Normalizations and misspecification in skill formation models

Joachim Freyberger

University of Bonn

September, 2022

Introduction

- In structural models researcher often distinguish between two types of restrictions: assumptions and normalizations.
- A normalization typically fixes one or more parameters of the model to achieve point identification.
- When is a restriction a normalization?
- A normalization typically
  - affects many parameters of the model, but
  - should be without loss of generality.
- Example: Probit model where

$$Y = 1(\beta_1 + \beta_2 X \geq U),$$

$U \mid X \sim N(\mu, \sigma^2)$ and we set $\mu = 0$ and $\sigma = 1$.

Introduction

- A restriction might not affect parameters of interests, such as marginal effects in the probit model.
- I define a restriction to be a normalization if it does not affect the identified set of a function of interest.
- Implications:
    - The definition typically holds for some functions, but not others.
    - Whether a restriction is a normalization depends on the model.
    - If a normalization achieves point identification, then the function should be identified without the restriction.

Motivating example: skill formation models

- Estimate production functions of skills or other latent variables.
  - How do skills evolve over time?
  - What is the best timing of interventions?
- Data only contains noisy measures of skills without natural scales.
- Identification and estimation often proceed in two steps:
  - Identify the joint distribution of skills from the measurements.
  - Use that distribution to identify the production function.
- First step requires scale and location restrictions, but the production function imposes additional parametric assumptions.
- Are these restrictions (1) over-identifying and (2) normalization with respect to which parameters and counterfactuals?

Introduction

- I derive the identified sets without normalizations for different production functions under baseline assumptions, which shows:
    - Without additional restrictions, the model is not identified, but many important features are.
    - Additional restrictions needed depends on the production function.

Introduction

- I derive the identified sets without normalizations for different production functions under baseline assumptions, which shows:
  - Without additional restrictions, the model is not identified, but many important features are.
  - Additional restrictions needed depends on the production function.
- Trans-log: standard restrictions are not over-identifying.

$\Rightarrow$ Select an element of the identified set.

$\Rightarrow$ Parameters and certain counterfactuals are not invariant.

Introduction

- I derive the identified sets without normalizations for different production functions under baseline assumptions, which shows:
  - Without additional restrictions, the model is not identified, but many important features are.
  - Additional restrictions needed depends on the production function.
- Trans-log: standard restrictions are not over-identifying.

  $\Rightarrow$ Select an element of the identified set.

  $\Rightarrow$ Parameters and certain counterfactuals are not invariant.
- CES: standard scale restrictions are over-identifying.

  $\Rightarrow$ Estimates are biased - results depend on units of measurement.

  $\Rightarrow$ Need less restrictions and a more flexible estimator.

Outline

In this talk:

- Introduction
- Simplified skill formation models
  - Trans-log
  - CES
- Conclusion

Additional results in the paper:

- Normalization definition and examples
- Full skill formation model and technical details.
- Monte Carlo simulations

Literature

- **Normalizations:**
  - Factor models: Anderson and Rubin (1956), Madansky (1964), Cunha and Heckman (2008), Williams (2020), ...
  - General discussions: Matzkin (1994, 2007), Lewbel (2019), ...
  - Invariance: Aguirregabiria and Suzuki (2014), Komarova, Sanches, and Junior (2017), Kalouptsidi, Scott, and Souza-Rodrigues (2020) ...
  - Estimation: Hamilton, Waggoner, and Zha (2007), Chiappori, Komunjer, and Kristensen (2015), Gao and Li (2019), ...
- **Skill formation:**
  - Cunha and Heckman (2008), Cunha, Heckman, and Schennach (2010)
  - Applications: Helmers and Patnam (2011), Fiorini and Keane (2014), Attanasio, Meghir, and Nix (2019), Aucejo and James (2019), ...
  - Specification issues: Agostinelli and Wiswall (2016, 2017, 2020), Del Bono, Kinsler, and Pavan (2020). ...

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$\begin{aligned}
\theta_{t+1} &= f(\theta_t, I_t, \delta_t, \eta_{\theta,t}) && t = 0, \ldots, T-1 \\
Z_{\theta,t,m} &= \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} && t = 0, \ldots, T, m = 1,2,3 \\
Z_{I,t,m} &= \mu_{I,t,m} + \lambda_{I,t,m} \ln I_t + \varepsilon_{I,t,m} && t = 0, \ldots, T-1, m = 1,2,3 \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q && \\
\ln I_t &= \beta_{0,t} + \beta_{1,t} \ln \theta_t + \beta_{2,t} \ln Y_t + \eta_{I,t} && t = 0, \ldots, T-1
\end{aligned}$$

- We observe $Z_{\theta,t,m}$, $Z_{I,t,m}$, Y_t, and Q, but not θ_t and I_t.
- Other unobervables: $\eta_{\theta,t}$, $\varepsilon_{\theta,t,m}$, $\varepsilon_{I,t,m}$, $\eta_{I,t}$, and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, $\mu_{I,t,m}$, $\lambda_{I,t,m}$, β_t, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$
\begin{aligned}
\theta_{t+1} &= f(\theta_t, I_t, \delta_t, \eta_{\theta,t}) && t = 0, \ldots, T-1 \\
Z_{\theta,t,m} &= \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} && t = 0, \ldots, T, m = 1,2,3 \\
Z_{I,t,m} &= \mu_{I,t,m} + \lambda_{I,t,m} \ln I_t + \varepsilon_{I,t,m} && t = 0, \ldots, T-1, m = 1,2,3 \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q && \\
\ln I_t &= \beta_{0,t} + \beta_{1,t} \ln \theta_t + \beta_{2,t} \ln Y_t + \eta_{I,t} && t = 0, \ldots, T-1
\end{aligned}
$$

- We observe $Z_{\theta,t,m}$, $Z_{I,t,m}$, Y_t, and Q, but not θ_t and I_t.
- Other unobervables: $\eta_{\theta,t}$, $\varepsilon_{\theta,t,m}$, $\varepsilon_{I,t,m}$, $\eta_{I,t}$, and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, $\mu_{I,t,m}$, $\lambda_{I,t,m}$, β_t, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$
\begin{aligned}
\theta_{t+1} &= f(\theta_t, I_t, \delta_t) && t = 0, \ldots, T-1 \\
Z_{\theta,t,m} &= \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} && t = 0, \ldots, T, m = 1, 2, 3 \\
Z_{I,t,m} &= \mu_{I,t,m} + \lambda_{I,t,m} \ln I_t + \varepsilon_{I,t,m} && t = 0, \ldots, T-1, m = 1, 2, 3 \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q
\end{aligned}
$$

- We observe $Z_{\theta,t,m}$, $Z_{I,t,m}$, and Q, but not θ_t and I_t.
- Other unobervables: $\varepsilon_{\theta,t,m}$, $\varepsilon_{I,t,m}$, and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, $\mu_{I,t,m}$, $\lambda_{I,t,m}$, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$\begin{aligned}
\theta_{t+1} &= f(\theta_t, I_t, \delta_t) && t = 0, \ldots, T-1 \\
Z_{\theta,t,m} &= \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} && t = 0, \ldots, T, m = 1,2,3 \\
Z_{I,t,m} &= \mu_{I,t,m} + \lambda_{I,t,m} \ln I_t + \varepsilon_{I,t,m} && t = 0, \ldots, T-1, m = 1,2,3 \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q
\end{aligned}$$

- We observe $Z_{\theta,t,m}$, $Z_{I,t,m}$, and Q, but not θ_t and I_t.
- Other unobervables: $\varepsilon_{\theta,t,m}$, $\varepsilon_{I,t,m}$, and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, $\mu_{I,t,m}$, $\lambda_{I,t,m}$, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$\theta_{t+1} = f(\theta_t, I_t, \delta_t) \qquad t = 0, \ldots, T-1$$

$$Z_{\theta,t,m} = \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} \qquad t = 0, \ldots, T, m = 1, 2, 3$$

$$Q = \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q$$

- We observe $Z_{\theta,t,m}$, I_t, and Q, but not θ_t.
- Other unobervables: $\varepsilon_{\theta,t,m}$ and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$\theta_{t+1} = f(\theta_t, I_t, \delta_t) \qquad t = 0, \ldots, T-1$$

$$Z_{\theta,t,m} = \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} \qquad t = 0, \ldots, T, m = 1, 2, 3$$

$$Q = \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q$$

- We observe $Z_{\theta,t,m}$, I_t, and Q, but not θ_t.
- Other unobervables: $\varepsilon_{\theta,t,m}$ and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$\theta_{t+1} = f(\theta_t, I_t, \delta_t) \qquad t = 0, \ldots, T-1$$

$$Z_{\theta,t} = \mu_{\theta,t} + \lambda_{\theta,t} \ln \theta_t + \varepsilon_{\theta,t} \qquad t = 0, \ldots, T$$

$$Q = \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q$$

- We observe $Z_{\theta,t,m}$, I_t, and Q, but not θ_t.
- Other unobervables: $\varepsilon_{\theta,t}$ and η_Q.
- Parameters: δ_t, $\mu_{\theta,t}$, $\lambda_{\theta,t}$, and α.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- Simplified model:

$$
\begin{aligned}
\theta_{t+1} &= f(\theta_t, I_t, \delta_t) && t = 0, \ldots, T-1 \\
Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t && t = 0, \ldots, T \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q
\end{aligned}
$$

- We observe $\{Z_t\}_{t=0}^{T}$,$\{I_t\}_{t=0}^{T-1}$ and Q, but not $\{\theta_t, \varepsilon_t\}_{t=0}^{T}$ and η_Q.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- Simplified model:

$$\begin{aligned} \theta_{t+1} &= f(\theta_t, I_t, \delta_t) & t &= 0, \ldots, T-1 \\ Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t & t &= 0, \ldots, T \\ Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q \end{aligned}$$

- We observe $\{Z_t\}_{t=0}^{T}$,$\{I_t\}_{t=0}^{T-1}$ and Q, but not $\{\theta_t, \varepsilon_t\}_{t=0}^{T}$ and η_Q.
- Trans-log:

$$\ln \theta_{t+1} = a_t + \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} \ln \theta_t \ln I_t$$

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- Simplified model:

$$\begin{aligned} \theta_{t+1} &= f(\theta_t, I_t, \delta_t) & t &= 0, \dots, T-1 \\ Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t & t &= 0, \dots, T \\ Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q \end{aligned}$$

- We observe $\{Z_t\}_{t=0}^{T}$,$\{I_t\}_{t=0}^{T-1}$ and Q, but not $\{\theta_t, \varepsilon_t\}_{t=0}^{T}$ and η_Q.
- Trans-log:

$$\ln \theta_{t+1} = a_t + \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} \ln \theta_t \ln I_t$$

- Constant Elasticity of Substitution (CES):

$$\theta_{t+1} = \left(\gamma_{1t} \theta_t^{\sigma_t} + \gamma_{2t} I_t^{\sigma_t}\right)^{\frac{\psi_t}{\sigma_t}}$$

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- Simplified model:

$$\begin{aligned} \theta_{t+1} &= f(\theta_t, I_t, \delta_t) && t = 0, \ldots, T-1 \\ Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t && t = 0, \ldots, T \\ Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q \end{aligned}$$

- We observe $\{Z_t\}_{t=0}^{T}$,$\{I_t\}_{t=0}^{T-1}$ and Q, but not $\{\theta_t, \varepsilon_t\}_{t=0}^{T}$ and η_Q.
- Trans-log:

$$\ln \theta_{t+1} = a_t + \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} \ln \theta_t \ln I_t$$

- Constant Elasticity of Substitution (CES):

$$\theta_{t+1} = \left(\gamma_{1t} \theta_t^{\sigma_t} + \gamma_{2t} I_t^{\sigma_t}\right)^{\frac{\psi_t}{\sigma_t}}$$

- Nonparametric model in the paper.

Skill formation models

- The model consists of

$$\begin{aligned} \theta_{t+1} &= f(\theta_t, I_t, \delta_t) && t = 0, \ldots, T-1 \\ Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t && t = 0, \ldots, T \\ Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q \end{aligned}$$

- To isolate the main issues, I assume that $(\varepsilon_0, \ldots, \varepsilon_T, \eta)$ are independent of $(I_0, \ldots, I_{T-1}, \theta_0, \ldots, \theta_T)$ and that the joint distribution is known.

- Easy to show: The joint distribution of

$$(\mu_0 + \lambda_0 \ln \theta_0, \ldots, \mu_T + \lambda_T \ln \theta_T, \alpha_0 + \alpha_1 \ln \theta_T)$$

  is identified conditional on $(I_0, \ldots, I_{T-1})$, but δ_t, μ_t, and λ_t are not.

- Characterize identified set and discuss additional restrictions.

Trans-log production function

- First consider

$$\begin{aligned}
\ln\theta_{t+1} &= a_t + \gamma_{1t}\ln\theta_t + \gamma_{2t}\ln I_t + \gamma_{3t}\ln\theta_t\ln I_t \\
Z_t &= \mu_t + \lambda_t\ln\theta_t + \varepsilon_t \\
Q &= \alpha_0 + \alpha_1\ln\theta_T + \eta_Q
\end{aligned}$$

- Main idea: Rewrite the model in terms of $\ln\tilde{\theta}_t = \mu_t + \lambda_t\ln\theta_t$.

Trans-log production function

- First consider

$$\begin{aligned} \ln\theta_{t+1} &= a_t + \gamma_{1t}\ln\theta_t + \gamma_{2t}\ln I_t + \gamma_{3t}\ln\theta_t \ln I_t \\ Z_t &= \mu_t + \lambda_t \ln\theta_t + \varepsilon_t \\ Q &= \alpha_0 + \alpha_1 \ln\theta_T + \eta_Q \end{aligned}$$

- Main idea: Rewrite the model in terms of $\ln\tilde{\theta}_t = \mu_t + \lambda_t \ln\theta_t$.

- We can only identify certain combinations of parameters, such as

$$\frac{\lambda_{t+1}}{\lambda_t}\gamma_{1t}, \quad \lambda_{t+1}\left(\gamma_{2t} - \frac{\mu_t}{\lambda_t}\gamma_{3t}\right), \quad \frac{\lambda_{t+1}}{\lambda_t}\gamma_{3t}$$

- We cannot distinguish between changes in the quality of the measurements (λ_{t+1}/λ_t) and changes in the technology (γ_{1t}).

- Even when $\gamma_{3t} = 0$, we can only identify $\lambda_{t+1}\gamma_{2t}$.

Trans-log production function

- When

$$\ln\theta_{t+1} = a_t + \gamma_{1t}\ln\theta_t + \gamma_{2t}\ln I_t + \gamma_{3t}\ln\theta_t\ln I_t$$

three additional assumptions are commonly used in the literature.

Assumption 2

$\mu_0 = 0$ and $\lambda_0 = 1$.

Assumption 3

$a_t = 0$ and $\gamma_{1t} + \gamma_{2t} + \gamma_{3t} = 1$ for all $t = 0, \ldots, T-1$.

Assumption 4

$\lambda_t = \lambda_{t+1}$ and $\mu_t = \mu_{t+1}$ for all $t = 0, \ldots, T-1$.

Trans-log production function

- All parameters are point identified under the baseline assumption and either
  - Assumption 2 and 3 or
  - Assumption 2 and 4
- Restrictions are not overidentifying and observationally equivalent.
- Many estimated parameters are not invariant to the restriction $\lambda_0 = 1$ or changes in the units of measurement of the data, including
  - the elasticity of investment and
  - counterfactuals that depend on the level of skills, such as investment sequences that maximize $E(\theta_T)$.

Invariant features

- Some features that are identified without Assumptions 2 – 4 are:
  - $F_{\ln \theta_{t+1}}(a_t + \gamma_{1t} \ln Q_\alpha(\theta_t) + \gamma_{2t} \ln I + \gamma_{3t} \ln Q_\alpha(\theta_t) \ln I)$
  - $P(Q \leq q \mid I_0, \ldots, I_{T-1}, \theta_0 = Q_\alpha(\theta_0))$
- With these features, we can calculate:
  - Investment sequence that maximizes the skill rank in the final period or a feature of the distribution of the adult outcome.
  - Calculate skill ranks or distributions of Q for different counterfactual investments to compare means, variances, and heterogeneous effects.

CES production function

- Consider

$$\begin{aligned} \theta_{t+1} &= \left(\gamma_{1t}\theta_t^{\sigma_t} + \gamma_{2t} I_t^{\sigma_t}\right)^{\psi_t/\sigma_t} \\ Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t \\ Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q \end{aligned}$$

- Many parameters are point identified, including $\{\lambda_t, \sigma_t\}_{t=0}^{T-1}$.
- Point identification under various restrictions:
  - Fix μ_0 and use age-invariant μ_t or $\gamma_{1t} + \gamma_{2t} = 1$.
  - Parameters depend on the set of restrictions and μ_0.
- Setting $\lambda_0 = 1$ yields inconsistent estimators.
- Same features as before are identified under the baseline assumption.
- Features can be estimated using a more flexible estimator.

Monte Carlo simulations

- Setup adapted from Attanasio, Meghir, and Nix (2019) with $T = 2$ and

$$\theta_{t+1} = A_t \left(\gamma_t \theta_t^{\sigma_t} + (1-\gamma_t)\, I_t^{\sigma_t}\right)^{\frac{1}{\sigma_t}} \exp(\eta_{\theta,t})$$
$$\ln I_t = \beta_{1t} \ln \theta_t + \beta_{2t} \ln Y + \eta_{I,t}$$

- Three (linear) measurements each for $\ln \theta_t$ and $\ln I_t$ and the loading of the first measure is equal to 1 in all periods.

- Simulate $(\ln \theta_0, \ln Y)$ from a normal mixture and generate skills and investment recursively.

- Estimation in AMN proceeds in three steps:
  (0) Fit a normal mixture distribution using all measures.
  (1) Estimate distr. of skills, investments, and Y using restrictions.
  (2) Take a sample and estimate the production function by NLLS.

Monte Carlo simulations

- In this setting $\frac{\lambda_{I,t,1}}{\lambda_{\theta,t,1}}$ and $\frac{\lambda_{\theta,t+1,1}}{\lambda_{\theta,t,1}}$ are identified.
    - Only need to normalize $\lambda_{I,0,1}$, which I set to 1.
    - It is common to set $\lambda_{\theta,t,1} = 1$ and $\lambda_{I,t,1} = 1$ for all t.
- To estimate the scales, we need to relax the CES specification and use

$$\theta_{t+1} = \left(\gamma_{1t}\theta_t^{\sigma_t} + \gamma_{2t} I_t^{\frac{\sigma_t}{\lambda_{I,t,1}}} \right)^{\frac{1}{\sigma_t}}$$

- Generated data by

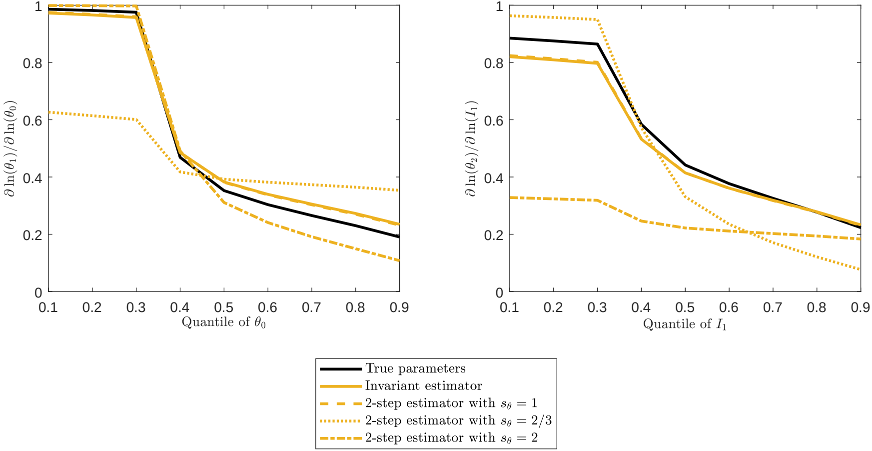
$$\tilde{Z}_{\theta,t,1} = \log(\theta_t) + \varepsilon_{\theta,t,1}$$

  but estimate the model using

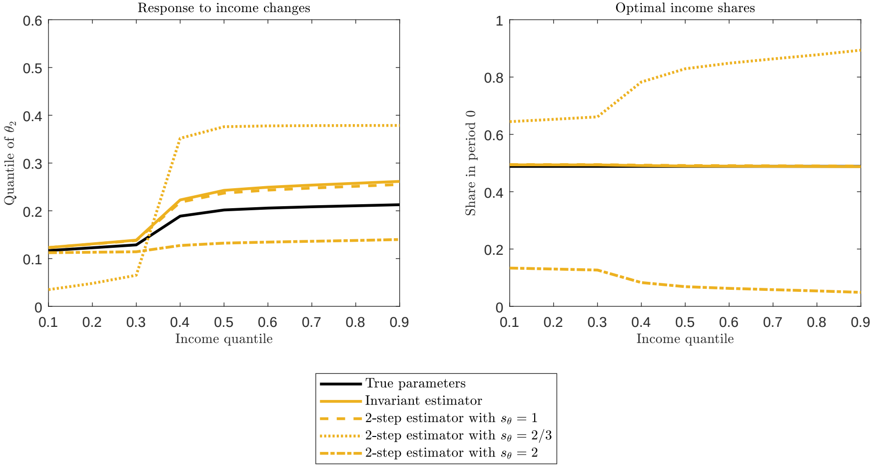
$$Z_{\theta,t,1} = s_\theta \tilde{Z}_{\theta,t,1} = s_\theta \log(\theta_t) + s_\theta \varepsilon_{\theta,t,1}.$$

- Parameters will be affected but with my estimator invariant features, including elasticities in this setting, are not.

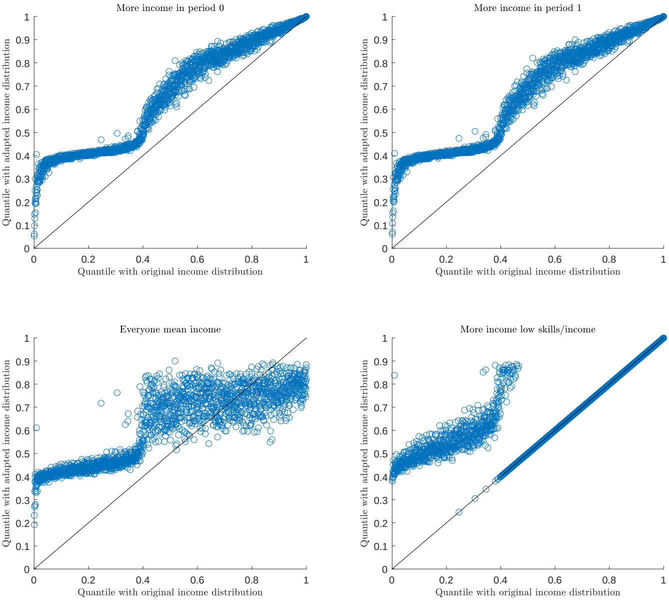
Elasticities



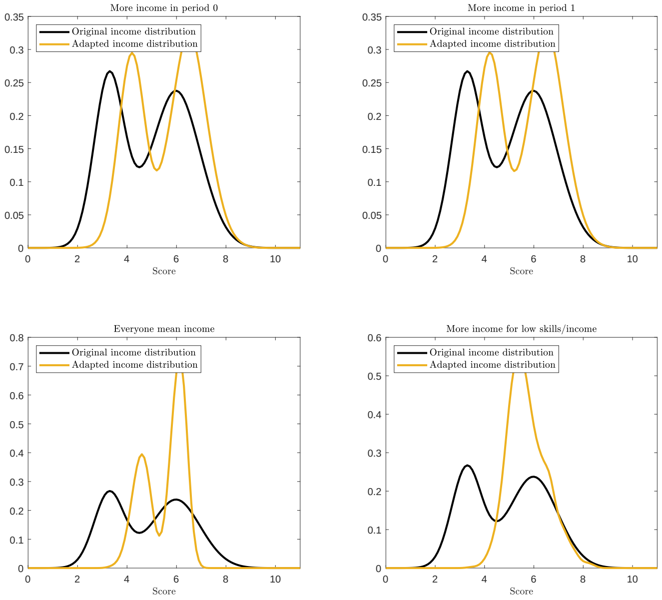
Optimal investment



Different investment strategies



Different investment strategies



Conclusion

- The paper provides a formal definition of a normalization.
  - Whether or not a restriction is a normalization depends on the model specification and the object of interest.
  - Researchers should argue that a restriction is truly a normalization.
- In skill formation models, seemingly innocuous restrictions are not normalizations and can affect parameters and counterfactuals.
  - Simply changing the units of measurements can yield ineffective investment strategies and misleading policy recommendations.
  - Key features are invariant to these restrictions, are identified under weaker assumptions, and provide robust policy implications.

Skill formation models

- Let θ_t denote skills at time t and let I_t be investment at time t.
- I consider the model:

$$\begin{aligned}
\theta_{t+1} &= f(\theta_t, I_t, \delta_t, \eta_{\theta,t}) && t = 0, \ldots, T-1 \\
Z_{\theta,t,m} &= \mu_{\theta,t,m} + \lambda_{\theta,t,m} \ln \theta_t + \varepsilon_{\theta,t,m} && t = 0, \ldots, T, m = 1, 2, 3 \\
Z_{I,t,m} &= \mu_{I,t,m} + \lambda_{I,t,m} \ln I_t + \varepsilon_{I,t,m} && t = 0, \ldots, T-1, m = 1, 2, 3 \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q && \\
\ln I_t &= \beta_{0,t} + \beta_{1,t} \ln \theta_t + \beta_{2,t} \ln Y_t + \eta_{I,t} && t = 0, \ldots, T-1
\end{aligned}$$

- We observe $Z_{\theta,t,m}$, $Z_{I,t,m}$, Y_t, and Q, but not θ_t and I_t.
- Other unobervables: $\eta_{\theta,t}$, $\varepsilon_{\theta,t,m}$, $\varepsilon_{I,t,m}$, $\eta_{I,t}$, and η_Q.
- Parameters: δ_t, $\mu_{\theta,t,m}$, $\lambda_{\theta,t,m}$, $\mu_{I,t,m}$, $\lambda_{I,t,m}$, β_t, and α.

Normalization

- $\theta_0 \in \Theta$ is the true parameter and Θ is the parameter space.
- Let Z contain all observed random variables, such as Y and X.
- Model with $\theta \in \Theta$ generates a joint distribution $P(Z, \theta)$.
- $\theta_1, \theta_2 \in \Theta$ are observationally equivalent if $P(Z, \theta_1) = P(Z, \theta_2)$.
- The identified set for θ_0 is

$$\Theta_0 = \{\theta \in \Theta : P(Z, \theta) = P(Z, \theta_0)\}.$$

- The identified set for $g(\theta_0)$ is

$$\Theta_{g_0} = \{g(\theta) : \theta \in \Theta_0\}.$$

- $g(\theta_0)$ could be point identified even if θ_0 is not.

Normalization

- A normalization is a restriction of the form $\theta \in \Theta_N$, where $\Theta_N \subseteq \Theta$ is a known set.
- Typically, $\Theta_0 \cap \Theta_N \neq \Theta_0$ and $\theta_0 \notin \Theta_N$.

Normalization

- A normalization is a restriction of the form $\theta \in \Theta_N$, where $\Theta_N \subseteq \Theta$ is a known set.
- Typically, $\Theta_0 \cap \Theta_N \neq \Theta_0$ and $\theta_0 \notin \Theta_N$.

Definition

The restriction $\theta \in \Theta_N$ is a *normalization* with respect to $g(\theta_0)$ if for any $\theta_0 \in \Theta$

$$\{g(\theta) : \theta \in \Theta_0 \cap \Theta_N\} = \{g(\theta) : \theta \in \Theta_0\}.$$

Normalization

- A normalization is a restriction of the form $\theta \in \Theta_N$, where $\Theta_N \subseteq \Theta$ is a known set.
- Typically, $\Theta_0 \cap \Theta_N \neq \Theta_0$ and $\theta_0 \notin \Theta_N$.

Definition

The restriction $\theta \in \Theta_N$ is a *normalization* with respect to $g(\theta_0)$ if for any $\theta_0 \in \Theta$

$$\{g(\theta) : \theta \in \Theta_0 \cap \Theta_N\} = \{g(\theta) : \theta \in \Theta_0\}.$$

- Typically, $\Theta_0 \cap \Theta_N$ is a singleton. The definition then requires that $g(\theta_0)$ is point identified, even without a normalization.

Simple example

- Consider the probit model

$$Y = 1(\beta_{0,1} + \beta_{0,2} X \geq U),$$

where $var(X) > 0$, $U \mid X \sim N(\mu_0, \sigma_0^2)$ and $\sigma_0^2 > 0$.

- We have $\theta_0 = (\beta_{0,1}, \beta_{0,2}, \mu_0, \sigma_0)'$ and $Z = (Y, X)$ and since

$$P(Y = 1 \mid X = x) = \Phi\left(\frac{\beta_{0,1} - \mu_0}{\sigma_0} + \frac{\beta_{0,2}}{\sigma_0} x\right),$$

we get

$$\Theta_0 = \left\{\theta \in \mathbb{R}^4 : \frac{\beta_1 - \mu}{\sigma} = \frac{\beta_{0,1} - \mu_0}{\sigma_0} \text{ and } \frac{\beta_2}{\sigma} = \frac{\beta_{0,2}}{\sigma_0}\right\}.$$

- Common restriction is $\mu = 0$ and $\sigma = 1$ so that $\Theta_N = \mathbb{R}^2 \times 0 \times 1$ and

$$\Theta_0 \cap \Theta_N = \left(\frac{\beta_{0,1} - \mu_0}{\sigma_0}, \frac{\beta_{0,2}}{\sigma_0}, 0, 1\right).$$

- Normalization with respect to marginal effects, but not $(\beta_{0,1}, \beta_{0,2})$.

Asssumptions

Assumption 1

- (a) $(\varepsilon_0, \ldots, \varepsilon_T, \eta_Q)$ are independent of $(I_0, \ldots, I_{T-1}, \theta_0, \ldots, \theta_T)$.
- (b) The joint distribution of $(\varepsilon_0, \ldots, \varepsilon_T, \eta_Q)$ is known and the cf of $(\varepsilon_0, \ldots, \varepsilon_T, \eta_Q)$ only has isolated zeros.
- (c) $\lambda_t > 0$ for all t.

Trans-log production function

- First consider

$$\ln \theta_{t+1} = a_t + \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} \ln \theta_t \ln I_t.$$

and define $\ln \tilde{\theta}_t = \mu_t + \lambda_t \ln \theta_t$.

Trans-log production function

- First consider

$$\ln \theta_{t+1} = a_t + \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} \ln \theta_t \ln I_t.$$

and define $\ln \tilde{\theta}_t = \mu_t + \lambda_t \ln \theta_t$.

- Here I set a_t, μ_t, and α_0 to 0 and focus on the scale issue.

Trans-log production function

- First consider

$$\ln \theta_{t+1} = \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} \ln \theta_t \ln I_t.$$

and define $\ln \tilde{\theta}_t = \lambda_t \ln \theta_t$.

- Here I set a_t, μ_t, and α_0 to 0 and focus on the scale issue.

Trans-log production function

- First consider

$$\ln\theta_{t+1} = \gamma_{1t}\ln\theta_t + \gamma_{2t}\ln I_t + \gamma_{3t}\ln\theta_t\ln I_t.$$

  and define $\ln\tilde{\theta}_t = \lambda_t\ln\theta_t$.

- Here I set a_t, μ_t, and α_0 to 0 and focus on the scale issue.

- We can then rewrite the production function in terms of $\tilde{\theta}_t$ and get

$$\begin{aligned} \ln\tilde{\theta}_{t+1} &= \tilde{\gamma}_{1t}\ln\tilde{\theta}_t + \tilde{\gamma}_{2t}\ln I_t + \tilde{\gamma}_{3t}\ln\tilde{\theta}_t\ln I_t \\ Z_t &= \ln\tilde{\theta}_t + \varepsilon_t \\ Q &= \tilde{\alpha}_1\ln\tilde{\theta}_T + \eta_Q \end{aligned}$$

  where

$$\tilde{\gamma}_{1t} = \frac{\lambda_{t+1}}{\lambda_t}\gamma_{1t}, \quad \tilde{\gamma}_{2t} = \lambda_{t+1}\gamma_{2t}, \quad \tilde{\gamma}_{3t} = \frac{\lambda_{t+1}}{\lambda_t}\gamma_{3t}, \quad \tilde{\alpha}_1 = \frac{\alpha_1}{\lambda_T}$$

Anchor

- All results are qualitatively identical when we anchor the skills at Q.
  - Let $Q = \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q$.
  - Impose $\alpha_0 = 0$ and $\alpha_1 = 1$ instead of $\mu_0 = 0$ and $\lambda_0 = 1$.

CES production function

- Consider

$$
\begin{aligned}
\theta_{t+1} &= \left(\gamma_{1t}\theta_t^{\sigma_t} + \gamma_{2t} I_t^{\sigma_t}\right)^{\psi_t/\sigma_t} \\
Z_t &= \mu_t + \lambda_t \ln \theta_t + \varepsilon_t \\
Q &= \alpha_0 + \alpha_1 \ln \theta_T + \eta_Q
\end{aligned}
$$

- Define $\tilde{\theta}_t = \exp(\mu_t)\theta_t^{\lambda_t}$ or $\ln \tilde{\theta}_t = \mu_t + \lambda_t \ln \theta_t$.

- We can then rewrite

$$
\begin{aligned}
\tilde{\theta}_{t+1} &= \left(\tilde{\gamma}_{1t}\tilde{\theta}_t^{\sigma_t/\lambda_t} + \tilde{\gamma}_{2t} I_t^{\sigma_t}\right)^{\lambda_{t+1}\psi_t/\sigma_t} \\
Z_t &= \ln \tilde{\theta}_t + \varepsilon_t \\
Q &= \tilde{\alpha}_0 + \tilde{\alpha}_1 \ln \tilde{\theta}_T + \eta_Q
\end{aligned}
$$

where the transformed parameters have complicated expressions.

CES production function

Theorem 3

Suppose Assumption 1 holds.

a. $\{\lambda_t, \sigma_t\}_{t=0}^{T-1}$ and $\{\psi_t\}_{t=0}^{T-2}$ are point identified.
b. Some complicated functions are point identified as well.
c. Observationally equivalence of any two sets of parameters.

Corollary 2

Suppose Assumption 1 holds.

a. $\{\mu_t\}_{t=0}^{T-1}$, $\{\gamma_{1t}, \gamma_{2t}\}_{t=1}^{T-2}$ are identified if $\gamma_{1t} + \gamma_{2t} = 1$ and $\mu_0 = 0$.
b. $\{\mu_t\}_{t=0}^{T-1}$, $\{\gamma_{1t}, \gamma_{2t}\}_{t=1}^{T-2}$ are identified if $\mu_t = \mu_{t+1}$ and $\mu_0 = 0$.
c. Full identification if λ_T is identified, e.g. if $\lambda_t = \lambda_{t+1}$ or $\psi_t = 1$.

CES production function

- In the trans-log case, setting $\lambda_0 = 1$ or $\lambda_0 = 10$ yields different estimated parameters, but observationally equivalent models.
- In the CES case, λ_0 is point identified and we cannot fix it.
- When we set $\lambda_0 = 1$, which is a standard restriction, multiplying all measures by a constant affects everything.
- The exact bias depends on how the parameters are estimated.
- To obtain point identification, we need to fix μ_0 and this restriction affects certain features, depending on the other assumptions.
- Again, anchoring yields similar conclusions.
- Same features as before are identified under Assumption 1 only.

Invariant features for the CES production function

Theorem 4

Suppose Assumption 1 holds. Then the following features are identified.

1. $F_{\ln \theta_{t+1}}((\gamma_{1t}\theta_t^{\sigma_t} + \gamma_{2t}I^{\sigma_t})^{\psi_t/\sigma_t})$
2. $P(Q \leq y \mid I_0, \ldots, I_{T-1}, \theta_0 = Q_\alpha(\theta_0))$
3. $\int P(Q \leq y \mid I_0, \ldots, I_{T-1}, \theta_0 = \theta) f_{\theta_0}(\theta) d\theta$
4. Sequences of investment that maximize known strictly increasing functions of $\ln \theta_T$ subject to $\theta_0 = Q_\alpha(\theta_0)$ and $\sum_{t=1}^{T-1} I_t = c$.
5. Sequences of investment that maximize linear functions of $E(\ln \theta_T)$ or $E(\theta_T)$ subject to $\sum_{t=1}^{T-1} I_t = c$.