Modeling Heterogeneous Risk-Taking Behavior in Route Choice: A Stochastic Dominance Approach

Xing Wu and Marco Nie

Department of Civil and Environmental Engineering
Northwestern University

Transportation Center of Northwestern University
April 14, 2011
Outline

1. Research background
2. Theory of Stochastic Dominance (SD)
3. Relationship with existing models
4. Finding SD-admissible paths
5. Optimal Path Problems with SD Constraints
6. Reliability-based Traffic Assignment Problem
7. Future Studies
Motivation

When should you leave home, and which route should you take, if you need to drive to an important appointment, such as catching a flight or a job interview?
Route travel time is random

(a) Interstate 94/90 from Chicago (Ohio St.) to Ohare International Airport (source: Google Map)

(b) Travel Time Distribution for that corridor during morning rush hour (6-10 AM)

Travel times vary from as low as about 15 minutes to as long as 80 minutes in the morning peak period (6 - 10 AM).

If a traveler wishes to capture the flight on time with a 90% chance, 48 minutes have to be budgeted for travel, over 50% more than the mean travel time (31 minutes).
Route travel time is random

(a) Interstate 94/90 from Chicago (Ohio St.) to O hare International Airport (source: Google Map)

(b) Travel Time Distribution for that corridor during morning rush hour (6-10 AM)

Route choice is similar to the choice of stock!

<table>
<thead>
<tr>
<th>Performance</th>
<th>Route</th>
<th>Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route time</td>
<td>Price</td>
<td></td>
</tr>
<tr>
<td>Personal</td>
<td>Money</td>
<td></td>
</tr>
<tr>
<td>Make best</td>
<td>Make</td>
<td></td>
</tr>
<tr>
<td>use of time</td>
<td>maximize profit</td>
<td></td>
</tr>
</tbody>
</table>

Travel times vary from as low as about 15 minutes to as long as 80 minutes in the morning peak period (6 - 10 AM).

If a traveler wishes to capture the flight on time with a 90% chance, 48 minutes have to be budgeted for travel, over 50% more than the mean travel time (31 minutes).
What we mean by "making the best use of time"?

- Minimize the time budget to arrive on-time or earlier with a desired probability - determined by the perceived importance of a trip.
- Extra time one is willing to reserve to ensure the desired probability - determined by his/her risk-taking preference
  - Risk aversion
  - Ruin aversion
Background

What we mean by "making the best use of time"?

- Minimize the time budget to arrive on-time or earlier with a desired probability - determined by the perceived importance of a trip.
- Extra time one is willing to reserve to ensure the desired probability - determined by his/her risk-taking preference
  - Risk aversion
  - Ruin aversion

Basic assumptions

- Travel utility is a decreasing function of travel time
- Travelers aim to maximize the expected utility
Two classes of problems

- Routing problems: Finding the “optimal” routes
- Assignment problems: Predicting “equilibria” of traffic flows

A stochastic dominance (SD) approach

- SD is widely used in finance and economics to rank random variables;
- It captures the commonality in risk-taking preferences of decision-makers;
Stochastic Dominance (SD) Theory

**Basic idea**

The SD theory ranks random variables according to various orders of integration of their probability density function.

**Definition (First-order stochastic dominance (FSD) \( \succ_1 \))**

Let \( F_X \) and \( F_Y \) be the CDFs of random variables \( X \) and \( Y \). \( X \succ_1 Y \) if \( F_X(t) \geq F_Y(t) \) and \( F_X(t) > F_Y(t) \) for some \( t \).
Stochastic Dominance (SD) Theory

**Basic idea**

The SD theory ranks random variables according to various orders of integration of their probability density function.

**Definition (First-order stochastic dominance (FSD) $\succ_1$)**

Let $F_X$ and $F_Y$ be the CDFs of random variables $X$ and $Y$. $X \succ_1 Y$ if $F_X(t) \geq F_Y(t)$ and $F_X(t) > F_Y(t)$ for some $t$.

Cumulative probability

Travel time

$Y$ is dominated by $X$ in the first order.
Definition (Second order stochastic dominance (SSD) \( \succ_2 \))

\[ X \succ_2 Y \text{ if } \int_t^T F_X(w)dw \geq \int_t^T F_Y(w)dw, \forall t \text{ and } \int_t^T F_X(w)dw > \int_t^T F_Y(w)dw \text{ for some } t \]
Definition (Second order stochastic dominance (SSD) $\succ_2$)

$X \succ_2 Y$ if

$$\int_t^T F_X(w)dw \geq \int_t^T F_Y(w)dw, \forall t$$

and

$$\int_t^T F_X(w)dw > \int_t^T F_Y(w)dw$$

for some $t$

**Definition of SSD**

Area under the curve

$$\text{for any } t<T$$
SD Theory

Definition (Second order stochastic dominance (SSD) $\succeq_2$)

$X \succeq_2 Y$ if

$$\int_t^T F_X(w)dw \geq \int_t^T F_Y(w)dw, \forall t \text{ and }$$

$$\int_t^T F_X(w)dw > \int_t^T F_Y(w)dw \text{ for some } t$$

Definition (Third order stochastic dominance (SSD) $\succeq_3$)

$X \succeq_3 Y$ if

$$\int_t^T \int_{\tau}^T F_X(w)dw d\tau \geq \int_t^T \int_{\tau}^T F_Y(w)dw d\tau, \forall t \leq T$$

and

$$\int_t^T \int_{\tau}^T F_X(w)dw d\tau > \int_t^T \int_{\tau}^T F_Y(w)dw d\tau \text{ for some } t$$
Theorem (SD and expected utility)

A random variable $X$ dominates another random variable $Y$

1. in the first order, i.e, $X \succ_1 Y$, if and only if
   \[ E[U(X)] > E[U(Y)] \]
   for any $U$ such that $U' < 0$;

2. in the second order, i.e, $X \succ_2 Y$, if and only if
   \[ E[U(X)] > E[U(Y)] \]
   for any $U$ such that $U' < 0$, $U'' < 0$; and

3. in the third order, i.e, $X \succ_3 Y$, if and only if
   \[ E[U(X)] > E[U(Y)] \]
   for any $U$ such that
   $U' < 0$, $U'' < 0$, $U''' < 0$.

where $U(X)$ is the utility function of $X$.

A decision maker is insatiable if his/her utility function is a monotone function of the random variable.

- In route choice, $U' < 0$ means less travel time is always preferred (i.e., traveler is insatiable wrt travel time).

- If $X \succ_1 Y$, any insatiable decision maker would prefer $X$ to $Y$. 
SD and risk-taking preference

SSD and risk aversion

A decision maker is considered “risk-averse” if he/she always prefers the expectation of a random variable, i.e., $E[X]$, to $X$ itself.

Friedman and Savage (1948)

- According to Jensens’ inequality, the utility function $U(\cdot)$ satisfies the above condition if and only if it is concave, i.e., $U'' < 0$.

- $E[U(X)] > E[U(Y)]$ for any $U$ s.t. $U' < 0$, $U'' < 0$ \iff $X \succ_2 Y$.

- If $X \succ_2 Y$, any risk-averse decision maker would prefer $X$ to $Y$. 
SD and risk-taking preference

SSD and risk aversion

Under FSD, Path 2 is preferred when on-time arrival probability is within this range. The shaded area $A >$ the shaded area $B$, which implies that Path 1 dominates path 2 in the second order. That is, All risk-averse travelers would prefer path 1 to path 2.
SD and risk-taking preference

**TSD and ruin aversion**

Ruin-averse decision makers are “unwilling to accept a small, almost certain gain in exchange for remote possibilities of ruin” (Heyer 2001, Ullrich 2009)
TSD and ruin aversion

Ruin-averse decision makers are “unwilling to accept a small, almost certain gain in exchange for remote possibilities of ruin” (Heyer 2001, Ullrich 2009)

![Diagram showing probability density over travel time with two paths: path A and path B. Path A has a mean travel time of t₁ < t₂, and max travel time of t₃ << t₄. Path B is dashed and shows a higher peak at t₂ with rapid decline.]
SD and risk-taking preference

TSD and ruin aversion

Ruin-averse decision makers are “unwilling to accept a small, almost certain gain in exchange for remote possibilities of ruin” (Heyer, 2001; Ullrich, 2009)

- In route choice, ruin aversion is related to negative skewness
- \( E[U(X)] > E[U(Y)] \) for any \( U \) s.t. \( U' < 0, U'' < 0 \) and \( U''' < 0 \) \( \iff \) \( X \succ_3 Y \)
- If \( X \succ_3 Y \), any ruin-averse decision maker would prefer \( X \) to \( Y \)
Admissible Paths under SD

Notation

- $c_{ij}$ – travel time on link $ij$
- $p_{ij}(\cdot)$ – PDF of $c_{ij}$
- $k^{rs}$ – path $k$ from node $r$ to node $s$
- $\pi^{rs}_k$ – travel time on path $k^{rs}$
- $u^{rs}_k(\cdot)$ – CDF of $\pi^{rs}_k$
- $v^{rs}_k(\cdot)$ – inverse CDF of $\pi^{rs}_k$
- $K^{rs}$ – set of all paths from node $r$ to node $s$
Admissible Paths under SD

**FSD/SSD/TSD-admissible paths**

A path $k^{rs}$ is an FSD/SSD/TSD-admissible path if and only if no such a path $l^{rs} \in K^{rs}$ exists that $\pi_l^{rs} \succ_1 / \succ_2 / \succ_3 \pi_k^{rs}$.
A path $k^{rs}$ is an FSD/SSD/TSD-admissible path if and only if no such a path $l^{rs} \in K^{rs}$ exists that $\pi_l^{rs} \succ_1 \succ_2 \succ_3 \pi_k^{rs}$.

Path 1 is FSD-admissible
Path 2 is not. It is dominated by 1
Path 1 forms the pareto frontier

Both Path 1 and 2 are admissible
They together form the pareto frontier

All three paths are FSD-admissible
Path 3 does not contribute to the frontier, but it is not dominated by either 1 or 2.
Admissible Paths under SD

- The optimal path for any insatiable/risk-averse/ruin-averse traveler must be FSD/SSD/TSD-admissible.

- An FSD/SSD/TSD-admissible path may not be optimal for any insatiable/risk-averse/ruin-averse traveler.

- TSD-admissible path set $\subseteq$ SSD-admissible path set $\subseteq$ FSD-admissible path set
On-time arrival probability (OTP)/Percentile travel time (PTT)

Frank (1969), Mirchandani (1976), Miller-Hooks and Mahmassani (2003), Fan et al. (2005), Nie and Wu (2009)

\[
\min \left\{ v_{k}^{rs}(\alpha), \forall k^{rs} \in K^{rs} \right\}, \max \left\{ u_{k}^{rs}(b), \forall k^{rs} \in K^{rs} \right\}
\]

- An optimal path with respect to PTT or OTP must be FSD-admissible, but not vice versa.
On-time arrival probability (OTP)/Percentile travel time (PTT)

Frank (1969), Mirchandani (1976), Miller-Hooks and Mahmassani (2003), Fan et al. (2005), Nie and Wu (2009)

\[
\min \left\{ v_{rs}^k(\alpha), \forall k^rs \in K^{rs}\right\}, \max \left\{ u_{rs}^k(b), \forall k^rs \in K^{rs}\right\}
\]

- An optimal path with respect to PTT or OTP must be FSD-admissible, but not vice versa.
Effective travel time/travel time budget (TTB)

Hall (1983), (Sivakumar and Batta, 1994; Sen et al., 2001), (Uchida and Iida, 1993; Lo et al., 2006; Shao, Meng and Tam, 2006)

\[
\text{min } \left\{ \mathbb{E}(\pi_{rs}^k) + \lambda \sqrt{\text{Var}(\pi_{rs}^k)}, \forall k^{rs} \in K^{rs} \right\}
\]

- TTB is closely related to PTT in some special case, e.g., normal distribution. In this case,
  1. \( \lambda \) always corresponds to an on-time arrival probability \( \alpha \)
  2. \( \lambda > 0 \rightarrow \alpha > 0.5, \lambda = 0 \rightarrow \alpha = 0.5, \text{ and } \lambda < 0 \rightarrow \alpha < 0.5. \)

- In the general case, there is lack of one-to-one correspondence between \( \alpha \) and \( \lambda \).
Effective travel time/travel time budget (TTB)

Hall (1983), (Sivakumar and Batta, 1994; Sen et al., 2001), (Uchida and Iida, 1993; Lo et al., 2006; Shao, Meng and Tam, 2006)

\[
\min \left\{ E(\pi_k^{rs}) + \lambda \sqrt{\text{Var}(\pi_k^{rs})}, \forall k^{rs} \in K^{rs} \right\}
\]

- Minimum TTB may not be found from FSD-admissible paths
- An example:
  1. Random travel times \( X, Y \) and their distributions \( F_X(t) = t \) and \( F_Y(t) = t^2, t \in [0, 1] \)
Effective travel time/travel time budget (TTB)

Hall (1983), (Sivakumar and Batta, 1994; Sen et al., 2001), (Uchida and Iida, 1993; Lo et al., 2006; Shao, Meng and Tam, 2006)

$$\min \left\{ \mathbb{E}(\pi_k^{rs}) + \lambda \sqrt{\text{Var}(\pi_k^{rs})}, \forall k^{rs} \in K^{rs} \right\}$$

- Minimum TTB may not be found from FSD-admissible paths
- An example:
  1. Random travel times $X$, $Y$ and their distributions $F_X(t) = t$ and $F_Y(t) = t^2$, $t \in [0, 1]$
  2. $X \succ_1 Y$
Effective travel time/travel time budget (TTB)

Hall (1983), (Sivakumar and Batta, 1994; Sen et al., 2001), (Uchida and Iida, 1993; Lo et al., 2006; Shao, Meng and Tam, 2006)

\[
\min \left\{ \mathbb{E}(\pi_{k}^{rs}) + \lambda \sqrt{\text{Var}(\pi_{k}^{rs})}, \forall k^{rs} \in K^{rs} \right\}
\]

- Minimum TTB may not be found from FSD-admissible paths
- An example:
  1. Random travel times $X$, $Y$ and their distributions $F_X(t) = t$ and $F_Y(t) = t^2$, $t \in [0, 1]$
  2. $X \succ_1 Y$
  3. However, $Y$ may have a smaller TTB when $\lambda > 6$
Mean excess travel time (METT)

Chen and Zhou (2010). Also known as conditional value-at-risk (CVaR) or tail value-at-risk (Tail VaR)

\[
\min \left\{ v_{k}^{rs}(\alpha) + \frac{1}{1-\alpha} \int_{v_{k}^{rs}(\alpha)}^{T} [x - v_{k}^{rs}(\alpha)] f_{k}^{rs}(x) \, dx , \forall k^{rs} \in K^{rs} \right\}
\]
Mean excess travel time (METT)

Chen and Zhou (2010). Also known as conditional value-at-risk (CVaR) or tail value-at-risk (Tail VaR)

\[
\min \left\{ v^{rs}_k(\alpha) + \frac{1}{1-\alpha} \int_{v^{rs}_k(\alpha)}^{T} [x - v^{rs}_k(\alpha)] f^{rs}_k(x) \, dx, \forall k^{rs} \in K^{rs} \right\}
\]

**Property**

For a given probability \( \alpha \), there always exists an FSD-admissible path which gives the minimum METT.
Mean excess travel time (METT)

Chen and Zhou (2010). Also known as conditional value-at-risk (CVaR) or tail value-at-risk (Tail VaR)

\[
\min \left\{ v_{rs}^k(\alpha) + \frac{1}{1-\alpha} \int_{V_{rs}^k(\alpha)}^T [x - v_{rs}^k(\alpha)] f_{k}^{rs}(x) dx, \forall k^{rs} \in K^{rs} \right\}
\]

**Property**

- For a given probability \( \alpha \), there always exists an FSD-admissible path which gives the minimum METT.

- Hence, a minimum METT path can be identified by evaluating METT of all FSD-admissible paths.
Maximize expected utility with special functional forms


\[ U_{QUF}(x) = \beta_0 + \beta_1 x + \beta_2 x^2, \quad \beta_1 < 0, \beta_2 < 0 \]
\[ U_{LUF}(x) = \beta_0 + \beta_1 x, \quad \beta_1 < 0 \]
\[ U_{EUF}(x) = \beta_1 \exp(\beta_2 x + \beta_0), \quad \beta_1 < 0, \beta_2 > 0 \]

Relation with SD-admissible paths

- A traveler with any QUF is risk-averse but not ruin-averse, and will always select SSD-admissible paths;
Maximize expected utility with special functional forms


\[
U_{\text{QUF}}(x) = \beta_0 + \beta_1 x + \beta_2 x^2, \quad \beta_1 < 0, \quad \beta_2 < 0
\]
\[
U_{\text{LUF}}(x) = \beta_0 + \beta_1 x, \quad \beta_1 < 0
\]
\[
U_{\text{EUF}}(x) = \beta_1 \exp(\beta_2 x + \beta_0), \quad \beta_1 < 0, \quad \beta_2 > 0
\]

Relation with SD-admissible paths

- A traveler with any QUF is risk-averse but not ruin-averse, and will always select SSD-admissible paths;
- A traveler with any LUF will always select the path with least expected travel time;
Maximize expected utility with special functional forms


\[ U_{\text{QUF}}(x) = \beta_0 + \beta_1 x + \beta_2 x^2, \quad \beta_1 < 0, \ \beta_2 < 0 \]
\[ U_{\text{LUF}}(x) = \beta_0 + \beta_1 x, \quad \beta_1 < 0 \]
\[ U_{\text{EUF}}(x) = \beta_1 \exp(\beta_2 x + \beta_0), \quad \beta_1 < 0, \ \beta_2 > 0 \]

Relation with SD-admissible paths

- A traveler with any QUF is risk-averse but not ruin-averse, and will always select SSD-admissible paths;
- A traveler with any LUF will always select the path with least expected travel time;
- A traveler with any EUF is ruin-averse, and will always select TSD-admissible paths.
Mean-variance rule


A path is non-dominated under the mean-variance rule if no such a path exists whose mean and variance are both smaller, the set of non-dominated paths denoted as $\Gamma_{MV}$.

- We have to enumerate all paths to identify $\Gamma_{MV}$

- If a traveler’s utility function is an QUF, the path with maximum expected utility obtained from $\Gamma_{MV}$ must be SSD-admissible.

- Generally, the SSD-admissible path set is different from $\Gamma_{MV}$. 
Mean-variance rule


A path is non-dominated under the mean-variance rule if no such a path exists whose mean and variance are both smaller, the set of non-dominated paths denoted as $\Gamma_{MV}$.

Example

1. $X$, $Y$ and $F_X(t) = t$, $F_Y(t) = t^2$, $t \in [0, 1]$
Mean-variance rule


A path is non-dominated under the mean-variance rule if no such a path exists whose mean and variance are both smaller, the set of non-dominated paths denoted as $\Gamma_{MV}$.

1. Example

   - $X, Y$ and $F_X(t) = t$, $F_Y(t) = t^2$, $t \in [0, 1]$

   - $X \succ_1 Y \Rightarrow X \succ_2 Y$
Mean-variance rule


A path is non-dominated under the mean-variance rule if no such a path exists whose mean and variance are both smaller, the set of non-dominated paths denoted as $\Gamma_{MV}$.

- **Example**
  1. $X, Y$ and $F_X(t) = t, F_Y(t) = t^2, t \in [0, 1]$
  2. $X \succ_1 Y \Rightarrow X \succ_2 Y$
  3. However, $E(X) > E(Y)$ and $\text{Var}(X) < \text{Var}(Y)$
Mean-variance rule


A path is non-dominated under the mean-variance rule if no such a path exists whose mean and variance are both smaller, the set of non-dominated paths denoted as $\Gamma_{MV}$.

Example

1. $X, Y$ and $F_X(t) = t$, $F_Y(t) = t^2$, $t \in [0, 1]$

2. $X \succ_1 Y \Rightarrow X \succ_2 Y$

3. However, $E(X) > E(Y)$ and $\text{Var}(X) < \text{Var}(Y)$

4. Both are non-dominated and should belong to $\Gamma_{MV}$
Summary of the relationship with existing models
Properties of SD-admissible paths

**Theorem**

1. FSD/SSD/TSD-admissible paths must be acyclic;
2. Subpaths of any FSD/SSD/TSD-admissible paths must also be FSD/SSD/TSD-admissible.
Properties of SD-admissible paths

**Theorem**

1. *FSD/SSD/TSD-admissible paths must be acyclic*;

2. *Subpaths of any FSD/SSD/TSD-admissible paths must also be FSD/SSD/TSD-admissible*.

- Property (1) implies that the number of FSD/SSD/TSD-admissible paths is finite.
Properties of SD-admissible paths

**Theorem**

1. *FSD/SSD/TSD-admissible paths must be acyclic;*

2. *Subpaths of any FSD/SSD/TSD-admissible paths must also be FSD/SSD/TSD-admissible.*

- Property (1) implies that the number of FSD/SSD/TSD-admissible paths is finite.

- Property (2) implies SD-admissible paths satisfy Bellman’s principle of optimality (so we can use dynamic programming!)
A Label-correcting algorithm

0 Initialize the scan list $Q$ with an empty path.
1 Select the first path from $Q$, and delete it from $Q$.
2 Extending the path along its predecessor node $i$ to create a new path.
   2.1 Calculate the distribution of the new path using convolution.
   2.2 Check SD for all paths originating at node $i$: if any of the existing path dominates the new path, go back to Step 2; otherwise, delete all paths that are dominated by the new path, insert the new path into $Q$.
3 If $Q$ is empty, stop; otherwise go to Step 1.
A Label-correcting algorithm

0 Initialize the scan list \( Q \) with an empty path.
1 Select the first path from \( Q \), and delete it from \( Q \).
2 Extending the path along its predecessor node \( i \) to create a new path.
   2.1 Calculate the distribution of the new path using convolution.
   2.2 Check SD for all paths originating at node \( i \): if any of the existing path dominates the new path, go back to Step 2; otherwise, delete all paths that are dominated by the new path, insert the new path into \( Q \).
3 If \( Q \) is empty, stop; otherwise go to Step 1.

Complexity

- Not a polynomial algorithm
- Heuristics exist that promise pseudo-polynomial complexity.
- Practical performance in transportation problems seems satisfactory.
Numerical example 1

| Path 1: 1->2->6->8->7->18 | Path 2: 1->2->6->8->16->18 |
| Path 3: 1->3->4->11->10->16->18 | Path 4: 1->3->4->5->6->8->7->18 |
| Path 5: 1->3->4->5->6->8->16->18 | Path 6: 1->3->4->5->9->8->7->18 |
| Path 7: 1->3->4->5->9->8->16->18 | Path 8: 1->3->4->5->9->10->16->18 |
| Path 9: 1->3->12->11->10->16->18 | Path 10: 1->2->6->8->16->19->18 |
| Path 11: 1->3->4->11->10->15->19->18 | Path 12: 1->3->4->11->10->16->19->18 |
| Path 15: 1->3->4->5->9->10->16->19->18 | Path 16: 1->3->4->5->9->8->16->19->18 |
| Path 17: 1->3->4->5->6->8->16->19->18 | Path 18: 1->3->12->11->10->15->19->18 |
| Path 19: 1->3->12->11->10->16->19->18 | Path 20: 1->3->12->11->14->15->19->18 |
| Path 21: 1->3->12->13->24->21->20->18 | Path 22: 1->3->12->13->24->21->22->20->18 |

<table>
<thead>
<tr>
<th>Expressway</th>
<th>Arterial Road</th>
<th>Local street</th>
<th>Expressway</th>
<th>Arterial Road</th>
<th>Local street</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>Var.</td>
<td>4</td>
</tr>
</tbody>
</table>
### Numerical example 1

<table>
<thead>
<tr>
<th>Route choice model</th>
<th>Path ID</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSD-admissible</td>
<td>3, 8, 9, 11, 21</td>
<td></td>
</tr>
<tr>
<td>SSD-admissible</td>
<td>3, 8, 9, 21</td>
<td></td>
</tr>
<tr>
<td>TSD-admissible</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Mean-variance rule</td>
<td>3, 8, 9, 21</td>
<td></td>
</tr>
<tr>
<td>METT-optimal</td>
<td>3, 9, 21</td>
<td>with $\alpha = 0.01, 0.02, \cdots, 1$</td>
</tr>
<tr>
<td>Least expected time</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Least variance</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>TTB-optimal</td>
<td>9, 21</td>
<td>with five different $\lambda$</td>
</tr>
<tr>
<td>QUF-optimal</td>
<td>21</td>
<td>with five different sets of $\beta$</td>
</tr>
<tr>
<td>EUF-optimal</td>
<td>21</td>
<td>with five different sets of $\beta$</td>
</tr>
</tbody>
</table>
Numerical example 1

[a. Frontiers for $\alpha > 0$]

[b. Frontiers $\alpha > 0.9$]
Chicago regional networks
### Numerical example 2: Computational performance

<table>
<thead>
<tr>
<th>Network</th>
<th>CPU time (second)</th>
<th>FSD</th>
<th>SSD</th>
<th>TSD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ave. $\Gamma_{FSD}$</td>
<td>Max. $\Gamma_{FSD}$</td>
<td>Ave. $\Gamma_{SSD}$</td>
<td>Max. $\Gamma_{SSD}$</td>
</tr>
<tr>
<td>CK</td>
<td>0.819</td>
<td>2.341</td>
<td>10.200</td>
<td></td>
</tr>
<tr>
<td>CMAP</td>
<td>162.57</td>
<td>2.27</td>
<td>42.25</td>
<td></td>
</tr>
<tr>
<td>CK</td>
<td>0.711</td>
<td>1.719</td>
<td>5.200</td>
<td></td>
</tr>
<tr>
<td>CMAP</td>
<td>38.23</td>
<td>1.45</td>
<td>7.75</td>
<td></td>
</tr>
<tr>
<td>CK</td>
<td>0.475</td>
<td>1.559</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CMAP</td>
<td>23.79</td>
<td>1.27</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

- **CK**: 3,000 links (Chicago sketch)
- **CMAP**: 40,000 links (Chicago regional)
The theory of stochastic dominance (SD) provides a coherent approach to modeling heterogenous risk-taking behavior in route choice.

The optimal solution to existing reliability-based route choice models may be interpreted using the SD theory, and be found from corresponding admissible paths.

General dynamic programming can be employed to generate SD-admissible paths.

Finding SD-admissible paths is computational viable even for large transportation networks.
## Optimal Path Problems with SD Constraints

### Motivation

- Routing decisions may involve other objectives in addition to travel time reliability.

- Alternative objectives: fuel consumption, emissions, penalty associated with non-punctual arrival.

- One way to integrate these objectives with the primary one, i.e., the minimization of travel time, is to introduce SD constraints.
Optimal Path Problems with SD Constraints

**Motivation**

- Routing decisions may involve other objectives in addition to travel time reliability.
- Alternative objectives: fuel consumption, emissions, penalty associated with non-punctual arrival.
- One way to integrate these objectives with the primary one, i.e., the minimization of travel time, is to introduce SD constraints.

**Two choices of constraints**

- Optimize against the set of SD-admissible paths
- Optimize against the set of paths that dominate a benchmark path (stemming from the desire to find a stock portfolio that beats stock market index)
Optimal Path Problems with SD Constraints

Motivation

- Routing decisions may involve other objectives in addition to travel time reliability.
- Alternative objectives: fuel consumption, emissions, penalty associated with non-punctual arrival.
- One way to integrate these objectives with the primary one, i.e., the minimization of travel time, is to introduce SD constraints.

Two types of constraints

- Optimize against the set of SD-admissible paths
- Optimize against the set of paths that dominate a benchmark path (stemming from the desire to find a stock portfolio that beats stock market index)
Optimal Path Problems with SSD Constraints

SSD-constrained optimal path problem

\[
\begin{align*}
\min & \quad \alpha_1 \sum_{a \in A} z_a(x_a) + \alpha_2 f \left( \sum_{a \in A} c_a x_a \right) \\
\text{s.t.} & \quad \sum_{a \in I(i)} x_a - \sum_{a \in O(i)} x_a = d_i, \quad \forall i \in \mathcal{N} \\
& \quad \sum_{a \in A} c_a x_a \geq 2 \pi_{rs}^k, \quad x_a \in \{0, 1\} \quad \forall a \in A.
\end{align*}
\]

- Both link and path costs are considered.
- The second constraint ensures all risk-averse travelers prefer the feasible path to the benchmark.
Case of non-punctual arrival penalty

Cost of path: penalty of early and late arrival

\[
f \left( \sum_{a \in A} c_a x_a \right) = \sum_{\theta} P_\theta [\beta e^-_\theta + \gamma e^+_\theta]
\]

\[
e^-_\theta = [\tau_0 - \sum_{a \in A} c_a(\theta) x_a]_+; \quad e^+_\theta = [\sum_{a \in A} c_a(\theta) x_a - \tau_0]_+
\]

- \(c_a(\theta)\) is the travel time on link \(a\) at realization \(\theta \in \Theta\); \(\Theta\) is a set of realizations of link travel times; and \(P_\theta\) is the probability of realization \(\theta\).
- The carrier may impose a penalty cost on late arrival, or on both late and early arrival.
- \(\beta \geq 0\) and \(\gamma \geq 0\) are parameters associated with early and late arrival penalty.
Case of non-punctual arrival penalty

An IP transformation

\[
\begin{align*}
\min & \sum_{\theta} P_{\theta} [\beta e_\theta^- + \gamma e_\theta^+] \\
\text{s.t. } & e_\theta^- \geq \tau_0 - \sum_{a \in A} c_a(\theta) x_a, \ \forall \theta \\
& e_\theta^+ \geq \sum_{a \in A} c_a(\theta) x_a - \tau_0, \ \forall \theta \\
& \sum_{a \in I(i)} x_a - \sum_{a \in O(i)} x_a = d_i, \ \forall i \in \mathcal{N} \\
& \sum_{a \in A} c_a(\theta) x_a - s(\theta, \eta) \leq \eta, \ \forall \theta \in \Theta, \ \eta \in \Phi \\
& \sum_{\theta} P_{\theta} s(\theta, \eta) \leq \sum_{\theta} P_{\theta} [(\pi_{k}^{rs}(\theta) - \eta)_+] , \ \forall \eta \in \Phi \\
& e_\theta^- , e_\theta^+ \geq 0, \ \forall \theta \in \Theta, \ s(\theta, \eta) \geq 0, \ \forall \theta \in \Theta, \eta \in \Phi \\
& x_a \in \{0, 1\}, \ \forall a \in A
\end{align*}
\]
Case of non-punctual arrival penalty: solution methods

**IP/LP techniques**

- The problem can be solved with existing LP/IP solvers such as CPLEX.
- This may not be fully tractable because
Case of non-punctual arrival penalty: solution methods

IP/LP techniques

- The problem can be solved with existing LP/IP solvers such as CPLEX.

- This may not be fully tractable because
  1. The transformed problem may be very large even for small instances.
  2. Cyclic solutions may not be fully excluded.
Case of non-punctual arrival penalty: solution methods

**IP/LP techniques**

- The problem can be solved with existing LP/IP solvers such as CPLEX.
- This may not be fully tractable because
  1. The transformed problem may be very large even for small instances.
  2. Cyclic solutions may not be fully excluded

**Dynamic programming**

- The problem can be also solved based on general dynamic programming
Case of non-punctual arrival penalty: solution methods

**IP/LP techniques**
- The problem can be solved with existing LP/IP solvers such as CPLEX.
- This may not be fully tractable because
  1. The transformed problem may be very large even for small instances.
  2. Cyclic solutions may not be fully excluded

**Dynamic programming**
- The problem can be also solved based on general dynamic programming
  1. First generate only those paths that are not dominated by the benchmark path;
  2. Then identify the optimal path from the above set.
Reliability-based traffic assignment problem

Hall (1983), Uchida and Iida (1993),(Lo and Tung, 2003; Lo et al., 2006), Shao, Lam and Tam (2006), Lam et al. (2008)

Introduction

- Link travel time is treated as an independent random variable, whose distribution is endogenously determined from the distribution of the road capacity and link performance function.
Reliability-based traffic assignment problem

Hall (1983), Uchida and Iida (1993), (Lo and Tung, 2003; Lo et al., 2006), Shao, Lam and Tam (2006), Lam et al. (2008)

Introduction

- Link travel time is treated as an independent random variable, whose distribution is endogenously determined from the distribution of the road capacity and link performance function.

- The route travel time is also a random variable whose distribution is obtained by convolving the distributions of its member links.
Reliability-based traffic assignment problem

Hall (1983), Uchida and Iida (1993), (Lo and Tung, 2003; Lo et al., 2006), Shao, Lam and Tam (2006), Lam et al. (2008)

Introduction

- Link travel time is treated as an *independent* random variable, whose distribution is endogenously determined from the distribution of the road capacity and link performance function.

- The route travel time is also a random variable whose distribution is obtained by convolving the distributions of its member links.

- Travelers are assumed to have perfect knowledge of the route travel time distributions, and choose best routes to arrive at the destination with their desired probability $\alpha$. 
Percentile user equilibrium

Notation

- $\mathcal{N}$ and $\mathcal{A}$: sets of nodes and links, respectively;
- $\mathcal{W} \subset \mathcal{N}^2$: set of OD pairs
- $\mathcal{K}_w$: set of routes between OD pair $\mathcal{W}$;
- $\mathcal{K} = \bigcup_{w \in \mathcal{W}} \mathcal{K}_w$: set of all routes;
- $f^w_{k}^{\alpha}$: flow for class $\alpha$ travelers from OD pair $w$ on route $k$
- $\xi^w_{k}^{\alpha}$: percentile travel time for class $\alpha$ travelers from OD pair $w$ on route $k$
- $\Delta_{ak} = 1$ if link $a$ is on route $k$ and $\Delta_{ak} = 0$ otherwise;
- $\Lambda_{wk} = 1$ if route $k \in \mathcal{K}_w$ and $\Lambda_{wk} = 0$ otherwise.
Percentile user equilibrium

Characterization

The equilibrium conditions imply that any used route has the identical and minimum percentile route travel time for users with same $\alpha$, i.e.,

$$f_k^{w\alpha} > 0 \rightarrow \xi_k^{w\alpha} = \pi^{w\alpha} ; \xi_k^{w\alpha} \geq \pi^{w\alpha}, \forall k, w, \alpha; f \in \Omega$$
Percentile user equilibrium

Characterization

- The equilibrium conditions imply that any used route has the identical and minimum percentile route travel time for users with same \( \alpha \), i.e.,

\[
 f_{k}^{\omega \alpha} > 0 \rightarrow \tilde{\zeta}_{k}^{w \alpha} = \pi^{w \alpha} ; \zeta_{k}^{w \alpha} \geq \pi^{w \alpha} , \forall k, w, \alpha ; f \in \Omega
\]

Evaluation of percentile travel time

- \( \tilde{\zeta}_{k}^{\omega} = \sum a \Delta_{ak} \tilde{t}_{a} = \sum a \Delta_{ak} g(x_{a}, \tilde{c}_{a}, \beta_{a}) \)

- The distribution of \( \tilde{\zeta}_{k}^{w} \) can be evaluated by convolution

\[
 F_{\tilde{\zeta}_{k}^{w}}(y) = \int_{0}^{y} v_{a}(z) F_{\tilde{\zeta}_{k}^{w}}(y - z) dz,
\]

- \( \alpha \)-percentile travel time on route \( k^{w} \): \( \zeta_{k}^{w \alpha} = F_{\tilde{\zeta}_{k}^{w}}^{-1}(\alpha) \)

- The derivative of \( \zeta_{k}^{w \alpha} \) may be approximated using convolution.
Formulation

As a VI problem

The problem of finding percentile UE route flows may be formulated as a variational inequality (VI) problem as follows: find $f^* \in \Omega$, such that

$$\langle \bar{\zeta}(f^*), f - f^* \rangle \geq 0, \forall f \in \Omega$$

where $\bar{\zeta}(f)$ is the percentile route travel time corresponding to a route flow pattern $f$. 
Solution algorithm

A gradient project algorithm

0: Initialization.

1: Column generation. Generate, for each O-D pair, optimal paths with respect to any $\alpha$. If they are not already in the path set, add them into the set.

2: Flow and distribution update. Update link flows, the PDF of all link travel times, and the PDF of all link travel time derivatives.

3: Equilibrate flow on each O-D pair $w$ and each user class $\alpha$.

3: If converged, stop; otherwise go back to Step 1.
Solution algorithm

A gradient project algorithm

0: Initialization.

1: Column generation. Generate, for each O-D pair, optimal paths with respect to any $\alpha$. If they are not already in the path set, add them into the set.

2: Flow and distribution update. Update link flows, the PDF of all link travel times, and the PDF of all link travel time derivatives.

3: Equilibrate flow on each O-D pair $w$ and each user class $\alpha$.

3: If converged, stop; otherwise go back to Step 1.

Column generation

- Generate FSD-admissible paths and the corresponding Pareto frontier.
- Select the optimal path for each $\alpha$ from the frontier.
Future Studies

### SD theory

- Derive tighter results for the relationship between the SD approach and the existing routing models.
- For those that are known to be inconsistent with SD (e.g., general TTB, mean-variance rule), can we bound approximation errors?
- Can we integrate the SD with robust optimization?
Future Studies

**SD theory**
- Derive tighter results for the relationship between the SD approach and the existing routing models.
- For those that are known to be inconsistent with SD (e.g., general TTB, mean-variance rule), can we bound approximation errors?
- Can we integrate the SD with robust optimization?

**Optimal path problems with SD constraints**
- Accommodate other SD rules, and consider multiple SD constraints (e.g., multiple benchmark) and O-D pairs.
- Incorporate other objectives (emission, fuel consumptions).
- More efficient solution algorithms (sampling techniques, DP-based heuristics).
Future studies

Reliability-based assignment

- Incorporate higher-order SD - Addressing heterogenous risk-taking in assignment.
- Develop efficient algorithms for real networks.
- Extend to network design problems or consider multiple objectives

More practical aspects

- Analyzing and evaluating reliability performance of existing transportation systems
- Reliability routing in transit networks
- Emerging data sources (consumer GPS data, cellular data etc.)


