

# Recreational Homes' Impact on Agricultural Land Use<sup>a</sup>

Charles Towe and Zhenshan Chen<sup>1</sup>

## Abstract

When non-residents purchase agricultural properties, the land use decision can make farmland operate below potential, still allowing for tax credits. We empirically investigate how non-resident ownership affects the agricultural land-use decisions in upstate New York. A difference-in-difference matching approach shows a causal link between purchases by non-residents and a loss of 11% of acreage to a lower-productivity use. A generalization shows this conversion counts for one-seventh of the decreased agricultural land in intensive uses in similar counties. Perhaps a simple opportunistic use of the tax-credit criteria, this phenomenon contradicts the policy's objective and might impose other consequences on rural communities.

*Keywords:* Agricultural Tax Exemption, Land Use, Policy Evaluation, Propensity Score Matching

*JEL:* Q15, H23

---

<sup>1</sup> Towe: Associate Professor, Department of Agricultural and Resource Economics, University of Connecticut, Storrs, CT.

Email: charles.towe@uconn.edu. Chen: Postdoctoral Associate, Department of Agricultural Economics, Mississippi State University, Starkville, MS. Email: zc352@msstate.edu. This is a pre-print copy.

Farmland protection has always been an important issue in the US. One of the central policies to protect farmland is preferential property tax policies, which lowers farmers' land carrying costs and serve the purpose of slowing down farmland conversion, especially in suburban areas where farmland is subject to considerable development pressure. However, in the suburban-rural adjacent areas (far exurban) that are not within the commuting distance to the urban core, the development pressure is low. Still, the pursuit of amenities by non-residents can alter agricultural operations to land uses of higher recreational values. More importantly, this non-resident owner can exploit the same economic preferential tax policies. Non-resident farmland ownership can result in land-use changes that targeting amenity preservation and tax minimization instead of agricultural profit, leading to decreased farmland in production and other consequences ranging from the rural job market, the fiscal soundness of the local government, and even a land supply effect for local food production. This study is the first to examine non-resident ownership's impact on agricultural land use empirically.

Recent studies suggest that rural-urban migration continues at pace, but rural areas offering natural amenities also have been increasingly experiencing housing pressure (Castle, Wu, and Weber 2011; Lichter and Brown 2014). One result of the latter trend is that farmland is increasingly in possession of non-resident owners<sup>a</sup>. This non-resident ownership is most likely to exist in the urban-rural adjacent areas that are close to an affluent urban area in the context to be drivable but too distant to commute on a daily basis. We define these urban-rural adjacent areas around a major city as the "second-home shed." The formation of a second-home shed relies on the fact that wealth and population are centralized in the urban area, but so are certain disamenities (e.g., noise and congestion). Many urban residents choose to spend their weekends and short vacations outside of urban areas. Leveraging this rising trend, the vacation rental industry (e.g., Airbnb) has been surging in these areas, which provides reliable investment value to the rural properties that further puts a focus on the amenity value of land-use decisions.<sup>b</sup> Combining these factors with the

preferential tax policies, metropolitan dwellers find purchasing homes in the second-home shed a feasible and increasingly popular option.

In the second-home shed, preferential tax policies may promote lower productivity agricultural uses by offering tax credits to operations doing little agriculture in practice. In our study area, upstate New York, non-resident owners have multiple ways to qualify for an agricultural exemption. For instance, they can rent to a local agricultural enterprise that already qualifies for the agricultural tax credit. This lessee farmer is required to file paperwork stating this land is an active part of their operation (being farmed) but is not explicitly required to earn income from the leased land.<sup>c</sup> As such, the incentive structures in place encourage rental of the land but do not encourage full agricultural usage of the land. Suppose a farm purchase is for a recreational home and the land is leased to a neighboring farmer, the non-resident owner will get the rural amenity as well as a significant property tax credit. The lease rate can be close to or even zero. The lessee farmer does not necessarily operate the land into its full agricultural potential, often simply baling low-quality bedding hay once per season.<sup>d</sup> This can happen for many reasons, including capacity limitations, the low cost of the land, or even at the behest of the non-resident owner who wishes to preserve pastoral views provided by a mostly dormant field. Thus, the land from these leased non-resident-owned farms tends to be underutilized compared to the full agricultural potential (State Board of Equalization and Assessment 1991), without mentioning the unleased parcels. Here we will focus on hay as the tax minimizing land use and further discuss prime farmland being cut as hay as a mismatched field in terms of quality and use.<sup>e</sup>

The literature studying farmland policies and rural-urban relationships has predominately focused on urban sprawl and farmland conversion (Brueckner and Kim 2003; Livanis et al. 2006; Irwin and Bockstael 2007; Liu and Lynch 2011). As suggested by Polyakov and Zhang (2008), preferential tax policies not only slow down development but also significantly affect rural land-use change between cropland and forest. However, they did not explicitly show evidence that the current tax policy might have a downside. Moreover, few have formally studied non-resident

ownership and land-use changes within the farmland due to data limitations and the thinness of the market. This paper contributes to the literature as the first paper utilizing annual land cover datasets and a causal inference framework to investigate the impact of non-resident ownership on farmland utilization in rural counties experiencing little direct urban influence. By addressing the downside of existing property tax policies in the emerging new rural-urban relationship, this article also adds to the debate around the rural-urban interdependence framework presented by Castle, Wu, and Weber (2011) and Wu, Weber, and Partridge (2017).

In this paper, we identify the land-use change in farmland due to being purchased by a non-resident owner. We focus on an area of upstate New York, where pressure in the form of non-resident agricultural ownership exists. The major identification problem is that recreational homes are not randomly assigned since the second-home transactions are the result of an endogenous process. In fact, many of the amenities that make good agricultural land happen also to be sought by second-home buyers. Moreover, finding a valid instrument that impacts the purchase decision and not the subsequent land use decision is a virtual impossibility. Thus, we rely on a rich dataset of observable attributes and a difference-in-difference matching (Heckman, Ichimura, and Todd 1997; Smith and Todd 2005) approach to deal with the endogeneity problem as well as the measurement errors.

While agriculture in the Northeast has struggled for years, even decades, the surge in interest of local foods has made farming viable again for farms close to urban centers. In this market where agriculture is undergoing a revitalization, our results suggest that non-resident owners are converting land from other agricultural cropping usages to hay at a much higher rate than existing residents, of whom many are farmers. Specifically, if an agricultural parcel in our primary study period (2001 to 2010) and area (Columbia County, NY) is purchased by a non-resident owner and sold to a resident owner, the agricultural land cover proportion decreases at least 0.11, and hay proportion would increase at least 0.11. The assumptions for difference-in-difference matching estimates are satisfied, and we see similar results in a neighboring county with similar geo-

locational amenities (Sullivan County). Moreover, a rough estimation of the total impact across similar counties with strong second-home markets suggests the underutilized farmland brought by non-resident ownership accounts for more than one-seventh of the dropped (from 2002 to 2012) total acreage of farmland in intensive uses in the agricultural second-home counties. From the outset and to be clear, it is undoubtedly within the landowners' choice set to engage in land uses that maximize their utility, but not necessarily their profit. The use-value assessment (i.e., considers soil type and productivity as a major component) in agricultural tax policies intends to encourage agriculture use commensurate with land quality. When a landowner benefits from the tax policy but engages in a lower productivity use than the land can support, it is effectively shifting the county's fiscal burden onto other residents. These urban second homeowners are wealthier than the other county residents. As such, this serves to exacerbate existing rural and urban financial inequalities.

The rest of the paper proceeds as follows. Section 1 introduces a conceptual framework to illustrate the economic incentive underlying second-home farmland underutilization. Section 2 presents the data and variables. Section 3 explains the approach to identify recreational homes on farmland and the empirical strategies to solve the identification problems. Section 4 presents the results of analyses and tests. Section 5 discusses the generalization to other second home counties and estimates the broader impact. The last section concludes and explores future works.

## 1. A Conceptual Framework

Housing markets function in a way that the highest bidder generally wins the right to purchase, and the property market in the second-home shed still functions properly in this sense. In this section, we begin with a brief conceptual framework to explain why the non-resident owners would underutilize their farmland. In our context, underutilization is the deviation from for-profit farmers' optimal land-use decisions based on historical land-use decisions.

All farmland owners (farmers or non-resident owners) choose a division of their land to maximize their utility<sup>f</sup>, which consists of two parts: the value of amenities ( $V(\cdot)$ ) and outside good

( $x$ ). This maximization problem is constrained by a budget balance condition and the total acreage of the disposable land ( $l_i$ ). If we simplify the model so that the agricultural land division is between unmanaged hay ( $l_{Hi}$ ) and agricultural uses excluding unmanaged hay ( $l_{Ai}$ ), this maximization problem can be formulated as:

$$\begin{aligned} & \max_{l_{Ai}, l_{Hi}} U(x, V(l_{Ai}, l_{Hi})) \\ \text{s. t. } & Y_i + p_A f_A(l_{Ai}; S_i) + p_H f_H(l_{Hi}; S_i) = p_x x + (C_A(S_i) - D_i t_A) l_A + (C_H(S_i) - D_i t_H) l_H; \\ & l_{Ai} + l_{Hi} = l_i. \end{aligned} \quad (1)$$

In Eq. (1), and the rest of this paper,  $i$  and  $t$  denotes parcel  $i$  and period  $t$ ,  $Y_i$  denotes income from outside the farm,  $f(\cdot)$  represents production function, and  $S_i$  denotes parcel-specific attributes other than acreage.  $p_A$ ,  $p_H$ , and  $p_x$  represent the prices for agricultural products, hay products, and the outside good, respectively.  $C_A(S_i)$  and  $C_H(S_i)$  are total costs per acre of agricultural production and hay production.  $C_H(S_i) < C_A(S_i)$ , since hay is a minimally invasive land use as it requires no fencing, no irrigation, and minimal land preparation. Prices are varying across time but identical to all farms within our study area, but costs ( $C$ ), linear in acreage, are related to parcel-specific attributes  $g$ .  $t_A$  and  $t_H$  are tax credits per acre of agricultural production and hay production.  $D_i = 1$ , if the parcel meets the minimum requirement of agricultural assessment, and  $D_i = 0$  otherwise. The following equation can be derived from the first-order conditions:

$$p_A \frac{\partial f_A}{\partial l_{Ai}} - C_A(S_i) + D_i t_A + \frac{U_2}{U_1} p_x \frac{\partial V}{\partial l_{Ai}} = p_H \frac{\partial f_H}{\partial l_{Hi}} - C_H(S_i) + D_i t_H + \frac{U_2}{U_1} p_x \frac{\partial V}{\partial l_{Hi}}, \quad (2)$$

where  $U_2 = \frac{\partial U}{\partial V}$ , and  $U_1 = \frac{\partial U}{\partial x}$ .

Three features might differentiate farmers and non-resident owners and hence may make their solutions different. First, farmers and non-resident owners may have different marginal utilities ( $U_1^R \neq U_1^F$  and  $U_2^R \neq U_2^F$ , where superscript  $R$  and  $F$  denote recreational owner and farmer, respectively). Second, farmers almost always meet the criteria of agricultural assessment ( $D_i^F = 1$ ), while non-resident owners do not necessarily meet the criteria. Third, non-resident owners'

production function might be different from farmers'<sup>1</sup>. *Appendix A* derives solutions for special cases, which shows that the optimal land division of farmers is different from that of non-residents in each case. However, it is not theoretically clear that nonresidents' optimal hay usage is always higher than farmers' ( $l_{Hi}^R > l_{Hi}^F$ ) in normal cases. In a special case where farmers only derive utility from production and non-residents' production function is flat, it is likely that  $l_{Hi}^R > l_{Hi}^F$ .

In a word, non-resident-owned agricultural properties will be underutilized in the sense that the land use management will deviate from resident farmers' optimal and profit-maximizing management. This deviation is a combined result of the different valuation of aesthetic values and different farming capabilities. The empirical question is whether the observed hay proportion of non-resident owned farms is larger than what it would be with resident owners ( $l_{Hi}^R > l_{Hi}^F$ ) and the size of this deviation. Relying on these theoretical predictions, we choose land cover proportion across time as the outcome, which has two advantages. First, the land cover data are derived from datasets combining a large-scale survey and satellite images, which are less influenced by attrition bias and more reliable than survey data. Second, hay, identified as the most favorable land cover for non-resident owners by the literature (e.g., State Board of Equalization and Assessment, 1991), is a distinguishable land cover in available datasets. A finding of hay intensification suggests that non-resident owners tend to underutilize their farmland.

## 2. Data

Before delving into the data, we first briefly introduce the empirical analysis to provide a context. We focus on an area of upstate New York where pressure in the form of agricultural non-resident ownership exists. After the initial steps to identify non-resident owners, we proceed to investigate whether and how much the land use is different between non-resident owners and regular farmers. Using a rich set of observational data, we employ a difference in difference

---

<sup>1</sup> A likely situation is that nonresidents have relatively limited resource (i.e., time, equipment, or skills) to increase crop productivity.

matching (DID-matching) approach with regression adjustment, which addresses the selection problems from time-invariant unobservables (i.e., selection bias of second home). The primary outcome of this study is agricultural land use. We conduct several tests to show our sample satisfies the assumptions of DID-matching and to check the robustness of the estimated effects.

Since the methodology largely focuses on solving the problems lying in the available data, we will introduce it in detail after this data section. Data for this study can be divided into three general types: geographic data, tax parcel data, and county-level information.

*Geographic data.* Geographic data contain spatial information about land cover, soil, temperature, precipitation, elevation, towns, populated areas, roads, conservation areas, and agricultural districts. The land cover data are from the NLCD data by the United States Geological Survey (USGS). Specifically, we use NLCD 2001, NLCD 2006, and NLCD 2011<sup>h</sup>, which are the ones where New York State's land cover is available. To match the time frame of the tax parcel data (ending 2014), we get additional land cover data (CDL) from National Agricultural Statistics Service (USDA NASS)<sup>i</sup>, including land cover information for 2002 and from 2008 to 2014. The soil data are downloaded from the Soil Survey Geographic Database (SSURGO)<sup>j</sup>. The agricultural district data are acquired from the Cornell University Geospatial Information Repository (CUGIR)<sup>k</sup>. All the other datasets come from Geospatial Data Gateway from the National Resources Conservation Service<sup>l</sup>.

*Tax parcel data.* These datasets contain two types of data: cadastral data and transaction data. Cadastral data are from New York GIS Clearinghouse<sup>m</sup>, including tax parcel point files with detailed property information from 2004 to 2010, and a tax parcel polygon file with detailed information for 2014 or 2015 (varies across counties). The transaction data come mainly from the Economic Research Service, United States Department of Agriculture (USDA ERS), and contain property transactions in New York from 1999 to 2008. With these transaction data, we can get detailed information about the landowners and the second-home transactions. The temporal variations allow us to extend the ownership status (resident or non-resident) back to 2001.

*County information data.* These data consist of commuting data, population and income data, community survey data (ACS), county business pattern data (CBP), and operation with direct sale data. Commuting information comes from the Census online datasets: Journey to work<sup>n</sup>, which are estimations based on large-scale surveys, containing the number of workers in commuting flow from county to county. The commuting information is employed to build a variable indicating non-resident ownership, as described in the following section. Other datasets in this part are mainly from online datasets provided by the US Census Bureau, and these national-level data are used to estimate the total impact.

*The Outcome.* The available data provide for a feasible quasi-experimental approach to show how the land use of an agricultural property might differ between a non-resident owner and a farmer. This subsection discusses the specification of the outcome variable. As a result of the conceptual framework, the outcome of interest is naturally the land-use decisions, of which the corresponding measurements in the geographic datasets are the land cover proportions of agricultural parcels. Examples of how the land cover layers interact with parcel boundaries are shown in Figure 1. We classify land covers into six types: agricultural land (used for agriculture except for hay, including land in crops, vegetables, fruits, tree nuts, and pasture land), hay, forest, wetland, vacant land, and developed land. We calculate the proportion of these land cover types across time on each parcel and use them as the outcome variables. Further, we define mismatched hay as hay on high-quality farmland having alternative higher-profit options<sup>o</sup>.

[placeholder – Figure 1 about here]

*Covariates.* The set of covariates naturally consists of those independent variables influencing agricultural land-use decisions. A subset of these observables (e.g., those variables determining the rural amenity of a parcel) are very likely to be correlated with the property's second-home status, and omitting them in the analysis would cause selection bias. By combining geographic datasets and tax parcel datasets, we get a comprehensive set of characteristics of a parcel, including parcel acreage, total assessed value, age of the house, house square footage, distance to the primary road,

whether the parcel is divided by a road, streams or ponds on the parcel, distance to population centers, elevation, average minimum temperature, erosion percentage, excess water percentage, soil limit percentage (SSURGO classification), land capability, prime farmland percentage, certified agricultural district. Though not exhaustive, these covariates can cover most of the relevant attributes of a parcel when it comes to land use and second-home transaction decisions.

### 3. Methodology

The main hypothesis we wish to test requires narrowing the study location to areas popular for vacation properties and identifying non-resident owners in these areas. First, we define the second-home shed of a major city, New York City (NYC). Then, we identify the agricultural counties in the second-home shed. Finally, we discuss how to identify the non-resident owners in these agricultural second-home counties.

#### **Identifying the Study Area and Non-resident Owners**

Conceptually, the second-home shed is an area close enough to a wealthy urban area to be suitable for regular short trips but far enough not to be commutable daily, and it is an area where ample amenities exist, and the rural land market experiences pressure from urban dwellers<sup>p</sup>. New York City is a major metropolitan area that imposes huge influence on surrounding rural areas; we present Figure 2 to illustrate the location of its second-home shed. The approach underlying Figure 2 begins with a basic assumption: if the tax documents are sent to NYC<sup>q</sup>, we assume that the landowner's primary residence is in NYC. This approach identifies the NYC-dweller-owned properties in Upstate New York. From Figure 2, we can see that a large share of the properties owned by NYC dwellers is located within 200km but outside the 100km buffer of NYC, where the density of NYC homeowners is even higher than what's in the 50km - 100km range. Therefore, we assume that the second-home shed of NYC is between its 100km buffer and 200km buffer.

[placeholder – Figure 2 about here]

We then identify the agricultural second-home counties in this second-home shed. Strictly speaking, not all the properties owned by urban dwellers in Panel A of Figure 2 are recreational homes since some of the owners may commute to NYC and have the tax bills delivered to their businesses. Panel B of Figure 2 illustrates the numbers of workers in the commuting flow to NYC by counties, where the numbers of homeowners from NYC are also included for comparison. The histogram suggests that Dutchess County and Putnam County, both of which are partly within the 100km radius of NYC, have many commuters to the city but not as many homeowners from the city. However, counties like Columbia and Sullivan have many more homeowners from the city than commuters to the city, indicating that a fair number of people own properties here but do not consider these properties as primary residences. Hence, Columbia, Ulster, Sullivan, Delaware, and Greene appear to be candidate second-home counties of NYC.<sup>r</sup> Since our focus is on agricultural land use, we then narrow down the location of interest to agriculturally important counties. In New York, agricultural districts, where law encourages the use of land for farming, have been established since 1971 and certified regularly; we take it as an indicator of agricultural importance.<sup>s</sup> Appendix Figure G1 shows the wide geographic dispersion of agricultural districts, where Columbia, Sullivan, and Delaware have large acreage in agricultural districts, while Greene and Ulster are mostly occupied by the Catskill Mountains. Thus, the non-resident ownership, preferential agricultural taxes, and high amenity values combine in Columbia County, Sullivan County, and Delaware County. We conduct our impact evaluation for Columbia County first, then apply it to Sullivan<sup>t</sup> to verify its robustness.

Finally, we investigate individual properties to find those agricultural properties owned by non-residents<sup>u</sup>. This involves some complexity since non-resident owners are not necessarily from NYC (see Appendix C for detailed information). If the tax bill of a property is consecutively sent to a zip code that is not commutable, then we treat it as a second home (see detailed second home zone for Columbia in the appendix Figure G2). And if this property is used for agricultural purposes recently in the tax parcel datasets (since 1999), we classify the property as a non-resident-

owned agricultural property. While this approach generates an indicator for non-resident owned farmland, two issues in this approach and a question concerning land cover data quality may make our estimates lower bounds of the true effects:

First, with this approach, we may identify fewer non-resident owners than there are in reality. We exclude nearby non-resident owners as these short-range commutes are convenient enough to manage the farm, but this doesn't necessarily mean they actively operate the farm. This approach is conservative and includes the most credible second-home owners. Since we omit some real non-resident owners or mistakenly assign them into the control group (farmers), our estimates would be a lower bound of the true effects of non-resident owners on land use.

Second, the assumption that non-commutable tax bills indicate agricultural recreational homes may fail when a big agricultural enterprise from the nearby urban area owns a large acreage of for-profit farmland. In this situation, the underlying assumption that non-resident owners are not optimizing agricultural profit may fail. In our data, we identify only a few extremely large parcels owned by non-residents, and empirical evidence does not show this as a major concern (Table 3). However, if this issue were pervasive, it does bias our estimates *downward* because these profit-motivated non-resident owners would presumably operate similarly to a resident farmer.

Third, due to limitations in available land cover data, we combine data from two sources, the National Land Cover Database (NLCD) and the Cropland Data Layer (CDL). These land cover data are used to construct the outcome variable, and the estimation strategy is capable of addressing the random technical differences between these two sources as these differences are not systematically related to non-resident parcels. While economists apply these datasets in empirical studies (e.g., Hendricks, Smith, and Sumner, 2014), problems have been raised toward the quality of these datasets (e.g., Irwin and Bockstael 2007; Wickham et al. 2013) but these are context-sensitive. A representative of the criticism in economic research, Irwin and Bockstael (2007) suggest that the NLCD data are not accurate in a target raster cell (30m by 30m) if surrounding cells are of different land cover types<sup>v</sup>. This issue matters most when the dataset is used to identify

low-density development or rather small landscape changes. This problem does not influence our results, since identifying low-density developments usually involves land cover detection of around .05 acres, which is one-quarter of a cell, while the farms we study are mostly above the scale of ten cells<sup>w</sup>. Even if this issue did exist (i.e., the owner turns a tiny portion of land from crops to hay, and the change is non-detectable because of limits in data quality), it would cause the estimated effect to be a lower bound of the true effect.

### **Identification problems**

The theory presented in the second section suggests that underutilization of non-resident-owned farmland naturally comes from non-residents' utility-maximizing decisions, given their flexible financial constraints and recreational amenity orientation. Our empirical work intends to quantify the agricultural land-use changes caused by this non-resident farmland ownership. The identification issues can be classified into two types: selection bias due to the nonrandom assignment of non-residents and measurement errors due to data limitation.

The non-resident ownership is not randomly assigned to the farmland: non-resident buyers choose which property to buy, and farmers choose to sell or not. It turns out that agricultural parcels with certain attributes are more likely to be owned by non-residents, and these attributes are also correlated with the land use outcome. Methods ignoring this type of selection issue tend to yield inconsistent estimates. Given that there are no persuasive instrumental variables available (exogenous attributes influencing second-home transactions but not land use decisions), matching estimators are good choices, especially when the attained variables can adequately describe both the selection process and land use outcome (Blundell, Dearden, and Sianesi 2005). Due to the relatively short time range of the data, we assume that if the parcel boundaries do not change throughout this period, we do not expect that the area, farmland quality, geography, precipitation, temperature, elevation, assessed values, or other attributes vary<sup>x</sup>. We employ a propensity score matching (PSM) estimator based on observable parcel attributes to deal with the selection-on-observable problem.

Besides the selection-on-observables problem, reverse causation and selection-on-unobservables are relevant issues in this context. For example, the non-resident buyers are very likely to make purchase decisions based on prior land use, and the land-use variable is usually autocorrelated. If second-home buyers prefer farmland with more land in hay, the land cover proportion of hay is likely to be higher in the non-resident purchased parcels even if non-resident owners behave the same way as a farmer might. In this situation, the causal relation cannot be fully recovered by simply implementing regression or matching estimator methods without considering selection on prior outcomes. Moreover, a similar problem happens when the non-resident buyers prefer some time-invariant unobservables, and these unobservables happen to be correlated with land-use outcomes. The reverse causation could be considered as a special case of the selection-on-unobservable problem, and both are usually addressed, in the literature, with a difference-in-difference (DID) approach when these unobservables are time-invariant.

Strategies to deal with the selection problems assume the parcels stay the same throughout the study period. But the tax assessment data show that some agricultural parcels are splitting up or merging<sup>y</sup>, which makes it hard to address the selection issue as suggested above. We also consider this as a major measurement error, since the acquired parcel boundaries cannot reflect these changes and hence the resulting parcel attributes would be wrongly computed. The analysis only includes parcels having the same acreage and assessed values throughout the study period to avoid these problems brought by changing parcels.

Another kind of measurement error might exist considering two different groups of land cover data we use: one group is from NLCD by USGS, and the other group is from CDL by USDA. Even though they are using the same satellite data<sup>z</sup>, the measurement standard by two different groups of people across time may be different, which is suggested in Wickham et al. (2013). For example, one major difference is that, in CDL 2002, there's only alfalfa hay denoted and we cannot differentiate other hay from grassland<sup>aa</sup>, while in later land cover datasets, alfalfa and other types of hay are clearly noted (thus CDL 2002 is not used in the analysis). Moreover, some measurement

accuracy issues might persist in all datasets and bias the estimates. The measurement errors can be addressed by the DID-matching approach since the time-invariant biases (persisting errors) will be taken out while the technical differences between different datasets are not likely to be dependent on non-resident ownership (though there might be an attenuation effect since more noises are included).

### **Identification strategy**

To avoid major measurement errors and deal with the selection problems, we firstly exclude the parcels with experience of ‘major’ divisions or mergers. Taking Columbia County as an example, we do this by purging those parcels with an unbalanced panel of ownership information or with vastly changing acreage on record<sup>bb</sup>. A balanced panel for 813 stable agricultural parcels is acquired after this process<sup>cc</sup>. We assume that the split or merged parcels are influenced by non-resident owners in a similar way to resident owners. With this approach, we get a less biased estimator for avoiding major identification problems and still achieve a lower bound estimation for the population impact, as suggested above.

We first consider a fixed-effects model<sup>dd</sup> to solve the selection problems mentioned above. However, the limitation in the available assessment data (2004-2010) and land cover information (data available for 2001, 2002, 2006, and 2008-2014) results in a final panel including complete information for only 2006, 2008, and 2010. Also, since the land cover does not usually change immediately, the changes in homeownership between 2006 and 2008 may not be revealed by the land cover in 2008 but would be in 2010. Thus, we end up with a DID approach that uses the outcome in a way that accommodates the time-lagged nature of land cover change and requires less data.

The main analysis combines the DID approach and a matching estimator to deal with the endogeneity problem as well as the remaining measurement errors. We take those parcels newly purchased (between 2001 and 2010) by non-resident owners as the treated group, and parcels owned by residents, who are more likely to be actively farming throughout the period, as the

control group. The average treatment effects on treated (ATT) are estimated based on a propensity score matching approach, in which the scores are generated through a logit model linking the treatment dummy (1 if the farmland was purchased by a non-resident from a farmer in the study period, 0 otherwise) to various parcel attributes. Intuitively, by this process, we compare the land-use changes in the treated group throughout the study period with that in the control group, given that these paired parcels have nearly identical time-invariant attributes. This difference-in-difference matching estimator (Heckman, Ichimura, and Todd 1997, DID-matching hereafter) is the preferred estimator among many matching estimators (Smith and Todd 2005). It deals with the endogeneity of non-resident ownership in our context (selection on time-invariant observables and unobservables), while in practice, it also performs better when the study integrates different datasets.

The validity of the DID-matching estimator requires the three assumptions: *the stable unit treatment value assumption (SUTVA)*, *conditional parallel trends assumption (unconfoundedness or ignorable assignment)*, and *the Common support (overlap) assumption*. With Appendix B and Appendix figure G3 reporting the details of these assumptions in our context, we discuss *the Common support (overlap)* assumption here since it is usually the one with the most concerns in empirical work. *the Common support* assumption essentially suggests that for every non-resident parcel purchase (a treated observation), there must exist a farmer-owned parcel (in the control group) having roughly the same observed attributes, and hence the same propensity score. This assumption is usually implemented loosely in empirical applications, for example, it is sometimes turned into a requirement of the same domain of the propensity scores for different groups (as in Dehejia and Wahba 1999, 2002)<sup>ee</sup>. Huber, Lechner, and Wunsch (2013), based on carefully conducted simulations, suggest that trimming observations with too much weight (i.e. with weak support) is important for all estimators, and that radius matching (see Dehejia and Wahba 2002) combined with regression performs best overall.

On top of the DID-matching strategy, we also employ a regression adjustment approach to address the leftover discrepancies in covariates. Average nearest neighbor matching estimators assign the same weight to the chosen number of nearest neighbors, regardless of their distance in propensity score from the treated observation, which may be consistent on a large enough sample under some circumstances (Abadie and Imbens 2006), but not likely to work here<sup>ff</sup>. Other pure matching estimators are also proven to perform worse than matching and regression adjustment estimators (Abadie and Imbens 2011). Moreover, in a relatively small sample, non-exact matching leads to a potential bias through the leftover discrepancies in the matched sample (Caliendo et al. 2008), and regression adjustment would reduce this bias by utilizing the difference in the covariates. Thus, we follow the recommendation by Rubin (1973) of combining matching with regression adjustment (RA), which is employed by many empirical studies (e.g., Heckman, Ichimura, and Todd 1998; Rubin and Thomas 2000; and Lawley and Towe 2014).

Therefore, the DID-matching estimators will identify the ATT in this form:

$$\widehat{ATT} = E\{ E[Y_{1,it'} - Y_{1,it} | \mathbf{Z}_{1,i}] - E[Y_{0,it'} - Y_{0,it} | \mathbf{Z}_{0,i}] | D_i = 1 \}, \quad (6)$$

where  $Y_{1,it}$  and  $Y_{0,it}$  denote the outcomes for observation  $i$  at time  $t$  prior to treatment for treated and control observations, respectively.  $Y_{1,it'}$  and  $Y_{0,it'}$  denotes the post-treatment outcome.  $D_i$  is a dummy denoting the treatment status with  $D_i = 1$  denoting the treated observations (non-resident owned parcels), and  $\mathbf{Z}_i$  is a vector of time-invariant parcel attributes. Moreover, the regression adjustment estimators to be presented will additionally adjust this ATT by the differences in covariates between the matched pairs and can be interpreted by:

$$\widehat{ATT}_{RA} = E\{ E[Y_{1,it'} - Y_{1,it} | \mathbf{Z}_{1,i}] - E[Y_{0,it'} - Y_{0,it} | \mathbf{Z}_{0,i}] | D_i = 1 \} - E[\boldsymbol{\beta}_0(\mathbf{Z}_{1,i} - \mathbf{Z}_{0,i})], \quad (7)$$

where  $\boldsymbol{\beta}_0$  is the vector of coefficients resulted from a regression of outcome on  $\mathbf{Z}_i$ , which is conducted merely on matched controls.

#### 4. Results and Discussion

Second-home markets around NYC are increasingly active in recent decades, as the property price in urban areas ascends to historical highs. Since 2001, non-resident ownership of farmland has increased by 22.7% in Columbia County, 29.8% in Sullivan County, and has remained stable in Ulster County which has maintained rates hovering around 65% demonstrating a saturated market. Appendix table G1 shows these full results.

##### **Non-residents' impact on farmland**

We first apply our estimation strategies to Columbia County, NY, where a second-home market is active and expanding. Relying on the conceptual framework, regressing land cover proportions on parcel attributes may give us insights into the different management strategies between farmers and non-residents. Preliminary regression models and details are displayed in Appendix D. We show separate OLS regressions for resident owners and non-resident owners, and it is clear that their land cover management reactions to different parcel attributes are quite different.

To get the proposed ATT, we proceed to the DID-matching approach and transform the panel data into a cross-sectional dataset. We divide the 813 transacted agricultural parcels into four groups: transacted from resident to non-resident during the study period, transacted from non-resident to resident during the same period, owned by residents throughout the study period, and owned by non-resident throughout the study period.

[Table 1 about here]

In the main specification, the focused treatment group consists of parcels transacted from resident to non-resident in the study period, which will be denoted as treatment group 1, and the control group includes parcels owned by residents. We get 49 agricultural parcels<sup>gg</sup> in treatment group 1, and 609 in the control group. To construct the outcome, given that the study period is from 2002 to 2010<sup>hh</sup>, we take differences for each parcel between the land use proportions of 2012<sup>ii</sup> and 2001. The summary statistics are available in Table 1. The sample covariates are not balanced,

suggesting that non-residents do seek different parcels from an average property buyer. Parcels in the treatment group have a higher average land proportion in agriculture and lower in hay before 2002.

A logit model is used to generate propensity scores (shown in Appendix Figure G4) employed in the DID-matching estimator. Logit results and covariates are shown in the appendix table G2. These variables are time-invariant, influencing simultaneously treatment and outcome, and not influenced by the anticipation of being treated. We exclude the prior land use from this logit model because, in a DID-matching estimation, controlling on pre-treatment outcomes is biased and equivalent to simple matching (Chabé-Ferret 2010).<sup>jj</sup>

Following the idea of trimming observations with weak support (Huber, Lechner, and Wunsch 2013), for one treated, we require six control observations within a propensity score radius of 0.01. During this process, 3 out of 49 treated are trimmed<sup>kk</sup>. Appendix table G3 and Figure G5 show the balance tests and distributions before and after radius matching resampling (n=6). We can conclude that this radius matching approach generally ensures the overlap of covariates between the treatment and control group, improving from the unbalanced raw sample.

[Table 2 about here]

Table 2 shows the results of the OLS estimator (as baseline), radius matching estimators, and regression-adjusted matching estimators with different numbers of nearest neighbors restricted by a caliper of .01. From Table 2, we can see with the full sample, the treatment (non-resident) effects estimated by OLS is -.09 (significant at 95% level) for agricultural land cover proportion change and .07 (90% level) for hay. All matching estimators suggest higher impacts than the OLS estimator does (larger than .11, negative, and significant for agricultural usage change; larger than .1, positive, and significant for hay change), which points out the differences made by mitigating selection bias. Comparing to regression adjusted estimators, average nearest neighbor matching generally estimates a smaller effect size and a higher standard error, which is due to three issues: the leftover discrepancies in covariates, unevenly distributed counterpart controls for

treated observations near boundary (with scores very close to maximum or minimum scores), and the failure of bootstrapping standard errors in matching estimators (Abadie, Imbens 2008). In the regression adjustment specifications, the first two sources of potential bias are addressed, and robust standard errors (Abadie, Imbens 2006) are implemented, which results in estimates of over .11 significantly positive impact in hay change, robust to a different number of neighbors. Regression adjustment with  $n=3$  gives the highest ATT (-.13 for agricultural, .13 for hay, both significant at 95% level), while regression adjustment with  $n=6$  gives the smallest ATT (-.12 for agricultural, .11 for hay, both significant at 95% level). Propensity score tests<sup>11</sup> suggest that the matching works better when the number of nearest neighbors rises. Thus, we take the results with  $n=6$  as the least biased ATTs. Comparing to regular residents having identical farmland, on average in Columbia County, NY, the non-resident owners bring at least a decrease of .12 in farmland proportion used for agricultural (except hay), and an increase of .11 in farmland proportion used for hay.

The thinness of the market we are studying and the identification issues limit the sample size of our treated group, which brings one concern in this empirical context: the estimated results might be subject to the influence from the outliers. Thus, it's important to show the results are similar on different subsamples. Under consideration that very large agricultural parcels carry higher opportunity costs and hence are less likely to be recreational second homes, we conduct further estimation with "less or equal to" restrictions on acreage. The results are displayed in Table 3, where the cut-off criteria are set to 70 and 90 acres. In general, the regression adjustment estimators for the smaller parcels consistently produces larger ATTs in agricultural use change, for example, for the group of acreage less than or equal to 90 acres, the estimated ATT (with  $n=3$ ) equals to -.15 for agriculture (significant at 90% level) and .14 for hay (not significant at 90% level, but very close). This is consistent with the story that recreation-oriented landowners tend to decrease agricultural usage while keeping the land in hay to attain the agricultural assessment.

[Table 3 about here]

## Placebo test and alternative hypothesis test

To prove the validity of the estimated average treatment effects, as stated above, we need three conditions: SUTVA, conditional parallel trends, and common support. Since SUTVA is not a major concern in our context, and the common support condition is enhanced by the caliper matching process, the conditional parallel trends condition is the only obstacle left to validate the ATT assumptions. To test the conditional parallel trends assumption, we usually compare the treatment group's pre-treatment land-use trend with the control group's corresponding land use trend, conditional on a comprehensive set of covariates. One tricky thing here is: we only have one year's land cover (2001) before the first relevant second-home transaction, which is not enough for a trend test. To deal with this, we take all such parcels that are transacted from residents to non-residents from 2006 to 2010 as the new treatment group and estimate placebo DID-matching ATTs with land cover changes before the actual treatment (from 2001 to 2006) as the outcome.

We expect zero treatment effects from this test since there is no treatment happening in the treatment group. Table 4 shows the placebo test results: 4 treated are purged by radius matching; even though OLS gives significant results to the opposite direction of the estimated effects, all adjusted matching estimators give small insignificant trend differences. In any event, there are no statistically significant differences in the prior conditional trends between the treatment group and control group across all the matching specifications, and the insignificant opposite-direction trend differences further guarantee our estimated ATTs as consistent lower bounds. These results do not only secure the parallel trend condition but also rule out the possibility that random errors prior to the treatment dominate (due to the small sample size) the estimated effects.

[Table 4 about here]

Now that the fitness of the three conditions to our dataset is tested, we are confident that the differences in land cover trends are caused by the second-home transactions. But are these effects led by the fact that non-resident owners are underutilizing farmland instead of the discontinuity in management associated with the transactions or ownership changes? To rule out this alternative

hypothesis, we conducted the same DID-matching approach on another group of parcels (treatment group 2), which includes all parcels that transacted from non-residents to residents from 2002 to 2010. The results are shown in Appendix E Table E1, with no significant effects for all estimators across all land cover types, so it appears that the increase of farmland proportion used for hay is due primarily to the non-resident ownership and not simply the transition of ownership.

Based on the results of our estimation and tests, if an agricultural parcel in Columbia County is sold to a non-resident owner instead of a resident, the agricultural land cover proportion change during 2001-2012 will decrease at least .11, and the hay proportion change during 2001-2012 will increase at least .11. Moreover, non-residents tend to produce hay on farmland with high capability and greater potential profit in another use. Further, this percentage is the estimate on recent transactions and relatively quick changes in land use, and it does not measure land-use changes for second-home residents before the beginning of the study period.

#### **A robustness check: non-residents' impact in Sullivan**

As discussed previously and illustrated in the appendix table G1, Columbia and Sullivan are similar when it comes to the non-resident owned ratio of farmland, while the agricultural parcels in Ulster look like a saturated second-home market. Since there are not many new non-resident owners in Ulster, we are only able to apply the DID-matching approach to Sullivan to get some ideas of the robustness and generalizability of estimated impacts for Columbia. Although Sullivan has a similar proportion of non-resident owners as Columbia, it is not geographically close to Columbia: the Hudson River and the Catskill Mountains are between them. Thus, similar results for Sullivan would suggest that the underlying behavior of non-resident owners is not county-specific, and the negative impact of non-resident ownership generally exists at least in emerging second-home counties.

Appendix table G4 shows the DID-matching estimates for Sullivan, and the results are generally similar to the Columbia results: the impacts on agricultural usage and hay are sizable and significant. Since the sample size in Sullivan is even smaller, we increase the matching radius

to .15 and place more trust in the specification with more nearest neighbors. Based on the adjusted matching estimations with  $n=6$ , we claim that compared to regular residents, on average, the non-resident owners bring at least .9 less (statistically significant at 90% level) agricultural land use, .9 more (significant at 90% level) hay, and .11 more (statistically significant at 95% level) mismatched hay to farmland in Sullivan County, NY. Also, like Columbia, the Sullivan data satisfy the SUTVA condition, the common support condition, and the conditional parallel trends condition (test results are available upon request).

As a conclusion, based on the DID-matching approach, we get very similar average treatment effects of non-resident owners on farmland in Columbia and Sullivan, where agricultural second-home markets are emerging. We take the land cover change on resident-owned parcels as the counterfactual outcome for parcels newly owned by non-residents and show that the differences in land cover change are caused by the non-resident owners' land cover management behavior, which is quite different from the farmers'. Specifically, during the study period (2002-2010), if an agricultural parcel is owned by a non-resident instead of a resident, its agriculture proportion change will drop, on average, by about .1, and its hay proportion will rise by about .1.

## 5. Discussion and Total Impact Estimation

We have identified the non-resident owners' impacts on agricultural land cover in two second-home counties of NYC, but one might wonder how prevalent this situation is in the country at large and what the potential magnitude of the aggregated impact is. We assume other major urban areas with considerable wealth and population should also be a supply of second-home owners, who also generate similar land cover trends to what is estimated above. To find urban-adjacent areas most like Columbia County, we employ a distance measurement incorporating Mahalanobis distances and a few structural assumptions, which are described fully in Appendix F. For each county, this procedure considers its distance to all surrounding metropolitan areas and utilizes a comprehensive set of metropolitan characteristics (Table F1). It generates a ranking list with those counties most similar to Columbia ranked first. Not surprisingly, we find the most similar counties

along the Northeast corridor from DC to Boston. Many of these counties are known second home spots and they all fall under agricultural assessment tax provisions. To quantify an estimate of the total impact of non-resident owners on agriculture in this region, we use the estimate of .1 underutilized farmland acreage by non-residents to the top 4% of the ranking list.

We use the acreage of land in farms<sup>mm</sup> in each selected county from NASS data as the measure of the total area of agricultural land, and the acreage of land in production (harvested cropland) is used as the comparison. Table 5 summarizes the rough estimation of total underutilized acreage under different scenarios. Assuming that the selected counties in the ranking list have a similar situation as Columbia (20% of agricultural parcels are owned by non-residents), the total acreage of underutilized land in farms would add up to at least 45,393 acres in 2012, which is about 5% of the total acreage of harvested cropland in the same counties. If these counties have reached a status of saturated second-home market (65% non-residents), the total underutilized acreage by non-residents will be 147,527 acres, which is about 16% of the total acreage of harvested cropland.

[Table 5 about here]

More importantly, one should note that the decrease of harvested land acreage is much larger than the decrease of land in farms, and this discrepancy suggests that the harvested proportion of land in farms is dropping across time. Given the definition of harvest cropland, this type of loss suggests land-use changes within farms. In Table 5, the estimated increase in underutilized acreage from non-residents (11,348) is comparable to the farmland loss reported by NASS, regardless of whether we compare it to the decreased acreage of farmland in production (-70,211), or the decreased acreage of land in farms (-33,415). These facts illustrate the potentially important role of farmland underutilization in decreasing agricultural production, given that the underutilization estimated here is defined in just one way, within non-resident-owned farms, it is most likely a lower-bound estimate. Moreover, this kind of underutilization problem does not just suggest agricultural production loss but also links to misallocation issues: the numbers shown here also indicate the total amount of sub-optimal agricultural tax benefits that might be given out.

Though not the direct focus of this study, the underutilization of peri-urban farmland could have negative impacts from other perspectives. One example is its impact on the ever-popular “local” food supply (Hardesty 2008; Darby et al. 2008). Lands prime for local food production<sup>nn</sup> lie in the same belt of distance as the lands in the second-home shed. Urban dwellers buy recreational homes on farmland and pose a threat to the production of local foods, while these local products are highly valued by these very landowners (Thilmany, Bond, and Bond 2008). Moreover, agricultural land being underutilized may have social spillover effects among rural residents - they potentially pay a higher tax rate to balance out the tax credits offered to the non-resident owners.

## 6. Conclusion

Our analysis indicates a considerable amount of farmland is underutilized by non-resident owners. This is an interesting finding and is a perfectly acceptable market outcome. However, under the incentives created by the flexible and unbinding criteria for an agricultural assessment, this underutilized land is still eligible for a tax credit. Upon enrolling in the credit, the broader base of rural taxpayers is effectively subsidizing this parcel's underutilization. Based on results from a regression-adjusted DID-matching approach, we find patterns in second-home counties of NYC suggesting that, compared to farmers, non-resident owners of farmland place a smaller proportion of their land in crops and a larger proportion in hay. These results should be taken as lower bounds of the true impacts of non-resident owners on agricultural land use because the way to identify the non-resident owners is rather conservative, omits long-term non-resident owners, and some identified non-resident owners might be actual farming enterprises.

A great deal of attention has been paid to farmland conversion, but few studies have considered the problem at the intensive margin of the agricultural parcels. Incentivized by the agricultural tax credits that can be easily attained, non-farmers can operate farmland for primarily non-farming objectives at a societal expense.<sup>oo</sup> This phenomenon, likely led by unbinding tax-credit criteria, is a contradiction to the policy's objective and might impose other consequences on rural communities. Admittedly, non-farmer owners could provide certain social benefits by converting

crop production to conservation land covers, and in practice, these social benefits could be paid back to the owners by programs like the Conservation Reserve Program (CRP).<sup>pp</sup> However, in our study area and similar counties, CRP is perceived to target specific lands with issues limiting agricultural production (e.g., erosion) and its enrollment of farmland is very limited. This likely suggests that converting farmland to conservation covers is generally not perceived as a socially optimal behavior, while such behaviors would be incentivized by agricultural tax credit nonetheless.

Future work can be done in several directions: the first is to generalize the impact evaluation to more detailed outcomes, for example, getting farmers' perception and valuation of nearby non-resident owned farmland, or classifying land cover into more detailed types indicating productivity (like high-quality hay and low-quality hay); a second direction is to generalize the impact evaluation to other counties of interest; finally and more broadly, the impacts on the supply of local food and the supply of available farmland to new and beginning farmers producing local products is an important policy outcome to examine.

## **Acknowledgment**

We thank participants at the AAEA Summer meetings (2017 and 2018) and seminar participants at the University of Connecticut for many helpful comments and suggestions as this work progressed through many stages and data improvements. This work was supported by the USDA National Institute of Food and Agriculture, Hatch project 1007175.

## Reference

- Abadie, A., & Imbens, G. W. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74(1): 235-267.
- Abadie, A., & Imbens, G. W. 2008. On the failure of the bootstrap for matching estimators. *Econometrica* 76(6): 1537-1557.
- Abadie, A., & Imbens, G. W. 2011. Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics* 29(1): 1-11.
- Alix-Garcia, J. M., Sims, K. R., & Yañez-Pagans, P. 2015. Only one tree from each seed? Environmental effectiveness and poverty alleviation in Mexico's Payments for Ecosystem Services Program. *American Economic Journal: Economic Policy* 7(4): 1-40.
- Augurzky, B., & Schmidt, C. M. 2001. The propensity score: A means to an end.
- Bertram-Huemmer, V., & Kraehnert, K. 2017. Does index insurance help households recover from disaster? Evidence from IBLI Mongolia. *American Journal of Agricultural Economics* 100(1): 145-171.
- Blundell, R., Dearden, L., & Sianesi, B. 2005. Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168(3): 473-512.
- Bryant, W. R. 1976. The Qualification of Rented Land for Farm Value Assessment in New York (No. 184816). Cornell University, Department of Applied Economics and Management.
- Brueckner, J. K., & Kim, H. A. 2003. Urban sprawl and the property tax. *International Tax and Public Finance* 10(1): 5-23.
- Caliendo, M., & Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1): 31-72.
- Castle, E. N., Wu, J., & Weber, B. A. 2011. Place orientation and rural-urban interdependence. *Applied Economic Perspectives and Policy* 33(2): 179-204.

Chabé-Ferret, S. 2010. To control or not to control? Bias of simple matching vs difference-in-difference matching in a dynamic framework. In 10th world congress of the Econometric Society, Shanghai, February (pp. 17-21).

Chabé-Ferret, S. 2015. Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. *Journal of Econometrics* 185(1): 110-123.

Chabé-Ferret, S., & Subervie, J. 2013. How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching. *Journal of Environmental Economics and Management* 65(1): 12-27.

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96(1): 187-199.

Darby, K., Batte, M. T., Ernst, S., & Roe, B. 2008. Decomposing local: a conjoint analysis of locally produced foods. *American Journal of Agricultural Economics* 90(2): 476-486.

Dehejia, R. H., & Wahba, S. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94(448): 1053-1062.

Dehejia, R. H., & Wahba, S. 2002. Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics* 84(1): 151-161.

Hardesty, S. D. 2008. The growing role of local food markets. *American Journal of Agricultural Economics* 90(5): 1289-1295.

Heckman, J. J., Ichimura, H., & Todd, P. E. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* 64(4): 605-654.

Heckman, J. J., Ichimura, H., & Todd, P. 1998. Matching as an econometric evaluation estimator. *The Review of Economic Studies* 65(2): 261-294.

Hendricks, N. P., Smith, A., & Sumner, D. A. 2014. Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics* 96(5): 1469-1491.

Huber, M., Lechner, M., & Wunsch, C. 2013. The performance of estimators based on the propensity score. *Journal of Econometrics* 175(1): 1-21.

Irwin, E. G., & Bockstael, N. E. 2007. The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. *Proceedings of the National Academy of Sciences* 104(52): 20672-20677.

Jackson-Smith, D. B., & Jensen, E. 2009. Finding farms: comparing indicators of farming dependence and agricultural importance in the United States. *Rural Sociology* 74(1): 37-55.

Lawley, C., & Towe, C. 2014. Capitalized costs of habitat conservation easements. *American Journal of Agricultural Economics* 96(3): 657-672.

Lichter, D. T., & Brown, D. L. 2014. The new rural-urban interface: Lessons for higher education. *Choices* 29(1): 1-6.

Liu, X., & Lynch, L. 2011. Do agricultural land preservation programs reduce farmland loss? Evidence from a propensity score matching estimator. *Land Economics* 87(2): 183-201.

Livanis, G., Moss, C. B., Breneman, V. E., & Nehring, R. F. 2006. Urban sprawl and farmland prices. *American Journal of Agricultural Economics* 88(4): 915-929.

Nickerson, C., Morehart, M., Kuethe, T., Beckman, J., Ifft, J., & Williams, R. 2012. Trends in US farmland values and ownership.

Polyakov, M., & Zhang, D. 2008. Property tax policy and land-use change. *Land Economics* 84(3): 396-408.

Rubin, D. B. 1973. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*: 185-203.

Rubin, D. B., & Thomas, N. 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* 95(450): 573-585.

Smith, J. A., & Todd, P. E. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125(1): 305-353.

State Board of Equalization and Assessment. 1991. Conversion of New York Farmland Under Agricultural Assessment to Non-Farm Use. Retrieved from [https://www.tax.ny.gov/pdf/publications/orpts/conversion\\_ny\\_farm\\_under\\_agassess\\_nonfarm.pdf](https://www.tax.ny.gov/pdf/publications/orpts/conversion_ny_farm_under_agassess_nonfarm.pdf).

Thilmany, D., Bond, C. A., & Bond, J. K. 2008. Going local: Exploring consumer behavior and motivations for direct food purchases. *American Journal of Agricultural Economics* 90(5): 1303-1309.

Tully, T., Stowe, S. "The Wealthy Flee Coronavirus. Vacation Towns Respond: Stay Away" *New York Times*. March 27, 2020. Page 16. Print.

Watson, D. F., & Philip, G. M. 1985. A refinement of inverse distance weighted interpolation. *Geo-processing* 2(4): 315-327.

Wickham, J. D., Stehman, S. V., Gass, L., Dewitz, J., Fry, J. A., & Wade, T. G. 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment* 130: 294-304.

Wu, J., Weber, B. A., & Partridge, M. D. 2017. Rural-Urban Interdependence: A Framework Integrating Regional, Urban, and Environmental Economic Insights. *American Journal of Agricultural Economics* 99(2): 464-480.

## Tables

Table 1. Summary Statistics for Columbia

Variables	Farmland transferred to Non-residents		Farmland stay owned by residents		T-test/Total
	Mean	SD.	Mean	SD.	P( T > t )
Agricultural ratio before treatment	.1720	.2654	.1468	.2394	.4829
Hay ratio before treatment	.4291	.2829	.4740	.2857	.2896
Forest before treatment	.2742	.2758	.2509	.2502	.5336
Agricultural ratio change	-.0027	.3019	.0913	.2772	.0237**
Hay ratio change	-.1334	.3235	-.2080	.2840	.0805*
Mismatch ratio change	-.0996	.2744	-.1565	.2586	.1406
Forest ratio change	.1478	.1172	.1309	.1346	.3934
Vacant ratio change	.0486	.0639	.0424	.0680	.5349
Wetland ratio change	-.0650	.1063	-.0588	.1180	.7196
Developed ratio change	.0047	.0266	.0022	.0520	.7426
Acreage	82.97	85.16	74.30	81.75	.4766
Ln (Total assessed value)	11.78	1.321	11.61	1.321	.3974
Ln (Assessed land value)	11.52	1.072	11.32	1.139	.2316
Stream or pond in parcel	.4693	.5042	.3941	.4891	.3013
Divided by road	.0204	.1429	.0805	.2722	.1274
Close to primary road	.4694	.5042	.3842	.4868	.2405
Ln (Distance to Hudson city)	9.657	.5705	9.764	.5549	.1958
Ln (Average elevation)	4.723	.5935	4.621	.5985	.2481
Ln (Elevation range)	3.457	.8094	3.128	1.005	.0262**
Annual precipitation	41.87	1.028	41.74	1.466	.5571
Erosion percentage	.4437	.3201	.3702	.2983	.0996*
Excess water percentage	.1922	.1854	.2330	.2433	.2512
Soil limit percentage	.3265	.3402	.3637	.3239	.4403
Prime farmland percentage	.6288	.2837	.6468	.2779	.6631
No. of obs.	49		609		658

*Note:* T-test significance level: \* p<0.1, \*\* p<0.05, \*\*\* p<0.01.

Table 2.DID-matching results: ATT for Columbia County

Outcome: change in land cover proportion from 2001 to 2012				Treatment: transactions from farmers to non-residents <sup>I</sup>			
Estimators <sup>II</sup>	OLS coefficient	Radius match 1 with n=3, caliper=.01	Regression Adjustment with matched sample 1	Radius match 2 with n=5, caliper=.01	Regression Adjustment with matched sample 2	Radius match 3 with n=6, caliper=.01	Regression Adjustment with matched sample 3
ATT on Ag <sup>III</sup> change	-.0896** (.0411 <sup>IV</sup> )	-.1258** (.0637)	-.1208** (.0578)	-.1105** (.0581)	-.1132** (.0570)	-.1141** (.0557)	-.1168** (.0569)
ATT on Hay change	.0724* (.0420)	.1293* (.0668)	.1293** (.0620)	.1090* (.0633)	.1139* (.0609)	.1075* (.0607)	.1124* (.0606)
ATT on mismatch change	.0479 (.0381)	.0842 (.0573)	.0837 (.0529)	.0638 (.0551)	.0709 (.0533)	.0638 (.0518)	.0716 (.0531)
PS-test(B/R)		41.0/.91	41.0/.91	27.5/1.13	27.5/1.13	23.8/1.00	23.8/1.00
N treated: on support(all)	49	46(49)	46(49)	46(49)	46(49)	46(49)	46(49)
N control: matched(all)	609	104(609)	104(609)	155(609)	155(609)	179(609)	179(609)
N total	658	150(658)	150(658)	201(658)	201(658)	225(658)	225(658)

*Notes:* I. Treatment: farmland transacted from residents to non-residents between 2002 and 2010; Control: farmland owned by residents throughout the study period. II. We employ propensity matching with the radius set to .01. To implement the common support condition, we exclude 3 observations in the treatment group who does not satisfy that: each treated has at least 6 nearest neighbors in the control group within the radius .01. We carry out 2 kinds of estimators based on matching with a different number of near neighbors (n=3,5,6): radius matching (RM) and regression adjustment with the matched sample (RA). III. Ag denotes agricultural land cover except Hay. IV. Bootstrapping (2000 replications for RMs) or robust (for RAs) standard errors are in parentheses; \* p<0.1, \*\* p<0.05, \*\*\* p<0.01.

Table 3. DID-matching results: ATT for Columbia County with sample limited by Acreage

Outcome: change in land cover proportion from 2001 to 2012      Treatment: transactions from farmers to non-residents <sup>I</sup>									
	RA with radius matched sample (n=3, r=.01)			RA with radius matched sample (n=5, r=.01)			RA with radius matched sample (n=6, r=.01)		
Less than/equal to (acres) <sup>II</sup>	∞	90	70	∞	90	70	∞	90	70
ATT on Ag <sup>III</sup> change	-.1208** (.0578 <sup>IV</sup> )	-.1523* (.0858)	-.1530* (.0832)	-.1132** (.0570)	-.1482* (.0846)	-.1250 (.0845)	-.1168** (.0569)	-.1371* (.0858)	-.1284 (.0850)
ATT on Hay change	.1293** (.0620)	.1431 (.0894)	.1032 (.0942)	.1139* (.0609)	.1377 (.0894)	.0786 (.0889)	.1124* (.0606)	.1378 (.0911)	.0843 (.0906)
ATT on Mismatch change	.0837 (.0529)	.0568 (.0822)	.0514 (.0816)	.0709 (.0533)	.0581 (.0802)	.0282 (.0775)	.0716 (.0531)	.0591 (.0817)	.0341 (.0793)
PS-test(B/R)	41.0/.91	43.6/.38	46.7/1.04	27.5/1.13	39.8/.51	34.0/.92	23.8/1.00	43.2/.51	40.3/.47
N treated: on support(all)	46(49)	28(32)	29(32)	46(49)	28(32)	29(32)	46(49)	28(32)	29(32)
N control: matched(all)	104(609)	67(401)	64(345)	155(609)	102(401)	98(345)	179(609)	115(401)	114(345)
N total	150(658)	95(433)	93(377)	201(658)	130(433)	127(377)	225(658)	143(433)	143(377)

Notes: I. Treatment: transactions from farmer to non-resident happen between 2002 and 2010; Control: farmer owner throughout the whole period.

II. Extra-large agricultural parcels are less likely to be recreational second homes, so we can see estimation with large parcels (>90 acres) tend to give smaller ATTs. While when we further restrict the sample to 70 acres (29 treated), the increasing variances dominate the changes.

III. Ag denotes agricultural land cover except Hay.

IV. Robust standard errors are in parentheses; \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Table 4. Placebo test results-ATT for Columbia<sup>I</sup>

Outcome: change in land cover proportion from 2001 to 2006		Treatment: transactions from farmers to non-residents during 2006-2010 <sup>II</sup>		
Estimator <sup>II,III</sup>	OLS	Regression	Regression	Regression
	coefficient	Adjustment with matched sample (n=4, caliper=.01)	Adjustment with matched sample (n=5, caliper=.01)	Adjustment with matched sample (n=6, caliper=.01)
ATT on Ag <sup>IV</sup> change	.0169* (.0087 <sup>V</sup> )	.0159 (.0146)	.0145 (.0151)	.0140 (.0151)
ATT on Hay change	-.0195** (.0083)	-.0196 (.0136)	-.0183 (.0140)	-.0178 (.0140)
ATT on mismatch change	-.0142** (.0063)	-.0141 (.0090)	-.0127 (.0094)	-.0127 (.0093)
PS-test(B/R)	-	37.5/.89	30.3/.94	30.5/.85
N treated: on support(all)	47	44(47)	44(47)	44(47)
N control: matched(all)	609	142(609)	167(609)	191(609)
N total	656	186(656)	211(656)	235(656)

Notes: I. We do not claim anticipation effects since we do not get consistently significant placebo test results like Chabé-Ferret and Julie (2013).

II. Control: farmer owner throughout the whole period. We employ propensity matching with caliper set to .01.

III. To implement the common support condition, we exclude 3 observations in the treatment group who does not satisfy that: each treated has at least 6 nearest neighbors in the control group within the radius .01. We carry out regression adjustment with radius matched samples (RA) on different number of near neighbors (n=4,5,6). All matching estimators give us small insignificant trend differences in the opposite direction of the estimated effects. Even though OLS gives significant results, we do not claim anticipation effects considering the selection issue.

IV. Ag denotes agricultural land cover except for Hay.

V. Robust standard errors in parentheses; \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

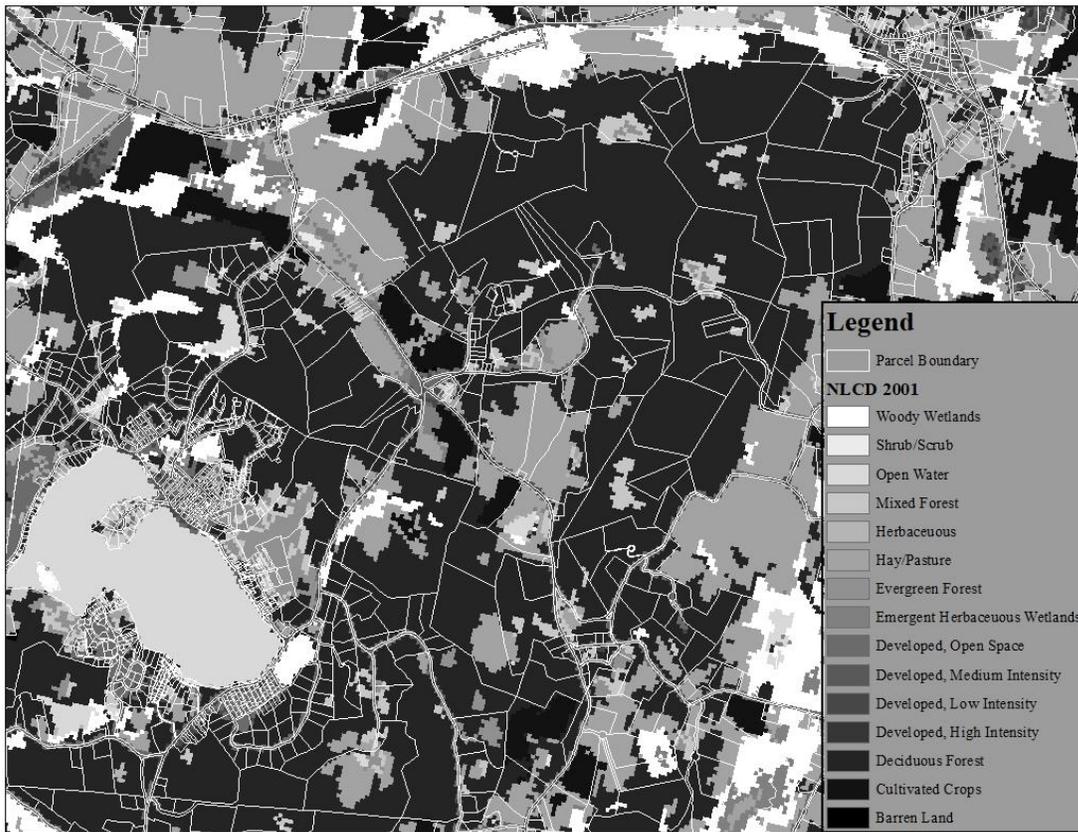
Table 5. Estimation of the total impacts of non-resident owners

Applied counties: top 4% (27 counties) NASS 2002 data as the baseline		2002	2012	Change from 2002 to 2012
Total Acreage underutilized if	15% of the agricultural parcels owned by non-residents, 5% to be owned by non-residents (emerging second-home market, i.e. Columbia)	34,045	45,393	11,348
	65% of the agricultural parcels owned by non-residents (Saturated second- home market, i.e. Ulster)	147,527	<147,527	<0
Total Acreage from NASS	Land in production (harvesting field crops/hay/vegetables)	986,639	916,428	-70,211
	Total land in farms	2269,543	2236,128	-33,415

*Notes:* Outcomes are lower bound estimations of total underutilized acreage land in farms in applicable counties if farmland is owned by non-residents instead of farmers. The applicable counties are selected based on the ranking list generated in Appendix F. The decrease of “total land in farms” is closely related to farmland conversion, which is usually used to explain the decrease in agricultural production. The decrease of “land in production” incorporates both farmland conversion and farmland underutilization.

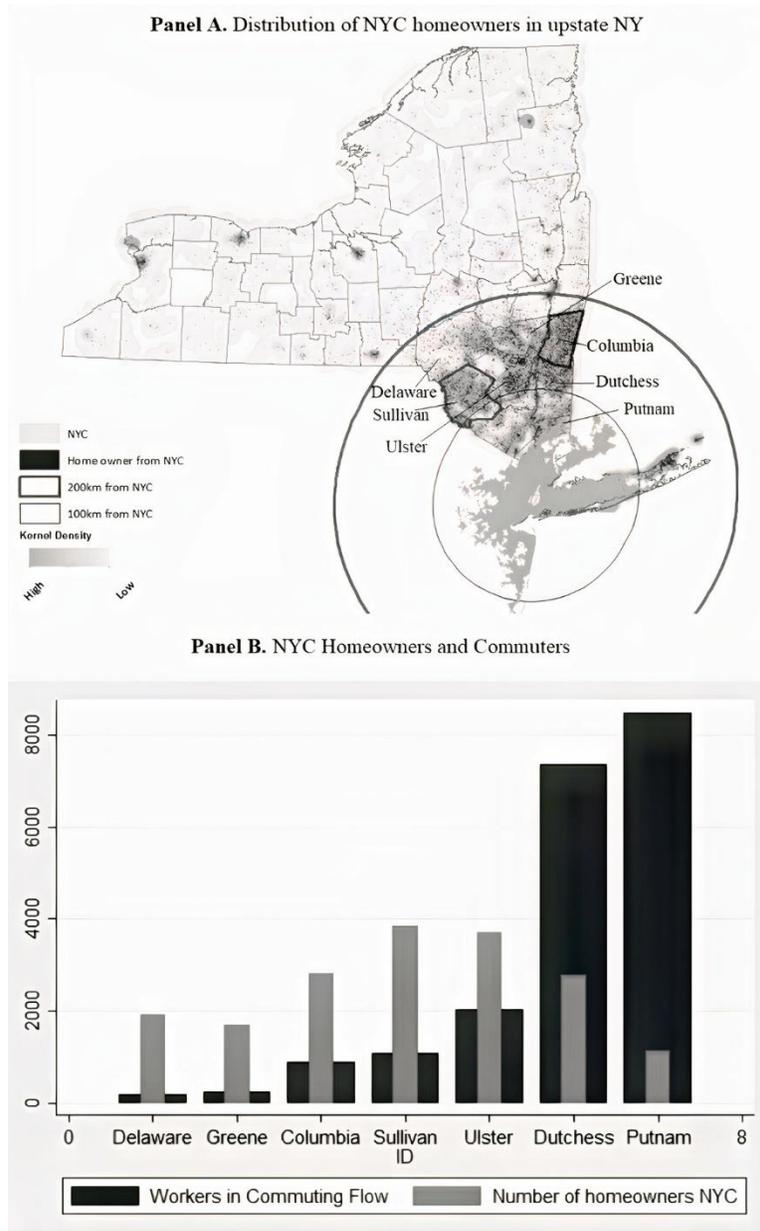
## Figures

Figure 1. An Example of Land Cover and Parcel Layers



*Note:* CDL data are not used as examples, but they have the same resolution as the NLCD data.

Figure 2. The Second-home Shed of New York City



*Note:* Data comes from New York State tax records and the Journey to Work survey.

---

<sup>a</sup> In this paper, non-resident owner is generally a synonym to recreational or second-home owner. A similar concept is “non-operators” in Nickerson et.al (2012).

<sup>b</sup> The COVID pandemic has induced similar migrations from dense urban areas.

<sup>c</sup> See § 301-1, § 301-2, and § 301-4 (page 28) in New York Agricultural District law: <http://www.agriculture.ny.gov/ap/agsservices/25-AA.pdf>. Also, see Bryant (1976).

<sup>d</sup> They primarily use it for low-quality hay in the New York, according to State Board of Equalization and Assessment, 1991.

<sup>e</sup> The data limit a full examination of a plot that is actively managed for hay (with seeding and nutrient application) from a plot that is unmanaged and simply cut, often only once per season, to produce a low-quality feed product or bedding.

<sup>f</sup> We choose a utility-maximizing framework instead of a profit-maximizing one since the non-residents are not necessarily maximizing profit.

<sup>g</sup> This is to say, costs of inputs other than land vary across different farms. More specifically, farms with different conditions need different amounts of other inputs.

<sup>h</sup> Specifically, the 2011 edition of NLCD data. Check <https://www.mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend>.

<sup>i</sup> See <https://nassgeodata.gmu.edu/CropScape/>.

<sup>j</sup> The SSURGO database contains information about soil, which is collected by the National Cooperative Soil Survey. See [http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2\\_053627](http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627).

<sup>k</sup> See <http://cugir.mannlib.cornell.edu>.

<sup>l</sup> See <https://datagateway.nrcs.usda.gov/>.

<sup>m</sup> See <http://gis.ny.gov>.

<sup>n</sup> See <https://www.census.gov/topics/employment/commuting.html>.

<sup>o</sup> This analysis is based on yield data from SSURGO and price data from NASS. Having alternative higher-profit options means the operators have high opportunity costs to plant hay on these farms.

<sup>p</sup> Figure F1 in Appendix F illustrates the pressure imposed by major metropolitan areas on rural land markets.

<sup>q</sup> Since tax bills are important, we assume that it is not the case that landowners have their tax bills sent to addresses other than their primary residences or offices (and their offices are close enough to be commutable from the primary residences). These assumptions will be reconsidered below to identify individual non-resident owners. This figure is based on landowners’ billing addresses for tax documents in 2010 and where the tax documents are sent is obviously based on the billing addresses.

<sup>r</sup> This assumption is confirmed in many articles documenting NYC residents movements during the COVID-19 pandemic, see <https://www.nytimes.com/2020/03/25/nyregion/coronavirus-leaving-nyc-vacation-homes.html>. (Tully and Stowe, 2020)

---

<sup>s</sup> In Jackson-Smith and Jensen (2009), agricultural importance was formally defined and calculated at county level. According to this definition, Columbia County and Sullivan County are agriculturally important.

<sup>t</sup> The analysis for Delaware County cannot be conducted for now due to data limitation.

<sup>u</sup> We confirmed individual parcels use the Agricultural Tax Credit using current county tax rolls.

<sup>v</sup> Also, Wickham et al., 2013 conducts an accuracy assessment of NLCD 2006 data, and suggests that “the high overall agreement rate was driven by the large proportion of area of no change”.

<sup>w</sup> The 5<sup>th</sup> percentile of the sample acreage is about 2.2 acres, which means 95% of the investigated farms include at least 10 cells.

<sup>x</sup> At least, the changes in these attributes would not be big enough to influence the second-home transaction.

<sup>y</sup> We can observe that some parcels with the same parcel ID get quite different acreage across time.

<sup>z</sup> The data are from the Landsat program, which is the longest-running enterprise for acquisition of satellite imagery of Earth.

<sup>aa</sup> Though the econometrics are supposed to deal with random errors, this issue is crucial enough to be one of the reasons not to use CDL 2002 in our main analysis.

<sup>bb</sup> By unselecting parcels with ‘vastly changing acreage’, we mean the absolute value of percentage change in acreage is smaller than or equal to 5%.

<sup>cc</sup> This panel for ownership information ranges from 2001 to 2010. Specifically, we have data for these years: 2001, 2002, 2004, 2006, 2008, 2010, among which the 2001 and 2002 ownership information is imputed based on sales data.

<sup>dd</sup> With the sample of stable parcels, an ideal approach to solve our identification problems would be a fixed effects panel model based on pre-matched sample as in Alix-Garcia, Sims, and Yañez-Pagans (2015), where the pre-matching deals with the selection on observable problem by ensuring ample covariate overlap, and fixed effects mitigate the biases brought by time-invariant unobservables and measurement errors.

<sup>ee</sup> To address the problem of lacking overlap, Crump et al. (2009) developed a systematic approach and suggested a simple rule of discarding propensity scores outside [.1, .9], but this strategy is not applicable in this study since it will delete too many treated observations.

<sup>ff</sup> In a relatively small sample, non-exact matching leads to a potential bias through the leftover discrepancies in the matched sample (Caliendo et.al 2008).

<sup>gg</sup> Other papers using the similar sample size and similar method include: Lawley and Towe (2014), Bertram-Huemmer and Kraehnert (2017).

<sup>hh</sup> The study period lasts from 2002 to 2010 for two reasons: it is consistent with the 8-year cycle of the agricultural district certification, hence would result in a reasonable average estimate over time; it is also consistent with the temporal range of available data, allowing us to get the land cover before the first possible transaction and still have a while after the latest transaction.

<sup>ii</sup> This allows 2 years’ tolerance for the land cover change to take place, while not including too much confounding effect from later unobserved ownership changes.

<sup>jj</sup> We do not employ a simple matching controlling pre-treatment outcome since DID-matching generally performs better according to previous studies (Smith and Todd, 2005; Chabé-Ferret, 2015). We also exclude variables not vital for outcome and not explaining treatment at all, including average minimum temperature, average distance

---

along roads to nearest business districts, square footage of the house, and history of the house. They will increase the variance of estimated propensity scores and potentially cause problems in a sample with limited size (Augurzky and Schmidt, 2001).

<sup>kk</sup> This does not vary when we set the trimming criterion to 3, 4, or 5 neighbors within a radius of 0.01.

<sup>ll</sup> These tests are conducted with the “pctest” command in Stata, which calculates several measures (the t-test, the standardized percentage bias, and the variance ratio) of the balancing of the specified variables.

<sup>mmm</sup> According to NASS definition, a farm is “any place from which \$1,000 or more of agricultural products were produced and sold, or normally would have been sold, during the year.” (See Farms and Land in Farms 2017 Summary by USDA NASS) And land in farms consists of land used for all kinds of operation, consistent with the land cover types investigated in this study.

<sup>nn</sup> Local food is defined as produced within 161 km (100 mile) by “locavores”, the local food movement.

<sup>oo</sup> The agricultural tax credits function as an incentive here for two reasons. First, the reason the properties are in our sample is that they were historically agricultural lands in or before the study period, which means they are qualified for and likely to receive agricultural tax credits. Second, switching to hay but not nonagricultural usage (e.g., natural grassland) suggests that reducing management effort and acquiring natural amenity are not the whole purpose of the land use switch, the precise target is to achieving acceptable levels of these two goals while maintaining low land holding costs. While this second point suggests nonresidents are making rational decisions balancing benefits and costs, the tax credits function essentially as the incentive.

<sup>pp</sup> The Conservation Reserve Program (CRP), providing \$82 payment on average nationally (2020), is perceived to be a competing option in this context. Although NY state CRP rate is typically lower than the national average, it’s reasonable to say that CRP is an alternative program (at least in some cases) to agricultural assessment in lowering land holding costs. However, CRP is generally not the choice in our studied region for a few reasons. 1. The studied properties are primarily agricultural land and do include many CRP properties (possibly some newly enrolled CRP properties). Also, official figures confirm that CRP acres are limited in the investigated counties (in 2014, based on [https://conservation.ewg.org/crp\\_regions.php?fips=36000&regionname=NewYork](https://conservation.ewg.org/crp_regions.php?fips=36000&regionname=NewYork), Columbia 449 acres, Sullivan 41 acres). 2. In our studied region, if a farmer switch from agriculture to CRP, he needs to pay higher property tax, while the CRP payment may (usually do) not offset the discrepancy. 3. It’s fair to say that CRP holds a relatively high bar in approval for enrollment in our studied region - CRP is perceived to target specific lands with erosion issues and mostly in the Midwest. Therefore, in practice, the preferential tax policy is the program of choice in order to lower farmland holding costs for non-residents in our study region.