Assessing Small-Sample Error in Labor Mobility Statistics: Evidence from a Simulated Two-Sector Stochastic Stock-and-Flow Model

Nabil Ben Hassoun

A Thesis in the Department of Economics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Economics) at
Concordia University
Montréal, Québec, Canada

July 2025

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the Thesis prepared

By: Nabil Ben Hassoun

Entitled: Assessing Small-Sample Error in Labor Mobility Statistics: Evidence from a Simulated Two-Sector Stochastic Stock-and-Flow Model

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Economics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

__________________________ Examiner
Dr. Christian Sigouin

__________________________ Supervisor
Dr. Damba Lkhagvasuren

Approved by: __________________________
Dr. Christian Sigouin
Graduate Program Director

Date: ____________ __________________________
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Assessing Small-Sample Error in Labor Mobility Statistics:
Evidence from a Simulated Two-Sector Stochastic Stock-and-Flow Model

Nabil Ben Hassoun

This paper analyzes the measurement error arising from limited sample sizes in the estimation of gross and net mobility rates using simulated data and a two-sector stochastic stock-and-flow model. It further investigates how the magnitude of these errors changes with increasing sample size and in response to reductions in the volatility of underlying stochastic shocks. The results demonstrate that sampling variability in mobility statistics declines substantially with increasing sample size, but the rate of improvement diminishes beyond a certain point. I identify a sample size of 10,000 observations as a critical point beyond which the marginal reduction in standard error becomes negligible.

Contents

List of Figures

List of Tables

1 Introduction

Labour mobility plays a critical role in the growth and structural transformation of economic sectors and geographic regions. Standard macroeconomic theory suggests that labour supply adjusts in response to sectoral and regional demand shifts, facilitating the expansion of growing industries and mitigating labour shortages. Using US data, Davis and Haltiwanger (2014) show that a lack of labour market flexibility, or fluidity, can have a negative impact on employment, productivity and real wages. Engbom (2022) shows, using OECD data, that workers in countries with higher labour mobility experience higher wage growth through higher human capital accumulation. The reallocation of workers across sectors is thus a key driver of economic adaptability and long-term growth.

Obtaining reliable statistics on worker reallocation and labor mobility is challenging due to the high costs of conducting and maintaining large panel data studies. The relatively small sample sizes of the panel datasets available to researchers pose difficulties in accurately estimating mobility statistics. This paper addresses these challenges by analyzing the measurement error that arises from limited sample sizes in the estimation of gross and net mobility rates. Specifically, it seeks to answer the following research question: *"What measurement errors can researchers expect when using different sample sizes —* 10^2*,* 10^3*,* 10^4*, and* 10^5 *— to estimate gross and net mobility rates within two-sector stochastic stock-and-flow models?"*

Mobility rates provide valuable insight into labour market dynamics, capturing both geographic mobility (e.g., interprovincial migration) and sectoral mobility (e.g., transitions from manufacturing to services).[1] Economists analyze these rates to better understand workers' mobility decisions in response to changing labour market conditions and broader economic fluctuations. Gross mobility rates serve as an indicator of an economy's ability to reallocate workers efficiently, while net mobility rates reveal sectoral or regional disparities by identifying which industries or geographic areas experience net gains or losses in labour.

Sectoral mobility rates provide a measure of how effectively the labour market reallocates workers in response to shifts in economic activity and changes in demand for goods and services. A high degree of sectoral mobility suggests that workers are able to transition smoothly between industries, whereas rigidities in mobility may indicate barriers such as skill mismatches, labour market frictions, or institutional constraints.

Geographic mobility rates capture the movement of workers across provinces or regions, offering insight into interprovincial labour market trends. These metrics help policymakers

1. Some argue that geographic mobility and sectoral mobility are decisions made simultaneously by workers; for further information on this topic see Osberg, Gordon, and Lin (1994)

assess the impact of region-specific economic events, fiscal policies, and regulatory changes. Provincial governments, in particular, can leverage geographic mobility data to evaluate the effectiveness of local economic policies and identify regions that may require targeted interventions to address labour shortages or outmigration trends. Gomme et al. (2019) provide an analysis of labour mobility between the canadian provinces, with a focus on the wage differences between movers and stayers.

A complete understanding of labor mobility requires both *gross mobility*, which captures the total volume of worker flows, and *net mobility*, which isolates directional imbalances. As emphasized by Lkhagvasuren (2014) and Coen-Pirani (2010), analyzing these measures in tandem provides a clearer picture of labor market dynamics. Foundational work by Davis and Haltiwanger (1992) formalizes these metrics and highlights their interpretive value.

In empirical research, mobility statistics are typically estimated from panel datasets such as the PSID, or NLSY which have limited sample sizes—especially after disaggregation.[2] In such contexts, sampling variability may induce substantial measurement error, undermining the reliability of estimated mobility rates. This has been aknowledged in past research, for example, Kambourov and Manovskii (2008) state: "As opposed to the level of gross mobility, the levels of net mobility found in the PSID should be interpreted with caution because of the relatively small PSID sample size."

This paper quantifies the impact of sample size on the precision of these estimates. The analysis is based on a simulated two-sector stochastic stock-and-flow model that abstracts from richer microfoundations—such as search frictions, endogenous decision-making, and matching inefficiencies—but retains the key dynamic structure needed to track worker flows. The simplified environment adopted here allows for transparent and tractable analysis of statistical error.

By simulating the model across a range of sample sizes and computing standard mobility metrics, the paper evaluates how sampling error varies with sample size and identifies thresholds beyond which additional data yield limited gains in precision. This paper demonstrates that sampling variability in mobility statistics declines substantially with increasing sample size, but the rate of improvement diminishes beyond a certain point. I identify $L_t = 10^4$ as a critical point beyond which the marginal reduction in standard error becomes negligible.

This research contributes to the literature by providing empirical guidance for researchers using small or moderately sized datasets in order to estimates mobility statistics. This contribution adds precision to the link between empirical data and multiple theoretical models used in labor economics today, especially those relying on the Roy (1951) framework or modelling persistent sectoral shocks as in Lucas and Prescott (1974). The big picture nature

2. Panel Study of Income Dynamics (PSID), National Longitudinal Survey of Youth (NLSY)

of the analysis makes the results applicable to models that estimate labor mobility statistics whether it be geographical, sectoral, in-and-out of employment, between occupations, or job-to-job mobility.

The remainder of the paper is organized as follows. Section 2 provides an overview of the literature on labor mobility as it relates to this research. Section 3 provides a brief outline of the analytic process used. Section 4 introduces the model and setup, while Section 5 outlines the simulation procedure. Section 6 discusses the data, the calibration targets, and the calibration procedure. Section 7 presents the quantitative analysis of the benchmark model, while Section 8 presents a numerical experiment for a special case. Finally, Section 9 discusses the findings, concludes, and suggests extensions for further research on the topic.

2 Literature review

This section reviews key theoretical and empirical contributions in the labor mobility literature, organized around four areas: sectoral selection, sectoral dynamics, and the distinction between gross and net mobility flows. Each area contributes to understanding the model and how workers reallocate across sectors and regions in response to individual, sectoral, and macroeconomic factors.

Roy (1951) introduces a foundational model in which individuals make labor mobility decisions based on their comparative productivity and potential earnings. Using a stylized framework with two sectors—a hunting village and a fishing village—he shows that individuals choose their occupation to maximize income, taking into account their idiosyncratic skill levels and prevailing output prices. Building on Roy's model, McLaughlin and Bils (2001) apply the framework to examine empirical patterns of interindustry mobility. They provide evidence of cyclical upgrading, consistent with the idea that during economic expansions, workers are better able to sort into sectors where they hold a comparative advantage. As a result, more workers transition into higher-wage industries, and those who switch tend to experience persistent wage gains. Moscarini (2001) builds on the foundational Roy (1951) model by introducing a dynamic framework that incorporates heterogeneous job quality and frictional labor markets. In this setting, workers engage in on-the-job search and transition between positions based on match quality. He demonstrates that, even in the absence of sectoral or aggregate productivity shocks, substantial excess mobility can emerge endogenously as a result of optimal worker behavior in the presence of match-specific heterogeneity. Moscarini and Vella (2008) examines the cyclicality of occupational mobility and job reallocation using empirical data. The paper builds on the theoretical foundations laid by Roy (1951), emphasizing self-selection and comparative advantage, and extends the framework developed in Moscarini (2001), which models excess reallocation as an outcome of match-specific heterogeneity and frictional labor markets. Lkhagvasuren (2014) develops a dynamic model of regional mobility that is both parsimonious and flexible. He builds on the work of McLaughlin and Bils (2001), which in turn is grounded in the self-selection framework of Roy (1951). Lkhagvasuren (2014) applies the extended model to geographical mobility data to investigate whether mobility patterns vary by educational attainment. His dynamic model incorporates persistent labor income shocks, unobserved ability, and moving costs.

Lucas and Prescott (1974) introduce a stylized model of labor reallocation across spatially distinct markets—or sectors—where workers decide whether to remain employed or search elsewhere based on prevailing market conditions. The model is static and analytically tractable only in equilibrium; while changes to model parameters reveal how the equilibrium allocation shifts, the model does not characterize the transitional dynamics or the path toward equilibrium. Rogerson (1987) extends the sectoral reallocation framework developed by Lucas and Prescott (1974) by introducing a two-period, two-sector version of the original model featuring permanent sectoral shocks. He demonstrates that this extended model can trace the time path of equilibrium allocations, offering insights into how the economy adjusts to various sectoral shocks. Jovanovic and Moffitt (1990) develop a dynamic, stochastic model of sectoral labor mobility that incorporates two primary sources of mobility: worker-employer mismatch and sectoral shocks. These shocks are transmitted to workers through the wages offered in each sector, introducing uncertainty that, in combination with match quality, drives individual mobility decisions.

Vanderkamp (1971) presents an empirical analysis of the determinants of net and gross worker flows within Canada, showing that gross flows substantially exceed net flows. The study also distinguishes between one-time and round-trip migration, demonstrating that return migration plays a significant role in shaping overall mobility patterns. Davis and Haltiwanger (1992) introduce a foundational distinction between gross and net employment flows, emphasizing that gross flows—comprising job creation and destruction—are bidirectional and substantially larger than net flows, often by an order of magnitude. Net employment change is shown to be the small residual resulting from the near-offsetting of these gross components. They also develop the concept of *excess reallocation*, defined as the portion of gross flows that exceeds what is necessary to accommodate net employment change, attributing it to adjustment frictions and inefficiencies in the labor market. Coen-Pirani (2010) examines patterns of interstate worker mobility in the United States and finds that gross flows substantially exceed net flows. The study challenges explanations that attribute mobility primarily to structural shifts, instead emphasizing the role of unobserved individual heterogeneity, worker-specific shocks and location-specific shocks in driving labor reallocation.

Measurement error has been acknowledged and studied in the literature. For example, Bergin (2011) and Shibata (2022) examine issues arising from classification errors, where survey respondents provide inaccurate or inconsistent answers.

Additionally, Blanchard and Katz (1992) use state-level panel data with a relatively small number of observations in their study of regional labor market adjustments in the United States, focusing on how employment, unemployment, and wages respond to regional shocks. They mention the challenges that using small panel data poses for their estimation. Their work does not explicitly study measurement error from small samples, but it does illustrate the potential limitations of using limited data to analyze mobility patterns. Similarly, Kambourov and Manovskii (2008) acknowledge the challenges posed by the small sample sizes of datasets such as the PSID as it relates to their study of worker mobility between industries and occupations. Dahl (2002) briefly touches on the issue in the context of his analysis of mover–stayer wage gaps, where he conducts simulations with 1,000 and 10,000 observations and selects the larger sample size as better suited for his research.

Considering the above, no study to date has rigorously and systematically analyzed the measurement error that arises specifically from small sample sizes — this is the gap my paper addresses. This paper simply aims to properly characterize these small sample size errors by confirming their presence and their magnitude, for different sample sizes, as well as their behavior as the sample size varies from 100 observations to 100,000 observations.

3 Outline of the analysis

This analysis investigates the measurement error—specifically, the variance—of estimates of labor mobility. The objective is to understand how sample size influences the precision of mobility statistics such as gross and net mobility.

I begin by taking an established benchmark mobility rate of approximately 6.78%. I then simulate four economies with population sizes $N \in \{100, 1{,}000, 10{,}000, 100{,}000\}$ over $T = 2{,}000$ periods. The economies feature sectoral shocks that follow a discretized AR(1) process, implemented using the Rouwenhorst method to approximate highly persistent shocks.

For each worker in each time period, I generate a draw from the standard uniform distribution. This draw is compared to the transition probability—which equals the mobility rate—to determine whether the worker changes sectors or remains in the current sector. I record the sectoral allocation of all workers over all T periods in a matrix, which serves as a proxy for panel survey data. From this simulated panel, I compute mobility statistics for each economy and time period: gross mobility, net mobility within each sector, and total net mobility. For each sample size, I calculate the mean and variance (standard error) of the mobility statistics across the T periods. This allows me to assess how sample size affects the dispersion and accuracy of mobility estimates.

The benchmark model excludes sectoral shocks to show the "theoretical" case, which closely matches analytic expectations and validates the simulation approach. The results provide practical guidance for empirical researchers by illustrating the expected bias and variability in mobility estimates at different sample sizes. They highlight how closely researchers can expect their estimates to approximate the true mobility rate and what variance they may encounter depending on their data's sample size.

4 Model

In this section, I introduce a two-sector stochastic stock-and-flow model set in discrete time, with a finite horizon and infinitely lived workers.[3] Following the framework of Roy (1951), and as in McLaughlin and Bils (2001) and Lkhagvasuren (2014), the model captures self-selection into sectors but also incorporates sectoral shocks inspired by the tradition of Lucas and Prescott (1974).

4.1 Setup

Let T denote the total number of periods and t represent a specific time period, where $t \in \{1, 2, 3, \ldots, T\}$. The economy consists of two sectors, indexed by $j \in \{0, 1\}$. This notation offers a convenient simplification: for a worker employed in sector j, the alternative sector is naturally given by $1 - j$, and vice versa.

Transition probabilities. Workers' transitions between sectors are governed by the transition probabilities $\lambda_{0,t}$ and $\lambda_{1,t}$, where $\lambda_{0,t}$ denotes the probability of a worker leaving sector 0 to join sector 1 at period t, and $\lambda_{1,t}$ represents the probability of a worker leaving sector 1 to join sector 0 at time t. These transition probabilities satisfy $0 \leq \lambda_{0,t} \leq 1$ and $0 \leq \lambda_{1,t} \leq 1$, ensuring that they remain within the unit interval.

Stocks. Let $L_{j,t}$ denote the size of sector j at time t, measured by the number of workers employed in that sector. By definition, the total workforce in the economy at time t is given by $L_t = L_{0,t} + L_{1,t}$, where L_t represents the aggregate number of workers in the economy at time t. This formulation defines the stock component of our stock-and-flow model.

3. For tractability, I abstract from worker utility maximization, wages, match quality, match-specific shocks, accrued human capital, unobserved ability, and moving costs.

Flows. To capture worker transitions between sectors, let $X_{j,t}$ represent the number of workers exiting sector j at time t and $M_{j,t}$ the number of workers entering sector j at time t. In a two-sector framework, these flows are governed by the identities $X_{j,t} = M_{1-j,t}$ and $M_{j,t} = X_{1-j,t}$, ensuring that all workers leaving one sector enter the other. This specification defines the flow component of our stock-and-flow model. Figure 1 provides a diagrammatical representation of worker flows.

4.2 Measures and variables

The stock variables, $L_{0,t}$ and $L_{1,t}$, evolve over time according to the following equation:

$$L_{j,t+1} = L_{j,t} + X_{1-j,t} - X_{j,t}, \quad \forall j \in \{0, 1\}. \tag{1}$$

Equation (1) states that the size of sector j at time $t + 1$ is determined by its size in the previous period, $L_{j,t}$, adjusted for net worker flows. Specifically, it decreases by the number of workers exiting the sector, $X_{j,t}$, and increases by the number of workers entering from the other sector, $X_{1-j,t}$.

4.2.1 Gross mobility

Gross mobility in sector j at time t, denoted $G_{j,t}$, is defined as the total number of worker transitions, comprising both workers leaving and entering the sector.[4] Formally, this is expressed as:

$$G_{j,t} = X_{j,t} + X_{1-j,t}. \tag{2}$$

The economy-wide *gross mobility rate*, in period t, denoted g_t, is defined as the ratio of gross flow mobility in either sector, $G_{j,t} = G_{1-j,t} = G_t$, to the total workforce, L_t. Formally, it is given by:

$$g_t = \frac{G_t}{L_t}. \tag{3}$$

4.2.2 Net mobility

Net mobility in sector j at time t, denoted $N_{j,t}$, is defined as the difference between the number of workers entering and leaving the sector during that period. Formally, this is expressed as:

$$N_{j,t} = X_{1-j,t} - X_{j,t}. \tag{4}$$

4. In a two-sector framework, it follows that $G_{j,t} = G_{1-j,t}$, since addition is commutative and every worker exiting one sector must enter the other.

The *net mobility rate* in sector j at time t, denoted $n_{j,t}$, is defined as:

$$n_{j,t} = \frac{N_{j,t}}{L_t}. \tag{5}$$

The economy-wide *net mobility rate* at time t, denoted n_t, is defined as:

$$n_t = \frac{1}{2} \sum_j \frac{|N_{j,t}|}{L_t} = \frac{1}{2} \sum_j |n_{j,t}|, \tag{6}$$

where the $\frac{1}{2}$ term adjusts for double counting, following the convention of Jovanovic and Moffitt (1990) and Davis and Haltiwanger (1992).

4.3 Sectoral shocks

Similarly to Lucas and Prescott (1974), I introduce sectoral shocks, denoted by z_t, which influence the transition probabilities λ_j, thereby affecting workers' sectoral mobility decisions. These shocks create asymmetries between sectors, leading to periods in which one sector is more attractive than the other. The sectoral shock z_t follows an autoregressive process of order one, AR(1), given by:

$$z_{t+1} = \mu + \rho z_t + \epsilon_t, \tag{7}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a normally distributed error term with mean zero and variance σ_ϵ^2.

4.4 Worker reallocation rule

At the start of each period, workers are each assigned a draw from the standard uniform distribution, $\mathcal{U}[0, 1]$. Each worker then evaluates their draw against the value of $\lambda_{j,t}$ and acts according to the following reallocation rule: Workers in sector j at time t will leave sector j, to join sector $1 - j$, if their draw from the standard uniform distribution, $y \sim \mathcal{U}(0, 1)$, is lesser than or equal to $\lambda_{j,t}$.

5 Simulation procedure

In this section I outline the simulation process used to generate the simulated panel data set. In order to simulate the mobility data, I generate a matrix containing all the worker draws from the uniform distribution, a mobility matrix that will be used to keep track of workers' locations for all time periods, and the sequence of simulated sectoral shocks.

In the first period, all workers are divided evenly between the two sectors, 0 and 1.

5.1 Worker draws

There is a draw from the standard uniform distribution, $y \sim \mathcal{U}(0,1)$, for each individual worker at each time period t which is generated and stored in the $\mathbf{U}$ matrix of size L_t by T.

$$\mathbf{U} = \begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} & \cdots & U_{1,T} \\ U_{2,1} & U_{2,2} & U_{2,3} & \cdots & U_{2,T} \\ U_{3,1} & U_{3,2} & U_{3,3} & \cdots & U_{3,T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{L_t,1} & U_{L_t,2} & U_{L_t,3} & \cdots & U_{L_t,T} \end{bmatrix}, \quad U_{i,t} \sim \mathcal{U}(0,1)$$

These draws will be compared to the applicable transition probability, $\lambda_{j,t}$, and will dictate if a specific worker will remain where he is or move to the other sector in the following time period, $t+1$.

5.2 Transition probabilities

The transition probabilities as introduced in the model setup are used by workers to determine wether they have to move sectors or remain in their current sector, according to the worker reallocation rule.

5.2.1 Benchmark model: Constant λ

Under this model specification, the values of $\lambda_{0,t}$ and $\lambda_{1,t}$ are constant and do not change over time such that:

$$\lambda_{0,t-1} = \lambda_{0,t} = \lambda_{0,t+1}, \quad \forall t. \tag{8}$$

Workers evaluate their draws against the applicable $\lambda_{j,t}$ value, which depends only on their current sector and not on the time period t.

5.2.2 Sectoral shock model: Time-dependent λ

In order to conduct the simulation, I discretize the continuous AR(1) process governing the economy's sectoral shocks using the Rouwenhorst (1995) method.[5] This procedure yields

5. An alternative for discretizing this AR(1) process would have been the method proposed by Tauchen (1986). However, based on the analyses and findings of Kopecky and Suen (2010) and Galindev and Lkhagvasuren (2010), I opt for the Rouwenhorst (1995) method, as it exhibits superior performance when modeling highly persistent shocks.

a discrete-time Markov chain approximating the z_t process, whose evolution over time is represented by the sequence:

$$\{z_t\}_{t=1}^{T} = (z_1, z_2, z_3, \dots, z_T). \tag{9}$$

I then construct the sequence of time-dependent transition rates, denoted $\lambda_{j,t}$, by adding or substracting the corresponding element of the $\{z_t\}_{t=1}^{T}$ sequence to λ_j as shown below:

$$\lambda_{j,t} = \begin{cases} \lambda_0 + z_t, & \text{if } j = 0, \\ \lambda_1 - z_t, & \text{if } j = 1. \end{cases} \tag{10}$$

This will cause the worker reallocation rule to become time-dependent, in addition to depending on the current sector (or location). The shock z_t positively affects sector 1 and negatively affects sector 0: subtracting z_t from λ_1 reduces the likelihood that a worker will leave sector 1 when $z_t > 0$. Conversely, adding z_t to λ_0 increases the likelihood of leaving sector 0, as the probability associated with λ_0, now denoted $\lambda_{0,t}$, rises with positive z_t. Negative values of z_t have the opposite effect on both $\lambda_{1,t}$ and $\lambda_{0,t}$. The parameter values are chosen such that $\lambda_{j,t}$ remains positive for all t, since any λ_j exceeds the absolute value of any realization of the z_t sequence.

5.3 Tracking workers over time

The $\mathbf{S}$ matrix, of dimensions $L_t \times T$, records the sector of each worker in each time period $t = 1, \dots, T$. Each element $s_{i,t}$ of $\mathbf{S}$ takes the value 0 or 1, indicating whether worker i belongs to sector 0 or sector 1 at time t.

$$\mathbf{S} = \begin{bmatrix} s_{1,1} & s_{1,2} & s_{1,3} & \cdots & s_{1,T} \\ s_{2,1} & s_{2,2} & s_{2,3} & \cdots & s_{2,T} \\ s_{3,1} & s_{3,2} & s_{3,3} & \cdots & s_{3,T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{L,1} & s_{L,2} & s_{L,3} & \cdots & s_{L,T} \end{bmatrix}, \quad s_{i,t} \in \{0, 1\}.$$

The $\mathbf{S}$ matrix thus captures the full simulated history of worker-sector allocations and serves as a proxy for survey data, which will subsequently be used to compute mobility statistics.

5.4 Initiating the simulation

In the initial period, $t = 1$, workers are evenly distributed between the two sectors, 0 and 1. The workers then evaluate their draws from the $\mathbf{U}$ matrix for time $t = 1$ with the value of $\lambda_{j,1}$ and move or stay according to the reallocation rule. This reassignment takes place at the precise moment between the end of period $t = 1$ and the beginning of period $t = 2$.

6 Data and calibration

This section presents the key empirical facts, describes the data sources and references, and outlines the calibration procedure, targets, and results.

6.1 Empirical data

The benchmark mobility rate used to generate the simulated data, 6.78%, is taken from Auray et al. (2024). They compute this rate using two data sources. The first combines the PSID annual waves (1968–1997) with the Retrospective Occupation-Industry Supplemental Data Files (covering 1968–1980). The second consists of sectoral and aggregate employment and productivity series from Bureau of Labor Statistics (2020a, 2020b).

6.2 Calibration targets

The calibration procedure targets three empirical moments. The first is the *gross mobility* rate, set at 6.78%. The second and third are the *volatility* and *persistence* of sectoral employment, which receive equal weight in the calibration. These two moments, also from Auray et al. (2024), have values of 0.0129 (volatility) and 0.6193 (persistence). The targeted moments are summarized in Table 2.

6.3 Calibration procedure

The gross mobility rate is calculated as the mean gross mobility rate over all time periods $t = 1, \ldots, T$. To measure volatility and persistence, I apply the Hodrick–Prescott filter with a smoothing parameter of 100, which attenuates short-term fluctuations and yields more representative estimates of these moments. Using the filtered data, persistence is measured as the first-order serial autocorrelation, and volatility as the standard deviation. Finally, to calibrate these three moments, I adjust the parameters of the Rouwenhorst (1995) discretization method. The calibrated parameters are presented in Table 3.

7 Quantitative analysis

In this section, I simulate the benchmark model, with time-invariant transition rates, where $\lambda_{j,t} = \lambda_{j,t+1} = \lambda_j$ for all t, such that λ_j remains constant over the entire time horizon.

7.1 Benchmark model: Constant λ_j

I conduct simulation experiments using the benchmark model, setting $T = 2{,}500$ and evaluating four sample sizes: $L_t = \{10^2, 10^3, 10^4, 10^5\}$. For each specification, I compute gross and net mobility rates, along with their expected values over time and corresponding standard deviations. Table 4 reports the expected values and standard deviations (SD), all scaled by 10^2, for the gross mobility rate g_t, the sectoral net mobility rates $n_{0,t}$ and $n_{1,t}$, and the total net mobility rate n_t, across sample sizes $L_t \in \{100,\ 1{,}000,\ 10{,}000,\ 100{,}000\}$. Table 5 presents the corresponding percentage changes in standard deviations, capturing the precision gains associated with increasing sample size. Figure 2 visualizes the expected values of the mobility rates, while Figure 3 depicts the corresponding standard deviations.

For the gross mobility rate, the expected values $\mathbb{E}_t(g_t) \times 10^2$ are 6.74, 6.80, 6.79, and 6.78 for $L_t = 10^2$, 10^3, 10^4, and 10^5, respectively, with corresponding standard deviations $\mathrm{SD}(g_t) \times 10^2$ of 2.41, 0.73, 0.23, and 0.073. The standard deviation decreases by approximately 69.73% from $L_t = 10^2$ to 10^3, followed by a 67.95% reduction from 10^3 to 10^4, and a further 68.87% from 10^4 to 10^5. This pattern reflects diminishing marginal gains in estimation precision as sample size increases.

For the sectoral net mobility rates $n_{0,t}$ and $n_{1,t}$, the expected values $\mathbb{E}_t(n_{j,t}) \times 10^2$ are approximately zero across all sample sizes, consistent with the symmetric structure of sectoral flows. The corresponding standard deviations $\mathrm{SD}(n_{j,t}) \times 10^2$ are 2.49, 0.80, 0.25, and 0.081 for $L_t = 10^2$, 10^3, 10^4, and 10^5, respectively. This corresponds to a 67.77% reduction in standard deviation from 10^2 to 10^3, followed by a 68.54% drop from 10^3 to 10^4, and a 67.89% reduction from 10^4 to 10^5.

For the total net mobility rate n_t, the expected values $\mathbb{E}_t(n_t) \times 10^2$ are 2.01, 0.67, 0.211, and 0.068 for the four respective sample sizes. The corresponding standard deviations $\mathrm{SD}(n_t) \times 10^2$ are 1.52, 0.47, 0.148, and 0.047. These values imply a 69.35% reduction in standard deviation from 10^2 to 10^3, a 68.24% reduction from 10^3 to 10^4, and a 68.33% drop from 10^4 to 10^5. Figure 2 shows that $\mathbb{E}_t(n_t)$ declines steadily with increasing sample size and eventually converges to a value near zero. This convergence is made possible by the equality and constancy of the transition rates λ_j, which ensure that worker flows between the two sectors are symmetric and reciprocal. As a result, the overestimation problem introduced by the absolute value operator in equation (6) is substantially mitigated.

8 Further analysis

This section provides a quantitative analysis of the sectoral shock model, which incorporates stochastic transition rates and accounts for time-dependent fluctuations in workers' mobility behavior.

8.1 Sectoral shock model: Time-dependent $\lambda_{j,t}$

I conduct simulation experiments using the sectoral shock model, setting $T = 2{,}500$ and evaluating four sample sizes: $L_t = \{10^2, 10^3, 10^4, 10^5\}$. For each specification, I compute gross and net mobility rates, along with their expected values over time and corresponding standard deviations. The resulting gross and net mobility rates, along with their expected values and standard deviations, are reported in Table 6, while Figure 4 provides a graphical depiction of the expected values of the mobility rates. Table 7 complements this by reporting the percentage reductions in standard deviations, thereby quantifying the precision gains attributable to increasing the sample size.

The expected gross mobility rate, $\mathbb{E}_t(g_t) \times 10^2$, is 6.74 for $L_t = 10^2$, 6.81 for $L_t = 10^3$, 6.79 for $L_t = 10^4$, and 6.78 for $L_t = 10^5$. The corresponding standard deviation, $\mathrm{SD}(g_t) \times 10^2$, declines from 2.41 at $L_t = 10^2$ to 0.73 at $L_t = 10^3$, 0.24 at $L_t = 10^4$, and 0.08 at $L_t = 10^5$. The standard deviation thus decreases by approximately 69.7% when increasing the sample size from 10^2 to 10^3, by an additional 67.71% from 10^3 to 10^4, and by 68.2% from 10^4 to 10^5. This pattern illustrates diminishing marginal reductions in standard errors as the sample size grows.

For sector 0, the expected net mobility rate, $\mathbb{E}_t(n_{0,t}) \times 10^2$, is 0.0080 for $L_t = 10^2$, 0.0000 for $L_t = 10^3$, –0.0003 for $L_t = 10^4$, and –0.0007 for $L_t = 10^5$. The corresponding standard deviation, $\mathrm{SD}(n_{0,t}) \times 10^2$, exhibits a clear downward trend with increasing sample size, declining from 2.49 at $L_t = 10^2$ to 0.95 at 10^3, 0.61 at 10^4, and 0.57 at 10^5. These reductions represent a 62.01% decline in standard deviation from 10^2 to 10^3, followed by a further 35.94% decrease from 10^3 to 10^4. From 10^4 to 10^5, the standard deviation decreases marginally by 6.04%, indicating that the gains in precision taper off at larger sample sizes.

The expected net mobility rate for sector 1, $\mathbb{E}_t(n_{1,t}) \times 10^2$, closely mirrors that of sector 0, but with opposite signs, reflecting the symmetric structure of worker flows between the two sectors. Likewise, the standard deviations $\mathrm{SD}(n_{1,t}) \times 10^2$ are identical to those of sector 0 across all sample sizes, further reinforcing the model's structural symmetry.

For the total net mobility rate, $\mathbb{E}_t(n_t) \times 10^2$ is 2.04 for $L_t = 10^2$, 0.81 for $L_t = 10^3$, 0.58 for $L_t = 10^4$, and 0.58 for $L_t = 10^5$. The corresponding standard deviation, $\mathrm{SD}(n_t) \times 10^2$, declines from 1.51 at $L_t = 10^2$ to 0.56 at $L_t = 10^3$, 0.28 at $L_t = 10^4$, and 0.19 at $L_t = 10^5$. These figures imply a 63.23% reduction in standard deviation from 10^2 to 10^3, a 48.89% decline from 10^3 to 10^4, and a further 32.51% decrease from 10^4 to 10^5, indicating consistent—though diminishing—gains in statistical precision as sample size increases.

These results align with theoretical expectations, demonstrating that larger sample sizes reduce standard errors significantly at lower sample levels, but with diminishing returns as the sample size continues to grow. The marginal gains in precision become less pronounced beyond $L_t = 10^4$, suggesting that further increasing the sample size may yield limited benefits in a setting that includes sectoral shocks. This pattern is further illustrated in Figure 5, which plots the standard deviations of the estimated variables against sample sizes $L_t = \{10^2, 10^3, 10^4, 10^5\}$. The downward slope of the standard deviation curves becomes noticeably flatter between $L_t = 10^4$ and $L_t = 10^5$, reinforcing the idea of diminishing returns to precision as sample size increases beyond this threshold.

Figure 4 shows that $\mathbb{E}_t(n_t)$ declines as sample size increases but does not converge to zero. This persistence is due to the stochastic shocks and their effect on the overestimation problem introduced by the absolute value operator in equation (6). Negative sectoral mobility in one period cannot offset positive mobility in another, as both are transformed into positive values. Consequently, total net mobility remains strictly positive even in the absence of persistent structural shifts.

9 Discussion and closing remarks

This paper quantifies the sampling error in gross and net mobility statistics using a simulated two-sector stochastic stock-and-flow model.

In the benchmark model, with constant λ_j, the results are in line with theoretical expectations. Specifically, sampling error decreases as sample size increases. This is evident in the results: the gains in precision from increasing the sample size increase with no sign of diminishing returns even at $L_t = 10^4$ and $L_t = 10^5$, which is consistent with the behavior of a highly stylized model characterized by minimal to non-existent stochasticity and analytically tractable outcomes. With that in mind, the results arising from the benchmark model are not suprising and are consistent with standard econometric theory.

The sectoral shock model results also demonstrate that sampling variability in mobility statistics declines substantially as sample size increases, however, the rate of improvement diminishes beyond a certain point. I identify $L_t = 10^4$ as a critical point beyond which the marginal reduction in standard error becomes negligible under the specifications of the sectoral shock model with stochastic transition rates. These results do not contradict the findings of Dahl (2002), who found that his two-sector simulated model with 10,000 observations outperformed the one with only 1,000 observations in terms of precision.

My results show that researchers working with small sample data sets in two-sector stochastic stock-and-flow models will indeed encounter measurement errors in measuring mobility statistics and that the possible gains in precision they might accrue from using larger sample sizes can be determined using the results in Table 7. These results are particularly interesting for measures of the total net mobility rate, n_t, where the $\%\Delta\text{SD}(n_t)$ for moving from one sample size to the other are: -63.23%, -48.89%, and -32.51% for increases in sample size from 10^2 to 10^3, 10^3 to 10^4, and 10^4 to 10^5, respectively.

I believe these results may also be applicable to other, related models, such as extensions of the Roy (1951) framework or studies that model sectoral shocks following the approach of Lucas and Prescott (1974). I discuss more possible extensions to this research in the next section.

9.1 Further research

As discussed above, reproducing this analysis using variations of the Roy (1951) model could be interesting and useful for researchers working with such models. It could also be of interest to complicate this model by adding other shocks following the tradition of Lucas and Prescott (1974). Possible extensions could include incorporating match quality and introducing match-specific shocks. This would allow for the analysis of measurement errors in mobility statistics beyond those examined in this paper such as excess mobility and the share of excess mobility.

Other possible extensions of this work could involve introducing moving costs, accrued human capital, unobserved ability, or sector-specific skill accumulation. These features would make it possible to explore the measurement errors that may arise when mobility statistics are estimated using models that account for such worker- and match-specific characteristics.

10 Use of generative AI and AI-assisted tools

During the preparation of my thesis, I used **ChatGPT** for **help improving some of the wording and sentence structure of this paper**. After using this tool/service, I reviewed and edited the content as needed and take full responsibility for the content of my thesis.

References

Auray, Stéphane, David Fuller, Damba Lkhagvasuren, and Antoine Terracol. 2024. "Understanding the Role of a Persistent Match Shock in a Multi-Sector Economy."

Bergin, Adele. 2011. "Job Mobility and Measurement Error." PhD diss., National University of Ireland Maynooth.

Blanchard, Olivier, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 1992 (1): 1–75.

Bureau of Labor Statistics. 2020a. *Manufacturing Sector: Employment [PRS30006013], 1987:Q1 to 2020:Q3.* Retrieved from FRED, Federal Reserve Bank of St. Louis.

———. 2020b. *Nonfarm Business Sector: Employment [PRS85006013], 1987:Q1 to 2020:Q3.* Retrieved from FRED, Federal Reserve Bank of St. Louis.

Coen-Pirani, Daniele. 2010. "Understanding gross worker flows across U.S. states." *Journal of Monetary Economics* 57 (7): 769–784.

Dahl, Gordon B. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica* 70 (6): 2367–2420.

Davis, Steven J., and John C. Haltiwanger. 1992. "Gross job creation, gross job destruction, and employment reallocation." *The Quarterly Journal of Economics* 107 (3): 819–863.

Davis, Steven J, and John Haltiwanger. 2014. *Labor market fluidity and economic performance.* Technical report. National Bureau of Economic Research.

Engbom, Niklas. 2022. *Labor market fluidity and human capital accumulation.* Technical report. National Bureau of Economic Research.

Galindev, Ragchaasuren, and Damba Lkhagvasuren. 2010. "Discretization of Highly Persistent Correlated AR(1) Shocks." *Journal of Economic Dynamics and Control* 34 (7): 1260–1276.

Gomme, Paul, Gazi Mohammad Jamil, Tatyana Koreshkova, and Damba Lkhagvasuren. 2019. *An Analysis of the Ins and Outs of Migration within Canada.* Technical report. CIRANO.

Jovanovic, Boyan, and Robert Moffitt. 1990. "An estimate of a sectoral model of labor mobility." *Journal of Political Economy* 98 (4): 827–852.

Kambourov, Gueorgui, and Iourii Manovskii. 2008. "Rising occupational and industry mobility in the United States: 1968–1997." *International Economic Review* 49 (1): 41–79.

Kopecky, Karen, and Richard Suen. 2010. "Finite State Markov-Chain Approximations to Highly Persistent Processes." *Review of Economic Dynamics* 13 (3): 701–714.

Lkhagvasuren, Damba. 2014. "Education, mobility and the college wage premium." *European Economic Review* 67:159–173.

Lucas, Robert E., and Edward C. Prescott. 1974. "Equilibrium Search and Unemployment." *Journal of Economic Theory* 7 (2): 188–209.

McLaughlin, Kenneth J., and Mark Bils. 2001. "Interindustry Mobility and the Cyclical Upgrading of Labor." *Journal of Labor Economics* 19 (1): 94–135.

Moscarini, Giuseppe. 2001. "Excess worker reallocation." *Review of Economic Studies* 68 (3): 593–612.

Moscarini, Giuseppe, and Francis G. Vella. 2008. *Occupational Mobility and the Business Cycle.* NBER Working Paper w13819. National Bureau of Economic Research.

Osberg, Lars, Daniel Gordon, and Zhengxi Lin. 1994. "Interregional migration and interindustry labour mobility in Canada: A simultaneous approach." *Canadian Journal of Economics,* 58–80.

Rogerson, Richard. 1987. "An Equilibrium Model of Sectoral Reallocation." *Journal of Political Economy* 95 (4): 824–834.

Rouwenhorst, K Geert. 1995. "Asset pricing implications of equilibrium business cycle models." *Frontiers of business cycle research* 1:294–330.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3:135–146.

Shibata, Ippei. 2022. "Reassessing classification errors in the analysis of labor market dynamics." *Labour Economics* 78:102252.

Tauchen, George. 1986. "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics letters* 20 (2): 177–181.

Vanderkamp, John. 1971. "Migration Flows, Their Determinants and the Effects of Return Migration." *Journal of Political Economy* 79 (5): 1012–1031.

A Tables

Table 1: Transition Probabilities

	State at $t+1$	
State at t	Sector 0	Sector 1
Sector 0	$1-\lambda_{0,t}$	$\lambda_{0,t}$
Sector 1	$\lambda_{1,t}$	$1-\lambda_{1,t}$

Notes: This table presents the transition probabilities between sectors. Specifically, $\lambda_{0,t}$ denotes the probability of a worker transitioning from sector 0 to sector 1 at time t, while $\lambda_{1,t}$ represents the probability of moving from sector 1 to sector 0 at time t.

Table 2: Targeted and Simulated Moments

Moment	**Data**
Gross mobility (%)	6.78
Volatility of net mobility (%)	1.29
Persistence of net mobility ($\times$100)	61.93

Notes: This table reports the targeted mobility moments which serve as empirical benchmarks for model calibration.

Table 3: Parameters used in the AR(1) Discretization Process

Parameter	**Value**
Number of states in discretization (n_z)	2
Autoregressive coefficient (ρ)	0.325
Standard deviation of innovation (σ_ε)	0.0060665

Notes: This table reports the parameters of the AR(1) process discretized using the method of Rouwenhorst (1995), used in the calibration. The standard deviation refers to the innovation term ε_t.

Table 4: Simulation Results - Benchmark Model (Scaled by 10^{-2})

Simulated Variable	$L_t = 100$	$L_t = 1{,}000$	$L_t = 10{,}000$	$L_t = 100{,}000$
$\mathbb{E}_t(g_t) \times 10^2$	6.74	6.80	6.79	6.78
$\mathrm{SD}(g_t) \times 10^2$	2.41	0.73	0.23	0.07
$\mathbb{E}_t(n_{0,t}) \times 10^2$	0.00	0.00	0.00	0.00
$\mathrm{SD}(n_{0,t}) \times 10^2$	2.49	0.80	0.25	0.08
$\mathbb{E}_t(n_{1,t}) \times 10^2$	0.00	0.00	0.00	0.00
$\mathrm{SD}(n_{1,t}) \times 10^2$	2.49	0.80	0.25	0.08
$\mathbb{E}_t(n_t) \times 10^2$	2.01	0.67	0.21	0.07
$\mathrm{SD}(n_t) \times 10^2$	1.52	0.47	0.15	0.05

Notes: This table presents the expected values ($\mathbb{E}_t$) and standard deviations (SD) of the gross mobility rate (g_t) and net mobility rates ($n_{0,t}$, $n_{1,t}$, and n_t), scaled by 10^{-2}, for different sample sizes L_t.

Table 5: % change in Standard Deviations - Benchmark Model

Simulated Variable	$L_t = 100 \to 1{,}000$	$L_t = 1{,}000 \to 10{,}000$	$L_t = 10{,}000 \to 100{,}000$
$\%\Delta\mathrm{SD}(g_t)$	-69.73%	-67.95%	-68.87%
$\%\Delta\mathrm{SD}(n_{0,t})$	-67.77%	-68.54%	-67.89%
$\%\Delta\mathrm{SD}(n_{1,t})$	-67.77%	-68.54%	-67.89%
$\%\Delta\mathrm{SD}(n_t)$	-69.35%	-68.24%	-68.33%

Notes: This table presents the percentage changes in standard deviations (SD) for the gross mobility rate (g_t) and net mobility rates ($n_{0,t}$, $n_{1,t}$, and n_t) as sample sizes increase from $L_t = 100$ to $L_t = 100{,}000$.

Table 6: Simulation Results - Sectoral Shock Model (Scaled by 10^{-2})

Simulated Variable	$L_t = 100$	$L_t = 1{,}000$	$L_t = 10{,}000$	$L_t = 100{,}000$
$\mathbb{E}_t(g_t) \times 10^2$	6.74	6.81	6.79	6.78
$\text{SD}(g_t) \times 10^2$	2.41	0.73	0.24	0.07
$\mathbb{E}_t(n_{0,t}) \times 10^2$	0.00	0.00	0.00	0.00
$\text{SD}(n_{0,t}) \times 10^2$	2.49	0.95	0.61	0.57
$\mathbb{E}_t(n_{1,t}) \times 10^2$	0.00	0.00	0.00	0.00
$\text{SD}(n_{1,t}) \times 10^2$	2.49	0.95	0.61	0.57
$\mathbb{E}_t(n_t) \times 10^2$	2.04	0.81	0.58	0.58
$\text{SD}(n_t) \times 10^2$	1.51	0.56	0.28	0.19

Notes: This table presents the expected values ($\mathbb{E}_t$) and standard deviations (SD) of the gross mobility rate (g_t) and net mobility rates ($n_{0,t}$, $n_{1,t}$, and n_t), scaled by 10^{-2}, for different sample sizes L_t.

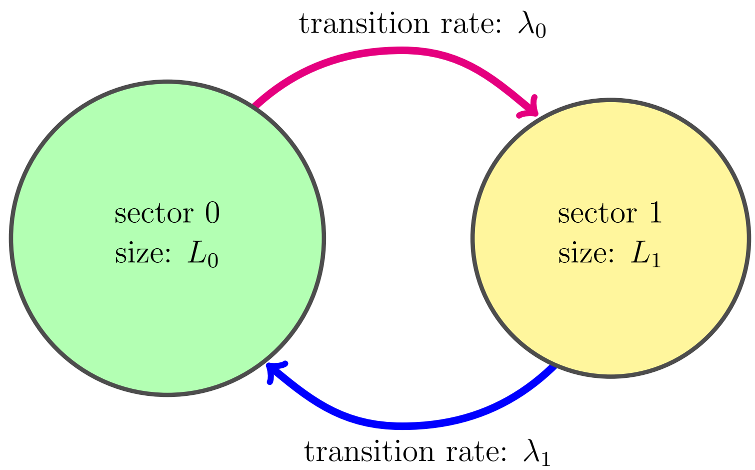
Table 7: % change in Standard Deviation - Sectoral Shock Model

Simulated Variable	$L_t = 100 \rightarrow 1{,}000$	$L_t = 1{,}000 \rightarrow 10{,}000$	$L_t = 10{,}000 \rightarrow 100{,}000$
$\%\Delta\text{SD}(g_t)$	-69.7%	-67.71%	-68.2%
$\%\Delta\text{SD}(n_{0,t})$	-62.01%	-35.94%	-6.04%
$\%\Delta\text{SD}(n_{1,t})$	-62.01%	-35.94%	-6.04%
$\%\Delta\text{SD}(n_t)$	-63.23%	-48.89%	-32.51%

Notes: This table presents the percentage changes in standard deviations (SD) for the gross mobility rate (g_t) and net mobility rates ($n_{0,t}$, $n_{1,t}$, and n_t) as sample sizes increase from $L_t = 100$ to $L_t = 100{,}000$.

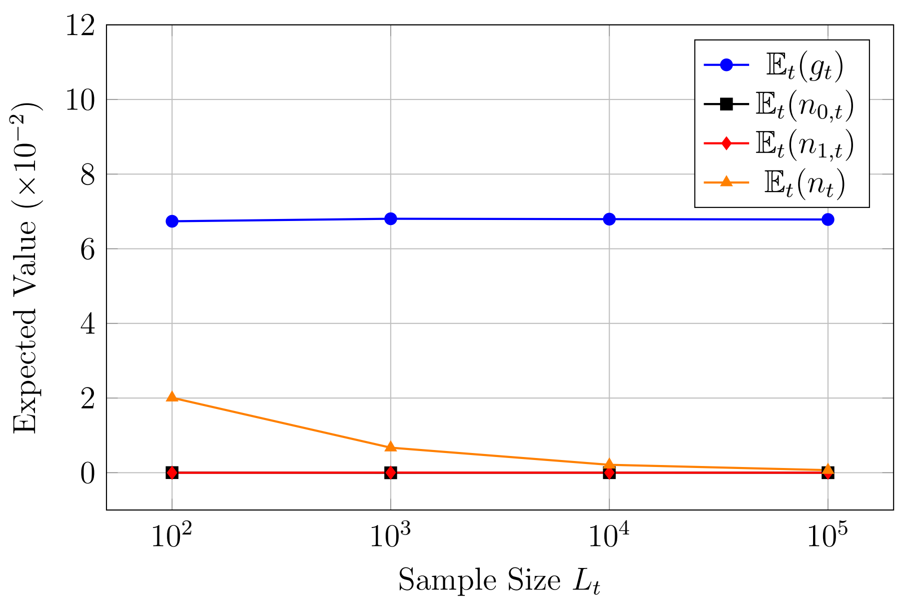
B Figures

Figure 1: Stocks and Flows



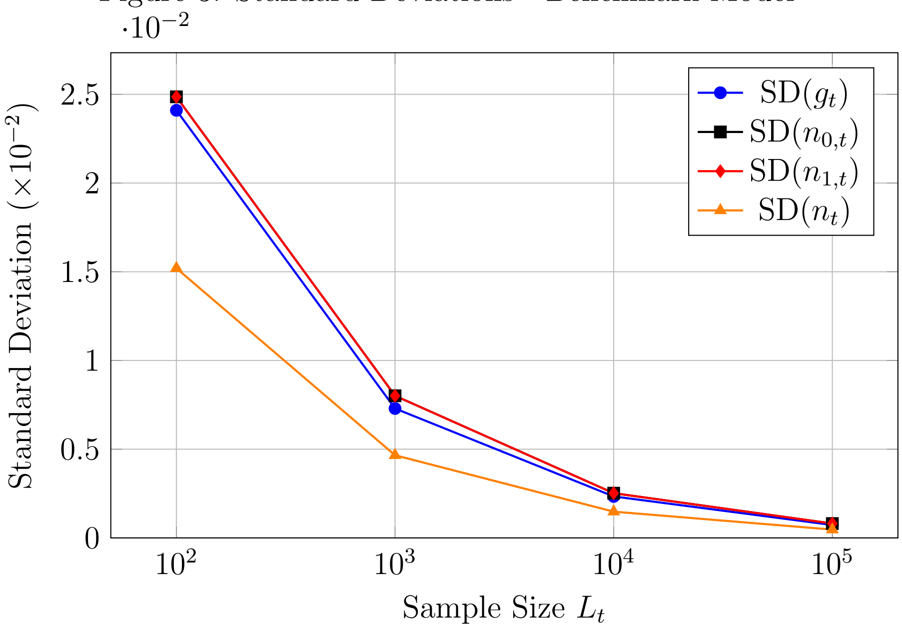
Notes: This figure illustrates a two-sector stock-and-flow model, where arrows represent worker flows between sectors. In this framework, λ_j, where $j \in \{0, 1\}$ and $0 < \lambda_0, \lambda_1 \leq 1$, denote the *outflow rates*, indicating the proportion of workers transitioning from sector j to sector $1 - j$ in each period.

Figure 2: Expected Values - Benchmark Model



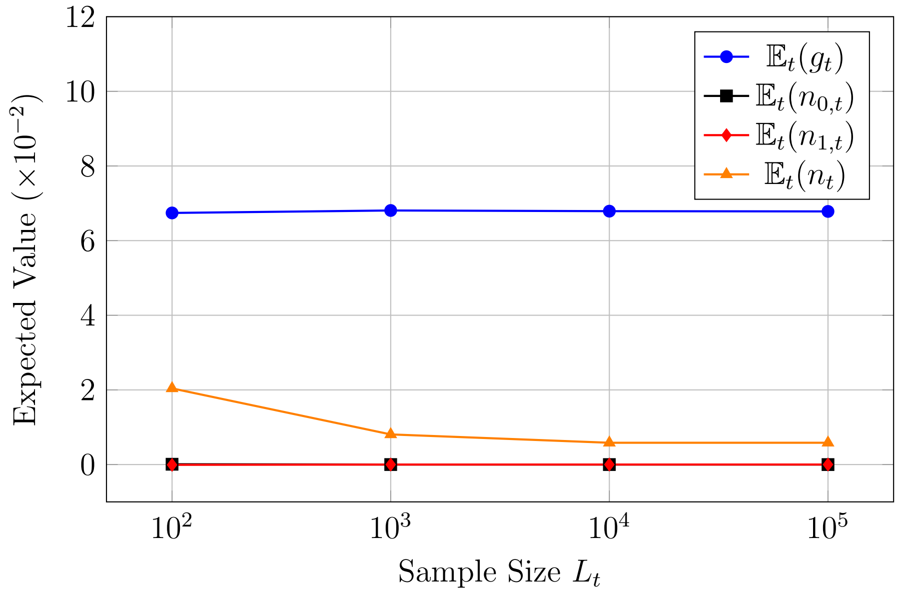
Notes: This figure presents the mean values of the gross mobility rate (g_t) and net mobility rates ($n_{0,t}$, $n_{1,t}$, and n_t), scaled by 10^{-2}, across different sample sizes L_t. In this figure, $\mathbb{E}_t(n_t)$ declines steadily with increasing sample size and eventually converges to a value near zero. This convergence is made possible by the equality and constancy of the transition rates λ_j, which ensure that worker flows between the two sectors are symmetric and reciprocal. As a result, the overestimation problem introduced by the absolute value operator in the net mobility formula is substantially reduced.

Figure 3: Standard Deviations - Benchmark Model



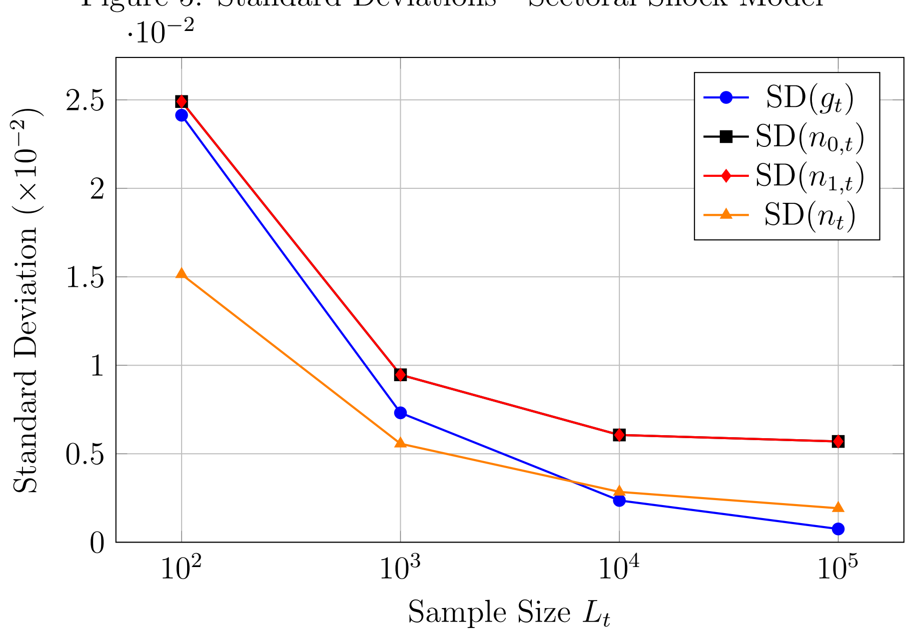
Notes: This figure displays the standard deviations of gross and net mobility rates across varying sample sizes L_t, with all values scaled by 10^{-2}. The results align closely with theoretical predictions, consistent with the model's highly stylized structure and minimal stochastic variation.

Figure 4: Expected Values - Sectoral Shock Model



Notes: This figure presents the mean values of the gross mobility rate (g_t) and net mobility rates ($n_{0,t}$, $n_{1,t}$, and n_t), scaled by 10^{-2}, for different sample sizes L_t. The figure shows that $\mathbb{E}_t(n_t)$ declines as sample size increases but does not converge to zero. This persistence is due to the stochastic shocks and their effect on the overestimation problem introduced by the absolute value operator in equation (6). Negative sectoral mobility in one period cannot offset positive mobility in another, as both are transformed into positive values. Consequently, total net mobility remains strictly positive even in the absence of persistent structural shifts.

Figure 5: Standard Deviations - Sectoral Shock Model



Notes: This figure illustrates the standard deviations of gross and net mobility rates at different sample sizes L_t. Values are scaled by 10^{-2}.