APPENDIX A. ADDITIONAL TECHNICAL DETAILS

In this appendix we provide additional details on steps used in fitting multiresolution models and deriving basketball metrics from EPV estimates.

A.1 Time-Varying Covariates in Macrotransition Entry Model

As revealed in (8), the hazards $\lambda_j^\ell(t)$ are parameterized by spatial effects (ξ_j^ℓ and $\tilde{\xi}_j^\ell$ for pass events), as well as coefficients for situation covariates, $\boldsymbol{\beta}_j^\ell$. The covariates used may be different for each macrotransition j, but we assume for each macrotransition type the same covariates are used across players ℓ.

Among the covariates we consider, `dribble` is an indicator of whether the ballcarrier has started dribbling after receiving possession. `ndef` is the distance between the ballcarrier and his nearest defender (transformed to $\log(1+d)$). `ball_lastsec` records the distance traveled by the ball in the previous one second. `closeness` is a categorical variable giving the rank of the ballcarrier's teammates' distance to the ballcarrier. Lastly, `open` is a measure of how open a potential pass receiver is using a simple formula relating the positions of the defensive players to the vector connecting the ballcarrier with the potential pass recipient.

For $j \leq 4$, the pass event macrotransitions, we use `dribble`, `ndef`, `closeness`, and `open`. For shot-taking and turnover events, `dribble`, `ndef`, and `ball_lastsec` are included. Lastly, the shot probability model (which, from (10) has the same parameterization as the macrotransition model) uses `dribble` and `ndef` only. All models also include an intercept term. As discussed in Section 4.1, independent CAR priors are assumed for each coefficient in each macrotransition hazard model.

A.2 Player Similarity Matrix $\mathbf{H}$ for CAR Prior

The hierarchical models used for parameters of the macrotransition entry model, discussed in Section 4.1, are based on the idea that players who share similar roles for their respective teams should behave similarly in the situations they face. Indeed, players' positions (point guard, power forward, etc.) encode their offensive responsibilities: point guards move and distribute the ball, small forwards penetrate and attack the basket, and shooting guards get open for three-point shots. Such responsibilities reflect spatiotemporal decision-making tendencies, and therefore informative for our macrotransition entry model (7)–(8).

Rather than use the labeled positions in our data, we define position as a distribution of a player's location during his time on the court. Specifically, we divide the offensive half of the court into 4-square-foot bins (575 total) and count, for each player, the number of data points for which he appears in each bin. Then we stack these counts together into a $L \times 575$ matrix (there are $L = 461$ players in our data), denoted $\mathbf{G}$, and take the square root of all entries in $\mathbf{G}$ for normalization. We then perform non-negative matrix factorization (NMF) on $\mathbf{G}$ in order to obtain a low-dimensional representation of players' court occupancy that still reflects variation across players (Miller et al. 2013). Specifically, this involves solving:

$$\hat{\mathbf{G}} = \underset{\mathbf{G}^*}{\operatorname{argmin}}\{D(\mathbf{G}, \mathbf{G}^*)\}, \text{ subject to } \mathbf{G}^* = \left(\underset{L\times r}{\mathbf{U}}\right)\left(\underset{r\times 575}{\mathbf{V}}\right) \text{ and } U_{ij}, V_{ij} \geq 0 \text{ for all } i, j, \tag{A.1}$$

where r is the rank of the approximation $\hat{\mathbf{G}}$ to $\mathbf{G}$ (we use $r = 5$), and D is some distance function; we use a Kullback-Liebler type

$$D(\mathbf{G}, \mathbf{G}^*) = \sum_{i,j} G_{ij} \log\left(G_{ij}/G^*_{ij}\right) - G_{ij} + G^*_{ij}.$$

The rows of $\mathbf{V}$ are non-negative basis vectors for players' court occupancy distributions (plotted in Figure 8) and the rows of $\mathbf{U}$ give the loadings for each player. With this factorization, $\mathbf{U}_i$ (the ith row of $\mathbf{U}$) provides player i's "position"—a r-dimensional summary of where he spends his time on the court. Moreover, the smaller the difference between two players' positions, $||\mathbf{U}_i - \mathbf{U}_j||$, the more alike are their roles on their respective teams, and the more similar we expect the parameters of their macrotransition models to be a priori.

Formalizing this, we take the $L \times L$ matrix $\mathbf{H}$ to consist of 0s, then set $H_{ij} = 1$ if player j is one of the eight closest players in our data to player i using the distance $||\mathbf{U}_i - \mathbf{U}_j||$ (the cutoff of choosing the closest eight players is arbitrary). This construction of $\mathbf{H}$ does not guarantee symmetry, which is required for the CAR prior we use, thus we set $H_{ji} = 1$ if $H_{ij} = 1$. For

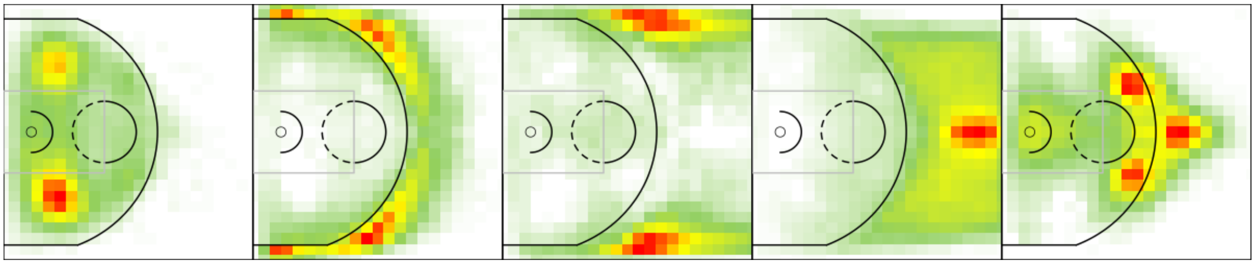

Figure 8: The rows of $\mathbf{V}$ *(plotted above for* $r = 5$*) are bases for the players' court occupancy distribution. There is no interpretation to the ordering.*

instance, LeBron James' "neighbors" are (in no order): Andre Iguodala, Harrison Barnes, Paul George, Kobe Bryant, Evan Turner, Carmelo Anthony, Rodney Stuckey, Will Barton, and Rudy Gay.

A.3 Basis Functions for Spatial Effects ξ

Recalling (13), for each player ℓ and macrotransition type j, we have $\xi_j^\ell(\mathbf{z}) = \sum_{i=1}^d w_{ji}^\ell \phi_{ji}(\mathbf{z})$, where $\{\phi_{ji}, i = 1, \ldots, d\}$ are the basis functions for macrotransition j. During the inference discussed in Section 4, these basis functions are assumed known. They are derived from a pre-processing step. Heuristically, they are constructed by approximately fitting a simplified macrotransition entry model with stationary spatial effect for each player, then performing NMF to find a low-dimensional subspace (in this function space of spatial effects) that accurately captures the spatial dependence of players' macrotransition behavior. We now describe this process in greater detail.

Each basis function ϕ_{ji} is itself represented as a linear combination of basis functions,

$$\phi_{ji}(\mathbf{z}) = \sum_{k=1}^{d_0} v_{jik}\psi_k(\mathbf{z}), \tag{A.2}$$

where $\{\psi_k, k = 1, \ldots, d_0\}$ are basis functions (as the notation suggests, the same basis is used for all j, i). The basis functions $\{\psi_k, k = 1, \ldots, d_0\}$ are induced by a triangular mesh of d_0 vertices (we use $d_0 = 383$) on the court space $\mathbb{S}$. In practice, the triangulation is defined on a larger region that includes $\mathbb{S}$, due to boundary effects. The mesh is formed by partitioning $\mathbb{S}$ into triangles, where any two triangles share at most one edge or corner; see Figure 9 for an illustration. With some arbitrary ordering of the vertices of this mesh, $\psi_k : \mathbb{S} \rightarrow \mathbb{R}$ is the unique function taking value 0 at all vertices $\tilde{k} \neq k$, 1 at vertex k, and linearly interpolating between any two points within the same triangle used in the mesh construction. Thus, with this basis, ϕ_{ji} (and consequently, ξ_j^ℓ) are piecewise linear within the triangles of the mesh.

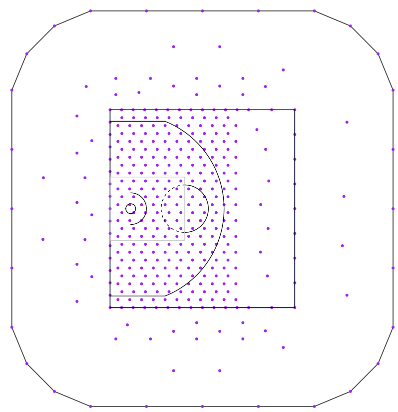

Figure 9: Triangulation of $\mathbb{S}$ *used to build the functional basis* $\{\psi_k, k = 1, \ldots, d_0\}$*. Here,* $d_0 = 383$.

This functional basis $\{\psi_k, k = 1, \ldots, d_0\}$ is used by Lindgren et al. (2011), who show that it can approximate a Gaussian random field with Matérn covariance. Specifically, let $x(\mathbf{z}) = \sum_{k=1}^{d_0} v_k \psi_k(\mathbf{z})$ and assume $(v_1 \ \ldots \ v_k)' = \mathbf{v} \sim \mathcal{N}(0, \mathbf{\Sigma}_{\nu,\kappa,\sigma^2})$. The form of $\mathbf{\Sigma}_{\nu,\kappa,\sigma^2}$ is such that the covariance function of x approximates a Matérn covariance:

$$\text{Cov}[x(\mathbf{z}_1), x(\mathbf{z}_2)] = \boldsymbol{\psi}(\mathbf{z}_1)' \mathbf{\Sigma}_{\nu,\kappa,\sigma^2} \boldsymbol{\psi}(\mathbf{z}_2) \approx \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} (\kappa ||\mathbf{z}_1 - \mathbf{z}_2||)^\nu K_\nu(\kappa ||\mathbf{z}_1 - \mathbf{z}_2||), \quad \text{(A.3)}$$

where $\boldsymbol{\psi}(\mathbf{z}) = (\psi_1(\mathbf{z}) \ \ldots \ \psi_{d_0}(\mathbf{z}))'$. As discussed in Section 4.2, the functional basis representation of a Gaussian process offers computational advantages in that the infinite dimensional field x is given a d_0-dimensional representation, as x is completely determined by $\mathbf{v}$. Furthermore, as discussed in Lindgren et al. (2011), $\mathbf{\Sigma}^{-1}_{\nu,\kappa,\sigma^2}$ is sparse ((A.3) is actually a Gaussian Markov random field (GMRF) approximation to x), offering additional computational savings (Rue 2001).

The GMRF approximation given by (A.2)–(A.3) is actually used in fitting the microtransition models for offensive players (5). We give the spatial innovation terms μ^ℓ_x, μ^ℓ_y representations using the ψ basis. Then, as mentioned in Section 4.3, (5) is fit independently for each player in our data set using the software R-INLA.

We also fit simplified versions of the macrotransition entry model, using the ψ basis, in order to determine $\{v_{jik}, k = 1, \ldots, d_0\}$, the loadings of the basis representation for ϕ, (A.2). This simplified model replaces the macrotransition hazards (8) with

$$\log(\lambda_j^\ell(t)) = c_j^\ell + \sum_{k=1}^{d_0} u_{jk}^\ell \psi_k(\mathbf{z}^\ell(t)) + \mathbf{1}[j \leq 4] \sum_{k=1}^{d_0} \tilde{u}_{jk}^\ell \psi_k(\mathbf{z}_j(t)), \tag{A.4}$$

thus omitting situational covariates ($\boldsymbol{\beta}_j^\ell$ in (8)) and using the ψ basis representation in place of ξ_j^ℓ. Note that for pass events, like (8), we have an additional term based on the pass recipient's location, parameterized by $\{\tilde{u}_{jk}^\ell, k = 1, \ldots, d_0\}$. As discussed in Section 4.3, parameters in (A.4) can be estimated by running a Poisson regression. We perform this independently for all players ℓ and macrotransition types j using the R-INLA software. Like the microtransition model, we fit (A.4) separately for each player across $L = 461$ processors (each hazard type j is run in serial), each requiring at most 32GB RAM and taking no more than 16 hours.

For each macrotransition type j, point estimates $\hat{u}_{jk}^\ell$ are exponentiated[6], so that $[\mathbf{U}_j]_{\ell k} = \exp(\hat{u}_{jk}^\ell)$. We then perform NMF (A.1) on $\mathbf{U}_j$:

$$\mathbf{U}_j \approx \left(\underset{L \times d}{\mathbf{Q}_j}\right)\left(\underset{d \times d_0}{\mathbf{V}_j}\right). \tag{A.5}$$

Following the NMF example in Section A.2, the rows of $\mathbf{V}_j$ are bases for the variation in coefficients $\{u_{jk}^\ell, k = 1, \ldots, d_0\}$ across players ℓ. As $1 \leq k \leq d_0$ indexes points on our court triangulation (Figure 9), such bases reflect structured variation across space. We furthermore use these terms as the coefficients for (A.2), the functional basis representation of ϕ_{ji}, setting

[6] The reason for exponentiation is because estimates $\hat{u}_{jk}^\ell$ inform the log hazard, so exponentiation converts these estimates to a more natural scale of interest. Strong negative signals among the $\hat{u}_{jk}^\ell$ will move to 0 in the entries of $\mathbf{U}_j$ and not be very influential in the matrix factorization (A.5), which is desirable for our purposes.

$v_{jik} = [\mathbf{V}_j]_{ik}$. Equivalently, we can summarize our spatial basis model as:

$$\xi_j^\ell(\mathbf{z}) = [\mathbf{w}_j^\ell]'\boldsymbol{\phi}_j(\mathbf{z}) = [\mathbf{w}_j^\ell]'\mathbf{V}_j\boldsymbol{\psi}(\mathbf{z}). \tag{A.6}$$

The preprocessing steps described in this section—fitting a simplified macrotransition entry model (A.4) and performing NMF on the coefficient estimates (A.5)—provide us with basis functions $\phi_{ji}(\mathbf{z})$ that we treat as fixed and known during the modeling and inference discussed in Section 4.

Note that an analogous expression for (A.6) is used for $\tilde{\xi}_j^\ell$ in terms of $\tilde{\mathbf{w}}_j^\ell$ and $\tilde{\mathbf{V}}_j$ for pass events; however, for the spatial effect ξ_{s}^ℓ in the shot probability model, we simply use $\mathbf{V}_5$. Thus, the basis functions for the shot probability model are the same as those for the shot-taking hazard model.

A.4 Calculating EPVA: Baseline EPV for League-Average Player

To calculate the baseline EPV for a league-average player possessing the ball in player ℓ's shoes, denoted $\nu_t^{r(\ell)}$ in (19), we start by considering an alternate version of the transition probability matrix between coarsened states $\mathbf{P}$. For each player $\ell_1, \ldots, \ell_5$ on offense, there is a disjoint subset of rows of $\mathbf{P}$, denoted $\mathbf{P}_{\ell_i}$, that correspond to possession states for player ℓ_i. Each row of $\mathbf{P}_{\ell_i}$ is a probability distribution over transitions in $\mathcal{C}$ given possession in a particular state. Technically, since states in $\mathcal{C}_{\text{poss}}$ encode player identities, players on different teams do not share all states which they have a nonzero probability of transitioning to individually. To get around this, we remove the columns from each $\mathbf{P}_{\ell_i}$ corresponding to passes to players not on player ℓ_i's team, and reorder the remaining columns according to the position (guard, center, etc.) of the associated pass recipient. Thus, the interpretation of transition distributions $\mathbf{P}_{\ell_i}$ across players ℓ_i is as consistent as possible.

We create a baseline transition profile of a hypothetical league-average player by averaging these transition probabilities across all players: (with slight abuse of notation) let $\mathbf{P}_r = \sum_{\ell=1}^{L} \mathbf{P}_\ell / L$. Using this, we create a new transition probability matrix $\mathbf{P}_r(\ell)$ by replacing player ℓ's transition probabilities ($\mathbf{P}_\ell$) with the league-average player's ($\mathbf{P}_r$). The baseline (league-average) EPV at time t is then found by evaluating $\nu_t^{r(\ell)} = \mathbb{E}_{\mathbf{P}_{r(\ell)}}[X|C_t]$. Note that $\nu_t^{r(\ell)}$ depends only on the coarsened state C_t at time t, rather than the full history of the possession, $\mathcal{F}_t^{(Z)}$, as in ν_t (4). This "coarsened" baseline $\nu_t^{r(\ell)}$ exploits the fact that, when averaging possessions over the entire season, the results are (in expectation) identical to using a full-resolution baseline EPV that assumes the corresponding multiresolution transition probability models for this hypothetical league-average player.