ω_{it} conditional on k_{it-1} and $\{i_{it} > z\}$ first order stochastically dominates (FOSD) the distribution function of ω_{it} conditional on k_{it-1} and $\{i_{it} < z\}$*

$$F_{\omega_{it}}(t \mid k_{it-1}, i_{it} > z) \geqslant F_{\omega_{it}}(t \mid k_{it-1}, i_{it} < z),$$

for all $(t, z) \in \mathbb{R}_+^2$.
(iii) *The mean of ω_{it} conditional on k_{it-1} and $\{i_{it} > z\}$ is greater than the mean of ω_{it} conditional on k_{it-1} and $\{i_{it} < z\}$:*

$$\mathbb{E}[\omega_{it} \mid k_{it-1}, i_{it} > z] \geqslant \mathbb{E}[\omega_{it} \mid k_{it-1}, i_{it} < z], \tag{3.3}$$

for all z in the support of i_{it},

Proof. See Appendix A.

This proposition shows that weak monotonicity of investment in ω_{it} gives three stochastic orderings: (i) monotone likelihood ratio, (ii) first-order stochastic dominance, and (iii) mean ordering. The proof of this proposition shows that the weak monotonicity encompasses all the information given by the proxy variable. That is, MLRP holds if and only if investment is weakly monotone in productivity.

It is useful to compare the statements of this proposition with the invertibility condition in the proxy variable approach. When investment is invertible, and therefore is a perfect proxy, the ranking of the firm in investment equals to the ranking in productivity. This makes it possible to infer productivity using investment. In my model, investment is not a perfect proxy, so it is not possible to recover productivity from investment. However, by weak monotonicity, investment still provides information about productivity, so it becomes an imperfect proxy.

This proposition shows that an imperfect proxy can be used to order productivity stochastically, rather than deterministically. In particular, Proposition 3.2 says that when firms are grouped based on how much they invest, we can infer that high investment firms will be more productive than low investment firms, on average. The main idea for identification is to use these stochastic orderings in the form of moment inequalities to set identify the production function.

My first moment inequality derivation exploits the condition Proposition 3.2(iii). To see how to obtain a moment inequality, first note that the Markov assumption implies

$$\omega_{it} = g(\omega_{it-1}) + \zeta_{it},$$

where $\zeta_{it} = \omega_{it} - \mathbb{E}[\omega_{it} \mid \omega_{it-1}]$ and $\mathbb{E}[\zeta_{it} \mid I_{it-1}] = 0$ by construction. Also, the assumption that $P(\omega_{it} \mid \omega_{it-1})$ is stochastically increasing implies that $g(\omega_{it-1})$ is a monotone function. This representation of ω_{it} has been commonly used in the proxy variable approach for constructing moments. Substituting productivity into the production function yields:

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}. \tag{3.4}$$

This representation of the production function involves three error terms: innovation to productivity ζ_{it}, measurement error in labor η^l_{it}, and the transitory productivity shock ϵ_{it}. Let me define a function which takes data and parameter:

$$m(w_{it}, \tilde{\theta}) := y_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} \tag{3.5}$$

with $w_{it} = (k_{it}, l_{it})$ and $\tilde{\theta} = (\tilde{\theta}_l, \tilde{\theta}_k)$. Also let θ denote the vector of true parameter values. The next proposition presents a conditional moment inequality using Equation (3.4) and Proposition 3.2.

Proposition 3.3. *For all $z \in \mathcal{I}$ and $k_{it-1} \in \mathcal{K}$*

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} < z] \geqslant 0 \tag{3.6}$$

This proposition is the main identification result of the paper. Conditional on k_{it-2} if we compare two groups of firms, one with investment greater than z and one with investment lower than z, Equation (6.1) is satisfied at the true parameter values. The key conditions needed for this proposition are monotonicity of $g(\omega_{it-1})$ and the weak monotonicity of investment in productivity.

Another key condition for this proposition is that $(\zeta_{it}, \eta^l_{it}, \epsilon_{it})$ are orthogonal to the firm's information set at $t-1$. Recall that Proposition 3.2(iii) provides moment inequality in terms of ω_{it}. However, we can only recover $g(w_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}$ from the observed variables and parameters. Therefore, we need to account for $(\zeta_{it}, \eta^l_{it}, \epsilon_{it})$. The orthogonality condition allows me to achieve this, as $\zeta_{it} + \eta^l_{it} + \epsilon_{it}$ drop from the moment inequality in Equation (6.1) when we take conditional expectations.

Remark 3.1 (Comparison to Proxy Variable)**.** Single dimensional unobserved heterogeneity and strong monotonicity of investment in productivity allow OP to invert the investment function and recover productivity shock as

$$\omega_{it} = f_t^{-1}(i_{it}, k_{it}).$$

An invertible investment function means that one can control for unobserved productivity by conditioning on observables. My model relaxes the two necessary conditions for invertibility. First, I allow the investment function to be weakly monotone in productivity. Second, there are additional unobserved variables affecting investment. Therefore, we can no longer compare the productivity levels of two firms by comparing their investments under my assumptions.

Remark 3.2 (Conditioning on Two Investment Levels)**.** One might think that a moment inequality similar to Equation (3.3) holds, conditional on two different investment levels:

$$\mathbb{E}[\omega_{it} \mid k_{it-1}, i_{it} = z_1] \geqslant \mathbb{E}[\omega_{it} \mid k_{it-1}, i_{it} = z_2],$$

where $z_1 < z_2$. However, this inequality does not hold, as it is easy to find counterexamples. Therefore, it is crucial to group firms using a cutoff value in investment.

Remark 3.3 (Relation to Monotone Instrument and Imperfect Instrument Literature)**.** This paper is related to the literature on monotone instrumental variables (Manski (1997), Manski and Pepper (2000)). This literature assumes that the mean potential outcomes are ordered based on an observed variable, which is called a monotone instrument. In my model, investment can be considered a monotone instrument for productivity. The main difference of my model from the standard monotone instrumental variable approach is that the monotone instrument comes from within the model in this paper. My approach is also related to the 'imperfect instrument approach,' which assumes that the researcher has some prior information about the correlation between the endogenous variable and unobserved heterogeneity. This information is then used to construct moment inequalities. See, for example, Nevo and Rosen (2012) and Conley et al. (2012).

3.3 Identified Set

In this section, I characterize the identified set using the derived moment inequalities. Since I have conditional moment inequalities, the identified set is given by intersection bounds. Recall that the true parameter satisfies:

$$\mathbb{E}[y_{it} - \theta_k k_{it} - \theta_l l_{it} \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[y_{it} - \theta_k k_{it} - \theta_l l_{it} \mid k_{it-2}, i_{it-1} < z] \geqslant 0. \tag{3.7}$$

In what follows, I assume that all expectations are conditional on $(k_{it-2} = k)$, so I drop it from notation. Define the following variables indexed by z

$$I^h_{it}(z) := \mathbb{1}\{i_{it-1} > z\}, \qquad I^l_{it}(z) := \mathbb{1}\{i_{it-1} < z\}.$$

$I^h_{it}(z)$ equals one if a firm invests less than z and zero otherwise. The opposite is true for $I^l_{it}(z)$. I call firms with $I^h_{it}(z) = 1$ as high investment and $I^l_{it}(z) = 1$ as low investment. With some abuse of notation, I treat these variables as events when they are in the conditioning set. I can write the moment inequality in Equation (3.7) as:

$$\theta_k\big(\mathbb{E}[k_{it} \mid I_{it}^h(z)] - \mathbb{E}[k_{it} \mid I_{it}^l(z)]\big) + \theta_l\big(\mathbb{E}[l_{it} \mid I_{it}^h(z)] - \mathbb{E}[l_{it} \mid I_{it}^l(z)]\big) \leqslant \big(\mathbb{E}[y_{it} \mid I_{it}^h(z)] - \mathbb{E}[y_{it} \mid I_{it}^l(z)]\big).$$

This expression shows that the identified set depends on average capital and labor of 'high' and 'low' investment firms. For example, if average capital and labor do not vary with investment, then the identified set would not be informative. Next, I characterize the identified set. Define

$$a_l(z,k) := \mathbb{E}[l_{it} \mid I_{it}^h(z)] - \mathbb{E}[l_{it} \mid I_{it}^l(z)], \tag{3.8}$$

$$a_k(z,k) := \mathbb{E}[k_{it} \mid I_{it}^h(z)] - \mathbb{E}[k_{it} \mid I_{it}^l(z)], \tag{3.9}$$

$$a_y(z,k) := \mathbb{E}[y_{it} \mid I_{it}^h(z)] - \mathbb{E}[y_{it} \mid I_{it}^l(z)]. \tag{3.10}$$

Using these definitions, the moment inequality can be expressed as

$$a_k(z,k)\theta_k + a_l(z,k)\theta_l \leqslant y_l(z,k) \qquad \text{for all} \quad z > 0, \qquad k \in \mathcal{K}.$$

The identified set, conditional on k and z, is a region defined by a half-plane. Therefore, the identified set is the intersection of these half-planes.

Proposition 3.4 (Identified Set)**.** *Assume $\theta \in \tilde{\Theta}$, a compact parameter space. The identified set Θ is defined as the set of parameters that satisfy the conditional moment inequalities*

$$\Theta := \left\{\tilde{\theta} \in \tilde{\Theta} : \bigcap_{k\in\mathcal{K}} \bigcap_{z>0} a_y(z,k) - \tilde{\theta}_k a_k(z,k) - \tilde{\theta}_l a_l(z,k) \geqslant 0 \quad a.s.\right\},$$

and the identified set contains true parameter value $\theta \in \Theta$.

3.4 Moment Inequalities Using FOSD and MLRP

This subsection shows how to construct moment inequalities using MLRP and FOSD under further assumptions. These assumptions include restrictions on the distribution of unobservables.

Proposition 3.2 establishes that the distributions of productivity conditional on high and low investment satisfy MLRP and FOSD. However, when characterizing the identified set, I only used the mean ordering, a much weaker condition. This is because even though MLRP and FOSD hold for productivity, I can only recover $g(w_{it-1}) + \zeta_{it} + \eta_{it}^l + \epsilon_{it}$. Therefore, I need additional conditions that ensure that MLRP and FOSD are preserved when there are additive errors. The following two theorems present those conditions.

Theorem 3.1 (Shaked and Shanthikumar (2007))**.** *Let X_1 and X_2 be two independent random variables, and Y_1 and Y_2 be another two independent random variables. If $X_i \leqslant_{FOSD} Y_i$ for $i = 1, 2$*

then

$$X_1 + X_2 \leqslant_{FOSD} Y_1 + Y_2.$$

Theorem 3.2 (Shaked and Shanthikumar (2007)). *Let X_1, X_2 and Z be random variables such that X_1 and Z are independent and X_2 and Z are independent. If $X_1 \leqslant_{MLRP} X_2$ and Z has a log-concave probability density functions, then*

$$X_1 + Z \leqslant_{MLRP} X_2 + Z.$$

These two theorems suggest that in order for MLRP and FOSD to be preserved under convolutions I need: (i) an independence condition for FOSD and (ii) independence and log-concavity conditions for MLRP. Therefore, I next impose these conditions on the unobservables to derive moment inequalities using FOSD and MLRP.

3.4.1 Identified Set Using FOSD

As Theorem 3.1 suggests, I need to impose independence restrictions on unobservables to preserve FOSD ordering.

Assumption 3.1. *$(\eta_{it}, \zeta_{it}, \epsilon_{it})$ are jointly independent from $\mathcal{I}_{it-1}$.*

With this assumption $g(\omega_{it-1})$ becomes jointly independent from the rest of the unobservables, $\zeta_{it} + \epsilon_{it} + \eta_{it}$, conditional on $\mathcal{I}_{it-1}$. Therefore, MLRP for $g(\omega_{it-1})$ conditional on high and low investment is preserved for $g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it}$ conditional on high and low investment. Under this assumption, I can strengthen the mean ordering in Equation (3.3) to the first-order stochastic dominance ordering, and characterize the identified set accordingly.

Proposition 3.5 (Identified Set-FOSD). *Under Assumptions 2.1-2.6 and Assumption 3.1, the true parameter $\theta \in \tilde{\Theta}$ satisfies the following condition*

$$F_{m(w_{it},\theta)}(\cdot \mid k_{it-2} = k, I^h_{it}(z) = 1) - F_{m(w_{it},\theta)}(\cdot \mid k_{it-2} = k, I^l_{it}(z) = 1) \geqslant 0, \tag{3.11}$$

for all (k, z) in the support of (k_{it-2}, i_{it-1}) such that $0 < \mathbb{E}[I^h_{it}(z) \mid k_{it-2} = k] < 1$. The identified set (conditional on k and z) is

$$\Theta^{FOSD}_t = \{\widehat{\theta} \in \Theta : (3.11) \text{ holds with } \theta \text{ in place of } \widehat{\theta}\}.$$

The independence assumption is non-standard in production function models, but it is difficult to consider situations where mean independence holds and independence does not hold. Next, I specify the assumption that allows me to use the MLRP result in Proposition 3.2(i) for identification.

3.4.2 Identified Set Using MLRP

As Theorem 3.2 suggests, I need to impose independence and shape restrictions on the distributions of unobservables to be able to use MLRP.

Assumption 3.2. *$(\eta_{it}, \zeta_{it}, \epsilon_{it})$ are jointly independent from $\mathcal{I}_{it-1}$, and each variable in $\eta_{it}, \zeta_{it}, \epsilon_{it}$ has a log-concave probability distribution function.*

Under this assumption, I can strengthen the mean inequality in Proposition (**??**) to MLRP and characterize the identified set accordingly.

Proposition 3.6 (Identified Set-MLRP)**.** *Under Assumptions 2.1-2.6 and Assumption 3.2, the true parameter $\theta \in \tilde{\Theta}$ satisfies the following inequality:*

$$F_{m(w_{it},\theta)}(\cdot \mid a \leqslant m(w_{it},\theta) \leqslant b, k_{it-2} = k, I^h_{it}(z) = 1) - \\ F_{m(w_{it},\theta)}(\cdot \mid a \leqslant m(w_{it},\theta) \leqslant b, k_{it-2} = k, I^l_{it}(z) = 1) \geqslant 0 \tag{3.12}$$

for all (k, z) in the support of (k_{it-2}, i_{it-1}) such that $0 < \mathbb{E}[I^h_{it}(z) \mid k_{it-2} = k] < 1$ and for all (a, b) such that $a < b$. The identified set (conditional on k and z) is

$$\Theta^{MLRP}_t = \{\widehat{\theta} \in \Theta : (3.12) \text{ holds with } \widehat{\theta} \text{ in place of } \theta\}.$$

Most well known distributions ,such as those in the exponential family, satisfy the assumptions required for this proposition.

3.5 Discussion

An advantage of identification analysis under a wide range of assumptions is that we can see the role of each assumption in identification. For example, we can start with the most general model to impose as few restrictions as possible. If the estimated set is not informative, then a nested model can be considered to shrink the identified set. Also, comparing estimates from a nested and a nesting model would test the restrictions imposed by the nested model.

Note also that proxy variable specification is a special case of my framework. So, if the estimates set is not information, one can use the proxy variable approach to point identify the parameters. It is also worth noting that, the identified set under my assumptions does not have to include the point estimates obtained from proxy variable method. The reason is that under the proxy variable assumptions the model is over-identified. If my partial identification method uses overidentification restrictions, the point estimates might be excluded from the identified set. This would mean rejecting the proxy variable assumptions.

One may think that set identifying the production function parameters is not useful unless the set is tight. As in most set identification results, the informativeness of the identified set depends on the data and empirical example. However, as discussed above, there are other advantages of using my framework. Most importantly, since the standard OP approach is nested under my assumptions, there is no harm in starting with a more general model.

4 Estimation

The identification analysis generates conditional moment inequalities. It is convenient to turn conditional moment inequalities into unconditional ones for estimation. To achieve this, I integrate out k_{it-2} and define the propensity of high and low investment conditional on k_{it-2}.

First, I define the propensity score, which equals the probability that investment is greater than a cutoff z.

$$m(k_{it-2}, z) = \mathbb{E}[I_{it}^h(z) \mid k_{it-2}].$$

I define the moments using the propensity scores in the following ways.

$$\mathbb{E}[(y_{it} - \theta_k k_{it} - \theta_l l_{it}) \mid k_{it-2}, I_{it}^h(z)] = \mathbb{E}\left[\frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it}) I_{it}^h(z)}{m(k_{it-2}, z)} \mid k_{it-2}\right],$$
$$\mathbb{E}[(y_{it} - \theta_k k_{it} - \theta_l l_{it}) \mid k_{it-2}, I_{it}^h(z)] = \mathbb{E}\left[\frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it})(1 - I_{it}^h(z))}{(1 - m(k_{it-2}, z))} \mid k_{it-2}\right].$$

Integrating out k_{it-2}, the unconditional moment inequality can be written as:

$$\mathbb{E}\left[\frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it}) I_{it}^h(z)}{m(k_{it-2}, z)}\right] \geqslant \mathbb{E}\left[\frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it})(1 - I_{it}^h(z))}{1 - m(k_{it-2}, z)}\right].$$

Define

$$s^h(\theta, z) = \mathbb{E}\left[\frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it}) I_{it}^h(z)}{m(k_{it-2}, z)}\right], \qquad s^l(\theta, z) = \mathbb{E}\left[\frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it})(1 - I_{it}^h(z))}{1 - m(k_{it-2}, z)}\right].$$

Estimation can be carried out by testing the hypothesis $\theta^h(\theta, z) \geqslant s^l(\theta, z)$ and inverting the test. Specifically, for a given θ, test $s^h(\theta, z) \geqslant \theta^l(\theta, z)$ for all $z > 0$ and include θ in the identified set if the the null hypothesis fails to be rejected. The natural estimators for $s^h(\theta, z)$ and $s^l(\theta, z)$ are

$$\widehat{s}^h(\theta, z) = \sum_{i=1}^{N} \frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it}) I_{it}^h(z)}{\widehat{m}(k_{it-2}, z)}, \qquad s^l(\theta, z) = \sum_{i=1}^{N} \frac{(y_{it} - \theta_k k_{it} - \theta_l l_{it}) I_{it}^l(z)}{1 - \widehat{m}(k_{it-2}, z)}, \tag{4.1}$$

where $\widehat{m}(k_{it-2}, z)$ is an estimate of $m(k_{it-2}, z)$. These functions involve the propensity score, which is a nuisance function and needs to be estimated in the first stage. To make the estimation procedure more robust to an estimation error in the nuisance function, I can define the doubly robust moment functions. This requires other nuisance functions in the moment functions. Define

$$g^h(k,z,\theta) = \mathbb{E}[\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) \mid k_{it-2} = k, I^h_{it}(z)],$$
$$g^l(k,z,\theta) = \mathbb{E}[\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) \mid k_{it-2} = k, I^l_{it}(z)].$$

Using these functions, the doubly robust moments are

$$s^h_{db}(\theta,z) = g^h(k_{it-2},z,\theta) + \frac{I^h_{it}(z)\Big(\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) - g^h(k_{it-2},z,\theta)\Big)}{m(k_{it-2},z)},$$
$$s^l_{db}(\theta,z) = g^l(k_{it-2},z,\theta) + \frac{\big(1 - I^h_{it}(z)\big)\big((y_{it} - \theta_k k_{it} - \theta_l l_{it}) - g^l(k_{it-2},z,\theta)\big)}{1 - m(k_{it-2},z)}.$$

The sample analog of these moments can be obtained similar to Equation (4.1). Expectations of the doubly robust moments and the original moments equal to each other:

$$\mathbb{E}\big[s^h_{db}(\theta,z)\big] = \mathbb{E}\big[s^h(\theta,z)\big], \qquad \mathbb{E}\big[s^l_{db}(\theta,z)\big] = \mathbb{E}\big[s^l(\theta,z)\big].$$

Doubly robust moments have the property that if one of the nuisance functions is correct, then the moment is correct. Semenova (2017) studies moment inequality estimation with nuisance functions shows how to do inference when the nuisance functions are estimated using machine learning methods.

One can also consider using conditional moment inequalities to tighten the identified set instead of integrating out capital. For that estimation problem one can use many moment inequalities framework of Chernozhukov et al. (2018) or conditional moment inequality estimation framework of Andrews and Shi (2013).

5 Extensions

The approach developed in this paper can be extended to other forms of production functions. To give some examples, I discuss the application to gross production function and using materials as the proxy variable instead of labor. I also show how my model accommodates measurement error in capital.

5.1 Gross Production Function

In this subsection, I show how to extend my model to a gross production function. The estimation procedure remains the same, with an increase in the number of parameters.

A gross production function is given by

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \theta_m m_{it} + \omega_{it} + \eta^l_{it} + \epsilon_{it} \tag{5.1}$$

Similar to the main model, using Markov we can the production function as

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \theta_m m_{it} + g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}$$

All the proposition shown in this paper apply to this model since they do not depend on the functional form of the production function. Therefore, we can construct a moment function as

$$m(w_{it}, \tilde{\theta}) := y_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} - \tilde{\theta}_m m_{it}$$

with $w_{it} = (k_{it}, l_{it}, m_{it})$, $\tilde{\theta} = (\tilde{\theta}_l, \tilde{\theta}_k, \tilde{\theta}_m)$. Also let θ denote the vector of true parameter values.

Proposition 5.1. *For all $z \in \mathcal{I}$ and $k_{it-1} \in \mathcal{K}$*

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} < z] \geqslant 0$$

This moment inequality can be used to estimate the model parameters. The only difference is that there are more parameters to estimate, so the estimates might be less precise.

5.2 Using Materials as Proxy

Levinsohn and Petrin (2003) propose using materials instead of investment as a proxy for productivity shock, because investment is lumpy in the data. Even though my framework is robust to this problem, the model allows for using materials as a proxy instead of investment. For this, I need to replace Assumption 2.4 with the following assumption.

Assumption 5.1. *Firms' materials decision is given by*

$$m_{it} = m_t(k_{it-1}, \omega_{it}, \xi_{it}),$$

where ξ_{it} is a vector of unobserved random variables that affect firm's materials decision and it is assumed to be independent of ω_{it} conditional on k_{it-1}.

When materials is used as an imperfect proxy, the moment inequalities become

Proposition 5.2. *For all $z \in \mathcal{I}$ and $k_{it-1} \in \mathcal{K}$*

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, m_{it} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, m_{it} < z] \geqslant 0.$$

Moment function, $m(w_{it}, \theta)$, is defined in Equation (3.5). This moment inequality can be used to estimate the model parameters. This moment applied to a gross production function.

5.3 Identification with the Standard Timing Assumption

To accommodate the standard timing assumption, I replace Assumption 2.3 with the following assumption.

Assumption 5.2. *Capital accumulates according to*

$$k_{it} = \delta k_{it-1} + i_{it-1}.$$

This assumption implies that the amount of capital used for time t production is determined at time $t-1$. I also need to replace Assumption (2.4) with

Assumption 5.3. *Firms' investment decision is given by*

$$i_{it} = f_t(k_{it}, \omega_{it}, \xi_{it})$$

where ξ_{it} is a vector of unobserved variables that affect firm's investment decision and it is assumed to be independent of ω_{it} conditional on k_{it}.

These changes only affect the conditioning set in the moment inequalities. In particular I need condition on (k_{it-1}, i_{it-1}) instead of (k_{it-2}, i_{it-1}). So the moment inequality becomes:

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, i_{it-1} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, i_{it-1} < z] \geqslant 0,$$

where the moment function, $m(w_{it}, \theta)$, is defined in Equation (3.5). This moment inequality can be used to estimate the parameters. The estimation procedure remains the same.

5.4 Nonparametric Production Function

The identification strategy in this paper requires Cobb-Douglas functional form only to account for measurement errors. Thus, if I rule out measurement errors, I can use nonlinear production function that is known up to a finite dimensional parameter vector. To demonstrate this extension, consider the following model:

$$y_{it} = r(\theta, w_{it}) + \omega_{it} + \epsilon_{it}, \tag{5.2}$$

where $r(\theta, w_{it})$ is a known function up to the parameter vector θ. For this model, the moment function becomes:

$$m^*(w_{it}, \tilde{\theta}) := y_{it} - r(\tilde{\theta}, w_{it}). \tag{5.3}$$

With this moment function, the results for the main model can be applied to estimate Equation (5.2).

5.5 Measurement Error in Capital

My framework can also accommodate measurement error in capital, which is important because among all inputs, capital is most prone to measurement error.

Let η^k_{it} denote measurement error in investment. It is natural to model measurement error in capital using measurement error in investment because capital is accumulated through investment. Also, capital is often constructed from investment series in the data.[9] Let i^*_{it} denote true investment. The observed investment is given by:

$$\begin{aligned} i_{it} &= i^*_{it} + \eta^k_{it} \\ &= f_t(k_{it-1}, \omega_{it}, \xi_{it}) + \eta^k_{it}. \end{aligned}$$

Measurement error, η^k_{it}, can be included into f_t function as a part of ξ_{it} vector. Define $\xi^*_{it} = (\xi_{it}, \zeta^k_{it})$ and rewrite the investment function as

$$i_{it} = f_t(k_{it-1}, \omega_{it}, \xi^*_{it}).$$

With measurement error in investment, observed capital takes the form

$$k_{it} = k^*_{it} + \sum_{j=0}^{t} (1-\delta)^j \eta^k_{it},$$

where k^*_{it} is true capital. Substituting this into the production function yields

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + g(\omega_{it-1}) + \zeta_{it} - \Big(\sum_{j=0}^{t} (1-\delta)^j \eta^k_{it}\Big)\theta_k + \eta^l_{it} + \epsilon_{it}. \tag{5.4}$$

I can combine measurement errors in capital and labor as $\eta_{it} = -\Big(\sum_{j=0}^{t} (1-\delta)^j \eta^k_{it}\Big)\theta_k + \eta^l_{it}$. With the combined measurement error, production function becomes:

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + g(\omega_{it-1}) + \zeta_{it} + \eta_{it} + \epsilon_{it}. \tag{5.5}$$

This equation takes the same form I consider for my partial identification results. So if we assume that η^k_{it} be is independent of ω_{it} conditional on k_{it-1} and $\mathbb{E}[\eta_{it} \mid \mathcal{I}_{it-1}] = 0$, we can use the same moment inequalities derived in Equation (6.1).

Since measurement error in capital is an important problem, the literature has paid particular attention to measurement error in capital. Some examples are Hu et al. (2011), Collard-Wexler and De Loecker (2016), and Kim et al. (2016). These papers require an instrumental variable or another proxy variable to control for measurement error in capital and show point identification. In

[9] For example, the perpetual inventory method.

contrast, my method does not require another variable, but it gives a bound instead of a point.

6 Empirical Application

In this section, I apply my method to a panel production data from Turkey.

6.1 Data

The data for Turkey are provided by the Turkish Statistical Institute (TurkStat; formerly known as the State Institute of Statistics, SIS), which collects plant-level data for the manufacturing sector.[10] Periodically, Turkstat conducts the Census of Industry and Business Establishments (CIBE), which collects information on all manufacturing plants in Turkey. In addition, TurkStat conducts the Annual Surveys of Manufacturing Industries (ASMI) that covers all establishments with at least 10 employees. The set of establishments used for ASMI is obtained from the CIBE. In non-census years, the new private plants with at least 10 employees are obtained from the chambers of industry.

I use a sample from Annual Surveys of Manufacturing Industries, covering a period from 1983 to 2000. Data from a more recent period is available, but due to major changes in the survey methodology, it is not possible to link this dataset to the data from a more recent period. I limit the sample to only private establishments. I focus on the textile industry, which is the largest 3-digit industry in terms of the number of firms. My sample includes 1437 firms and 14271 year-firm observations.

The data includes gross revenue, investment, the book value of capital, materials expenditures and the number of production and administrative workers. The real value of annual output is obtained by deflating the plant's total annual sales revenues by an industry-specific price index. Material inputs include all purchases of intermediate inputs. I deflate the nominal value of total material input cost by each plant using the industry-level intermediate input price index. Finally, capital stock series is constructed using the perpetual inventory method where investment in new capital is combined with deflated capital from period $t-1$ to form capital in period t.

6.2 Empirical Model

For my empirical application, I consider a value-added production function.

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \omega_{it} + \epsilon_{it}.$$

[10] This dataset has been used by Levinsohn (1993) and Taymaz and Yilmaz (2015).

I maintain the standard timing assumption that capital is chosen one period in advance.[11] I use the following moment inequality for estimation

$$\mathbb{E}[y_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} \mid k_{it-1}, i_{it-1} > z] - \mathbb{E}[y_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} \mid k_{it-1}, i_{it-1} < z] \geqslant 0. \tag{6.1}$$

I follow the steps outlined in the estimation section to construct unconditional moment inequalities from this conditional moment inequality. In particular, I first calculate the empirical analog of $\widehat{s}^h_{db}(\theta, z)$ and $s^l_{db}(\theta, z)$ using

$$\widehat{s}^h_{db}(\theta, z) = \sum_{i=1}^{N} \widehat{g}^h_t(k_{it-1}, z, \theta) + \frac{I^h_{it}(z)\Big(\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) - \widehat{g}^h_t(k_{it-1}, z)\Big)}{1 - \widehat{m}_t(k_{it-1}, z)}, \tag{6.2}$$

$$\widehat{s}^l_{db}(\theta, z, \theta) = \sum_{i=1}^{N} \widehat{g}^l_t(k_{it-1}, z) + \frac{I^l_{it}(z)\Big(\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) - \widehat{g}^l_t(k_{it-1}, z)\Big)}{\widehat{m}_t(k_{it-1}, z)}, \tag{6.3}$$

where I estimate the nuisance functions $\widehat{m}_t$, $\widehat{g}^h_t$ and $\widehat{g}^l_t$ using random forest method in the first stage. Following Semenova (2017), I also employ cross-fitting, i.e, I estimate the nuisance functions in the first half of the sample and construct the moments, $\widehat{s}^h_{db}(\theta, z)$ and $\widehat{s}^l_{db}(\theta, z)$, using the second half of the sample. I then swap these samples to avoid loss of efficiency.

I test the moment inequality $s^h_{db}(\theta, z) \geqslant s^l_{db}(\theta, z)$ using the Chernozhukov et al. (2018) many moment inequalities framework with the empirical analogs given in Equations (6.2) and (6.3). I obtain 50 moment inequalities by choosing 10 different z values from the support and testing moment inequalities in 5 different periods (84-87, 88-90, 91-93, 94-96, 97-00) for each z value.[12] The estimated set is constructed by inverting this test, that is, estimated set includes all parameter values for which I fail to reject $s^h_{db}(\theta, z) \geqslant s^l_{db}(\theta, z)$.

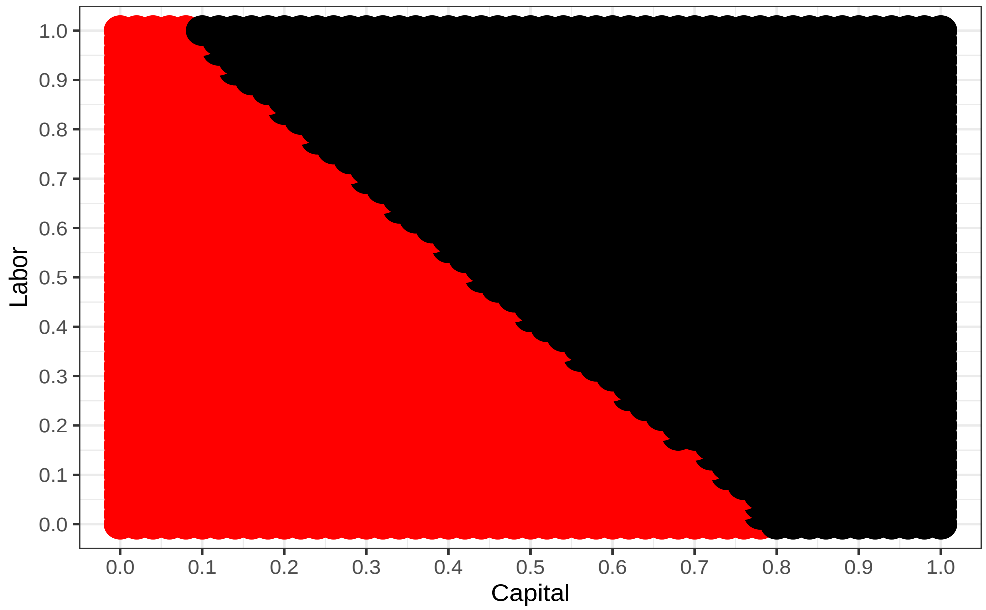
I present the estimation results in Figure 1, where the red region shows the estimated set (with 95% coverage). As suggested in the identification analysis of Section 3, the estimated set is a half-plane. It is more informative about θ_k than θ_l, as it excludes the large values of capital elasticity values. One can also combine this estimated set with a priori restrictions on the production function based on economic theory. For example, if we impose that the restriction that the production function is constant returns to scale, then we can conclude from the identified set that the elasticity of capital is less than 0.28.

Even though the identified set does not give a definite answer about the production function parameters, it suggests that less restrictive assumptions in the proxy variable framework is still informative about production technology. Also, the estimated set is not sharp, since the estimation

[11] I assume this because the capital series is constructed under this assumption using the perpetual inventory method.

[12] To calculate the critical values I use their multiplier bootstrap method.

Figure 1: Estimated Set



Note: The estimated of set for the value-added Cobb-Douglas production function parameters. The reported set shown in black covers the true parameter with 95% probability.

procedure does not use moment inequalities conditional on capital. Moreover, one can also use the moment inequalities derived from FOSD and MLRP restrictions under slightly stronger assumptions in Subsection 3.4. These estimations might give more informative estimated sets.[13]

7 Conclusion

This paper extends the production function estimation literature by relaxing the restrictive assumptions of the proxy variable approach and showing that the parameters remain partially identified. My model (i) allows for multi-dimensional unobserved heterogeneity, (ii) relaxes strict monotonicity to weak monotonicity, (iii) accommodates a more general timing assumption. Also, the method is robust to measurement errors in inputs, an important problem in production function estimation.

I accomplish this by using an 'imperfect proxy' variable for identification. An 'imperfect proxy' variable contains information about productivity, but it cannot directly be used to control for productivity as in the proxy variable approach. Instead, an imperfect proxy variables generates stochastic orderings of productivity distributions, which can be exploited for estimation. I show how to use this stochastic ordering in the form of moment inequalities to obtain bounds for production function parameters.

[13]I plan to look at the estimated sets under these assumptions in the future versions of the paper.

References

Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth. 2015. "Identification properties of recent production function estimators". *Econometrica* 83 (6): 2411–2451. (Cit. on p. 1).

Ackerberg, Daniel, Lanier Benkard, C., Berry, Steven, and Pakes, Ariel. 2007. "Chapter 63 Econometric Tools for Analyzing Market Outcomes". In *Handbook of Econometrics*, 6:4171–4276. Elsevier. (Cit. on p. 1).

Andrews, Donald WK and Shi, Xiaoxia. 2013. "Inference Based on Conditional Moment Inequalities". *Econometrica* 81 (2): 609–666. (Cit. on p. 16).

Blundell, Richard and Bond, Stephen. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions". *Econometric Reviews* 19 (3): 321–340. (Cit. on p. 2).

Chernozhukov, Victor, Chetverikov, Denis, and Kato, Kengo. 2018. "Inference on Causal and Structural Parameters Using Many Moment Inequalities". *The Review of Economic Studies* 86 (5): 1867–1900. (Cit. on pp. 16, 21).

Collard-Wexler, Allan and De Loecker, Jan. 2016. "Production Function Estimation with Measurement Error in Inputs". *Working Paper.* (Cit. on pp. 3, 19).

Conley, Timothy G, Hansen, Christian B, and Rossi, Peter E. 2012. "Plausibly exogenous". *Review of Economics and Statistics* 94 (1): 260–272. (Cit. on p. 11).

Cooper, Russell W and Haltiwanger, John C. 2006. "On the Nature of Capital Adjustment Costs". *The Review of Economic Studies* 73 (3): 611–633. (Cit. on p. 6).

De Loecker, Jan, Eeckhout, Jan, and Unger, Gabriel. 2018. "The Rise of Market Power and the Macroeconomic Implications". *Working Paper.* (Cit. on p. 6).

Gandhi, Amit, Navarro, Salvador, and Rivers, David. 2018. "On the Identification of Gross Output Production Functions". *Working Paper.* (Cit. on p. 2).

Goolsbee, Austan and Gross, David B. 2000. "Estimating Adjustment Costs with Data on Heterogeneous Capital Goods". *Working Paper.* (Cit. on p. 6).

Griliches, Zvi and Mairesse, Jacques. 1995. "Production Functions: the Search for Identification". *Working Paper 5067 National Bureau of Economic Research.* (Cit. on p. 2).

Hamermesh, Daniel S and Pfann, Gerard A. 1996. "Adjustment Costs in Factor Demand". *Journal of Economic Literature* 34 (3): 1264–1292. (Cit. on p. 6).

Hu, Yingyao, Huang, Guofang, and Sasaki, Yuya. 2011. "Estimating Production Functions with Robustness Against Errors in the Proxy Variables". *Working Paper.* (Cit. on pp. 3, 19).

Kim, Kyoo il, Petrin, Amil, and Song, Suyong. 2016. "Estimating Production Functions with Control Functions when Capital is Measured with Error". *Journal of Econometrics* 190 (2): 267–279. (Cit. on pp. 3, 19).

Levinsohn, James. 1993. “Testing the Imports-as-market-discipline Hypothesis”. *Journal of International Economics* 35 (1): 1–22. (Cit. on p. 20).

Levinsohn, James and Petrin, Amil. 2003. “Estimating Production Functions Using Inputs to Control for Unobservables”. *The Review of Economic Studies* 70 (2): 317–341. (Cit. on pp. 1, 17).

Manski, Charles F. 1997. “Monotone Treatment Response”. *Econometrica: Journal of the Econometric Society*: 1311–1334. (Cit. on pp. 3, 11).

Manski, Charles F and Pepper, John V. 2000. “Monotone Instrumental Variables: With an Application to the Returns to Schooling”. *Econometrica* 68 (4): 997–1010. (Cit. on pp. 3, 11).

Marschak, Jacob and Andrews, William H. 1944. “Random Simultaneous Equations and the Theory of Production”. *Econometrica* 12 (3): 143–205. (Cit. on p. 2).

Mundlak, Yair. 1961. “Empirical Production Function Free of Management Bias”. *Journal of Farm Economics* 43 (1): 44–56. (Cit. on p. 2).

Mundlak, Yair and Hoch, Irving. 1965. “Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions”. *Econometrica: Journal of the Econometric Society*: 814–828. (Cit. on p. 2).

Nevo, Aviv and Rosen, Adam M. 2012. “Identification with Imperfect Instruments”. *Review of Economics and Statistics* 94 (3): 659–671. (Cit. on p. 11).

Olley, G. Steven and Pakes, Ariel. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry”. *Econometrica* 64 (6): 1263. (Cit. on p. 1).

Semenova, Vira. 2017. “Machine Learning for Set-Identified Linear Models”. *Working Paper, arXiv:1712.10024*. (Cit. on pp. 16, 21).

Shaked, Moshe and Shanthikumar, J George. 2007. *Stochastic Orders*. Springer Science & Business Media. (Cit. on pp. 12, 13).

Taymaz, Erol and Yilmaz, Kamil. 2015. “Estimating Plant-level Marginal Costs and Mark-ups in the Turkish Manufacturing Industry”. *Middle East Technical University, ERC Working Paper*. (Cit. on p. 20).

A Proofs

A.1 Proof of Proposition 3.1

Proof. Under Assumption 2.2 productivity shock can be written as

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}$$

where $g(\cdot)$ is a monotone function as the distribution is stochastically increasing in ω_{it} by stochastic dominance condition in Assumption 2.2. Substituting this into Equation 3.2 and expanding the left hand side

$$\begin{aligned}
&\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big((g(\omega_{it-1}) - \omega_{it-1}) + \xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)^2\Big] + \mathbb{E}\Big[(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)^2\Big] + 2\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)\Big]
\end{aligned}$$

Expanding the right hand side similarly

$$\begin{aligned}
&\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1} + \xi_{it} + \epsilon_{jt} - \epsilon_{jt-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)^2\Big] + \mathbb{E}\Big[(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)^2\Big] + 2\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)\Big]
\end{aligned}$$

By the iid assumption the second expectations are equal to each

$$\mathbb{E}\Big[(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)^2\Big] = \mathbb{E}\Big[(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)^2\Big]$$

By the orthogonality of $(\epsilon_{it}, \epsilon_{it-1}, \xi_{it-1})$ to ω_{it-1} and iid assumption third expectations are equal to zero

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)\Big] = E\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)\Big] = 0$$

Therefore we need to show that

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)^2\Big] \leqslant \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)^2\Big]$$

Expanding both sides

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)^2\Big] = \mathbb{E}[g(\omega_{it-1})^2] + \mathbb{E}[\omega_{it-1}^2] - 2\mathbb{E}[g(\omega_{it-1})\omega_{it-1}]$$

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)^2\Big] = \mathbb{E}[g(\omega_{it-1})^2] + \mathbb{E}[\omega_{jt-1}^2] - 2\mathbb{E}[g(\omega_{it-1})\omega_{jt-1}]$$

The second and third moments are equal to each other by iid assumption

$$\mathbb{E}[g(\omega_{it-1})^2] = \mathbb{E}[g(\omega_{it-1})^2]$$
$$\mathbb{E}[\omega_{it-1}^2] = \mathbb{E}[\omega_{jt-1}^2]$$

So we need show that

$$\mathbb{E}[g(\omega_{it-1})\omega_{it-1}] \geqslant \mathbb{E}[g(\omega_{it-1})\omega_{jt-1}]$$

Observe that $\mathbb{E}[g(\omega_{it-1})\omega_{jt-1}] = 0$ again by iid assumption. For a random variable X, $\mathbb{E}[f(X)X] \geqslant 0$ for an increasing function f. Therefore

$$\mathbb{E}[g(\omega_{it-1})\omega_{it-1}] \geqslant 0$$

which gives the inequality in proposition.

$$\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \leqslant \mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big]$$

□

A.2 Proof of Proposition 3.2

Proof. Using Bayes rule for continuous random variables, I write the conditional probability distribution function of ω_{it} as (by changing the notation slightly)

$$f(\omega_{it} \mid k_{it-1}, i_{it} > z) = \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) f(\omega_{it} \mid k_{it-1}) f(k_{it-1})}{\Pr(i_{it} > z \mid k_{it-1}) f(k_{it-1})} = \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) f(\omega_{it} \mid k_{it-1})}{\Pr(i_{it} > z \mid k_{it-1})}$$

and similarly for $f(\omega_{it} \mid k_{it}, i_{it} < z)$. By taking the ratio of the two

$$\frac{f(\omega_{it} \mid k_{it-1}, i_{it} > z)}{f(\omega_{it} \mid k_{it-1}, i_{it} < z)} = \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})\Pr(i_{it} < z \mid k_{it-1})}{\Pr(i_{it} < z \mid k_{it-1}, \omega_{it})\Pr(i_{it} > z \mid k_{it-1})}$$
$$= \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})\Pr(i_{it} < z \mid k_{it-1})}{\big(1 - \Pr(i_{it} > z \mid k_{it-1}, \omega_{it})\big)\big(1 - \Pr(i_{it} < z \mid k_{it-1})\big)}$$

This function is increasing in $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$ if because numerator is increasing and denominator is decreasing in $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$. So if I can show that $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$ is weakly increasing in ω_{it} that imply

$$\frac{f(\omega_{it} \mid k_{it-1}, i_{it} > z)}{f(\omega_{it} \mid k_{it-1}, i_{it} < z)}$$

has the monotone likelihood ratio property. Therefore the next step is to show that $\Pr(i_{it} > z \mid k_{it}, \omega_{it})$ is increasing in ω_{it}. We write $\Pr(i_{it} > z \mid k_{it}, \omega_{it})$ as

$$\begin{aligned}\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) &= \int \mathbb{1}\{f(k_{it-1}, \omega_{it}, \xi_{it}) > z\} f(\xi_{it} | \omega_{it}, k_{it-1}) d\xi_{it} \\ &= \int \mathbb{1}\{f(k_{it-1}, \omega_{it}, \xi_{it}) > z\} f(\xi_{it} | k_{it-1}) d\xi_{it}\end{aligned}$$

where second line follows from Assumption 2.4. Since by assumption Assumption 2.5 $f(k_{i-1}, \omega_{it}, \xi_{it})$ is monotone in ω_{it}, I conclude that $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$ is increasing in ω_{it}. MLRP ratio implies the first order stochastic dominance and mean inequality, so other results follow. □

A.3 Proof of Proposition 5.2

Proof. From Equation (3.4) I can write the moment function at the true values of parameters as

$$m(w_{it}, \theta) = y_{it} - \theta_0 - \theta_k k_{it} - \theta_l l_{it} = g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it}$$

Substituting $m(w_{it}, \theta) = g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it}$ into Equation (6.1) we need to show that the following inequality holds

$$\mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it} \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it} \mid k_{it-2}, i_{it-1} < z] \geqslant 0$$

I proceed in two steps. First note that, by my assumptions ξ_{it}, ϵ_{it} and ηit are orthogonal to information set at $t-1$. Therefore we have

$$\begin{aligned}\mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it} \mid k_{it-2}, i_{it-1} > z] = \mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} > z] \\ \mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it} \mid k_{it-2}, i_{it-1} < z] = \mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} < z]\end{aligned}$$

To conclude the proof I need to show that

$$\mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} > z] \geqslant \mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} < z]$$

To see this, by proposition 2.1, we have a first order stochastic dominance relationship between ω_{it-1}'s conditional on k_{it-1} and investment is larger and smaller than a threshold

$$F_{\omega_{it-1}}(t \mid k_{it-2}, i_{it-1} > z) \geqslant F_{\omega_{it-1}}(t \mid k_{it-2}, i_{it-1} < z) \qquad \text{for all} \qquad t > 0$$

where $F_x(t|y)$ denotes probability distribution function of x conditional on y. Since $g(\omega)$ is a monotone function and stochastic order is preserved under monotone transformation this implies

$$F_{g(\omega_{it-1})}(t \mid k_{it-2}, i_{it-1} > z) \geqslant F_{g(\omega_{it-1})}(t \mid k_{it-2}, i_{it-1} < z) \qquad \text{for all} \qquad t > 0$$

which leads to the condition we wanted to show

$$\mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[g(\omega_{it-1} \mid k_{it-2}, i_{it-1} < z] \geqslant 0.$$

□