entry literature (see, e.g., Bresnahan and Reiss 1991), is a strong restriction. In the online Appendix we show that the key qualitative features of demand which we highlight in our descriptive analysis are present for both large and small towns, and in Appendix A we show sensitivity to removing the largest and smallest hinterland towns from the sample.

Advertising Game.—Our model of advertising competition draws heavily on the theoretical literature on competition in two-sided markets with multihoming (Armstrong 2002; Anderson, Foros, and Kind 2011; Ambrus, Calvano, and Reisinger 2013). Allowing for advertising competition is important because advertising accounted for the majority of newspaper revenue during the period we study.

The prediction of diminishing returns to duplicate impressions fits with narrative evidence from the period we study. It was common for advertisers to assess the duplication in readership across publications when considering where to place ads, and to consider duplicate impressions to the same household to be less valuable than unique impressions.[24] Indeed, these practices explain the existence of the readership surveys that we use for a portion of our analysis, which were typically sponsored by one or more local newspapers.

Our advertising model makes several important simplifying assumptions. First, we do not allow the quantity of ads to affect the utility of a newspaper to consumers. In contrast to the literature on broadcast media, for print media the evidence is mixed on consumers' valuations of advertising, with good empirical support for a positive value to readers in some settings (Bogart 1981; Sonnac 2000; Kaiser and Wright 2006; Kaiser and Song 2009). Implicitly, our approach follows Dertouzos and Trautman (1990) in assuming the consumer treats advertising and news content symmetrically. Second, we assume that advertisers' valuations are homogeneous and do not depend on consumer types (as they do, for example, in Chandra 2009).

We impose these restrictions because we do not have reliable cross-sectional data on advertising quantities and rates. The most important consequence of these assumptions is that all advertisers are served in equilibrium in the model, which means that the advertising side of the market is allocatively efficient even under imperfect competition. This is a strong assumption. However, we note that because display advertising rates are often negotiated individually, newspapers have substantial scope for price discrimination in rates, which means that imperfect competition may not lead to quantity restrictions relative to the first best.

We also assume that newspaper costs are independent of the number of ads printed. This is primarily for simplicity. Equilibrium advertising prices and quantities would be unchanged if we allowed for a per-reader cost of printing each advertiser's ad, provided the printing cost per reader is less than a_l.

For ease of exposition, we assume that circulation prices are fixed at the time firms set advertising rates. This is consistent with the fact that subscription prices are typically posted (and so adjust relatively infrequently), whereas advertising rates

[24] In his text on advertising campaigns, Martin (1921, p. 148) writes that "The same advertisement seen in two or three newspapers is certainly more effective than if seen in one, but some advertisers are convinced that it is not worth three times as much to have an advertisement seen in three papers, reaching largely the same readers, as to have it seen in one."

are often negotiated. However, because advertising prices do not affect consumer demand, the equilibria of the game we study are equivalent to those of a game with simultaneous choice of both circulation and advertising prices.

Entry and Affiliation Games.—Our model of entry and affiliation choice is shaped by three main considerations. First, to credibly identify the strategic interactions of firms which share a market environment, it is important to allow for unobservable shocks at both the market and individual firm level (Brock and Durlauf 2007; Aguirregabiria and Nevo 2013). Second, to take advantage of our data, we want the model to reflect the reality that entry in these markets happened sequentially, and that we know the order in which it occurred. Third, as a practical matter, we want to avoid complications related to estimating models with multiple equilibria.

To meet these goals and still maintain tractability, we combine a relatively rich incomplete information model of affiliation choice with a more stylized complete-information model of the entry stage in the style of Bresnahan and Reiss (1991). In the affiliation stage, firms' decisions are based on both idiosyncratic and market-level unobservables. They condition on the choices of past movers, and take account of the way their decisions will affect those who come later. In the entry stage, by contrast, firms do not yet know their idiosyncratic shocks, and so they are all symmetric. Equilibrium is determined by a simple optimality condition, though it still requires numerical integration over the market-level shock as in Mazzeo (2002).

An important departure from Bresnahan and Reiss (1991) is that we allow the distribution of fixed costs to depend on market size. We do this because newspapers' fixed investments, notably editorial costs, are endogenous to the quality of the newspaper and hence to the size of the market served (Berry and Waldfogel 2010). In Section VII we report evidence that our estimates of the fixed costs of newspapers of different size are a good match to the data. In Appendix A we show that our findings are robust to allowing a more flexible dependence of the distribution of fixed costs on population.

Among the many restrictive assumptions our model embeds, two are particularly important. First, we assume that entry decisions precede affiliation decisions, with firms only learning their cost shocks ξ_{jm} post-entry.[25] This is clearly an abstraction, and is at odds with our preferred interpretation of the shocks ξ_{jm} as reflecting owners' political preferences. Substantively, we expect the main costs of this assumption to be that our model does not allow entry deterrence incentives to affect affiliation choices, and that our model does not allow selection of owners into like-minded markets. The benefits are that we simplify computation dramatically, since we need only backward induct through the sequential game for the set of actual entrants rather than the full set of potential entrants, and that our estimates are not sensitive to assumptions about the set of potential entrants. The latter is particularly important in our case, since we have almost no evidence on the nature of this pool or the way it varies across markets.

[25] The assumption that agents learn their private information after entry is common in the literature on auctions (Levin and Smith 1994; Bajari and Hortaçsu 2003). A related assumption is that firms do not know their order in the affiliation choice game at the time they enter. This is purely for technical convenience as it allows us to characterize the equilibrium of the entry game succinctly with equation (9). If we instead assumed that firms chose affiliations in the order in which they entered, conditions on the payoffs of the marginal entrant would no longer be sufficient.

Second, we approximate a dynamic entry process by a static model. Although we capture some aspects of the dynamics by making affiliation choice sequential, we abstract from the reality that one entrant typically operates in the market for a substantial time before the next entrant arrives, that firms do exit, and that different entrants face different demand conditions when choosing affiliations. The most obvious challenge in moving to a dynamic model is the computational difficulty of allowing for both market-level and firm-level unobservables (Aguirregabiria and Nevo 2013).

V. Demand Estimation

We estimate the parameters of equation (2) by maximum likelihood using circulation data from hinterland towns. We assume that measured circulation $\hat{Q}_{jt}$ of newspaper j in town t is equal to $q_{jt} S_t \zeta_{jt}$, where q_{jt} is the share of households purchasing newspaper j, S_t is the number of households in town t, and ζ_{jt} is a measurement error with $\log \zeta_{jt} \sim N\left(0, \sigma_\zeta^2\right)$, i.i.d. across newspapers and towns.

To implement the spatial identification strategy outlined in Section IIIC, we assume that the share ρ_t of consumers in town t with $\theta = R$ is unobserved and may be correlated within the pairs of neighboring towns defined in Section IC. Specifically, we assume that $\rho_t = \text{logit}^{-1}(\text{logit}(Z_t) + \nu_t)$, where Z_t is the observed Republican vote share in t's county and ν_t is a normally distributed unobservable with mean μ_ν^{town} and standard deviation σ_ν^{town}. The logit transformation ensures that $\rho_t \in (0, 1)$. We assume that ν is correlated (and jointly normal) between pairs of neighboring towns t and t', but independent across pairs, with the within-pair correlation restricted to match that of the observable Z:

$$\frac{\text{cov}(\nu_t, \nu_{t'})}{\text{var}(\nu_t)} = \frac{\text{cov}\left(\text{logit}(Z_t), \text{logit}(Z_{t'})\right)}{\text{var}\left(\text{logit}(Z_t)\right)}. \tag{10}$$

The assumption that the spatial correlation of unobservables is equal to that of observables is intermediate between two extremes: perfect correlation, in which case observably equivalent neighboring towns cannot have systematically different circulation patterns, and no correlation, in which case observably equivalent neighboring towns must have orthogonal circulation.

Our model of newspaper entry is appropriate for headquarter markets, but not necessarily for hinterland towns, where newspaper availability was often determined by the decisions of news dealers and other independent newspaper agents. To flexibly account for the endogeneity of the choice set to town ideology ρ_t, we adopt a reduced form in which $\Pr(\tau_{jt} = R) = \text{logit}^{-1}\left(\mu_\rho^0 + \mu_\rho^1 \,\text{logit}(\rho_t)\right)$, where μ_ρ^0 and μ_ρ^1 are parameters to be estimated. In our main estimates, we treat the number of newspapers J_t available in town t as nonstochastic. In Appendix A we show that our results are robust to modeling J_t as a random variable whose distribution depends on ρ_t and the size of the town S_t, and to allowing more flexibility in the dependence of affiliations on ρ_t.

As in the descriptive analysis in Section III, we use as our dependent measure the difference between the mean log circulation of Republican newspapers and the mean log circulation of Democratic newspapers in each town t. We do this to scale out variation in population, which is likely to be poorly measured.

In addition to the dependent measure, the econometrician observes Z_t and the sets $\mathcal{J}_t^R$ and $\mathcal{J}_t^D$ of Republican and Democratic papers available in town t, respectively. Given some true ideology ρ_t, the conditional likelihood of the data for town t is

$$(11)\quad L_t(\rho_t) \;=\; \frac{1}{\tilde{\sigma}_t}\,\phi\left(\frac{1}{\tilde{\sigma}_t|\mathcal{J}_t^R|}\sum_{j\in\mathcal{J}_t^R}\log\left(\frac{\hat{Q}_{jt}}{q_{jt}}\right) - \frac{1}{\tilde{\sigma}_t|\mathcal{J}_t^D|}\sum_{j\in\mathcal{J}_t^D}\log\left(\frac{\hat{Q}_{jt}}{q_{jt}}\right)\right)\Pr(\boldsymbol{\tau}_t|\rho_t, J_t),$$

where $\phi()$ denotes the standard normal PDF and $\tilde{\sigma}_t = \sigma_\zeta\sqrt{1/|\mathcal{J}_t^R| + 1/|\mathcal{J}_t^D|}$. The unconditional log likelihood of the observed data is

$$(12)\qquad \ln L \;=\; \sum_{(t,t')}\ln\int_{\rho_t,\rho_{t'}} L_t(\rho_t)\,L_{t'}(\rho_{t'})\;dF^{town}(\rho_t,\,\rho_{t'}|Z_t,\,Z_{t'}),$$

where $F^{town}()$ is the conditional joint distribution of ρ_t and $\rho_{t'}$ and the sum is taken over all pairs of neighboring towns. For towns that do not have at least one paper of each affiliation, the circulation portion of the likelihood $\phi(\cdots)/\tilde{\sigma}_t$ is unity; these towns contribute to identification only via $\Pr(\boldsymbol{\tau}_t|\rho_t, J_t)$.

We introduce additional data moments to complete identification of our model. Using our cost and revenue data, we calibrate the marginal cost MC and the monopoly advertising revenue per reader a_h to match their sample analogues in monopoly newspaper markets with $Z_m \in [0.45, 0.55]$. For any candidate value of the other parameters of the model, we choose the price coefficient α and the utility shifter $\underline{\beta}$ so that the predicted average price and circulation per household of monopoly newspapers in a market with $\rho = 0.5$ matches the observed average price and circulation per household of monopoly newspapers in markets with $Z_m \in [0.45, 0.55]$.[26] We also choose the substitution parameter Γ_d so that the predicted overlap in readership in a market with equal shares of Republicans and Democrats, one paper of each affiliation, and average prices, matches the average overlap in readership among different-affiliation newspapers in our readership survey data. In Appendix A we present evidence on the sensitivity of our estimates to changes in the empirical moments used in calibration.

We estimate the remaining parameters $\left\{\overline{\beta},\,\Gamma_s,\,\sigma_\zeta,\,\mu_\nu^{town},\,\sigma_\nu^{town},\,\mu_\rho^0,\,\mu_\rho^1\right\}$ by maximizing equation (12).[27]

A. *Identification*

Fixing the affiliations of available newspapers, the correlation shown in Table 3 between the relative demand for Republican newspapers and the observed fraction Republican Z_t identifies $\overline{\beta}$ relative to $\underline{\beta}$. The share of households reading the newspaper then pins down the levels of $\overline{\beta}$ and $\underline{\beta}$. Given these two parameters, observed

[26] Discounts to subscribers mean that circulation revenue per copy may be below posted subscription prices. We compute the average discount as the average ratio of subscription price to annual circulation revenue, and apply this discount to all subscription prices to compute the effective price of each newspaper.

[27] We approximate the integral in the likelihood using sparse grid integration with Gaussian kernel and accuracy 3 (Heiss and Winschel 2008; Skrainka and Judd 2011). In the online Appendix, we present estimates of the model in which we reduce and increase the accuracy by 1.