Economic Values of Ecosystem Service: Evidence from Bees

Xiaomeng Cui, Yuanning Liang, and Zimao Xiao*

March 26, 2026

Abstract

Ecosystem services are essential to economic activity, yet their social value is difficult to quantify since markets rarely internalize the externalities they generate. This paper provides causal evidence on the economic value of ecosystem services by studying bees, the primary pollinators in modern agriculture. Assembling comprehensive data on US agriculture and bee colonies, we construct a county-by-year measure of local bee services that combines variation in regional bee supply, crop-specific pollination demand, and local crop structure, and link it to revealed local economic outcomes. To address endogeneity, we develop a novel instrumental-variables strategy that exploits the mass seasonal movements of commercial bees, using machine-learning methods subject to ecological constraints to trace exogenous colony shocks across space. Our IV estimates indicate that the economic value of an additional colony is approximately $5,271, an order of magnitude larger than the observed rental price. The implied aggregate value of US bee colonies exceeds $14 billion annually, highlighting large unpriced benefits in ecosystem service provision and the potential gains from conservation and coordination policies.

Keywords: externality, ecosystem service, pollination, bee, agriculture

*Cui: Institute for Economic and Social Research (IESR), Jinan University, Guangzhou, China (email: cuixiaomeng@jnu.edu.cn). Liang: School of Economics, Peking University, Beijing, China (email: ynliang@pku.edu.cn). Xiao: London School of Economics, London, UK (email: xiaozimaofm3@gmail.com). The authors declare no sources of research support and no financial relationships or other potential conflicts of interest.

1 Introduction

Economic production often relies on natural inputs that are essential yet imperfectly governed by markets. A prominent example is ecosystem services, which are natural processes that often enter production functions and generate value without explicit prices (Costanza et al., 1997). Among these, bee pollination is a textbook case of a positive production externality (Meade, 1952; Bator, 1958). As dominant pollinator for many high-value crops that underpin the global food system, bees generate benefits that spill over across crops, producers, and regions.[1] But these values are not fully revealed by market transactions and remains poorly measured (Baylis et al., 2021).

A widely cited claim that "honeybee pollination adds more than $15 billion annually to US agriculture" has been repeatedly invoked in USDA reports and federal policy discussions. For example, the Obama Administration's 2014 Presidential Memorandum relied on this figure to justify coordinated federal action on pollinator health and habitat. Yet this estimate is based on simple accounting exercises using outdated data, and has been subject to continuous debate.[2] Credible, causal estimates are therefore essential for policy. This need is particularly salient given recent OMB guidance, which emphasizes that changes in ecosystem services should be evaluated within standard benefit-cost analysis frameworks used for regulatory decision-making (Office of Management and Budget, 2024; Fenichel et al., 2025).

Obtaining such estimates is empirically challenging. Even when markets for pollination services exist, observed prices fail to internalize the broader value that pollinators provide given multiple sources of externalities involved . This requires using revealed preference methods to infer the non-market value of pollinators (Freeman III et al., 2014). However, simply associating socioeconomic outcomes with pollinator abundance may introduce significant bias, as both could be simultaneously affected by unobserved factors. These challenges are further exacerbated by the limited availability and quality of ecological observations on pollinators.[3]

[1] Bees are the most economically valuable pollinators worldwide (Watanabe, 1994). Roughly three-quarters of the world's food crops rely to some extent on insect pollination (Klein et al., 2007). For many high-value crops such as almonds, avocados, berries, cocoa, and coffee, yields and quality would be sharply reduced without sufficient pollination services (Siopa et al., 2024).

[2] This estimate first appeared in Morse and Calderone (2000). The headline valuations of this type have then been questioned by later studies (e.g., Winfree et al., 2011; Hanley et al., 2015), and received strong pushback from industry stakeholders like Bayer.

[3] Environmental scientists note that the lack of data over space and time has severely constrained the study on pollinators (Kremen et al., 2011).

In this paper, we address these challenges and provide the first causal estimate on the economic value of bee pollination services. Guided by an economic model, we clarify the major sources of externalities involved that would create a wedge between the social value of bees and private rental fees observed in pollination markets. First, because bees forage across farm boundaries, colonies rented by one farm can raise neighboring farms' output, generating spillovers that private contracts do not internalize. Second, beekeeper investments in colony health reduce the shared burden of disease and other biological stress, so private incentives to maintain colonies fall short of the social optimum. Third, colony deployment and conservation can improve broader landscape conditions, including pollinator abundance, floral resources, and ecosystem resilience, that raise productivity more generally but remain outside market prices.

Making it possible to estimate the economic value of bees, we compile several datasets covering the contiguous US, including comprehensive county-level panels on multiple aspects of local agriculture and the most reliable panel data on seasonal bee colonies over a decade. Using these data, we construct a county-by-year exposure measure of local bee services that combines three sources of variation: (i) the state-by-quarter variation in the availability of bees, (ii) the crop-by-quarter variation in pollination demand, and (iii) the county-by-crop variation in local pollination dependency. Linking differential local bee services to a set of revealed local agricultural outcomes in a panel fixed effect estimation allows us to infer the economic benefits provided by bees.

To address potential endogeneity from non-random distribution of local bee services, we employ a novel instrumental variable strategy that exploits the mass seasonal movements of pollination bees across states. We trace external shocks to bee abundance in any origin state that eventually arrive in a destination state, providing exogenous variation in local bee services. Lacking systematically monitored data on actual bee movements, we address such data challenge by recovering state-to-state seasonal movement flows using a machine-learning approach from patterns of historical bee stock subject to ecological constraints. We show that our algorithm predicts robust migration patterns of pollination bees, similar to those in existing USDA documentations.

Our IV estimates indicate that a one standard deviation (SD) increase in bee services raises net farm income by 1.11 SD. This effect corresponds to approximately $5,271 per additional bee colony, more than 25 times the colony rental price. This wedge reflects large positive externalities jointly shaped by multiple channels illustrated in our economic model. Further decomposition

suggests that, from the lens of farmers, these gains arise from increases in farm income rather than cost reductions. Combining our estimate with the total number of honeybee colonies in the United States implies an aggregate economic value of bee pollination exceeding $14 billion, equivalent to more than 10% of the total value of US agriculture. Although this aggregate value is similar in magnitude to the earlier headline figure, our estimate is derived from a causal framework and reflects contemporary conditions rather than an accounting-based extrapolation from older data.

This paper builds on the literature on quantifying economic values of ecosystem services by focusing on bee pollination service, a unique setting distinct from the existing literature. First, pollination is a critically important component of the global food supply chain and its sustainability directly matters for human survival. This distinguishes from previous studies on the environmental and health consequences of disrupting certain wildlife species such as bats, vultures, and sparrows (e.g., Frank, 2024; Frank and Sudarshan, 2024; Frank et al., 2025). Second, the production externality of bees is spatially mobile due to mass seasonal movements, different from the value of ecosystem services provided by immobile resources such as trees in existing studies (e.g., Taylor and Druckenmiller, 2022; Xing et al., 2023; Han et al., 2024). Third, even though well-developed markets for pollination service exist and in theory they may partially internalize the narrow bilateral externality between farmers and beekeepers (Meade, 1952; Bator, 1958; Cheung, 1973), we show empirically that bees remain substantially underpriced relative to their social value.

This paper makes methodological contribution to the causal estimation of externalities, which has been a central task in public and environmental economics (Chay and Greenstone, 2005; Edlin and Karaca-Mandic, 2006; Hornbeck, 2010; Keiser and Shapiro, 2019, etc.). In particular, we provide the first causal estimate on the economic value of pollination service, getting over the data-limitation hurdle through employing a novel instrumental-variable strategy powered by machine learning subject to ecological constraints. By successfully recovering bee movement paths despite limited data availability, we also show an example of how machine-learning tools can be used to establish critical context-specific data linkages, adding a useful case to economists' toolbox. (Athey and Imbens, 2019).

This paper also has important and immediate policy implications. Our estimates, comparing against the market-revealed rental prices, reflect the exact magnitude of positive externality of bees. These estimates provide a key input required under the 2023 revision of OMB Circular A-4

for benefit–cost analysis of regulations affecting ecosystem services (Fraas et al., 2025; Liscow, 2025). They also help inform the scale and targeting of conservation incentives, including the prioritization of habitat restoration within the Conservation Reserve Program (CRP), where land-use decisions compete across multiple environmental objectives (Otto et al., 2018). More broadly, our results offer a quantitative basis for comparing investments in pollinator habitat against alternative conservation and agricultural support policies, enabling more efficient allocation of public funds toward activities with higher social returns (Baylis et al., 2021).

2 Background

Bee pollination in the United States. Pollination, the process of transferring pollens from the male part (the anther) to the female part (the stigma) of a flower, is crucial for enabling fertilization and production for many food crops. Pollination service was historically provided by the nature, through wind and wild animals. Driven by the increasing demand for pollination, commercial markets have emerged to provide rental services of domesticated pollinators. This phenomenon is epitomized by the honeybee industry in the United States, where mature markets for pollination services have been established since the 1980s (Baylis et al., 2021). In 2024, the total area pollinated by rented honeybees amounted to more than 1.7 million acres in the United States.[4]

Rental services and bee movement. Beekeepers rent out honeybee colonies for pollinating crops such as almonds, cherries, plums, apples during their blooming seasons. Unlike honey from traditional forage crops like alfalfa, honey from these pollination-dependent fruit and tree nut crops is often not palatable. Therefore, the rental services of honeybees become a more specialized business where the majority of beekeepers' revenues may come from collecting pollination fees from crop growers.[5] In recent years, the average pollination fee per colony is below $200.[6] Rental services of honeybee in the United States have a strong migratory nature. Beekeepers truck their hives long distances across multiple states, where the routes are shaped by the sequence of bloom-

[4]See USDA reports, e.g., https://ers.usda.gov/data-products/charts-of-note/chart-detail?chartId=112782.

[5]The latest USDA *Honey* report states that, in 2024, pollination income totaled 226 million while all other incomes from honey bees amounted to 51.3 million. Source: https://www.nass.usda.gov/Publications/Todays_Reports/reports/hony0325.pdf

[6]See USDA publication series of *Cost of Pollination* at https://usda.library.cornell.edu/concern/publications/d504rk335.

ing seasons of different crops in different regions. Among these movements, two major events exist: (i) an influx into California for pollinating almonds in the early spring, and (ii) an exodus toward Northern Plains for foraging in the summer (Bond et al., 2021).

Ecological stressors and bee colony loss. Similar to wild bees, domesticated bees are highly vulnerable to ecological stressors. Since the 1980s, US beekeepers have struggled with parasites and pathogens such as *Varroa destructor* mites and *Nosema* fungus that shorten bee lifespans and contribute to colony loss (Higes et al., 2008; Guzmán-Novoa et al., 2010). The impacts of bee colony losses due to local ecological stressors can be far-reaching given the migratory nature of pollination services in the United States. For instance, a famous media coverage by the PBS reported that the loss to US agricultural economy in 2006–2007 due to Colony Collapse Disorder (CCD) was projected between $8-12 billion.[7] Besides, the loss of natural habitat resulting from crop land expansion and deforestation present recurring challenges to providing forage to pollinators (Hellerstein et al., 2017).

Conservation efforts and ecosystem feedbacks. In the United States, pollination-bee conservation is undertaken by multiple actors. Beekeepers invest directly in colony health through Varroa monitoring and treatment, supplemental feeding, queen replacement, and related hive-management practices.[8] Growers and beekeeper–farmer partnerships also adopt pesticide-risk-reduction and pollinator-friendly management practices, while public agencies and NGOs, including USDA, NRCS, ARS, EPA, and the Honey Bee Health Coalition, support pollinator health through research, technical guidance, and cost-shared habitat programs such as pollinator plantings, flowering cover crops, buffers, and forage restoration.[9] These efforts can generate broader landscape and ecosystem benefits because improved forage and habitat support pollination services beyond a single farm, enhance resources for other beneficial insects, and can also improve soil health, reduce erosion, and improve water quality, (Kremen et al., 2007). At the same time, such conservation is widely thought to be underprovided relative to the social optimum, since the costs of habitat estab-

[7] CCD is an abnormal phenomenon started to spread in 2006, leading to significant hive abandonment and overwintering losses. The cause of CCD has been associated with a set of interacting factors involving exposure to certain agricultural chemicals, viruses, and pathogens, yet some specific mechanisms are still unknown (Baylis et al., 2021). The related media coverage was PBS Episode S26E1: *Silence of the Bees*; https://www.pbs.org/wnet/nature/silence-of-the-bees-introduction/38/

[8] For instance, see practical guidance in Honey Bee Health Coalition (2022, 2024).

[9] For instance, see technical report and practical guidance in National Honey Bee Health Stakeholder Conference Steering Committee (2013); U.S. Department of Agriculture, Natural Resources Conservation Service (2017).

lishment, altered management, and colony-health investments are borne privately, while many of the benefits spill across farms and landowners (Hellerstein et al., 2017).

3 Model

We provide a simple model to illustrate the sources of externalities that would create a wedge between the social and private values of bees even when a well-functioned pollination service market exists. We assume this market consists of N farms indexed by i and M beekeepers indexed by j, where farms demand pollination services by renting bee colonies from beekeepers at a market fee p per colony.

3.1 The Beekeeper's Problem

Beekeeper j chooses how many colonies to deploy for rental, $s_j \geq 0$, and how much to invest in colony health, $e_j \geq 0$. Deploying colonies requires transporting, preparing, and managing hives with increasing difficulty of scaling up deployment through a strictly convex cost function $\kappa(s_j)$ such that $\kappa' > 0$ and $\kappa'' > 0$. Conservation investment e_j encompasses practices such as nutritional supplementation and integrated pest management.

The health of beekeeper j's colonies is governed by a function $\phi(e_j, E_{-j})$, where $E_{-j} = \sum_{j' \neq j} e_{j'}$ is aggregate conservation investment by all other beekeepers. We assume that own investment improves colony health at a diminishing rate, and others' investment confers an additional benefit by reducing the shared burden of contagious colony diseases and other stress (i.e., $\phi_e > 0$, $\phi_{ee} < 0$, and $\phi_E > 0$), consistent with ecological understanding (Brosi et al., 2017). Beekeeper j also faces a strictly convex conservation cost $c(e_j)$ with $c' > 0$ and $c'' > 0$.

Taking the rental fee p and aggregate conservation E_{-j} as given, beekeeper j chooses colony supply s_j and conservation investment e_j to maximize profit:

$$\max_{s_j, e_j} \quad p s_j \phi(e_j, E_{-j}) - \kappa(s_j) - c(e_j). \tag{1}$$

The first-order conditions imply that her colony deployment and conservation efforts are deter-

mined by equating her private marginal revenues and marginal costs.

$$p\,\phi(e_j, E_{-j}) \;=\; \kappa'(s_j), \tag{2}$$

$$p\,s_j\,\phi_e(e_j, E_{-j}) \;=\; c'(e_j). \tag{3}$$

3.2 The Farmer's Problem

Farm i rents q_i colonies and combines them with a vector of other inputs $\mathbf{x}_i$ to produce output y_i sold at price $\bar{p}_y$. The farm's production technology is $y_i \;=\; f(\tilde{q}_i, \Lambda(Q,E), \mathbf{x}_i)$, where f is increasing and strictly concave in all arguments.[10] The first productive input is effective pollination, $\tilde{q}_i \;=\; q_i + \alpha Q_{-i}$, with $Q_{-i} = \sum_{k \neq i} q_k$ denoting colonies rented by all other farms and $\alpha \in [0,1)$ a spillover coefficient. Because honeybees forage over radii that routinely span farm boundaries, a share α of neighboring rental activity cross-pollinates farm i's crops at no cost to farm i.

The second productive input is a landscape health index $\Lambda(Q,E)$, defined over aggregate colony rental $Q = \sum_i q_i$ and aggregate conservation investment $E = \sum_j e_j$. This index captures broader ecosystem services, like wild pollinator abundance, floral diversity, and habitat resilience, that are distinct from the direct services of rented managed colonies and that enter every farm's production as a non-excludable public good, similar to the ecological framework in Kremen et al. (2007). We assume Λ is increasing and concave in both arguments and exhibits complementarity (i.e., $\Lambda_Q > 0$, $\Lambda_E > 0$, and $\Lambda_{QE} > 0$), so that a larger overall colony stock raises the marginal ecosystem return to conservation investment and vice versa.

Taking others' actions as given, farm i maximizes private profit by choosing the levels of colonies rented and other inputs:

$$\max_{q_i, \mathbf{x}_i} \quad \bar{p}_y \cdot f(\tilde{q}_i, \Lambda(Q,E), \mathbf{x}_i) \;-\; p\,q_i \;-\; \mathbf{w} \cdot \mathbf{x}_i, \tag{4}$$

[10] Agronomic studies have indicated input complementarity between pollination and other conventional inputs (Tamburini et al., 2019). But we acknowledge this complementarity does not necessarily create any externality if farmers are sufficiently rational and there is no information friction.

where $\mathbf{w}$ is the vector of input prices. The first-order condition for colony rental is

$$\bar{p}_y \left[\frac{\partial f_i}{\partial \tilde{q}_i} + \frac{\partial f_i}{\partial \Lambda} \Lambda_Q(Q,E) \right] = p. \tag{5}$$

The farmer equates the rental fee to two private marginal benefits: the direct managed-colony pollination effect $\bar{p}_y \frac{\partial f_i}{\partial \tilde{q}_i}$, and the landscape productivity feedback $\frac{\partial f_i}{\partial \Lambda} \Lambda_Q(Q,E)$, which arises because her marginal colony raises aggregate Q.

3.3 The Social Planner's Problem

A social planner chooses the allocations $\{q_i, \mathbf{x}_i\}_{i=1}^N$ and $\{s_j, e_j\}_{j=1}^M$ to maximize total surplus. Since the rental fee p is a transfer from farmers to beekeepers, it nets out of aggregate surplus. Total surplus is therefore gross farm revenue minus farm input costs minus beekeeper deployment and conservation costs:

$$\begin{aligned} \max_{\{q_i, \mathbf{x}_i, s_j, e_j\}} \quad & \sum_{i=1}^{N} [\bar{p}_y f(q_i + \alpha Q_{-i}, \Lambda(Q,E), \mathbf{x}_i) - \mathbf{w} \cdot \mathbf{x}_i] - \sum_{j=1}^{M} \left[\kappa(s_j) + c(e_j) \right] \\ \text{s.t.} \quad & \sum_{i=1}^{N} q_i \leq \sum_{j=1}^{M} s_j \phi(e_j, E_{-j}), \end{aligned} \tag{6}$$

with $\mu \geq 0$ the multiplier on the colony-supply constraint.

The planner's optimal colony allocation accounts for the full effect of q_i on all farms:

$$\bar{p}_y \frac{\partial f_i}{\partial \tilde{q}_i} + \Lambda_Q(Q,E) \sum_{k=1}^{N} \bar{p}_y \frac{\partial f_k}{\partial \Lambda} + \alpha \sum_{k \neq i} \bar{p}_y \frac{\partial f_k}{\partial \tilde{q}_k} = \mu. \tag{7}$$

An additional unit of q_i raises $\tilde{q}_i$ by one, raises $\tilde{q}_k$ for every $k \neq i$ by α through the cross-farm spillover, and raises Λ through aggregate Q at rate Λ_Q, which in turn raises every farm's output through f_Λ. Comparing with the private condition (5), the planner internalizes the two omitted terms: the direct cross-farm spillover ($\alpha \sum_{k \neq i} \bar{p}_y \frac{\partial f_k}{\partial \tilde{q}_k}$) and the landscape effect on other farms ($\Lambda_Q(Q,E) \sum_{k \neq i}^{N} \bar{p}_y \frac{\partial f_k}{\partial \Lambda}$). Since both omitted terms are strictly positive, it immediately implies that the shadow value exceeds the market price for bees (i.e., $\mu > p$).

We recognize that externalities originated from beekeepers' behavior also play an important

role. The planner's first-order condition for colony supply is

$$\mu\,\phi(e_j, E_{-j}) \;=\; \kappa'(s_j). \tag{8}$$

Comparing against the beekeeper's private condition (2), it implies that, with $\mu > p$, the planner would find it optimal to deploy strictly more colonies than the decentralized market equilibrium.

Regarding the optimal conservation investments, the planner accounts for all channels through which e_j affects overall welfare:

$$\mu\left[s_j\,\phi_e(e_j, E_{-j}) \;+\; \sum_{j'\neq j} s_{j'}\,\phi_E(e_{j'}, E_{-j'})\right] \;+\; \Lambda_E(Q,E)\sum_{i=1}^{N}\bar{p}_y\frac{\partial f_i}{\partial \Lambda} \;=\; c'(e_j). \tag{9}$$

An additional unit of e_j raises the effective services of beekeeper j's own colonies through ϕ_e, reduces the shared burden on all other beekeepers' colonies through ϕ_E, and improves the landscape index Λ through Λ_E, raising every farm's output through f_Λ. Even if $p = \mu$, the beekeeper's private condition (3) recovers only the first term in the bracket, while the burden-reduction benefit to other beekeepers and the landscape complementarity term are both absent. It also implies that beekeepers under-provide conservation as compared with the social optimal.

Since the socially optimal colony deployments and conservation investments are both higher than in the decentralized market equilibrium, the implied landscape health index Λ and the marginal products f_Λ and f_q would also be higher, contributing to the social value of bees as reflected in equation (7).

As summarized in Table 1, our model illustrates that cross-farm pollination spillover, cross-beekeeper burden reduction, and landscape health feedback are important sources of externalities that are not internalized in the pollination service markets, and they jointly contribute to the wedge between the social value and rental fee of bees.

Table 1: Sources of Externalities in the Pollination Market

Source	Omitted terms	Direction
Cross-farm pollination spillover:	$\alpha \sum_{k \neq i} \bar{p}_y \frac{\partial f_k}{\partial \tilde{q}_k}$	> 0
Cross-beekeeper burden reduction:	$\mu \sum_{j' \neq j} s_{j'} \phi_E(e_{j'}, E_{-j'})$	> 0
Landscape health feedback:	$\Lambda_Q(Q,E) \sum_{k \neq i} \bar{p}_y \frac{\partial f_k}{\partial \Lambda}$, $\Lambda_E(Q,E) \sum_{i=1}^{N} \bar{p}_y \frac{\partial f_i}{\partial \Lambda}$	> 0

4 Data

This section outlines the primary data sources and measures used in this research. We first provide the raw datasets collected from multiple sources. Next, we describe the county-by-year measurement of local bee pollination service constructed using these data.

Bee colony and loss. We use data from *Bee Colony Loss Quarterly Survey* published by USDA's National Agriculture Statistics Services (NASS). NASS mails out questionnaires to bee keepers about a week before the start of January, April, July, and October in each calendar year, inquiring information during the past quarter. The data available to us are quarter-by-state statistics since 2015, including the total number of bee colonies, the number of deadouts, and the proportion of bee colonies that are affected by various threats to health of bees.[11]

Farm operations. The Regional Economic Information System (REIS) from the Bureau of Economic Analysis (BEA) provides annual county-level data on agricultural revenues and expenses up to 2021. The variable of primary interest is net farm income, defined as the difference between gross farm income and total production expenses. Gross farm income contains (i) cash receipts from marketing (crops and livestock), (ii) government transfer payments, and (iii) other farm related income. Total production expenses include expenses on (i) feed, seed, and fertilizer, (ii) livestock purchases, (iii) petroleum products, (iv) hired farm labor, and (v) other miscellaneous.

Crop production. We obtain county-level crop-specific acreage from the US Census of Agriculture. We consider the following crops as pollinator-dependent: alfalfa for seed, almonds, apples,

[11] The threats include: (1) varroa mites; (2) other pests and parasites; (3) diseases; (4) pesticides; (5) other causes, including weather, starvation, insufficient forage, queen failure, hive damage/destroyed, etc.; (6) unknown causes.

apricots, avocados, blackberries, blueberries, cantaloupe, cherries, clover for seed, cranberries, cucumber, honeydew melon, kiwifruit, nectarines, oranges, peaches, pears, plums, pomegranate, pumpkin, raspberries, sunflowers and watermelon. This list is consistent with USDA reports and agronomic studies, and we attach pollinator-dependency (between 0-1) to each crop based on multiple ecological studies (Martin et al., 2019, e.g.,). We also search and hand-collect flowering period of each crop for categorizing the quarterly demand for pollination.

Other data. Daily data on average, maximum and minimum temperature, and total precipitation over 2015-2024 are obtained from PRISM historical weather data. We follow Schlenker and Roberts (2009) and aggregate 4km-by-4km pixel data to the county level, weighted by cropland areas. Daily data are aggregated to annual level by forming various 5°C temperature bins and total precipitation over the course of a calendar year. We also supplement a set of socioeconomic variables from the US Census of Agriculture, including county population, farmland area, farmland value, soil conditions (including soil salinity, soil erosion, the percent of soil that is clay or sand, soil moisture capability, and soil permeability), latitude and longitude, etc.

5 Empirical Strategy

5.1 Estimating the Economic Value of Bee Pollination Services

Bee pollination service to counties. The bee pollination service provided to a county is proportional to the state-level total number of bee colony, weighted by the county's crop area of all pollination-dependent crops. The rationale is that the relative bee colony services provided across the counties within a state should be driven by bee pollination demands from growing certain types of crops. The calculation is as follows:

$$\text{BeeService}_{c,t} = \sum_{q} \left[\frac{\sum_k \alpha_k \text{Area}^{\text{Pre}}_{k,c,q}}{\sum_{c \in s} \sum_k \alpha_k \text{Area}^{\text{Pre}}_{k,c,q}} \cdot \text{BeeColony}_{s,q,t} \right], \tag{10}$$

where $\text{Area}^{\text{Pre}}_{k,c,q}$ denotes the land area of pollination-dependent crop k in county c during quarter (season) q in 2012.[12] $\alpha_k \in (0,1)$ is the crop-specific relative pollination dependency drawn from Martin et al. (2019). Hence, $\alpha_k \text{Area}^{\text{Pre}}_{k,c,q}$ represents the effective area of land in a county that relies

[12] We use planted acres for row crops and acres bearing for tree crops, and use the census-reported acreage in year 2012 so that the acreage is unaffected by bee colony numbers during our sample period 2015-2024.

on bee pollination services.[13] $\text{BeeColony}_{s,q,t}$ is the state-by-quarter bee colony numbers from the USDA survey. $\text{BeeService}_{c,t}$ thus incorporates three layers of variation: (i) state-by-quarter variation of bee colony numbers; (ii) crop-by-quarter variation in flowering seasons that demand pollination services; (iii) crop-by-county variation in effective land area of pollination-dependent crops in the pre-analysis period.

Figure 1 illustrates the key sources of temporal and spatial variation in $\text{Bee Service}_{c,t}$. Panel (a) maps the average $\text{BeeService}_{c,t}$ across sample years. Panels (b)-(d) further decompose this measure into its underlying components. We proxy the county-level pollination demand using acreage shares of pollination-dependent crops, which vary systematically across seasons and locations. Panel (b) reports pollination-dependent crops by season with their pollination-dependency indices (see Appendix Table A.2), generating crop-by-quarter variation in $\text{BeeService}_{c,t}$. Panel (c) maps each county's share of pollination-dependent crop area within its state in the pre-analysis period, capturing cross-county differences in the spatial distribution of pollination-dependent crops. Panel (d) shows the season-to-season fluctuations of managed bee colonies in response to pollination demands, using California and North and South Dakota as an example. The patterns of bee movement underlying these fluctuations play a central role in our identification strategy, as described in the next section.

Estimation model. We begin with a panel fixed effect model to estimate the effect of pollination services on revealed economic outcomes. The regression equation is

$$y_{c,t} = \beta \cdot \text{BeeService}_{c,t} + \phi Z_{c,t} + (X_c \times \gamma_t)\Psi + \alpha_c + \gamma_t + \varepsilon_{c,t}, \tag{11}$$

where $y_{c,t}$ denotes an economic outcome in county c of year t, with net farm income as the primary outcome of interest. Weather controls Z_{ct} include temperature bins at 5°C intervals, annual cumulative precipitation, and its squared term. α_c and γ_t are county and year fixed effects, controlling for national trends in agricultural production and regional geographical conditions. Counties may experience differential time trends that jointly affect pollination demand and bee colony health. For example, population growth increases demand for pollination-intensive crops such as

[13] We show in Appendix Table A.10 that our results are robust to not using the crop-specific pollination dependency index.

almonds, while simultaneously intensifying human activity that may disrupt bee colonies. To flexibly account for such time-varying county-specific characteristics that affect agricultural production and bee colonies, we interact a vector of pre-period county characteristics X_c with year fixed effects γ_t. These characteristics include population, farmland area and value, acreage of pollination-dependent crops, and soil characteristics (salinity, erosion, moisture, permeability, sand and clay composition), along with latitude and longitude, separately. Standard errors are clustered at the county level. The coefficient of interest is β, capturing the effect of bee pollination service on agriculture outcomes.

5.2 Causal Identification

The panel fixed effect model may be subject to endogeneity concerns. One might worry that unobserved local farm factors, such as the use of pesticides, may increase agricultural output but harm bee colony health (Sanchez-Bayo and Goka, 2014; Goulson et al., 2015), and local beekeepers' compensating efforts (e.g., split bee hives) in response to negative bee colony shocks (e.g., CCD) could prevent losses of farm profits. In addition to these omitted-variable concerns, reverse causality may exist since the expansion of agricultural production may itself cause changes in the living environment of bees.

Our key explanatory variable, $\text{BeeService}_{c,t}$, is constructed using variation in county-level effective pollination-dependent areas and state-level bee colony numbers. Because the county-level areas are defined using pre-period data and time-invariant, the potential endogeneity most likely arises from the state-level colonies. To address this identification challenge, we employ a novel instrumental variable strategy that generates transported shocks to local bee service induced by non-local exogenous factors, powered by a combination of ecological knowledge and machine-learning techniques.

The movement of bee colony. Our instrumental variable strategy leverages the bee colony movements over the four seasons across states. Beekeepers move their bees across states over the seasons to pollinate crops with different blooming seasons and to find better forage for honey production and colony growth. As a result, in the spring, bees travel from a specific set of origin states where they wintering to another set of destination states to provide pollination services. As crops bloom

sequentially in the spring and summer, bees move interstates to provide pollination services to the corresponding crops. Figure 1 Panel (d) shows a contrasting pattern of bee colony residence between California and the Dakotas over the seasons, where bees are moved into California in the spring and moved out to the Dakotas after summer.

Our IV strategy exploits ecological shocks to bee colonies in the origin states before they arrive at the destination states, and use these as supply shocks to the bee colonies providing pollination services at the destination states. Ecological shocks, such as adverse weather, forage scarcity, or hive destruction, affect bee colony survival rates. The change in these shocks are plausibly exogenous to destination-specific factors, including the use of productive agricultural inputs, and local beekeepers' unobserved efforts.

A key empirical challenge is the lack of available data on the routes and quantities of seasonal bee movements. We develop a data-driven procedure that recovers the interstate bee movement flows across four seasons using machine-learning techniques. With these flows across origin-destination state pairs, we construct a predicted measure of transported bee loss due to ecological shocks, which is then used in 2SLS regressions to instrument local bee services.

Constructing the transported bee loss. To illustrate the procedure, we use California as an example of destination state to demonstrate how we construct the bee loss transported from its origin states. The same procedure is applied to all states in our dataset. Let O denote the set of potential origin states. For each origin state o, a fraction of bee colonies are moved to California. Without actual data on bee movement, we assume the fraction is proportional to (i) the inverse geographic distance between state o and California (proxy for travel cost), and (ii) the effective pollination-dependent acreage of crops in California (proxy for demand), relative to all destination states of origin o. A higher fraction therefore implies that ecological shocks affecting bee populations in origin state o are more strongly transmitted to California.

Using USDA's state-by-quarter data of bee colonies affected by ecological shocks, we construct the transported bee loss for California during its pollination seasons using the following formula:

$$\widehat{\text{BeeLoss}}_{\text{CA},q,t} = \sum_{o \in O} \text{EcolShock}_{o,q,t} \cdot \frac{w_{\text{CA},q}}{\sum_{d \in D} w_{d,q}}, \tag{12}$$

where $w_{d,q} \propto \text{distance}_{o\text{-}d}^{-1}$, and $w_{d,q} \propto \text{Area}_{d,q}^{\text{Pre}}$ for any destination state d (including CA).

In equation (12), $\text{EcolShock}_{o,q,t}$ is the number of adversely affected bee colonies by ecological shocks in origin state o in pollination season q of year t. $w_{d,q}$ is a composite weight for destination state d, proportional to (i) the inverse distance between states o and d ($\text{distance}_{o\text{-}d}^{-1}$), and (ii) the effective pollination acreage in state d ($\text{Area}_{d,q}^{\text{Pre}}$), where $\text{distance}_{o\text{-}d}$ and $\text{Area}_{d,q}^{\text{Pre}}$ have been min-max normalized to [0,1] to ensure comparability across states.[14] We consider a long list of specific functional forms of $w_{d,q}$, as summarized in Appendix Table A.6. Our baseline adopts a simple multiplicative specification where $w_{d,q} = \text{Area}_{d,q}^{\text{Pre}} \times \text{distance}_{o\text{-}d}^{-1}$, while the results are insensitive to alternative specifications (Appendix Table A.5).

The choice of origin states O which contribute to California bee services bears a bias-variance tradeoff. If we were to include a larger number of states as potential origins of bee inflow, it would help with the first stage relevance of the instrument because more data are used in the prediction. However, it could also introduce more noise into the prediction as the number of bee colonies from certain origins might be small or such movement only occurred in a single year. To address this tradeoff, we employ three regularization-based machine learning models to identify primary origin states whose lagged bee colony patterns most strongly predict bee populations in the destination, California.

Model 1 uses *Lasso* regression to select predictors of one- to four-quarter lags of bee colony stocks in all potential origin states o within 3,000 kilometers of California. The lagged predictors dynamically capture both immediate and delayed movement patterns, as beekeepers may move colonies across multiple quarters depending on pollination schedules and seasonal foraging. Model 2 uses *Group Lasso*, grouping the lagged predictors by USDA-defined ecological clusters for bees specifically, reflecting that states within the same ecological clusters are more likely to exhibit similar movement patterns.[15] Model 3 uses *Elastic Net* regression augmented with auxiliary variables capturing geographic distance, pollination-dependent crop acreage, and ecological cluster indica-

[14] The raw measure of $\text{distance}_{o\text{-}d}$ is the Euclidean distance between state centroids, and that of $\text{Area}_{d,q}$ is the pollination-dependency weighted total crop acreage in quarter q in the pre-analysis year 2012.

[15] Ecological clusters group states with similar climate, vegetation, and beekeeping patterns (Bond et al., 2021). For instance, Northern Great Plains states (ND, SD, MT, MN) keep colonies locally in summer for honey production from clover and sunflower, then move colonies to warmer states like California or Texas for winter pollination and overwintering. Clustering states this way leverages known movement patterns for more stable and interpretable analyses.

tors for each origin state, to improve the prediction. Details on the model specifications, as well as their strengths and limitations, are included in Appendix B.2.

Next, for each of the three models, we compute an origin-specific *prediction score* by summing the absolute values of all non-zero coefficients associated with origin state o across the four lags, and the auxiliary variables if included. Finally, we take the average of the prediction scores from the three models to compute a composite score for potential origin states to mitigate model-specific biases and ensure selected origins are robust across different modeling assumptions and statistical perspectives.[16] We then select eight states with the highest composite scores as California's primary origin states.

Bee interstate movement flows. Figure 2a maps the major interstate routes for California in our reconstructed bee movement network, showing both the primary pollination destinations of California-managed colonies and the principal states that supply colonies to California. The width of each flow reflects the relative volume of predicted colony movements between each pair of states. On the left side, where California is treated as the origin, the predicted flows show that California-managed colonies are most frequently moved to Washington, North Dakota, and Oregon, followed by other states in descending order of volume. On the right side, where California is the destination, our reconstruction identifies Texas, Washington, Oregon, Idaho, North Dakota, South Dakota, Montana, and Minnesota as the main eight out-of-state suppliers of colonies arriving in California.

The contrast between California's outbound and inbound flows reflects two features of the system. First, California hosts a large number of local beekeepers who keep their colonies within the state during California's pollination season, reducing the amount of interstate inflow needed. Second, these same beekeepers often ship their colonies to other states during those states' pollination seasons, contributing to larger outbound interstate flows.

Figure 2b extends the visualization of our prediction algorithm to map interstate movement patterns for all major beekeeping states. The diagram ranks states by the relative magnitude of their interstate colony movements. On the origin side, California, Texas, and Florida emerge as the

[16] For example, if Lasso arbitrarily selects Idaho over Oregon due to multicollinearity in their stock patterns, but both Group Lasso (via the Pacific Northwest ecological cluster) and Elastic Net (via geographic proximity captured by inverse distance) assign high scores to Oregon, the composite score will appropriately recognize Oregon's importance.

largest suppliers of colonies, reflecting their substantial commercial beekeeping industries and their role in supporting pollination cycles across the country. On the destination side, North Dakota, Washington, and California stand out as the primary recipients of incoming colonies, consistent with their extensive pollination and summer foraging demand. Our predicted flows align broadly with ecological expectations and USDA statistics (Bond et al., 2021).

We construct the predicted state-level bee loss outlined in equation (12) using shock-affected bee colony numbers in the eight selected major origin states for each destination state. To measure county-level exposure to bee loss, we allocate predicted state-level bee losses to counties using the spatial distribution of pollination-dependent crops, similar to the procedures in equation (10). The instrument is constructed as the first-differenced predicted county-level bee losses. We illustrate the complete workflow of IV construction in Appendix B.

6 The Ecosystem Service Value of Bee Pollination

6.1 Baseline Results

Table 2 Column (1) reports the estimated effect of bee pollination services on annual net farm income using a county-level panel fixed-effects regression based on equation (11). All specifications include flexible weather controls, county and year fixed effects, and baseline county characteristics interacted with year fixed effects. This panel fixed-effects framework exploits within-county, year-to-year variation in bee pollination services provided while controlling for time-invariant county characteristics and common shocks affecting both agricultural outcomes and bee populations. In our baseline specification, both the bee pollination service measure and the outcome variable are standardized. The estimates indicate a positive and statistically significant relationship: a one-standard-deviation increase in the number of bee colonies in a county is associated with a 0.094 standard deviation increase in net farm income.

While informative, this estimate may be attenuated due to endogeneity concerns discussed earlier. Our identification strategy exploits non-local adverse shocks to bee colonies in origin states, which reduce colonies available for transport and create exogenous variation in pollination services in destination counties. Table 2 Columns (2)–(4) present the 2SLS results, using the same set of controls and fixed effects as in equation (11). The first-stage results in Column (4) show a negative and statistically significant effect of transported bee losses on local bee pollination ser-

vices, consistent with fewer viable colonies in origin states leading to a reduced inflow of bees and, consequently, fewer colonies providing pollination services in destination counties. The Kleibergen–Paap F-statistic of 70.86 indicates a strong first stage. The reduced-form estimates in Column (3) similarly show that transported bee losses decrease local net farm income. Column (4) shows a positive and statistically significant causal effect of bee pollination services on net farm income. Specifically, a one-standard-deviation increase in the number of bee colonies in a county raises net farm income by 1.111 standard deviations.

6.2 Ecosystem Service Value

To contextualize our results, we convert the standardized estimate into the implied marginal increase in net farm income per bee colony using the following equation:

$$\widehat{\beta}_{2SLS} \times \frac{\sigma_{\text{Net Farm Income}_{ct}}}{\sigma_{\text{BeeService}_{ct}}} = 1.111 \times \frac{72{,}253}{15{,}230} \times \$1{,}000 \approx \$5{,}271 \text{ per bee colony}, \tag{13}$$

where $\widehat{\beta}_{2SLS}$ is the 2SLS estimate on annual net farm income from pollination services provided by bee colonies, instrumented using transported bee losses; $\sigma_{\text{Net Farm Income}_{ct}}$ and $\sigma_{\text{BeeService}_{ct}}$ are the standard deviations of annual county-level net farm income and annual county-level bee colonies, respectively.

US honey bee colonies totaled approximately 2.71 million on January 1, 2024, according to the USDA. Scaling our estimated effect on net farm income per bee colony by the national colony count yields an implied aggregate economic value of pollination services of roughly $14.28 billion. While our estimate reflects a marginal net income effect—capturing the profitability impact of an additional colony, the existing literature typically reports the total gross production supported by pollination services to be between $15 and $20 billion annually (Morse and Calderone, 2000; Johnson, 2010). The similarity in magnitude across these distinct metrics provides external validation of our empirical approach and underscores the central role of managed pollinators in sustaining US agricultural productivity. The average 2015 to 2024 inflation-adjusted US net farm income is around 150 billion, hence the economic value of bee pollination service contributes to 10% of all US agriculture value. This figure is broadly consistent with existing estimates that place the contribution of pollination services at roughly 1–2% of global GDP (Lippert et al., 2021).

Our estimate of the ecosystem service value of bees emphasize on the positive externalities that honey bees provide to the agricultural system but not yet fully account for in the production process. The value provided by bee pollination services is much larger than that suggested by the market pollination fee, which is around $250-$300 per bee colony. Therefore, the unpriced positive externality per bee colony is around $5,000 annually. Taking into account the 2.71 million total bee colony numbers for agricultural services in 2024, there is $13.5 billion agricultural production value from bee pollination services that is not yet priced in the market.

Our estimate also sheds light on the true economic cost of large-scale colony loss events. At prevailing market rental fees of $250–300 per colony, the market-implied cost of any given colony loss is nearly nineteen times smaller than our causal estimate of $5,271 — precisely because rental fees capture only the narrow bilateral transaction between farmers and beekeepers, leaving cross-farm pollination spillovers, cross-beekeeper burden reduction, and landscape health feedbacks unpriced. This gap offers a natural resolution to a longstanding puzzle: projected agricultural losses of $8–12 billion during the Colony Collapse Disorder episodes of 2006–2007 appeared difficult to reconcile with observed market rental prices, but are entirely consistent with our social value estimate applied to the scale of colony losses documented during that period.

6.3 Sensitivity Checks

Our causal analysis is robust across a series of sensitivity checks detailed in Appendix A. First, regarding the construction of the transported bee loss instrument, we show that results are stable across (1) alternative machine learning tuning parameters, including alternative weighting schemes of the three regularization-based machine learning algorithms, and eight to twelve numbers of origin states selected (Table A.3 and A.4), (2) fourteen alternative migration weighting schemes (Table A.5), and (3) exclusion of bee flows from geographically adjacent states (Table A.7).

Second, our timing assumptions regarding when bees are moved for pollination do not drive the results. Restricting transported bee losses to affect (1) all subsequent quarters after bees arrive (baseline), (2) only the concurrent arrival quarter, or (3) any pollination quarter yields estimates consistent with the baseline (Table A.8).

Third, the county-level bee service measure is robust to (1) alternative census years for crop acreage shares (Tables A.9) and (2) omitting crop-specific pollination dependency parameters (Ta-

ble A.10).

Fourth, our estimates are stable across regression specifications, including (1) excluding county-specific characteristics interacted with year fixed effects and (2) replacing to county and state-by-year fixed effects (Table A.11).

Finally, three falsification tests support our identification: (1) to rule out general agricultural trends as a confound, we show that bee colonies have no significant effect on non-pollinator-dependent crop acreage (Table A.12); (2) to confirm that our bee service measure captures specifically pollination-relevant variation, we replace pollinator-dependent crop shares with grain crop shares and find that the first stage collapses entirely (Table A.13); and (3) to rule out pesticide exposure as a confounding channel, we construct an alternative instrument based on neonicotinoid changes and find a near-zero first stage (Table A.14).

6.4 Other Agriculture Outcomes

Figure 3 illustrates the impact of bee pollination on various agricultural outcomes. It shows that bee pollination raises farm income substantially, while barely increasing costs — so almost all the revenue gain flows through to net income. Panel A presents the reduced-form effects of transported bee losses, while Panel B reports 2SLS estimates across different categories of agricultural income and expenses. Panel C translates these effects into the contribution of each category per bee colony, following the framework in equation (13).

Farm income. We observe positive responses across multiple farm-sector income metrics, using the same 2SLS regression framework in equation (11). To facilitate comparability across different outcomes, all variables—including the number of bee colonies—are standardized, allowing the estimated coefficients to represent standardized effects. Our result indicates that a one–standard-deviation increase in local bee pollination raises county-level gross farm income by 0.314 standard deviations. In monetary terms, one additional colony increases gross farm income by $5,814.

Disaggregating the income components reveals that cash receipts from farm marketing of crops and livestock rise by 0.268 standard deviations. Within this category, cash income from crops increases by 0.265 standard deviations, which suggests a direct role of bee pollination in enhancing crop yields, especially fruits and nuts which typically have higher market values, therefore, in-

creasing farm revenue. There is also a complementary increase in cash from livestock by 0.171 standard deviations consistent with an indirect channel through which bee pollination improves fodder crops such as alfalfa, thereby supporting livestock operations. In monetary terms, bee pollination increases cash receipts from marketing by $4,696, crops by $3,241, and livestock by $1,453.

Beyond direct agricultural marketing, receipts from other farm-related activities—including recreational services, forest product sales, and similar ventures—increase by 0.827 standard deviations, translating to $1,130. This broader response suggests that ecosystem services provided by bees stimulate multiple dimensions of rural economic activity, extending beyond direct crop production.

Government transfer payments also respond positively, though moderately, potentially reflecting policy incentives that encourage ecological or pollinator-friendly farming practices. However, excluding government payments from net farm income still yields a 1.004-standard-deviation increase, confirming that the main result is not driven by subsidy programs. Panel C also shows the relatively small monetary contribution from government payments.

Farm expenses. On the expense side, total production expenses show a small and statistically insignificant response to bee colonies (0.025 SD, $SE = 0.052$), equivalent to roughly $351 per additional colony, indicating that the income gains from bee pollination are not offset by rising costs. Breaking down components: material input expenses—including feed, fertilizer, seed, petroleum, livestock purchases—show a positive response of 0.372 SD, equivalent to $2,467 per colony, consistent with a scale complementarity as higher output calls for greater input use. Farm labor expenses exhibit a small reduction (-0.096 SD, SE=0.152, or -$339), suggesting limited substitution with manual pollination. Though the effect is insignificant as there is a small scale of manual pollination, which implies large agricultural production loss without bee pollination as possible adaptation to manual pollination is limited in the short run. All other miscellaneous operating expenses, such as accounting, advertising, or other overheads, decrease by 0.344 SD, approximately $1,788 per colony, potentially reflecting overhead efficiencies that accompany expanded production scale. Overall, the expense effects are modest and largely offset each other, indicating that increased bee pollination primarily boosts farm revenue rather than significantly altering total production costs.

7 Policy Implication

Our results carry direct implications for the design of policies governing pollinator conservation and agricultural ecosystem services. The positive externality from bee pollination implies that pollination markets systematically underprovide bee services relative to the social optimum. This wedge reflects three distinct externality channels identified in our model, each suggesting a corresponding policy lever.

First, because bees forage across farm boundaries, a share of the pollination benefits generated by any colony accrues to neighboring farms outside the rental contract. Our finding that farm labor expenses do not respond significantly to bee service confirms that manual pollination is not a viable margin of substitution, meaning this cross-farm spillover cannot be eliminated through private adaptation. This provides a rationale for subsidizing colony rental at the farm level, with the magnitude of the subsidy bounded above by the gap between social value and market price.

Second, beekeeper investments in colony health, such as Varroa monitoring and integrated pest management, reduce the shared burden of contagious disease across the broader beekeeper community, a benefit that private incentives fail to internalize. Our finding that the social value of colonies substantially exceeds market prices, even after accounting for direct crop and livestock income channels, is consistent with this conservation externality contributing meaningfully to the overall wedge. Policies that directly subsidize colony health investments by beekeepers are therefore likely to generate social returns exceeding their private returns by a similar order of magnitude.

Third, the complementarity between aggregate colony deployment and conservation investment in our model implies that habitat restoration programs generate higher marginal returns in areas with greater pollinator activity. This suggests that conservation programs such as the Conservation Reserve Program (CRP) should prioritize pollinator habitat in regions with high bee service intensity, such as the Northern Great Plains, rather than allocating resources uniformly across geographies.

More broadly, our causal estimate of $5,271 per colony provides a direct input for regulatory benefit–cost analysis under the 2023 revision of OMB Circular A-4, which now requires explicit accounting of ecosystem service values. Regulations affecting pollinator health — including pesticide approvals, land-use policies, and habitat mandates — have historically been evaluated against

market-implied colony values that, as our results demonstrate, capture less than 6% of the true social value. Our framework thus suggests that the benefits of pollinator-protective regulation have been systematically underestimated in prior policy analyses.

8 Conclusion

This paper quantifies the economic value of pollination services provided by managed honey bees in US agriculture. Ecosystem services such as animal pollination are fundamental to human well-being and agricultural production, yet their economic value is systematically underestimated due to unpriced spillovers, complementarities with wild pollinators, and interactions with other inputs. Combining county-level farm outcomes with detailed data on bee colony stocks, crop-specific pollination dependence, and spatial patterns of agricultural production, we construct a comprehensive measure of local pollination services and document a positive association between bee pollination and farm income.

To establish causality, we exploit exogenous variation in pollination services generated by interstate movement of bee colonies. Our instrumental variable strategy leverages non-local ecological shocks that affect bee colonies in origin states prior to transport, isolating plausibly exogenous supply-driven variation in local pollination services. To recover interstate movement patterns in the absence of direct flow data, we develop a data-driven approach that integrates ecological knowledge with machine learning techniques to identify primary origin states across seasons.

Our results show that a one-standard-deviation increase in bee colonies raises county-level net farm income by 1.11 standard deviations, implying an annual value of more than $5,000 per colony and an aggregate national value of approximately $14 billion. These estimates highlight that the welfare contribution of pollination services substantially exceeds observed market pollination fees. More broadly, this paper contributes to the literature on biodiversity and ecosystem services by providing causal, economy-wide evidence that pollinators are a productive input in agriculture and by offering an empirical framework that can be applied to other mobile ecosystem service providers.

References

Athey, Susan and Guido W Imbens, "Machine learning methods that economists should know about," *Annual Review of Economics*, 2019, *11* (1), 685–725.

Bator, Francis M., "The Anatomy of Market Failure," *The Quarterly Journal of Economics*, 1958, *72* (3), 351–379.

Baylis, Kathy, Elinor M Lichtenberg, and Erik Lichtenberg, "Economics of pollination," *Annual Review of Resource Economics*, 2021, *13* (1), 335–354.

Bond, Jennifer K, Claudia Hitaj, David Smith, Kevin Hunt, Agnes Perez, and Gustavo Ferreira, "Honey bees on the move: from pollination to honey production and back," 2021.

Brosi, Berry J, Keith S Delaplane, Michael Boots, and Jacobus C de Roode, "Ecological and evolutionary approaches to managing honeybee disease," *Nature Ecology & Evolution*, 2017, *1* (9), 1250–1262.

Chay, Kenneth Y and Michael Greenstone, "Does air quality matter? Evidence from the housing market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Cheung, Steven NS, "The fable of the bees: An economic investigation," *The Journal of Law and Economics*, 1973, *16* (1), 11–33.

Costanza, Robert, Ralph d'Arge, Rudolf De Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin Limburg, Shahid Naeem, Robert V O'neill, Jose Paruelo et al., "The value of the world's ecosystem services and natural capital," *Nature*, 1997, *387* (6630), 253–260.

Edlin, Aaron S and Pinar Karaca-Mandic, "The accident externality from driving," *Journal of Political Economy*, 2006, *114* (5), 931–955.

Fenichel, Eli P., Lydia Olander, and Heather Tallis, "Environmental and Ecosystem Services in Benefit–Cost Analysis," *Review of Environmental Economics and Policy*, June 2025, *19* (2), 238–247.

Fraas, Arthur G, John D Graham, Kerry Krutilla, Randall Lutter, Jason F Shogren, and W Kip Viscusi, "Improving distributional analysis in regulatory evaluation: an assessment of the 2023 Circular A-4," *Review of Environmental Economics and Policy*, 2025, *19* (2), 000–000.

Frank, Eyal and Anant Sudarshan, "The social costs of keystone species collapse: Evidence from the decline of vultures in India," *American Economic Review*, 2024, *114* (10), 3007–3040.

Frank, Eyal G, "The economic impacts of ecosystem disruptions: Costs from substituting biological pest control," *Science*, 2024, *385* (6713), eadg0344.

_ , Qinyun Wang, Shaoda Wang, Xuebin Wang, and Yang You, "Campaigning for extinction: Eradication of sparrows and the Great Famine in China," Technical Report, National Bureau of Economic Research 2025.

Goulson, Dave, Elizabeth Nicholls, Cristina Botías, and Ellen L Rotheray, "Bee declines driven by combined stress from parasites, pesticides, and lack of flowers," *Science*, 2015, *347* (6229), 1255957.

Guzmán-Novoa, Ernesto, Leslie Eccles, Yireli Calvete, Janine Mcgowan, Paul G Kelly, and Adriana Correa-Benítez, "Varroa destructor is the main culprit for the death and reduced populations of overwintered honey bee (Apis mellifera) colonies in Ontario, Canada," *Apidologie*, 2010, *41* (4), 443–450.

Han, Lu, Stephan Heblich, Christopher Timmins, and Yanos Zylberberg, "Cool cities: The value of urban trees," Technical Report, National Bureau of Economic Research 2024.

Hanley, Nick, Tom D. Breeze, Chris Ellis, and Dave Goulson, "Measuring the economic value of pollination services: Principles, evidence and knowledge gaps," *Ecosystem Services*, 2015, *14*, 124–132.

Hellerstein, Daniel, Claudia Hitaj, David Smith, and Amélie Davis, "Land use, land cover, and pollinator health: A review and trend analysis," 2017.

Higes, Mariano, Raquel Martín-Hernández, Cristina Botías, Encarna Garrido Bailón, Amelia V González-Porto, Laura Barrios, M Jesus Del Nozal, José L Bernal, Juan J Jiménez, Pilar García Palencia et al., "How natural infection by Nosema ceranae causes honeybee colony collapse," *Environmental Microbiology*, 2008, *10* (10), 2659–2669.

Honey Bee Health Coalition, *Tools for Varroa Management: A Guide to Effective Varroa Sampling & Control* The Keystone Policy Center on behalf of the Honey Bee Health Coalition, eighth ed. August 2022.

_ , *Honey Bee Nutrition: A Review and Guide to Supplemental Feeding* The Keystone Policy Center on behalf of the Honey Bee Health Coalition, first ed. January 2024.

Hornbeck, Richard, "Barbed wire: Property rights and agricultural development," *The Quarterly Journal of Economics*, 2010, *125* (2), 767–810.

III, A Myrick Freeman, Joseph A Herriges, and Catherine L Kling, *The measurement of environmental and resource values: theory and methods*, Routledge, 2014.

Johnson, Renée, *Honey bee colony collapse disorder*, Vol. 444, Congressional Research Service Washington, 2010.

Keiser, David A and Joseph S Shapiro, "Consequences of the Clean Water Act and the demand for water quality," *The Quarterly Journal of Economics*, 2019, *134* (1), 349–396.

Klein, Alexandra-Maria, Bernard E Vaissière, James H Cane, Ingolf Steffan-Dewenter, Saul A Cunningham, Claire Kremen, and Teja Tscharntke, "Importance of pollinators in changing landscapes for world crops," *Proceedings of the Royal Society B: Biological Sciences*, 2007, *274* (1608), 303–313.

Kremen, Claire, KS Ullman, and Robin W Thorp, "Evaluating the quality of citizen-scientist data on pollinator communities," *Conservation Biology*, 2011, *25* (3), 607–617.

_ , Neal M Williams, Marcelo A Aizen, Barbara Gemmill-Herren, Gretchen LeBuhn, Robert Minckley, Laurence Packer, Simon G Potts, T'ai Roulston, Ingolf Steffan-Dewenter et al., "Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change," *Ecology Letters*, 2007, *10* (4), 299–314.

Lippert, Christian, Arndt Feuerbacher, and Manuel Narjes, "Revisiting the economic valuation of agricultural losses due to large-scale changes in pollinator populations," *Ecological Economics*, 2021, *180*, 106860.

Liscow, Zachary, "Benefit–cost analysis of government spending," *Review of Environmental Economics and Policy*, 2025, *19* (2), 000–000.

Martin, Gabrielle, Colin Fontaine, Francesco Accatino, and Emmanuelle Porcher, "New indices for rapid assessment of pollination services based on crop yield data: France as a case study," *Ecological Indicators*, 2019, *101*, 355–363.

Meade, James E., "External Economies and Diseconomies in a Competitive Situation," *The Economic Journal*, 1952, *62* (245), 54–67.

Morse, Roger A and Nicholas W Calderone, "The value of honey bees as pollinators of US crops in 2000," *Bee Culture*, 2000, *128* (3), 1–15.

National Honey Bee Health Stakeholder Conference Steering Committee, "Report on the National Stakeholders Conference on Honey Bee Health," Technical Report, U.S. Department of Agriculture and U.S. Environmental Protection Agency, Alexandria, VA 2013. Conference held October 15–17, 2012; report released May 2, 2013.

Office of Management and Budget, "Guidance for Assessing Changes in Environmental and Ecosystem Services in Benefit-Cost Analysis," Technical Report, Executive Office of the President of the United States, Office of Management and Budget February 2024.

Otto, Clint RV, Haochi Zheng, Alisa L Gallant, Rich Iovanna, Benjamin L Carlson, Matthew D Smart, and Skip Hyberg, "Past role and future outlook of the Conservation Reserve Program for supporting honey bees in the Great Plains," *Proceedings of the National Academy of Sciences*, 2018, *115* (29), 7629–7634.

Sanchez-Bayo, Francisco and Koichi Goka, "Pesticide residues and bees–a risk assessment," *PloS One*, 2014, *9* (4), e94482.

Schlenker, Wolfram and Michael J Roberts, "Nonlinear temperature effects indicate severe damages to US crop yields under climate change," *Proceedings of the National Academy of Sciences*, 2009, *106* (37), 15594–15598.

Siopa, Catarina, Luísa G Carvalheiro, Helena Castro, João Loureiro, and Sílvia Castro, "Animal-pollinated crops and cultivars—a quantitative assessment of pollinator dependence values and evaluation of methodological approaches," *Journal of Applied Ecology*, 2024, *61* (6), 1279–1288.

Tamburini, Giovanni, Riccardo Bommarco, David Kleijn, Wim H van der Putten, and Lorenzo Marini, "Pollination contribution to crop yield is often context-dependent: A review of experimental evidence," *Agriculture, ecosystems & environment*, 2019, *280*, 16–23.

Taylor, Charles A and Hannah Druckenmiller, "Wetlands, flooding, and the clean water act," *American Economic Review*, 2022, *112* (4), 1334–1363.

U.S. Department of Agriculture, Natural Resources Conservation Service, *Pollinator Practices* USDA Natural Resources Conservation Service October 2017.

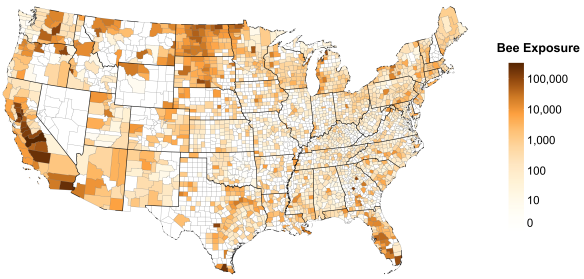
Watanabe, Myrna E, "Pollination worries rise as honey bees decline," *Science*, 1994, *265* (5176), 1170–1170.

Winfree, Rachael, Brian J. Gross, and Claire Kremen, "Valuing pollination services to agriculture," *Ecological Economics*, 2011, *71*, 80–88.

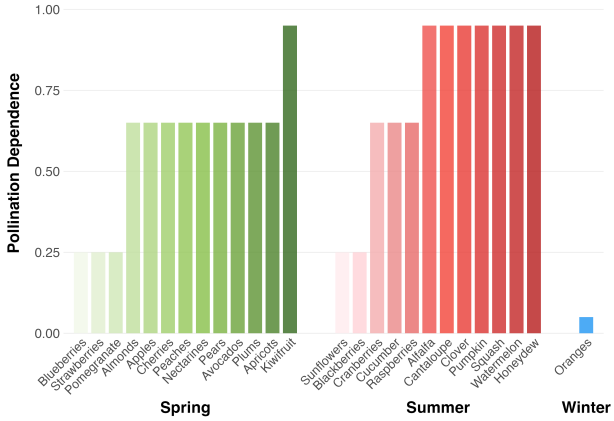
Xing, Jianwei, Zhiren Hu, Fan Xia, Jintao Xu, and Eric Zou, "Urban Forests: Environmental Health Values and Risks," Technical Report, National Bureau of Economic Research 2023.

Figure 1: Data Variation of Bee Service

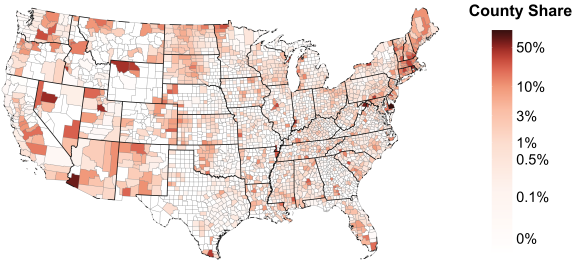
(a) Bee Service



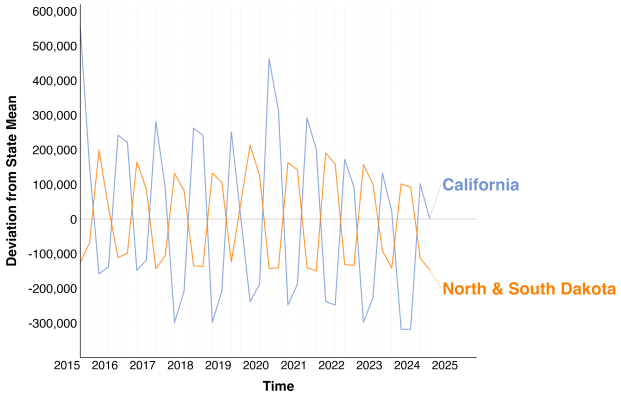
(b) Pollination Crops by Season



(c) Share of Pollination Crop Acreage



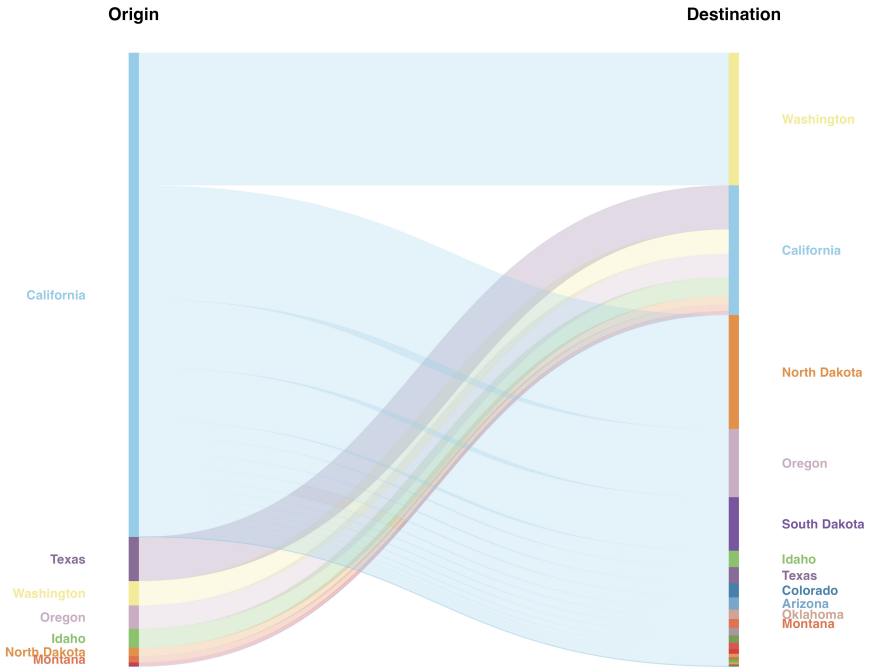
(d) Time Series of Bee Colonies by State



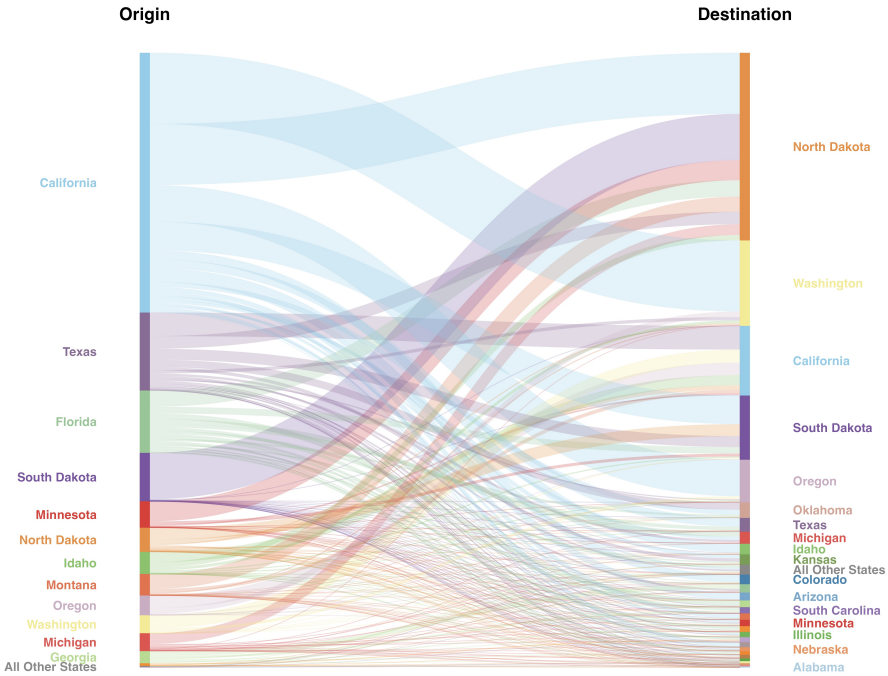
Notes: This figure illustrates spatial and temporal variation in county-level bee service (Bee Service$_{c,t}$). Panel (a) maps the average annual number of bee colonies in US counties during 2015–2024, constructed using equation (10) from USDA survey data. Panel (b) shows representative pollination-dependent crops by flowering season with crop-specific pollination dependence, drawn from Martin et al. (2019). Panel (c) displays each county's share of pollinator-dependent crop acreage, calculated as acreage of all such crops in the county relative to the total within its state, weighted by crop-specific pollination dependency. Panel (d) plots the seasonal time trends of total bee colonies for California versus North and South Dakotas.

Figure 2: Predicted Interstate Bee Movement Paths

(a) Inbound and Outbound Flows of California

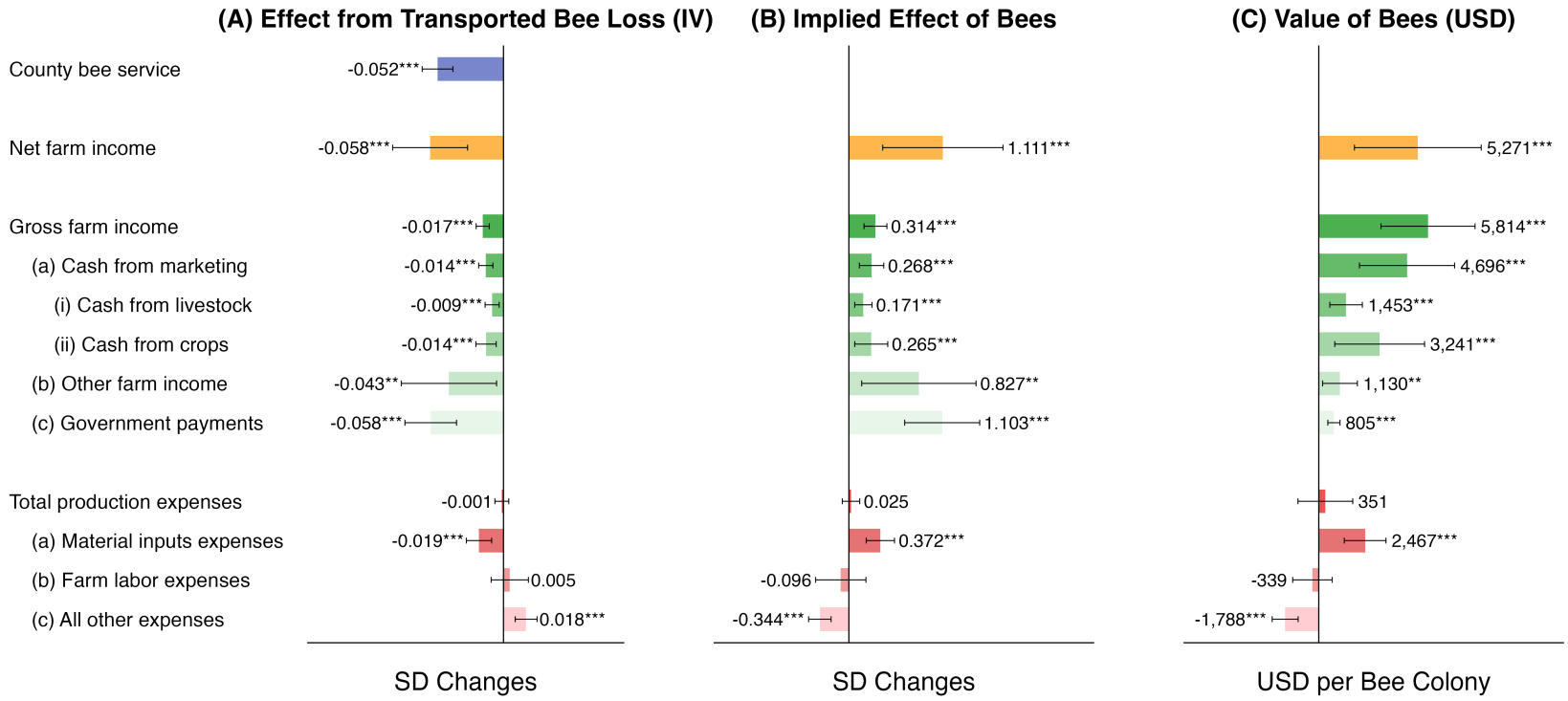


(b) All States Interstate Bee Movement Paths



Notes: This figure presents the average predicted interstate movement volumes of managed bee colonies across U.S. states, averaged over all years. The movement volume from origin state o to destination state d in quarter q and year t is defined as $\widehat{\text{BeeService}}_{o\text{-}d,q,t} = \text{BeeService}_{o,q,t} \cdot \frac{w_{d,q}}{\sum_{d \in D} w_{d,q}}$, where $w_{d,q}$ follows equation (12). Panel (a) shows California's main inbound and outbound flows, while Panel (b) displays all predicted interstate movement paths. Flow width indicates the predicted number of moved colonies, and states are ordered by total predicted flows. All movement paths are predicted using the machine-learning procedure described in Section 5.2.

Figure 3: Impact of Bee Pollination on Agricultural Outcomes



Notes: This figure shows the effects of bee pollination services on agricultural outcomes. Panels (A) report reduced-form impacts of transported bee loss on county-level agricultural outcomes. Panels (B) and (C) present the implied causal effects of bee service using 2SLS estimates. Coefficients are expressed in standard deviations in Panels (A)-(B), USD per bee colony in Panels (C) and measure how agricultural outcomes respond to changes in bee availability induced by ecological shocks to moved colonies. Bars display point estimates with 95% confidence intervals.

Table 2: Impacts of Bee Colonies on Farm Net Income

	Net Farm Income			County Bee Service
	(1)	(2)	(3)	(4)
County Bee Service	0.0937**	1.1110***	-	-
	(0.0407)	(0.3625)	-	-
Transported Bee Loss (IV)	-	-	-0.0582***	-0.0524***
	-	-	(0.0152)	(0.0062)
Specification	OLS	2SLS	Reduced-form	First-stage
Kleibergen-Paap F	-	70.86	-	-
Observations	13460	13460	13460	13460

Notes: Each column is a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) based on the data from the USDA's Quarterly Bee Colony Loss Survey. Both net farm income and county bee service are standardized. Column (1) reports the OLS regression following equation (11). Column (2) reports the 2SLS regression, where the county bee service is instrumented for using transported bee loss constructed following equation (12). Column (3) report estimates of net farm income with respect to the transported bee loss in a reduced form regression. Column (4) reports first stage regression, following $BeeStock_{ct} = \gamma \cdot IV_{ct} + \phi Z_{ct} + \psi (X_c \times \gamma_t) + \alpha_c + \gamma_t + \varepsilon_{ct}$. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix

A Sensitivity Checks

We organize our robustness checks into five groups: (1) the construction of the instrumental variable, (2) timing assumptions underlying bee movement, (3) the construction of the county-level bee service measure, (4) regression specification choices, and (5) placebo and falsification tests.

A.1 Robustness in IV construction.

The first group of checks addresses the construction of the transported bee loss instrument, which involves two main modeling choices: the machine learning procedure for identifying origin states, and the weighting scheme used to allocate origin-state ecological shocks to destination states.

Tuning of machine learning parameters. Regarding the bee movement pattern, we discuss the tuning of machine learning parameters when choosing the primary origin states that contribute to the destination state bee service. We assess sensitivity to alternative weighting schemes across the three algorithms and to different thresholds for selecting primary origin states. Specifically, we consider scenarios in which one algorithm receives a dominant weight of 60%, with the remaining two models each assigned a weight of 20%. We also vary the number of origin states selected from 4 to 12. Table A.3 and Table A.4 present the estimated first-stage, reduced-form, second-stage results along with the KP F-statistics under each alternative, following regression specification in equation (11). Our baseline specification is highlighted in bold. We find that the estimates remain stable across different weighting schemes and origin-state thresholds, suggesting that our main results are not sensitive to the specific tuning choices in the machine learning procedure.

Sensitivity to weighting schemes of bee movement. When modeling the interstate movement of bee colonies, we construct a weighting matrix that assigns greater weight to origin–destination pairs that are geographically proximate and have substantial acreage in pollinator-dependent crops. The baseline weighting scheme uses a simple multiplicative form combining the two forces, as shown in equation (12). We explore 13 alternative weighting schemes, including a gravity equation format, Cobb-Douglas specifications with varying elasticities, linear weighted averages, a log-linear balance, a bottleneck rule, and Sigmoid-based forms. Each functional form of weighting scheme captures a different behavioral or ecological assumption about feasibility of bee movement. All weights are normalized across destinations to ensure comparability. Table A.5 shows that our main results remain quantitatively stable across all weighting specifications, suggesting that our findings are not driven by any particular functional form used to represent bee movement. Table A.6 provides a detailed description of each scheme.

Excluding bees moved from neighboring states. To address concerns that common regional shocks—such as weather conditions—may simultaneously affect bee colonies in both origin and destination states, as well as agricultural outcomes, we exclude movement flows originating from geographically adjacent states. Specifically, when constructing the instrumental variable using equation (12), we set Bee $\text{Loss}_{o,q,t} = 0$ whenever the origin state o shares a border with the destination state d. For example, California receives bee colonies from eight primary origin states in our sample: Idaho, Minnesota, Montana, North Dakota, Oregon, South Dakota, Texas, and Washington. Among these states, only Oregon is geographically adjacent to California. Accordingly, we set Oregon's bee loss to zero when computing the transported bee loss to California. Appendix Table A.7 shows that the estimated effects remain broadly consistent with our baseline results.

A.2 Robustness to moving time assumptions.

The second group of checks addresses the assumptions we make about the timing of bee movement across seasons, which determine how ecological shocks in origin states are assigned to destination counties. Our baseline assumes that bees are moved during the destination's first pollination season in which any pollinator-dependent crop blooms, and that the transported shock is then allocated across subsequent quarters according to each county's pollination demand. We test two alternative timing assumptions.

First, we restrict the transported bee loss to affect only the concurrent quarter when bees arrive from origin states, rather than being allocated across quarters. Panel A of Appendix Table A.8 shows that the estimates remain highly consistent with our baseline findings.

Second, we relax the baseline assumption by allowing bees to be moved in any quarter during which at least one pollinator-dependent crop is in bloom. Panel B of Table A.8 shows that the estimated effects are qualitatively the same but larger in magnitude, consistent with capturing a more nuanced temporal distribution of bee movement activities and shock exposure when multiple pollination seasons are permitted.

A.3 Robustness in bee service measure construction.

The third group of checks addresses the construction of the county-level bee service measure, which depends on two auxiliary choices: the census year used to measure county crop acreage shares, and whether crop-specific pollination dependency parameters are applied.

Alternative census year for county crop acreage shares. We use 2012 as the baseline pre-treatment period to construct county-level pollination demand shares using data on bearing acreage of pollinator-dependent crops from the US Census of Agriculture, as our bee colony data begins in January 2015. To address potential year-specific measurement noise, we consider alternative

constructions using data from the 2007 and 2017 Censuses individually, as well as averages over the 2007–2012, 2007–2017, 2007–2022, 2012–2017, and 2012–2022 periods. Results reported in Table A.9 shows that the estimates remain relatively stable across these alternatives.

Omitting crop-specific pollination dependency parameters. Our baseline bee service measure weights each crop's contribution by its pollination dependency index α_k from Martin et al. (2019). Table A.10 reports results from a specification that omits these parameters, treating all pollinator-dependent crops as equally dependent. The main estimates remain stable, with a second-stage coefficient of 1.108, indicating that our findings are not sensitive to the specific dependency weights assigned to individual crops.

A.4 Robustness to regression specifications.

The fourth group of checks verifies that our estimates are stable across alternative regression specifications. First, Panel A of Appendix Table A.11 excludes the county-specific characteristics interacted with year fixed effects ($X_c \times \gamma_t$). The estimates remain consistent with the baseline, suggesting that the interacted controls do not drive our main results, though their inclusion improves precision. Second, we test for including county and state-by-year fixed effects that also accounts for state-specific, time-varying confounders–such as state-level conservation policies, agricultural subsidies, or localized ecological shocks–that may simultaneously influence bee abundance and agricultural productivity. Panel B of Appendix Table A.11 shows that results are comparable to our baseline specification.

A.5 Placebo and falsification tests.

The fifth group provides tests designed to yield null or collapsed results under our identification assumptions, helping rule out that our findings are driven by general agricultural trends, misspecified allocation from crop acreages, or local confounders.

Placebo: Non-pollinator-dependent crop acreage. We find no statistically significant effect of bee colonies on the planted acreage of non-pollinator-dependent crops—including corn, soybeans, cotton, sorghum, wheat, barley, oats, and rice. Table A.12 reports the results. This placebo test suggests that the channel through which pollination services affects net farm income operates specifically through improvements in pollinator-dependent crops, rather than through general agricultural development trends that would affect all crops simultaneously.

Falsification: Grain crop acreage as county allocation weights. As a direct test of the county-level bee service measure construction, we replace the pollinator-dependent crop acreage shares

— used in equation 10 to allocate state-level bee colonies to counties — with grain crop acreage shares instead. Because grain crops do not require pollination, counties with high grain acreage should not systematically attract more bee colonies, and this alternative measure should therefore carry no meaningful information about actual local pollination service. Table A.13 confirms that the first stage collapses entirely under this construction (KP F-statistic = 0.015), demonstrating that the predictive power of our baseline county bee service measure is specifically tied to the spatial distribution of pollinator-dependent agriculture rather than general agricultural activity.

Rule out pesticide as local confounder: Neonicotinoid-induced bee loss. We further assess whether an alternative instrument based on neonicotinoid pesticide exposure — rather than the ecological shocks underlying our baseline IV — can replicate our findings. Table A.14 shows that this alternative yields an extremely weak first stage (KP F-statistic = 0.19), indicating that neonicotinoid-driven bee losses do not generate sufficient variation in county-level bee service to serve as a valid instrument in our setting. This result also supports the exclusion restriction of our baseline instrument: the identifying variation we exploit through non-local ecological shocks is distinct from the neonicotinoid channel, which operates through local pesticide exposure and is therefore more likely to be correlated with local agricultural practices.

B Construction of Predicted Bee Loss IV

This appendix describes how we construct predicted exposure to bee losses at the county-year level. The construction starts from state-by-quarter ecological shocks and proceeds through three main steps: predict interstate propagation of shocks, select origin states through machine learning algorithms, and calculate county-by-year level exposure.

B.1 Predict interstate propagation of shocks.

The starting point of the analysis is a state-by-quarter panel of ecological shocks affecting managed bee colonies. These shocks include losses attributed to disease, pesticides, parasites, and other environmental factors. Let $\text{EcolShock}_{o,q,t}$ denote the ecological shock observed in origin state o during quarter q of year t. The raw data therefore have a state-by-quarter frequency.

Managed bee colonies are routinely transported across states to provide pollination services. As a result, ecological shocks occurring in origin states may affect pollination services in destination states that receive migrating bee colonies. We therefore construct predicted bee losses in destination states by allocating origin-state shocks using migration-based weights.

We use California as one destination state example to illustrate the procedure. Following this approach, predicted bee losses received by California in quarter q of year t are given by equation 12:

$$\widehat{\text{BeeLoss}}_{\text{CA},q,t} = \sum_{o \in O} \text{EcolShock}_{o,q,t} \cdot \frac{w_{\text{CA},q}}{\sum_{d \in D} w_{d,q}}. \tag{14}$$

Here O denotes the set of origin states supplying bees, and D denotes the set of potential destination states. The weight $w_{d,q}$ reflects the relative importance of destination state d in quarter q and is constructed using pollination-dependent crop area and geographic distance. Pollination demand occurs only during specific seasons determined by crop phenology. For each state we identify the primary pollination quarter (January, April, July, or October). Predicted bee losses are therefore matched to the pollination quarter of the destination state. As a result, for each state-year observation, predicted bee losses occur only during the pollination quarter of the destination state. Predicted losses in non-pollination quarters are set to zero by construction. In other words, although ecological shocks are observed for all four quarters at the origin level, the migration mapping assigns non-zero exposure to only one quarter for each destination state.

B.2 Machine learning algorithms for origin state selection

We employ three regularization-based machine learning models to identify primary origin states whose bee colony patterns most strongly predict bee populations in each destination state. The

fundamental logic is that predictive power reveals movement flows: if origin state o's historical colony levels systematically predict destination state d's current population, this indicates colonies move from o to d.

This approach exploits the time-lagged structure of interstate bee movement. For instance, if Texas has high colony stocks in winter quarter $t-1$, lower in spring t, and California shows elevated counts in spring quarter t during almond pollination, this lagged correlation is likely to suggest that colony flows from Texas to California. Similarly, if North Dakota's counts are high in summer quarter t but California's were high in spring $t-1$ and lower in summer t, this may be reflecting California-based colonies moving to North Dakota for summer foraging after completing California's pollination season.

By estimating regressions where destination state d's colony stock is the dependent variable and origin states' lagged stocks are predictors, we identify which origins most strongly explain destination colony availability. Before estimation, we weight each origin's lagged stocks by a factor combining inverse geographic distance and origin pollination-dependent crop acreage. This weighting reflects two economic realities. First, inverse distance captures transportation costs: geographically closer origins face lower costs and are more likely to send colonies. Second, origin crop acreage proxies for commercial beekeeping capacity and pollination service specialization. States with extensive pollination-dependent crops typically host larger commercial operations actively participating in interstate movement markets, whereas states with minimal such crops may rely more on local supply. The regression coefficients on lagged stocks thus provide a data-driven movement intensity measure accounting for both geographic feasibility and commercial capacity.

Model 1: Lasso-VAR for Individual origin Selection Rationale. The first model employs Lasso regression with vector autoregression structure. For each destination state d, we estimate:

$$\text{BeeStock}_{d,q} = \sum_{o \in O} \sum_{\ell=1}^{4} \lambda_{o\ell} \cdot \text{BeeStock}_{o,q-\ell} + \varepsilon_q \tag{15}$$

where $\text{BeeStock}_{d,q}$ denotes bee colonies in destination state d at quarter q, and $\text{BeeStock}_{o,q-\ell}$ represents lagged colony stocks from potential origin states o within 3,000 kilometers. To avoid mechanical self-prediction, we exclude destination state d's own lags from the predictor set.

Before estimation, we weight each origin's lagged stocks by $w_{od} = (1/\text{distance}_{od}) \times \text{croparea}_o$, where croparea_o is pollination-dependent crop acreage in origin state o. This weight is min-max normalized to $[0.1, 1]$ across all origins for a given destination to prevent extreme values while preserving relative magnitudes. The weighting prioritizes origins that are both geographically proximate (low transportation costs) and have substantial commercial beekeeping industries oriented toward out-of-state pollination (high crop acreage). For instance, when predicting Califor-

nia's population, Texas—with extensive pollination-dependent crops—receives higher weight than a similarly distant state with minimal commercial agriculture, reflecting Texas's greater capacity to supply colonies to interstate markets.

The Lasso penalty $\sum|\lambda_{o\ell}|$ shrinks irrelevant origin coefficients to zero, performing automatic feature selection. We use four quarters of lags ($\ell = 1,2,3,4$) to capture both immediate and delayed movement patterns, as beekeepers may move colonies across multiple quarters depending on pollination schedules and seasonal foraging opportunities.

We compute a origin-specific prediction score by summing absolute values of all non-zero coefficients associated with origin o across the four lags:

$$\text{PredictionScore}_o^{\text{Lasso}} = \sum_{\ell=1}^{4} |\lambda_{o\ell}| \tag{16}$$

This aggregation captures the total predictive contribution of origin state o, accounting for all temporal lags through which o's colony dynamics forecast destination population levels.

Strength. Lasso-VAR provides granular, individual-level selection. Each origin state and lag combination is evaluated independently based purely on predictive power, allowing the model to identify specific movement routes without structural assumptions. This data-driven approach excels when movement patterns are heterogeneous and origin states have distinct, independent relationships with destinations. The sparsity-inducing L_1 penalty ensures only origins with genuine predictive value receive non-zero coefficients.

Limitation. Lasso-VAR struggles with correlated predictors. When two or more origin states exhibit highly correlated colony patterns—common among geographically proximate states with similar climatic conditions and beekeeping industries—selection becomes unstable. Consider Idaho and Oregon, neighboring Pacific Northwest states sharing comparable agricultural seasons, foraging resources, and movement strategies. Their colony stocks tend to move in simultaneously, responding similarly to regional weather and market conditions.

Under such multicollinearity, Lasso typically selects only one state from the correlated group while shrinking others' coefficients to zero, even when both could be genuine origins. This arbitrary selection occurs because once Idaho's lagged stocks explain a portion of California's variation, Oregon's contribution appears redundant due to correlation, leading the penalty to exclude Oregon.

Model 2: Group Lasso with Ecological Clusters **Rationale.** The second model implements Group Lasso, imposing group structure on predictors based on USDA-defined ecological clusters for bees specifically, which is widely used in USDA data collection, analysis, and reports for bees. Regions may contain one or multiple states chosen to be representative of pollinated crops, forage,

and colonies in these locations (Bond et al., 2021).

The estimation equation remains:

$$\text{BeeStock}_{d,q} = \sum_{g} \sum_{o \in g} \sum_{\ell=1}^{4} \lambda_{o\ell} \cdot \text{BeeStock}_{o,q-\ell} + \varepsilon_q \tag{17}$$

but predictors are grouped by ecological cluster g, excluding destination state d's own lags to avoid self-prediction. As in Lasso-VAR, we weight each origin's lagged stocks by the product of inverse distance and origin crop acreage. The Group Lasso penalty $\lambda \sum_g \sqrt{p_g} \|\lambda_g\|_2$ encourages group-level sparsity, where p_g denotes group size and λ_g contains all coefficients within group g. For each ecological cluster, either all four lags from all states are retained together, or all are set to zero simultaneously. If the algorithm determines the Northern Great Plains region predicts California's population, it retains North Dakota, South Dakota, Montana, and all their lags as a block; conversely, if the region is uninformative, all are excluded together.

The rationale for ecological clustering stems from geographic and ecological similarity in beekeeping patterns. States within the same USDA-defined region—such as North Dakota, South Dakota, Montana and Minnesota in the Northern Great Plains—share similar climate, vegetation, and foraging conditions. Beekeepers face comparable seasonal constraints and economic incentives, leading to similar movement strategies. For example, Great Plains beekeepers typically keep colonies locally during summer for honey production from clover and sunflower, then move colonies to warmer states like California or Texas for winter pollination and overwintering. Treating ecologically similar states as a group leverages this structural information for more stable and interpretable selections aligned with known movement paths.

Strength. Group Lasso provides structured selection reflecting ecological and behavioral realities of bee movement. Rather than arbitrarily choosing between highly correlated neighboring states (as Lasso might), it makes decisions at the regional level, consistent with geographic patterns documented in USDA surveys (Bond et al., 2021). Second, it reduces sensitivity to idiosyncratic noise in individual state data. If Montana's colony numbers in one quarter are anomalously high due to measurement error or localized weather events, but the overall Great Plains region shows consistent patterns across North Dakota and South Dakota, Group Lasso's group-level evaluation mitigates such outlier impacts.

Limitation. Its inflexibility is problematic when only a subset of states within an ecological cluster are true origins. If Texas is a major origin for California but other southern states within the same group (Louisiana, New Mexico, Arizona) are not, Group Lasso might either include irrelevant states from that region or exclude Texas entirely, depending on whether overall cluster predictive power exceeds the selection threshold. This all-or-nothing constraint can lead to false positives (including irrelevant states from a partially important cluster) and false negatives (exclud-

ing relevant states when their cluster's average signal is weak). Additionally, beekeepers' movement decisions also depend on economic factors (pollination fees, honey prices) and regulatory environments (state regulations, agricultural policies) that may not perfectly map onto ecological zones.

Model 3: Elastic Net with Auxiliary Variables Rationale. The third model employs Elastic Net regression augmented with auxiliary variables capturing geographic, economic, and ecological determinants. Elastic Net combines L_1 (Lasso) and L_2 (Ridge) penalties:

$$\text{BeeStock}_{d,q} = \sum_{o} \sum_{\ell=1}^{4} \lambda_{o\ell} \cdot \text{BeeStock}_{o,q-\ell} + \sum_{o} \sum_{m} \delta_o^m \cdot \text{AuxVars}_o^m + \varepsilon_q \tag{18}$$

where AuxVars_o^m denotes auxiliary variables for origin o. As in previous models, each origin's lagged stocks (excluding destination d's own lags) are weighted by the product of inverse distance and origin crop acreage before estimation. The mixed penalty $\alpha\lambda\|\beta\|_1 + (1-\alpha)\lambda\|\beta\|_2^2$ balances feature selection via L_1 with coefficient shrinkage and stability via L_2. The parameter $\alpha \in [0,1]$ controls this trade-off: $\alpha = 1$ reduces to pure Lasso, $\alpha = 0$ to pure Ridge, and intermediate values blend both properties. We select α through time-series cross-validation, testing values $\{0.2, 0.5, 0.8\}$ to allow data to determine the optimal balance for each destination.

We augment the model with three auxiliary features for each origin state o: (1) inverse distance to destination, (2) origin crop acreage, and (3) ecological cluster indicators. While we already weight lagged stocks by distance and crop acreage, including these as separate predictors serves a distinct purpose. The weighting rescales each origin's lagged variables according to its characteristics; the auxiliary variables estimate whether these characteristics themselves directly predict destination populations beyond what weighted lagged stocks capture. For instance, even after controlling for all origins' weighted historical data, origin acreage can identify whether states with large commercial operations are more reliable suppliers independent of recent colony fluctuations.

Strengths. Elastic Net overcomes key limitations of both Lasso-VAR and Group Lasso. First, the L_2 penalty mitigates multicollinearity by retaining groups of correlated predictors with shrunken coefficients rather than arbitrarily selecting a single state, as occurs with Lasso. For example, when Texas and Louisiana display highly correlated colony-stock dynamics due to similar regional environments, Elastic Net can preserve both predictors with reduced magnitudes, reflecting their joint relevance. At the same time, the L_1 component continues to impose sparsity, ensuring that truly irrelevant origin states are excluded. Second, incorporating auxiliary variables improves prediction stability by adding information that is less sensitive to short-run fluctuations than raw colony counts.

Limitations. The primary trade-off is increased complexity and reliance on feature engineer-

ing. Performance depends critically on proper auxiliary variable specification. If important determinants for bee transports are omitted (e.g., state beekeeping regulations or pollination fee schedules) or included variables are poorly measured, the model could systematically favor or penalize certain origins based on incomplete information. Unlike Lasso-VAR, which relies purely on observed correlations, or Group Lasso, which uses established ecological classifications, Elastic Net's auxiliary variables require researchers to make explicit assumptions about which structural features matter for movement—assumptions that may not fully capture beekeeping decision complexities.

Ensemble Aggregation and Rationale After obtaining prediction scores from all three models, we compute a composite score for each origin-destination pair by taking a weighted average of min-max normalized scores:

$$\text{CompositeScore}_o = w_{\text{Lasso}} \cdot \text{Score}_o^{\text{Lasso}} + w_{\text{Group}} \cdot \text{Score}_o^{\text{Group}} + w_{\text{Enet}} \cdot \text{Score}_o^{\text{Enet}} \tag{19}$$

where each score is first min-max normalized to $[0,1]$ within each destination state to ensure cross-model comparability, and weights sum to one ($w_{\text{Lasso}} + w_{\text{Group}} + w_{\text{Enet}} = 1$). We then select the top K origins with highest composite scores for each destination.

The ensemble approach mitigates model-specific biases and uncertainties. If Lasso arbitrarily selects Idaho over Oregon due to multicollinearity in their stock patterns, but both Group Lasso (via the Pacific Northwest ecological cluster) and Elastic Net (via geographic proximity captured by inverse distance) assign high scores to Oregon, the ensemble composite will appropriately recognize Oregon's importance. This voting mechanism ensures selected origins are robust across different modeling assumptions and statistical perspectives, reducing the risk that origin identification is driven by idiosyncratic features of any single algorithm.

B.3 Calculate the county-by-year level exposure.

To measure local exposure, we allocate predicted state-level bee losses to counties using the spatial distribution of pollination-dependent crops. Using the Census of Agriculture, we compute acreage of pollination-dependent crops for each county and quarter.

For county c located in state d, we define the county-quarter share of pollination demand as

$$s_{c,d,q} = \frac{\text{CropArea}_{c,q}}{\sum_{c' \in d} \text{CropArea}_{c',q}}.$$

At the state level, the migration step assigns a predicted bee loss to the destination state's earliest pollination quarter. Let this quarter be denoted by q_d^*. Thus, each destination state-year receives a single predicted shock, $\widehat{\text{BeeLoss}}_{d,q_d^*,t}$, at the state level.

At the county level, however, pollination demand may persist beyond that earliest quarter. Some counties may have positive pollination demand only in later quarters, while others may have positive demand in both the earliest quarter and subsequent quarters. To account for this, we allocate the same state-level predicted shock to all county-quarters at or after the state's earliest pollination quarter, weighting by the county-quarter crop share.

Accordingly, county-quarter predicted bee losses are constructed as

$$\widehat{\text{BeeLoss}}_{c,q,t} = \mathbf{1}\{q \geq q_d^*\} \cdot \widehat{\text{BeeLoss}}_{d,q_d^*,t} \cdot s_{c,d,q}.$$

This implies that the same state-level shock is carried forward to all relevant county pollination quarters within the year. Quarters before the state's earliest pollination quarter receive no allocation, while quarters with zero county crop share also receive zero predicted exposure.

Next, we aggregate county-quarter predicted bee losses across quarters to the annual level:

$$\widehat{\text{BeeLoss}}_{c,t} = \sum_q \widehat{\text{BeeLoss}}_{c,q,t}.$$

The resulting annual measure captures total county exposure to the state's predicted bee loss over all quarters in which local pollination demand is present after the onset of the state's pollination season. For example, if a state's earliest pollination quarter is April, then the predicted state-level bee loss is allocated to county-level pollination activity in both April and July whenever the county has positive crop shares in those quarters.

To capture changes in bee availability driven by ecological shocks at source locations, we construct an instrumental variable using the year-to-year change in predicted bee losses:

$$\Delta\widehat{\text{BeeLoss}}_{c,t} = \widehat{\text{BeeLoss}}_{c,t} - \widehat{\text{BeeLoss}}_{c,t-1}.$$

This measure captures changes in predicted bee availability arising from ecological shocks in origin states and is used as an instrument for realized pollination services in the second-stage regression in equation 11.

Appendix Figures and Tables

Table A.1: Summary Statistics

	Mean	S.D.	Min	Median	Max	Observations
Panel A. Bees						
County bee service	2795.33	15230.10	0.00	68.63	352954.02	13460
Transported bee loss (IV)	125.02	859.08	0.00	1.17	32485.54	13460
Quarter-specific county share of pollination crop acreage	1.04	6.15	0.00	0.00	100.00	10768
County share of pollination crop acreage	0.35	1.09	0.00	0.03	19.80	13460
Panel B. Income						
Net farm income	22398.10	72253.08	-83007.15	7431.53	1917475.93	13460
Gross farm income	147769.72	281856.68	0.00	84308.86	6400450.69	13460
Cash from marketing	131033.91	267220.84	0.00	70303.95	6226065.62	13460
Cash from livestock	62392.57	129421.14	0.00	24449.65	2434482.60	13460
Cash from crops	68641.35	186632.24	0.00	28544.04	5037316.87	13460
Other farm income	16735.81	20808.88	0.00	10932.67	555951.86	13460
Government payments	6738.75	11119.11	0.00	3082.06	343058.22	13460
Panel C. Expenses						
Total production expenses	125371.62	216182.32	0.00	75991.18	4966902.68	13460
Material input expenses	57880.89	101069.58	0.00	30507.68	1883809.47	13460
Farm labor expenses	13173.94	53671.47	0.00	5146.75	1403689.19	13460
All other production expenses	54316.79	79252.54	0.00	36354.45	1759365.81	13460

Notes: The sample includes 2,692 U.S. counties. The county–quarter share of pollinator-dependent crop acreage (%) is defined as the share of acreage planted with crops that flower in a given quarter, calculated within each state. This measure is used in equation (10) to allocate state–quarter bee colony services and transported bee loss to counties. The county share of pollinator-dependent crop acreage (%) is calculated as the acreage of all such crops in the county relative to the total within its state on an annual basis (Figure 1c). Both measures are constructed using data from the 2012 U.S. Census of Agriculture.

Table A.2: Pollination-Dependent Crops and Their Flowering Periods, Survey Quarter Matching, and Dependency Parameters

Crop	Flowering Period	Quarter Matched	Pollination Dependency
Panel A. From Cost of Pollination Survey			
Almonds	Febuary - March	April (Jan 1 - Mar 31)	0.65
Alfalfa for seed	May - September	July (Apr 1 - Jun 30)	0.95
Apples	April - May	July (Apr 1 - Jun 30)	0.65
Blueberries	April - May	July (Apr 1 - Jun 30)	0.25
Cantaloupe	May - July	July (Apr 1 - Jun 30)	0.95
Cherries	March - April	April (Jan 1 - Mar 31)	0.65
Clover for seed	May - August	July (Apr 1 - Jun 30)	0.95
Cranberries	May - July	July (Apr 1 - Jun 30)	0.65
Cucumber	May - July	July (Apr 1 - Jun 30)	0.65
Oranges	January - March	April (Jan 1 - Mar 31)	0.05
Peaches	March - April	April (Jan 1 - Mar 31)	0.65
Nectarines	March - April	April (Jan 1 - Mar 31)	0.65
Pears	April - May	July (Apr 1 - Jun 30)	0.65
Pumpkin	June - August	July (Apr 1 - Jun 30)	0.95
Raspberries	May - June	July (Apr 1 - Jun 30)	0.65
Squash	June - July	July (Apr 1 - Jun 30)	0.95
Strawberries	April - May	July (Apr 1 - Jun 30)	0.25
Sunflowers	July - September	October (Jul 1 - Sep 30)	0.25
Watermelon	May - August	July (Apr 1 - Jun 30)	0.95
Panel B. Other Pollination-dependent Crops			
Avocados	Febuary - April	April (Jan 1 - Mar 31)	0.65
Plums	Febuary - April	April (Jan 1 - Mar 31)	0.65
Kiwifruit	Febuary - April	April (Jan 1 - Mar 31)	0.65
Pomegranate	March - June	April (Jan 1 - Mar 31)	0.25
Aprilicots	Febuary - March	April (Jan 1 - Mar 31)	0.65
Honeydew melon	May - July	July (Apr 1 - Jun 30)	0.95
Blackberries	May - June	July (Apr 1 - Jun 30)	0.25

Notes: This table lists representative pollination-dependent crops in U.S. agriculture, their typical flowering periods, and the corresponding nearest quarter matched in the USDA Cost of Pollination Survey. For example, crops flowering in February and March are assigned to the April quarter, while those flowering from May to July are assigned to the July quarter. Panel A includes crops explicitly covered by the survey, which reports pollination fees for these crops. Panel B lists additional pollination-dependent crops planted in the United States but not covered by the survey; these entries are manually collected and supplemented by the authors. The final column reports crop-specific pollination dependency parameters obtained from Martin et al. (2019). The Cost of Pollination Survey is mailed to beekeepers around January 1, April 1, July 1, and October 1 to collect information on bee colonies owned in the previous quarter. For example, the survey sent in January reflects bee colony conditions during October–December, and the April survey reflects January–March. Consequently, if a crop's flowering period is February–March, the corresponding survey quarter is April, which reports conditions for January–March.

Table A.3: Robustness of ML Weighting Scheme and Number of Source Selected - Part 1

Weighting Scheme	Number of Sources Selected	First-stage	Reduced-form	2SLS	KP F-stats
Panel A. Equal Contributions	4	-0.0411***	-0.0521***	1.2686***	71.04
33% Lasso, 33% Group Lasso, 33% ENet		(0.0049)	(0.0137)	(0.3833)	
	5	-0.0458***	-0.0552***	1.2063***	135.98
		(0.0039)	(0.0137)	(0.3186)	
	6	-0.0503***	-0.0562***	1.1174***	64.20
		(0.0063)	(0.0150)	(0.3741)	
	7	-0.0501***	-0.0556***	1.1088***	58.87
		(0.0065)	(0.0148)	(0.3757)	
	8	**-0.0524*** **	**-0.0582*** **	**1.1110*** **	**70.86**
		(0.0062)	**(0.0152)**	**(0.3625)**	
	9	-0.0542***	-0.0582***	1.0736***	80.22
		(0.0061)	(0.0152)	(0.3450)	
	10	-0.0557***	-0.0588***	1.0553***	84.16
		(0.0061)	(0.0153)	(0.3344)	
	11	-0.0557***	-0.0586***	1.0517***	82.55
		(0.0061)	(0.0154)	(0.3359)	
	12	-0.0524***	-0.0584***	1.1151***	64.59
		(0.0065)	(0.0156)	(0.3724)	
Panel B. Lasso Dominant	4	-0.0407***	-0.0519***	1.2749***	68.36
60% Lasso, 20% Group Lasso, 20% ENet		(0.0049)	(0.0136)	(0.3874)	
	5	-0.0460***	-0.0567***	1.2326***	128.37
		(0.0041)	(0.0141)	(0.3248)	
	6	-0.0501***	-0.0562***	1.1211***	63.17
		(0.0063)	(0.0150)	(0.3762)	
	7	-0.0503***	-0.0556***	1.1046***	60.03
		(0.0065)	(0.0148)	(0.3729)	
	8	-0.0514***	-0.0562***	1.0943***	66.40
		(0.0063)	(0.0146)	(0.3578)	
	9	-0.0542***	-0.0582***	1.0736***	80.22
		(0.0061)	(0.0152)	(0.3449)	
	10	-0.0558***	-0.0590***	1.0579***	84.33
		(0.0061)	(0.0153)	(0.3348)	
	11	-0.0557***	-0.0586***	1.0518***	82.52
		(0.0061)	(0.0154)	(0.3360)	
	12	-0.0524***	-0.0584***	1.1148***	64.67
		(0.0065)	(0.0156)	(0.3722)	

Notes: The dependent variables are standardized county bee exposure (first-stage) and net farm income (reduced-form and 2SLS). Panel A reports estimates using equal weights across the three algorithms when selecting the primary source states that contribute to a destination state's bee exposure, allowing the number of selected sources to range from 4 to 12. Panel B reports estimates when Lasso receives a 60% weight, with the remaining two algorithms each weighted at 20%. Baseline estimates from Table 2 are highlighted in bold red.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.4: Robustness of ML Weighting Scheme and Number of Source Selected - Part 2

Weighting Scheme	Number of Sources Selected	First-stage	Reduced-form	2SLS	KP F-stats
Panel A. Group Lasso Dominant 20% Lasso, 60% Group Lasso, 20% ENet	4	-0.0422*** (0.0042)	-0.0540*** (0.0136)	1.2786*** (0.3616)	99.08
	5	-0.0454*** (0.0039)	-0.0556*** (0.0137)	1.2254*** (0.3250)	132.38
	6	-0.0498*** (0.0063)	-0.0566*** (0.0150)	1.1368*** (0.3814)	62.07
	7	-0.0498*** (0.0066)	-0.0560*** (0.0148)	1.1251*** (0.3815)	57.50
	8	-0.0519*** (0.0063)	-0.0586*** (0.0152)	1.1282*** (0.3688)	68.52
	9	-0.0558*** (0.0060)	-0.0589*** (0.0151)	1.0551*** (0.3312)	87.91
	10	-0.0554*** (0.0061)	-0.0591*** (0.0153)	1.0669*** (0.3383)	82.45
	11	-0.0556*** (0.0061)	-0.0586*** (0.0154)	1.0527*** (0.3363)	82.55
	12	-0.0524*** (0.0065)	-0.0584*** (0.0156)	1.1150*** (0.3723)	64.60
Panel B. Elastic Net Dominant 20% Lasso, 20% Group Lasso, 60% Elnet	4	-0.0407*** (0.0049)	-0.0519*** (0.0136)	1.2744*** (0.3873)	68.39
	5	-0.0458*** (0.0039)	-0.0555*** (0.0138)	1.2133*** (0.3213)	134.45
	6	-0.0501*** (0.0063)	-0.0562*** (0.0150)	1.1210*** (0.3761)	63.18
	7	-0.0501*** (0.0065)	-0.0556*** (0.0148)	1.1091*** (0.3757)	58.93
	8	-0.0527*** (0.0062)	-0.0583*** (0.0154)	1.1056*** (0.3628)	73.30
	9	-0.0542*** (0.0061)	-0.0582*** (0.0152)	1.0736*** (0.3449)	80.22
	10	-0.0558*** (0.0061)	-0.0590*** (0.0153)	1.0580*** (0.3349)	84.26
	11	-0.0557*** (0.0061)	-0.0586*** (0.0154)	1.0518*** (0.3360)	82.52
	12	-0.0524*** (0.0065)	-0.0584*** (0.0156)	1.1151*** (0.3724)	64.59

Notes: The dependent variables are standardized county bee exposure (first-stage) and net farm income (reduced-form and 2SLS). Panel A reports estimates when Group Lasso receives a 60% weight, with the remaining two algorithms each weighted at 20%, when selecting the primary source states that contribute to a destination state's bee exposure; the number of selected sources is allowed to range from 4 to 12. Panel B reports estimates when Elastic Net receives a 60% weight, with the remaining two algorithms each weighted at 20%.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.5: Sensitivity to Different Migration Weighting Scheme

Migration Weight	First-stage	Reduced-form	2SLS	KP F-stats
(1) $w_{ij} = A_{ij} \cdot D_{ij}$	**-0.0524*** (0.0062)**	**-0.0582*** (0.0152)**	**1.1110*** (0.3625)**	**70.86**
(2) $w_{ij} = A_{ij} \cdot D_{ij}^{-2}$	-0.0518*** (0.0113)	-0.0656*** (0.0209)	1.2677** (0.5662)	21.02
(3) $w_{ij} = A_{ij}^{0.2} \cdot D_{ij}^{0.8}$	-0.0346** (0.0175)	-0.0368*** (0.0140)	1.0633 (0.7102)	3.93
(4) $w_{ij} = A_{ij}^{0.4} \cdot D_{ij}^{0.6}$	-0.0498*** (0.0150)	-0.0524*** (0.0162)	1.0523** (0.4947)	11.03
(5) $w_{ij} = A_{ij}^{0.5} \cdot D_{ij}^{0.5}$	-0.0539*** (0.0126)	-0.0559*** (0.0159)	1.0371** (0.4209)	18.31
(6) $w_{ij} = A_{ij}^{0.6} \cdot D_{ij}^{0.4}$	-0.0549*** (0.0106)	-0.0554*** (0.0147)	1.0093*** (0.3618)	26.82
(7) $w_{ij} = A_{ij}^{0.8} \cdot D_{ij}^{0.2}$	-0.0504*** (0.0088)	-0.0456*** (0.0115)	0.9054*** (0.2797)	32.99
(8) $w_{ij} = 0.4 \cdot A_{ij} + 0.6 \cdot D_{ij}$	-0.0362*** (0.0105)	-0.0355*** (0.0105)	0.9816** (0.4034)	12.00
(9) $w_{ij} = 0.5 \cdot A_{ij} + 0.5 \cdot D_{ij}$	-0.0370*** (0.0093)	-0.0372*** (0.0103)	1.0059*** (0.3731)	15.80
(10) $w_{ij} = 0.6 \cdot A_{ij} + 0.4 \cdot D_{ij}$	-0.0371*** (0.0085)	-0.0383*** (0.0103)	1.0339*** (0.3550)	19.24
(11) $w_{ij} = (A_{ij} D_{ij}) \cdot (A_{ij}^2 + D_{ij}^2 + \varepsilon)^{-1/2}$	-0.0615*** (0.0138)	-0.0670*** (0.0237)	1.0899** (0.5272)	19.88
(12) $w_{ij} = \min(A_{ij}, D_{ij})$	-0.0624*** (0.0143)	-0.0639*** (0.0242)	1.0250* (0.5269)	19.03
(13) $w_{ij} = \left(1 + e^{-5(A_{ij} - 0.5)}\right)^{-1} \left(1 + e^{-5(D_{ij} - 0.5)}\right)^{-1}$	-0.0300*** (0.0059)	-0.0478*** (0.0109)	1.5934*** (0.4494)	25.94
(14) $w_{ij} = (1 + e^{-10(A_{ij} - 0.5)})^{-1} (1 + e^{-10(D_{ij} - 0.5)})^{-1}$	-0.0264*** (0.0028)	-0.0389*** (0.0091)	1.4712*** (0.3488)	91.57

Notes: The dependent variables are standardized county bee service (first stage) and net farm income (reduced form and 2SLS). This table reports robustness to alternative weighting schemes that allocate ecological shocks from source states to destination states according to geographic proximity and pollination-dependent crop acreage. Row (1) uses a multiplicative weighting scheme; row (2) applies a gravity-style formulation; rows (3)–(7) use a Cobb–Douglas structure with different elasticities; rows (8)–(10) apply a linear weighted average with varying weights on acreage and distance; row (11) applies a log-linear balance; row (12) applies a bottleneck rule; and rows (13)–(14) use sigmoid-based weighting with different parameter choices. Baseline estimates from Table 2 are highlighted in bold red. See Appendix A for details.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.6: Weighting Schemes Used in Modeling Interstate Movement of Bee Colonies

Weighting Scheme	Formula	Notes
Multiplicative (Baseline)	Weight = distance factor $\times$ area factor	Migration is most likely when both proximity and demand are high. Low values in either sharply reduce the overall weight.
Gravity equation format	Weight = $(\text{distance factor})^2 \times$ area factor	This form amplifies the importance of geographic proximity by squaring the distance factor. Migration is heavily penalized for distant destinations even when demand (area) is high.
Cobb-Douglas	Weight = $\text{distance}^{\alpha} \times \text{area}^{1-\alpha}$	This formulation allows for flexible substitution: high demand can partially compensate for distance. We test various elasticities, including $\alpha = 0.2, 0.4, 0.5, 0.6, 0.8$.
Linear Weighted Average	Weight = $\alpha \cdot$ distance + $(1-\alpha) \cdot$ area	A convex combination that allows for an explicit trade-off between proximity and demand. We test for $\alpha = 0.2, 0.4, 0.5, 0.6, 0.8$.
Log-Linear Balance	Weight = $\frac{(distance \times area)}{\sqrt{distance^2 + area^2}}$	This form penalizes imbalanced pairs (e.g., close but low demand), and favors balanced combinations.
Bottleneck Rule	Weight = min(distance, area)	Migration is constrained by the weaker of the two factors. Even strong demand cannot compensate for long distances, and vice versa.
Sigmoid-Based	Weight = sigmoid(distance) $\times$ sigmoid(area)	Migration responds sharply only when both distance and demand exceed moderate thresholds, mimicking a threshold or tipping-point effect.

Notes: This table outlines the weighting schemes used in modeling interstate movement of bee colonies with detailed explanations.

Table A.7: Sensitivity to Excluding Migration from Neighbor States

	Net Farm Income			County Bee Service
	(1)	(2)	(3)	(4)
County Bee Service	0.0937**	0.3801**	-	-
	(0.0407)	(0.1816)	-	-
Transported Bee Loss (IV)	-	-	-0.0183**	-0.0482***
	-	-	(0.0088)	(0.0070)
Specification	OLS	2SLS	Reduced-form	First-stage
Kleibergen-Paap F	-	46.68	-	-
Observations	13460	13460	13460	13460

Notes: Each column is a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) without incorporating crop-specific pollination dependency α_k, based on data from the USDA's Quarterly Bee Colony Loss Survey. Both net farm income and county bee service are standardized. Column (1) reports the OLS regression following equation (11). Column (2) reports the 2SLS regression, where the county bee service is instrumented for using transported bee loss constructed following equation (12), after excluding transported bees originating from states that are geographically adjacent to the destination. Column (3) report estimates of net farm income with respect to the transported bee loss in a reduced form regression. Column (4) reports first stage regression, following $BeeStock_{ct} = \gamma \cdot IV_{ct} + \phi Z_{ct} + \psi(X_c \times \gamma_t) + \alpha_c + \gamma_t + \varepsilon_{ct}$. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.8: Sensitivity to Moving Time Assumptions

	Net Farm Income			County Bee Service
	(1) OLS	(2) 2SLS	(3) Reduced-form	(4) First-stage
Panel A. Only Affecting Concurrent Quarter				
County Bee Service	0.0937**	1.2306***	-	-
	(0.0407)	(0.3785)	-	-
Transported Bee Loss (IV)	-	-	-0.0449***	-0.0365***
	-	-	(0.0137)	(0.0063)
Kleibergen-Paap F	-	33.09	-	-
Observations	13460	13460	13460	13460
Panel B. Multiple Pollination Seasons				
County Bee Service	0.0937**	2.3988***	-	-
	(0.0407)	(0.5413)	-	-
Transported Bee Loss (IV)	-	-	-0.0553***	-0.0231***
	-	-	(0.0121)	(0.0028)
Kleibergen-Paap F	-	69.26	-	-
Observations	13460	13460	13460	13460

Notes: Each column reports a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) based on the data from the USDA's Quarterly Bee Colony Loss Survey. Both county bee service and net farm income are standardized. Panel A restricts the instrumented transported bee loss to affect only the concurrent quarter when bees arrive from source states (the first pollination season for each destination). Panel B allows bees to migrate across states in any quarter in which crops require pollination services. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.9: Sensitivity to Different County Crop Shares

Share	First-stage	Reduced-form	2SLS	KP F-stats
2007 Acreage	-0.0285***	-0.0402***	1.4109***	70.43
	(0.0034)	(0.0112)	(0.3957)	
2012 Acreage	**-0.0524*** **	**-0.0582*** **	**1.1110*** **	**70.86**
	(0.0062)	**(0.0152)**	**(0.3625)**	
2017 Acreage	-0.0439***	-0.0615***	1.4007***	23.37
	(0.0091)	(0.0165)	(0.4963)	
2007-2012 Avg Acreage	-0.0557***	-0.0550***	0.9879***	80.92
	(0.0062)	(0.0157)	(0.3433)	
2007-2017 Avg Acreage	-0.0484***	-0.0548***	1.1316**	19.20
	(0.0111)	(0.0148)	(0.4574)	
2007-2022 Avg Acreage	-0.0409***	-0.0299***	0.7319*	14.54
	(0.0107)	(0.0115)	(0.3793)	
2012-2017 Avg Acreage	-0.0440***	-0.0507***	1.1515**	15.22
	(0.0113)	(0.0137)	(0.4911)	
2012-2022 Avg Acreage	-0.0413***	-0.0374***	0.9058**	15.91
	(0.0103)	(0.0116)	(0.3845)	

Notes: The dependent variables are standardized county bee service (first stage) and net farm income (reduced form and 2SLS). This table reports robustness to alternative construction of baseline county shares of pollination-dependent crops. Row (1) uses acreage from the 2007 Census of Agriculture; row (2) uses 2012 acreage; row (3) uses 2017 acreage; row (4) uses the average of 2007 and 2012 acreage; row (5) uses the average of 2007–2017 acreage; row (6) uses the average of 2007–2022 acreage; row (7) uses the average of 2012–2017 acreage; and row (8) uses the average of 2012–2022 acreage. Baseline estimates from Table 2 are highlighted in bold red.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.10: Sensitivity to Omitting Crop-specific Pollination Dependency Parameters

	Net Farm Income			County Bee Service
	(1)	(2)	(3)	(4)
County Bee Service	0.1052**	1.1078***	-	-
	(0.0477)	(0.3127)	-	-
Transported Bee Loss (IV)	-	-	-0.0447***	-0.0404***
	-	-	(0.0107)	(0.0058)
Specification	OLS	2SLS	Reduced-form	First-stage
Kleibergen-Paap F	-	47.75	-	-
Observations	13460	13460	13460	13460

Notes: Each column is a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) without incorporating crop-specific pollination dependency α_k, based on data from the USDA's Quarterly Bee Colony Loss Survey. Both net farm income and county bee service are standardized. Column (1) reports the OLS regression following equation (11). Column (2) reports the 2SLS regression, where the county bee service is instrumented for using transported bee loss constructed following equation (12). Column (3) report estimates of net farm income with respect to the transported bee loss in a reduced form regression. Column (4) reports first stage regression, following $BeeStock_{ct} = \gamma \cdot IV_{ct} + \phi Z_{ct} + \psi (X_c \times \gamma_t) + \alpha_c + \gamma_t + \varepsilon_{ct}$. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.11: Sensitivity to Regression Specifications

	Net Farm Income			County Bee Service
	(1) OLS	(2) 2SLS	(3) Reduced-form	(4) First-stage
Panel A. Sensitivity to Interacted Controls				
County Bee Service	-0.0494	1.5072**	-	-
	(0.0541)	(0.5889)	-	-
Transported Bee Loss (IV)	-	-	-0.0414***	-0.0275***
	-	-	(0.0146)	(0.0078)
Kleibergen-Paap F	-	12.41	-	-
Observations	13460	13460	13460	13460
Panel B. Replace Year FE with State-by-year FE				
County Bee Service	0.0460	1.0011***	-	-
	(0.0397)	(0.3713)	-	-
Transported Bee Loss (IV)	-	-	-0.0549***	-0.0548***
	-	-	(0.0172)	(0.0064)
Kleibergen-Paap F	-	73.91	-	-
Observations	13460	13460	13460	13460

Notes: Each column reports a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) based on the data from the USDA's Quarterly Bee Colony Loss Survey. Both county bee service and net farm income are standardized. Panel A excludes interacted controls ($X_c \times \gamma_t$). Panel B replaces year fixed effects (γ_t) with state-by-year fixed effects (δ_{st}). Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.12: Impacts on Planted Acreage of Major Non-pollinator-dependent Crops

	Acreage of Non-pollinator-dependent Crops			County Bee Service
	(1)	(2)	(3)	(4)
County Bee Service	0.0075	0.0925	-	-
	(0.0289)	(0.1693)	-	-
Transported Bee Loss (IV)	-	-	-0.0051	-0.0555***
	-	-	(0.0097)	(0.0063)
Specification	OLS	2SLS	Reduced-form	First-stage
Kleibergen-Paap F		78.49	-	
Observations	9059	9059	9059	9059

Notes: Each column is a separate regression. Acreage of non-pollinator-dependent crops is the total planted acreage of non-pollinator-dependent crops—including corn, soybeans, cotton, sorghum, wheat, barley, oats, and rice. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) based on the data from the USDA's Quarterly Bee Colony Loss Survey. Both the acreage and county bee stocks are standardized. Column (1) reports the OLS regression following equation (11). Column (2) reports the 2SLS regression, where the county bee stock is instrumented for using transported bee loss constructed following equation (12). Column (3) report estimates of acreage of non-pollinator-dependent crops with respect to the transported bee loss in a reduced form regression. Column (4) reports first stage regression, following $BeeStock_{ct} = \gamma \cdot IV_{ct} + \phi Z_{ct} + \psi(X_c \times \gamma_t) + \alpha_c + \gamma_t + \varepsilon_{ct}$. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.13: Alternative Allocation Weights Based on Acreage of Grain Crops

	Net Farm Income			County Bee Service
	(1)	(2)	(3)	(4)
County Bee Service	0.0937**	-9.1814	-	-
	(0.0407)	(76.2692)	-	-
Transported Bee Loss (IV)	-	-	-0.0051**	0.0006
	-	-	(0.0025)	(0.0045)
Specification	OLS	2SLS	Reduced-form	First-stage
Kleibergen-Paap F	-	0.015	-	-
Observations	13460	13460	13460	13460

Notes: Each column is a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) based on the data from the USDA's Quarterly Bee Colony Loss Survey. Both net farm income and county bee service are standardized. Column (1) reports the OLS regression following equation (11). Column (2) reports the 2SLS regression, where the county bee service is instrumented for using transported bee loss constructed following equation (12). Column (3) report estimates of net farm income with respect to the transported bee loss in a reduced form regression. Column (4) reports first stage regression, following $BeeStock_{ct} = \gamma \cdot IV_{ct} + \phi Z_{ct} + \psi (X_c \times \gamma_t) + \alpha_c + \gamma_t + \varepsilon_{ct}$. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.14: Neonicotinoid-Exposure–Induced Transported Bee Losses as an Instrument for Bee Services

	Net Farm Income			County Bee Service
	(1)	(2)	(3)	(4)
County Bee Service	0.0812	1.2494	-	-
	(0.0494)	(2.7429)	-	-
Transported Bee Loss (IV)	-	-	-0.0063**	-0.0051
	-	-	(0.0028)	(0.0116)
Specification	OLS	2SLS	Reduced-form	First-stage
Kleibergen-Paap F	-	0.19	-	-
Observations	5871	5871	5871	5871

Notes: Each column is a separate regression. County bee service is the annual total number of bee colonies in a county, constructed using equation (10) based on the data from the USDA's Quarterly Bee Colony Loss Survey. Both net farm income and county bee service are standardized. Column (1) reports the OLS regression following equation (11). Column (2) reports the 2SLS regression, where the county bee service is instrumented for using transported bee loss constructed following equation (12). Column (3) report estimates of net farm income with respect to the transported bee loss in a reduced form regression. Column (4) reports first stage regression, following $BeeStock_{ct} = \gamma \cdot IV_{ct} + \phi Z_{ct} + \psi (X_c \times \gamma_t) + \alpha_c + \gamma_t + \varepsilon_{ct}$. Standard errors (in parentheses) are clustered at the county level.
Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.