Discussion of
“Productivity and Misallocation in General Equilibrium”
by David Baqae and Emmanuel Farhi

Alireza Tahbaz-Salehi
Northwestern University

NBER Monetary Economics Program Meeting
March 2018
Hulten (1978)

- In an efficient economy, the macro impact of a shock to industry \( i \) is determined by that industry’s Domar weight

\[
\frac{d \log \text{GDP}}{d \log A_i} = \frac{p_i y_i}{\text{GDP}}
\]

up to a first-order approximation.

- Extremely powerful result because
  
  (i) it is agnostic with respect to the microeconomic structure of the economy.
  (ii) can be calculated using observables.

- Two key qualifiers:
  
  • Efficiency is necessary: Hulten is a consequence of envelope theorem.
  • First-order approximations are not useful in non-linear economies.
This Paper

- A generalization of Hulten's Theorem to inefficient economies. how micro productivity and markup shocks shape aggregate outcomes

- A parametric model to relate various terms to structural parameters (micro elasticities, returns to scale, input-output linkages, etc.)

- **Key take-aways:**
  - Micro shocks impact macro outcomes via two channels: (1) a pure technology effect and (2) a reallocation effect.
  - The latter can be measured via changes in factor income shares.
  - Unlike efficient economies, all micro intricacies become important.
Key Idea

- Allocation matrix: $\mathcal{X}(A, \mu)$, where $\mathcal{X}_{ij} = \frac{x_{ij}}{y_i}$ be the allocation of inputs across various firms.
- GDP = $\mathcal{Y}(A, \mathcal{X})$

- Chain rule:

$$\frac{d \log GDP}{d \log A} = \frac{\partial \log \mathcal{Y}}{\partial \log A} \Delta \text{technology} + \frac{\partial \log \mathcal{Y}}{\partial \mathcal{X}} \frac{d \mathcal{X}}{d \log A} \Delta \text{allocative efficiency}$$

- If the initial allocation is efficient, then by the envelope theorem, the second term is equal to zero $\rightarrow$ Hulten’s Theorem (almost).
- In an inefficient economy, one needs to understand how the second term responds to shocks.
Framework

- $n$ goods, each produced by competitive producers using intermediate goods as well as $F$ factors that are inelastically supplied.

- Producers: constant-returns cost functions

$$\frac{1}{A_i} C_i \left( p_1, \ldots, p_n, w_1, \ldots, w_F \right) y_i$$

- Markups:

$$p_i = \frac{\mu_i C_i}{A_i}$$

- Final demand

$$Y = \max \ D(c_1, \ldots, c_n)$$

s.t. $$\sum_{i=1}^{n} p_i c_i = \sum_{f=1}^{F} w_f L_f + \sum_{i=1}^{n} \pi_i$$
Standard Objects

- Input-output matrix:
  \[ \Omega_{ij} = \frac{p_j x_{ij}}{p_i y_i} \]

- Final expenditure shares:
  \[ b_i = \frac{p_i c_i}{GDP} \]

- Domar weights:
  \[ \lambda_i = \frac{p_i y_i}{GDP} \]

- Leontief inverse:
  \[ \Psi = (I - \Omega)^{-1} = I + \Omega + \Omega^2 + \ldots \]

- Market-clearing:
  \[ \lambda_i = \sum_{k=1}^{n} b_k \Psi_{ki} \]
Standard Objects

- Input-output matrix:
  \[ \Omega_{ij} = \frac{p_j x_{ij}}{p_i y_i} \]

- Final expenditure shares:
  \[ b_i = \frac{p_i c_i}{GDP} \]

- Domar weights:
  \[ \lambda_i = \frac{p_i y_i}{GDP} \]

- Leontief inverse:
  \[ \Psi = (I - \Omega)^{-1} = I + \Omega + \Omega^2 + \ldots \]

- Market-clearing:
  \[ \lambda_i = \sum_{k=1}^{n} b_k \Psi_{ki} \]
Markup-Adjusted Objects

• Markup adjusted (“cost-based”) input-output matrix:

\[ \tilde{\Omega}_{ij} = \mu_i \Omega_{ij} \]

• Markup adjusted (“cost-based”) Leontief inverse:

\[ \tilde{\Psi} = (I - \tilde{\Omega})^{-1} \]

• “Cost-based” vs. “revenue-based” Domar weights:

\[ \lambda_i = \sum_{k=1}^{n} b_k \Psi_{ki} \]
\[ \tilde{\lambda}_i = \sum_{k=1}^{n} b_k \tilde{\Psi}_{ki} \]
Main Result

- Key result of the paper: apply envelope theorem to each producer and combine with chain rule:

\[
\frac{d \log Y}{d \log A} d \log A = \frac{\partial \log Y}{\partial \log A} d \log A + \frac{\partial \log Y}{\partial \mathcal{X}} d \mathcal{X}
\]

**Theorem**

*Pure technology effect:*

\[
\frac{d \log Y}{d \log A} d \log A = \sum_{k=1}^{n} \tilde{\lambda}_k d \log A_k
\]

*Resource reallocation:*

\[
\frac{\partial \log Y}{\partial \mathcal{X}} d \mathcal{X} = - \sum_{k=1}^{n} \tilde{\lambda}_k d \log \mu_k - \sum_{f=1}^{F} \tilde{\Lambda}_f d \log \Lambda_f.
\]

- Reduces to Hulten's theorem when \(\mu_k = 1\) for all \(k\).
Example 1: Horizontal Economy + Productivity Shocks

- Elasticity of substitution between various goods = $\theta$.

- “Pure” technology effect = $\lambda_k$
  - Holding the allocation of labor to each firm constant, a productivity shock to firm $k$ increases its output.
  - This in turn increases aggregate output.
Example 1: Horizontal Economy + Productivity Shocks

- Elasticity of substitution between various goods = $\theta$.

- “Pure” technology effect = $\lambda_k$
  - Holding the allocation of labor to each firm constant, a productivity shock to firm $k$ increases its output.
  - This in turn increases aggregate output.
Example 1: Horizontal Economy + Productivity Shocks

- Reallocation effect = \(-\lambda_k (\theta - 1) \left( \frac{\mu_k^{-1}}{\sum_i \lambda_i \mu_i^{-1}} - 1 \right)\)

- Shock to firm \(k\) reduces its price and increases its demand when \(\theta > 1\) via a substitution effect. Therefore, workers are reallocated to \(k\).
- When \(k\)'s markup is larger than average, the firm was too small from a social perspective → A positive shock to \(k\) improves the allocative efficiency.

- No change in allocative efficiency when \(\theta = 1\), as factor shares do not move.
Example 1: Horizontal Economy + Markup Shocks

- Reallocating effect: \( \lambda_k \theta \left( \frac{\mu_k^{-1}}{\sum_i \lambda_i \mu_i^{-1}} - 1 \right) \)

- \( \theta = 0 \): There are no reallocation effects, as the HH consumes a fixed quantity.

- \( \theta > 0 \): A markup shock increases \( k \)'s price, reduces its demand, and reallocates workers to other firms. Allocate efficiency would increase (decrease) depending on whether \( k \)'s markup was smaller (larger) than average markup.
Example 2: Vertical Economy

- Allocative efficiency:
  - No room for misallocation.
  - Productivity or markup shocks have no effect on allocative efficiency.

- Pure technology effect: \( \frac{d \log Y}{d \log A_k} = \tilde{\lambda}_k = 1 \)
  - Different from Hulten’s prediction due to double marginalization:
    \( \tilde{\lambda}_k > \lambda_k = \prod_{i=1}^{k-1} \mu_i^{-1} \).
So What?

- Besides comparative statistics, the non-parametric decomposition can be used for to measure changes in the economy’s allocative efficiency

\[
d \log Y - \tilde{\Lambda}' d \log L = \tilde{\lambda}' d \log A - \tilde{\lambda}' d \log \mu + \tilde{\Lambda}' d \log \Lambda
\]

as well as growth accounting.

- The parametric model can be used for measuring the level of allocative efficiency (among other applications).
Comment: Inefficient Economies and the Origin of Distortions

- The paper provides a general framework to handle distortions.
- The framework can be generalized to endogenous distortions and productivities by applying the chain rule one more time.
- But inefficiencies/distortions are only meaningful with respect to an efficiency benchmark.
- The right notion of efficiency cannot be decoupled from
  (a) the origin of the inefficiencies/wedges
  (b) the policy instruments available
Comment: Inefficient Economies and the Origin of Distortions

- The paper provides a general framework to handle distortions.
- The framework can be generalized to endogenous distortions and productivities by applying the chain rule one more time.
- But inefficiencies/distortions are only meaningful with respect to an efficiency benchmark.
- The right notion of efficiency cannot be decoupled from
  (a) the origin of the inefficiencies/wedges
  (b) the policy instruments available
Measuring Allocative Efficiency

- Typical measure for the *level* of allocative efficiency: distance to the frontier. 
  Restuccia and Rogerson (2008) and Hsieh and Klenow (2009)

- The efficiency benchmark used in the literature is the undistorted economy with no wedges.
  → Allocative inefficiency = increase in output if all wedges were eliminated.

- This is regardless of where the wedges come from (monopoly markups, taxes, financial frictions, etc.)
Measuring Allocative Efficiency: Information Frictions

• Suppose firms make production decisions under asymmetric information about the fundamentals (Angeletos, Iovino, and La’O, 2009)

• Asymmetry of information induces a wedge with respect to the complete information benchmark.

• Yet, the economy can still be constrained efficient → a planner who cannot transfer private information across firms would face the same exact wedges.

• This would be measured as misallocation, even though the economy is (constrained) efficient: there is no policy instrument that can improve upon the allocation.
Measuring Allocative Efficiency: Information Frictions

- Suppose firms make production decisions under asymmetric information about the fundamentals (Angeletos, Iovino, and La’O, 2009).

- Asymmetry of information induces a wedge with respect to the complete information benchmark.

- Yet, the economy can still be constrained efficient → a planner who cannot transfer private information across firms would face the same exact wedges.

- This would be measured as misallocation, even though the economy is (constrained) efficient: there is no policy instrument that can improve upon the allocation.
Changes in Allocative Efficiency

• Similar idea can be applied to changes in allocative efficiency.

\[
\log Y = \max_{\chi'} \log \mathcal{Y}(A, \chi') \\
\text{s.t.} \quad \log g(A, \chi') \geq 0
\]

• If the equilibrium is constrained efficient, we can reuse the envelope theorem, making the allocative efficiency terms second order again:

\[
\frac{d \log Y}{d \log A} = \frac{\partial \log \mathcal{Y}}{\partial \log A} + \eta \frac{\partial \log g}{\partial \log A}
\]

• Measuring either the (i) level or (ii) changes in allocative efficiency may require taking the origins of the wedges more seriously.
Changes in Allocative Efficiency

• Similar idea can be applied to changes in allocative efficiency.

\[ \log Y = \max_{\mathcal{X}} \log \mathcal{Y}(A, \mathcal{X}) \]
\[ \text{s.t. } \log g(A, \mathcal{X}) \geq 0 \]

• If the equilibrium is constrained efficient, we can reuse the envelope theorem, making the allocative efficiency terms second order again:

\[ \frac{d \log Y}{d \log A} = \frac{\partial \log \mathcal{Y}}{\partial \log A} + \eta \frac{\partial \log g}{\partial \log A} \]

• Measuring either the (i) level or (ii) changes in allocative efficiency may require taking the origins of the wedges more seriously.
Gains from Eliminating Markups

- The model can be used to estimate the gains from eliminating markups.
- Second-order approximation:

\[
\log(GDP^*) - \log(GDP) \approx \frac{1}{2} \sum_{i=1}^{n} \frac{d \log Y}{d \log \mu_i} \left( \frac{1 - \mu_i}{\mu_i} \right).
\]

- Using markup data and the calibrated model:

<table>
<thead>
<tr>
<th></th>
<th>2014/15</th>
<th>1997</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gutierrez &amp; Philippon</td>
<td>20%</td>
<td>3%</td>
</tr>
<tr>
<td>Lerner Index</td>
<td>17%</td>
<td>5%</td>
</tr>
<tr>
<td>De Loecker &amp; Eeckhout</td>
<td>35%</td>
<td>21%</td>
</tr>
</tbody>
</table>

- Two main observations:
  
  (a) Substantial increase in gains from eliminating markups.
  (b) Two order of magnitudes larger than Harberger's estimate (0.1%)
Comment: Gains from Eliminating Markups

\[
\log(GDP^*) - \log(GDP) \approx \frac{1}{2} \sum_{i=1}^{n} \frac{d \log Y}{d \log \mu_i} \left( \frac{1 - \mu_i}{\mu_i} \right)
\]

- Unlike the decomposition of the Solow residual, measuring the gains from eliminating markups requires the structural model:

\[
\frac{d \log Y}{d \log \mu_i} = -\bar{\lambda}_k - \sum_j (\theta_j - 1) \mu_j^{-1} \lambda_j \text{Cov} \left( \Psi_{(k)}, \Psi_L / \Lambda_L \right) - \lambda_k \frac{\Psi_{kL}}{\Lambda_L}
\]

- But the approximation above is exact only if the elasticities are close to 1 (Bigio and La’O, 2016).

- The estimation process seems to be internally inconsistent:
  
  (i) Uses structural elasticities when calculating \( d \log Y / d \log \mu_i \).
  
  (ii) Approximates the gains as if all production functions are Cobb-Douglas.
Comment: Gains from Eliminating Markups

• Calibrations for 2014/15:

<table>
<thead>
<tr>
<th></th>
<th>Benchmark</th>
<th>CD Counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>20%</td>
<td>4%</td>
</tr>
<tr>
<td>LI</td>
<td>17%</td>
<td>4%</td>
</tr>
<tr>
<td>DE</td>
<td>35%</td>
<td>7%</td>
</tr>
</tbody>
</table>

• Not sure how to interpret the large gap between the two calibrations.

• What if the much larger numbers of the benchmark (where the elasticities are $\epsilon = 0$, $\zeta = 8$) are driven by larger approximation errors?

• Since estimating the $d \log Y / d \log \mu_i$ requires the structural model anyways, why not estimate the gains from reducing markups also structurally?
Summary

- A generalization of Hulten's theorem to inefficient economies
- Applications:
  - comparative statics
  - growth accounting
  - measuring allocative efficiency
  - macro impacts of micro shocks
  - etc.

- The paper makes a strong argument for taking the microeconomic nature of the economy (input-output linkages, micro elasticities, returns to scale) seriously.
- Thinking about misallocation may also require taking the nature and origins of the “wedges” more seriously.