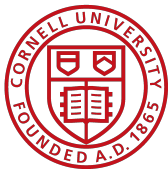


Bias from Aggregation?

Rethinking Occupational Transition & Diversification with
High-Frequency Data



Farnaz Safari

Why this matters for LSMS

- Transitions and diversification are core to resilience and livelihoods.
 - Transition: inter-temporal occupation change
 - Diversification: holding multiple occupations simultaneously
- Standard surveys likely overstate diversification by conflating it with transition, leading to biased welfare associations.
- Distinction matters
 - *Diversification* → Risk Management
 - *Frequent Transition* → Negative Coping (High Transition Cost)

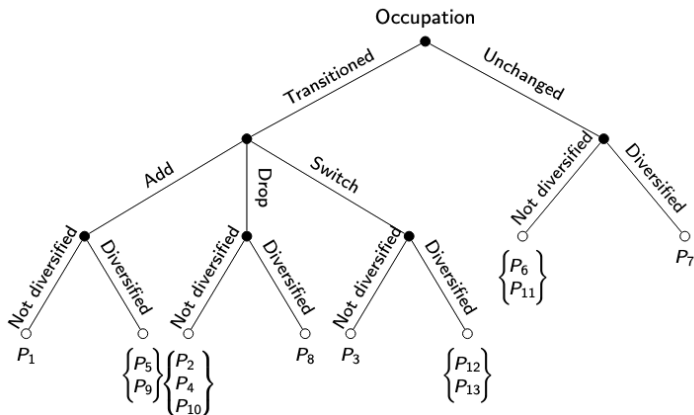
Research Questions

- Q1: How does high-frequency household data improve our understanding of livelihood adaptation by disentangling short-term transitions from moment-specific diversification? (Intertemporal Aggregation Bias)
- Q2: Is concurrent diversification driven by *multi-tasking within households*, or *specialization across members*? (Intra-household Aggregation Bias)
- Q3: What drives selection into different forms of occupational transition and diversification given household characteristics, contextual factors, and past choices? (Selection)
- Q4: What are the welfare associations of transition and diversification, and how might correcting the bias in diversification estimates clarify the inconclusive evidence on its benefits? (Welfare Links)

Data

- **Source:** Rapid Feedback Monitoring System (RFMS), 12 districts in rural Malawi, monthly in-person survey
- **Coverage:** 5,400 households observed monthly over 50 consecutive months (August 2020 - December 2024)
- **Core variables:**
 - Monthly and individual-level income activity participation **combines quarterly income modules + monthly information on some activities**
 - Monthly tracking of food security indicators (FCS, HHS, rCSI)
- **Additional modules:** Subjective shock reports; assets; credit & savings; labor supply; land use; demographics

Defining "Diversification" vs. "Transition"



Three survey design challenges:

1. Inter-temporal aggregation bias (transitions vs. concurrent diversification)
2. Intra-HH aggregation bias (multi-tasking vs. specialization)
3. Recall limitations and measurement error Graph

Q1: What high-frequency data reveal

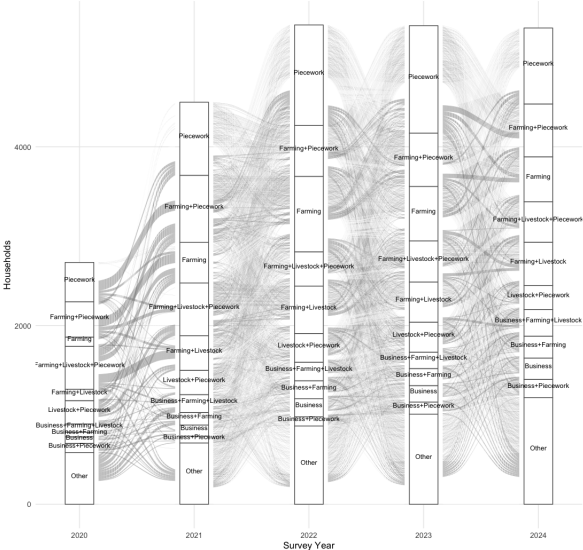
Intertemporal Aggregation Bias

ID	t	$t+1$	Description	Freq.	Perc. (%)
P1	o	d_j	Start activity after unemployment	887	0.41
P2	d_j	o	Lose the only income source	2,282	1.04
P3	d_j	d'_j	Switch activities	11,360	5.19
P4	d_j, d'_j	d_j	De-diversification through withdrawal	16,385	7.49
P5	d_j	d_j, d'_j	Diversification through adding	15,797	7.22
P6	d_j	d_j	Unchanged, undiversified	40,835	18.66
P7	d_j, d'_j	d_j, d'_j	Unchanged, diversified	67,973	31.06
P8	d_j, d'_j, d''_j	d_j, d'_j	De-diversified but still diversified	21,286	9.73
P9	o	d_j, d'_j	Start multiple activities after unemployment	316	0.14
P10	d_j, d'_j	o	De-diversification, become unemployed	397	0.18
P11	o	o	Remain unemployed	2,189	1.00
P12	d_j, d'_j	d_j, d''_j	Permanent + temporary activity, diversified	37,153	16.98
P13	d_j, d'_j	d''_j, d'''_j	Switch all activities while remaining diversified	1,998	0.91
Total			Observations from 5,400 households over 50 months	218,858	100.00

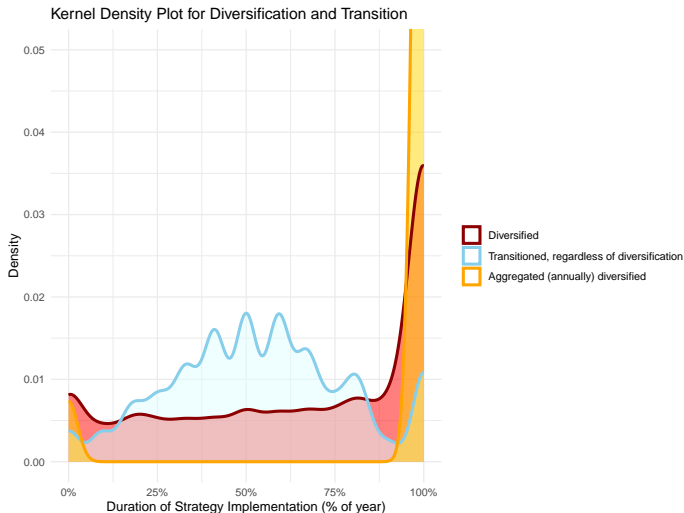
- In the short term, households are unlikely to switch activities solely for higher earnings; they weigh the net benefits, considering transition costs.
- **Over half of the sample engaged in diversification.**
- Frequent switching occurs as seasonality and shocks shift relative payoffs, altering feasible occupational opportunities over time, invisible in single annual snapshots.

Q1: Portfolio Transition

Transitions Across Yearly Portfolios (All Households)
 Top 10 portfolios kept



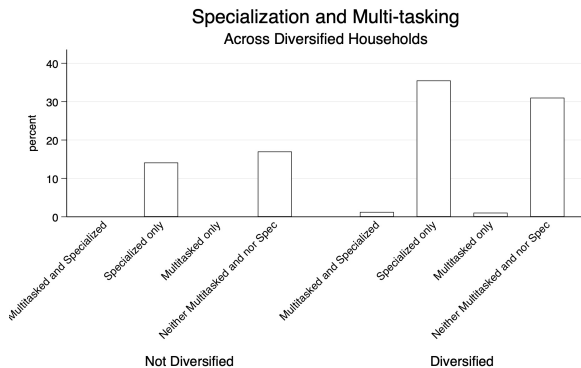
Q1: Inter-temporal aggregation bias



- Transitions cluster in roughly \sim half the year
- Concurrent diversification is *bimodal*: brief spells vs. persistent portfolios
- Annual “aggregated” indicator distorts patterns: Mass piles up at 100%, making it appear that households diversify all year \rightarrow **overstate** the **prevalence** and **persistence** of diversification

Q2: Multi-tasking within households

Intra-Household Aggregation Bias



Specialized and Multitasking Status	Frequency	Percent
Not specialized, no multi-tasking	671,250	84.98%
Specialized only	113,239	14.34%
Multi-tasking only	3,110	0.39%
Specialized and multi-tasking	2,286	0.29%
Total	789,885	100.00%

Table: Distribution of Specialized and Multitasking Households (Diversified Only)

Q3: Dynamics & Determinants

Selection

What drives transitions and diversification?

1. Changes in the choice set
2. Changes in payoffs

Path dependence: past choices reshape both payoffs and the available choice set. A dynamic specification is needed to capture how past decisions influence current strategy selection. Dynamic Specification

Estimated using the Arellano–Bond estimator. AB estimator

Outcomes: seasonal transition and diversification intensity and portfolio breadth. Outcome Variables

Q3: Dynamics & Determinants

Selection

	(1) Transition	(2) Adding	(3) Dropping	(4) Switching	(5) Diversification	(6) Portfolio Breadth
Transition Lag	0.122* (0.065)				-0.707*** (0.032)	
Adding Lag		0.951 (0.656)				
Dropping Lag			-0.161 (0.217)			
Switching Lag				1.211*** (0.274)		
Diversification Lag	-0.210*** (0.069)	0.129 (0.167)	0.202*** (0.046)	-0.718*** (0.268)	0.272*** (0.025)	-0.005** (0.003)
Portfolio Breadth Lag						0.735** (0.299)
TLU Lag	2.770 (2.309)	-5.897** (2.790)	-3.095*** (1.082)	20.930** (8.293)	15.690*** (1.341)	0.169*** (0.046)
Financial Access Lag	1.213** (0.484)	-0.793 (0.514)	0.140 (0.222)	3.269** (1.392)	4.513*** (0.428)	0.034*** (0.006)
Specialized Members	-4.603*** (0.487)	-1.174*** (0.446)	-0.680 (0.459)	-9.155*** (1.147)	4.365*** (0.417)	-0.057*** (0.014)
Observations	71533	71533	71533	71533	71533	71532

Note. Most outcomes pass diagnostic tests (AR(1) rejected, AR(2) not), except switching and diversification, which fail AR(2); diversification also fails Hansen ($p = 0.00$). Also, a coefficient above 1 (e.g., 1.211) is possible in a linear dynamic model; it is not a probability model and reflects strong persistence.

Q4: Benefits of diversification versus transition

Welfare Links

	(1) HHS	(2) HHS	(3) FCS	(4) FCS
L1 Transition	0.028*** (0.006)		-0.134*** (0.051)	
L1 Diversified	-0.037*** (0.011)	-0.036** (0.015)	0.538*** (0.097)	0.508*** (0.126)
L1 Portfolio Breadth	-0.007 (0.017)	-0.029 (0.020)	0.344** (0.154)	0.509*** (0.170)
TLU	-0.223*** (0.042)	-0.223*** (0.042)	1.258*** (0.399)	1.237*** (0.399)
Financial Access	-0.068*** (0.010)	-0.068*** (0.010)	1.022*** (0.092)	1.017*** (0.092)
Specialized Member	0.030*** (0.010)	0.030*** (0.010)	0.219*** (0.083)	0.219*** (0.083)
Observations	188,772	188,772	188,772	188,772
R ²	0.510	0.510	0.612	0.612
Controls included	Yes	Yes	Yes	Yes
FE included	Yes	Yes	Yes	Yes

Note. Fixed-effects regressions; SEs clustered at household level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Q4: Bias From Aggregation?

Concurrent vs. Annual

	(1) HHS	(2) HHS	(3) FCS	(4) FCS	(5) HHS annual	(6) FCS annual
Concurrent Diversification	-0.084*** (0.012)	-0.086*** (0.013)	1.190*** (0.106)	1.262*** (0.113)		
Annual Diversification (biased)					0.037 (0.030)	0.179 (0.265)
Transition	0.059*** (0.006)		-0.138*** (0.052)			
TLU	-0.224*** (0.043)	-0.222*** (0.043)	1.424*** (0.395)	1.376*** (0.395)	-0.172*** (0.038)	0.826*** (0.318)
Financial Access	-0.075*** (0.010)	-0.074*** (0.010)	1.165*** (0.089)	1.154*** (0.089)	-0.042*** (0.006)	0.800*** (0.062)
Specialized Member	0.035*** (0.009)	0.034*** (0.009)	0.081 (0.080)	0.094 (0.080)	-0.050*** (0.008)	0.213*** (0.072)
Multitasked Member	0.048** (0.019)	0.048** (0.019)	0.958*** (0.190)	0.935*** (0.190)	0.038*** (0.011)	0.644*** (0.104)
Observations	208,446	208,446	208,446	208,446	208,446	208,446
R ²	0.480	0.480	0.583	0.583	0.759	0.833
Shock Control	Yes	Yes	Yes	Yes	Yes	Yes
FE included	Yes	Yes	Yes	Yes	Yes	Yes

Note. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Summary & Conclusion

- **Both** concurrent diversification and transition are **widespread**.
- 15% of the **concurrent diversification** comes with long-term engagement, fostering **specialization**.
- Transition, diversification, and portfolio breadth are all **path dependent**.
- The presence of **specialized members** is associated with **greater diversification**, and **less transition**.
- **Diversification** is **positively associated** with food security, whereas **transition is negatively** associated.
- **Frequency of measure matters** when we talk about diversification. **Intertemporal bias attenuates** both the positive **correlation** with consumption and the negative correlation with hunger **toward zero**.

Thank You!

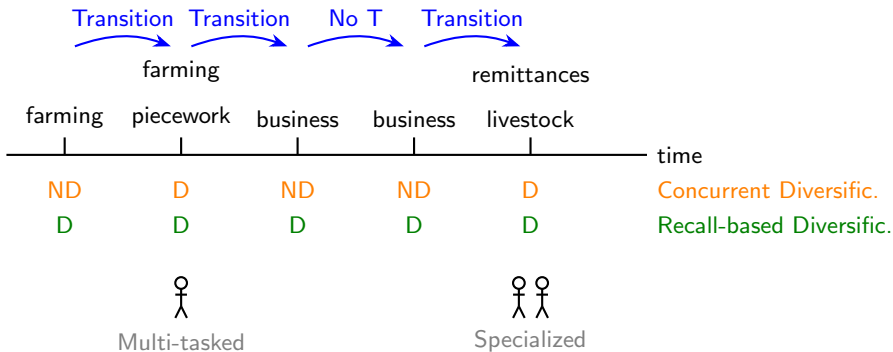
fsf6@cornell.edu

Takeaways for LSMS

- **Measurement matters** Temporal aggregation biases diversification upward and distorts welfare links.
- **Concurrent diversification** Use shorter recall / seasonal blocks and ask which activities are done at the same time.
- **Separate transitions from diversification** Track timing of starts/stops rather than only “ever in last 12 months”.
- Use individual-level modules to **distinguish member specialization from multi-tasking** and avoid intra-household aggregation bias.
- High-frequency evidence like this can also **benchmark** AI/ML tools trained on lower-frequency survey data.

Contributions from High Frequency Data

- *Captures seasonal dynamics:*
 - Seasonality drives fluctuations in occupational opportunities and payoffs, introducing uncertainty essential for understanding behavior and well-being.
- *Record moments beyond the mean:*
 - Welfare hinges on variance, not just annual averages.
 - Standard data smooths over shocks.
- Document non-stationarity and dynamics of decisions.
 - Dynamic processes with persistent effects.



Back

Dynamic Specification

$$y_{ist} = \alpha + \rho y_{i,s-1,t} + \underbrace{\beta \mathbf{X}_{ist}^{\text{exog}}}_{\text{Exogenous controls}} + \underbrace{\theta \mathbf{Z}_{i,s-1,t}^{\text{endo}}}_{\text{Endogenous controls (lagged)}} + \underbrace{\lambda_{st}}_{\text{Year} \times \text{Season FE}} + \underbrace{\eta_i}_{\text{HH FE}} + \varepsilon_{ist}.$$

$\mathbf{X}_{ist}^{\text{exog}} = [\text{HH size}_{ist}, \text{Head Charact.}_{ist}, \text{Dependency ratio}_{ist}, \text{Exogenous Shock Experienced}_{ist}]$,

$\mathbf{Z}_{i,s-1,t}^{\text{endo}} = [\text{TLU}_{i,s-1,t}, \text{Severe Coping Strategies}_{i,s-1,t}, \text{Financial access}_{i,s-1,t}, \text{Specialized Members}_{i,s-1,t}, \text{Multitasked Members}_{i,s-1,t}]$.

$y_{is,t} = \{D_{is}^{\text{conc}}, T_{is}, B_{is}\} \in [0, 1]$: Seasonal intensity [Back](#)

Key Outcomes

1. **Concurrent Diversification Intensity:** Share of months within a season when a household engages in multiple income sources.

$$D_{is}^{\text{conc}} = \frac{1}{M_{is}} \sum_{m \in s} \mathbf{1}\{\text{household } i \text{ has } \geq 2 \text{ income sources in } m\} \in [0, 1]$$

2. **Transition Intensity:** Frequency of adding, dropping, or switching primary occupations across adjacent months.

$$T_{is} = \frac{1}{M_{is}} \sum_{m \in s} \mathbf{1}\{a_{im} \neq a_{i,m+1}\} \in [0, 1]$$

3. **Occupational portfolio breadth:** within a season relative to local (EA) opportunities.

$$B_{is,t} = \frac{U_{is,t} - 1}{K_t^{(e(i),s)} - 1} \in [0, 1]$$

- $B_{is,t} = 0$: single source; $B_{is,t} = 1$: engage in all locally available occupations.
- High $B_{is,t}$ with low D_{is}^{conc} \Rightarrow frequent switching rather than concurrent diversification.

Arellano–Bond Estimator

Selection

Dynamic Panel Framework

- Addresses bias in dynamic panels with individual fixed effects:

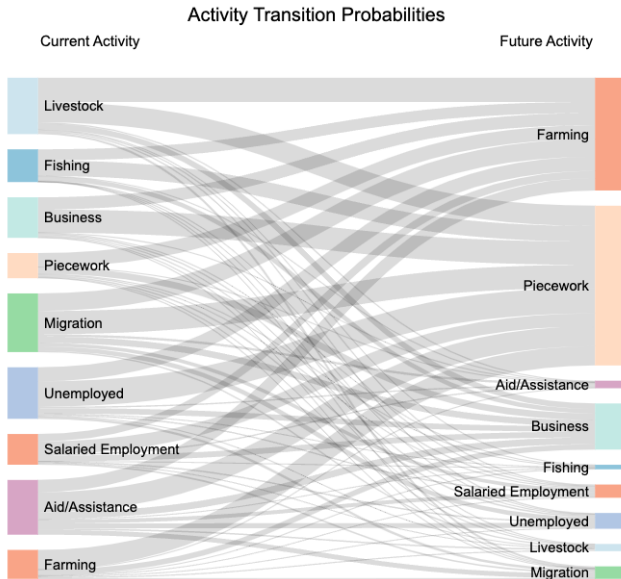
$$y_{is} = \rho y_{i,s-1} + \beta X_{is} + \eta_i + \varepsilon_{is}$$

- **Nickell bias** First-differencing removes η_i but induces correlation between $\Delta y_{i,s-1}$ and $\Delta \varepsilon_{is}$.
- **Endogeneity** Uses *internal instruments*—lagged levels of y and X to achieve consistent GMM estimation.
- **Heteroskedasticity** Two-step GMM applies a *robust weighting matrix* for efficient and consistent inference.

Contributions to Literature

- Adds to the literature on **Diversification** (Reardon, 1997; Barrett et al., 2000; Ellis, 2000b; Barrett et al., 2001a; Reardon et al., 2007), **Specialization** (Haggblade et al., 2005; Timmer et al., 2009; Losch et al., 2012; Davis et al., 2014; Bellon et al., 2020), **Seasonality** (Abay et al., 2017; Ma et al., 2017), and **income dynamics** (Lybbert et al., 2004; Macours et al., 2024)
- **Welfare impact of Adaptation** (Asfaw et al., 2019; Dedehouanou et al., 2020; Tesfaye et al., 2020; Antonelli et al., 2022; Khan et al., 2023)
- Adds to the debate on the benefits of using high-frequency data (Gourlay et al., 2021; Anderson et al., 2024; Joliffe et al., 2024)

Transition Profiles



$$P_{ij} = \Pr(a_{t+1} = a_j \mid a_t = a_i)$$

Diversification Profiles

Pooled LOR Heatmap

