Managing Components of Change in Cohort Change Ratio-Based Population Projections

(Related information: github.com/edyhsgr/CCRStable)

Florida State University Population Center Brown Bag
October 2020

Eddie Hunsinger
Transamerica

Thanks very much for inviting me to talk today – I'm excited for this opportunity to share this work on cohort change ratios, and managing components of change.

It's something that I think – at least parts of it if not all of it – could potentially be a helpful resource for researchers to work with in developing and interpreting population projections.

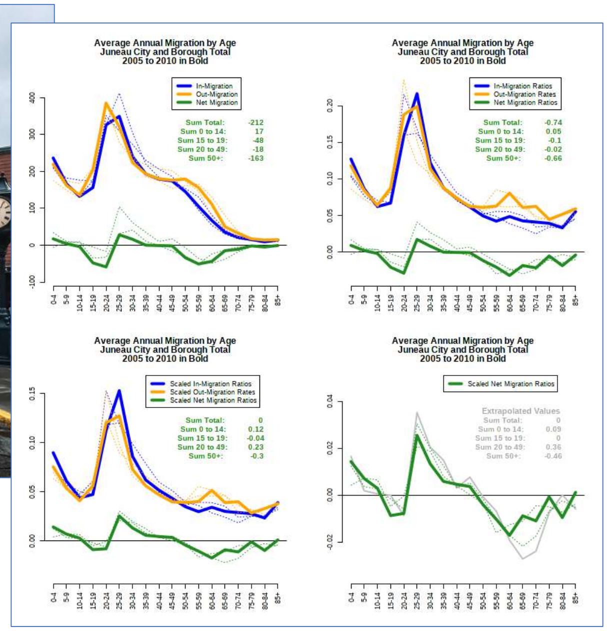
As I'll describe, I think in working from cohort change ratios we can tightly connect each of the components of change to change in age structures that we see, and also connect estimates and projections made by different methods – this can be helpful both for projections and the evaluation of estimates.

And in practical terms it may offer a useful method to adjust age structured models to match prescribed totals, which is something that may be helpful toward improved forecast accuracy.

And the method and associated code aim to offer information on stable population with modifiable components of change, which I think is interesting and worth exploring and communicating about for small areas.

Importantly, the indices for the components of change that are used would be well known and understood (or directly related to something well known or understood) by people involved in estimates and projections.

Downtown Juneau, Alaska. Photo taken in June 2016 by Flickr user Lee Coursey. Flickr.com.



edyhsgr.github.io

So, this presentation is about age structured projection models that we use so often in demography… whether with the cohort-component method, or the Hamilton-Perry (cohort change ratio) method, or grade progression ratios… That is, where we age the population forward in time, knowing adjustments are from births, deaths, or some form of migration into or out of the population.

The big value in age structured models is of course ever-important and valuable age detail… Age is such a useful resource for understanding so many things – I think offhand of labor, education, forms of income, childbearing, co-residence, mortality… I don't think there's consensus that the added detail of age structured models or other – even very sophisticated models – improves the accuracy of total population projections for small areas, but there's clear value in the detail of course, and projections give us some information on age structure that can be quite reliable (I think of broad trends for school enrollment and working age and senior populations in particular).

This photo is of Franklin Street in Downtown Juneau Alaska, and the graph is of in, out, and net migration data – 2000 to 2005 in bold, 2005 to 2010 and 2010 to 2015 are dotted – for Juneau, based on administrative record data (PFD applications) (data that David Howell, Eric Sandberg, and I worked on).

My initial interest in cohort change ratios and stable populations for small areas actually came from a phone call with someone in Juneau Alaska, when I was working in Anchorage for the State of Alaska. Though it's seen big changes in age structure, Juneau's population has hovered around 30,000 people for decades – in the early 2010s there was a bit of a surge in people in their 20s (this was in-step with aging, but there were also some net migration gains), and there was interest in this and a subsequent downturn and what would be next… and we opined for a bit about the long-term future, and he noted the idea of Juneau just tending toward being more and more like the US – just reasonably imagining a very open system…

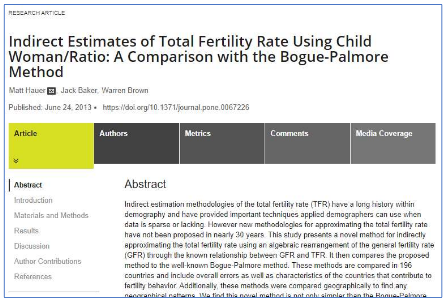
SO, I had this great (I'll say Wondrous!) opportunity to share with him that because of the consistency in age profiles of people moving to and from Juneau over time – which is something that so many localities have – there was a paper I think out of Wisconsin years several years back that I recall calls them "signature profiles" – because of this consistency over time, a better model might be stable population that we can see from something like cohort component projections (that is, age structured models, or something using projection matrices) (and he thought that was very neat). And after that I thought it would be neat to make something for it, and to share/communicate about it with data users.

Stable population can be a relatively challenging part of demography but it can also be considered and managed in simple terms… [Long term age structure and growth, from rates alone – no effect of current age structure…]

Also over the last several years there's been a lot of interest around use of Cohort Change Ratios in the state and local demography community… some that comes to mind in particular:
-Projections for all states by the University of Virginia's Weldon Cooper Center
-Projections by age, sex, and race for all counties by Matt Hauer, published in Nature
-Some entities taking Hamilton-Perry as their *method of choice* for their population projection program

-Also a full textbook Cohort Change Ratios and Their Applications by Jack Baker, David Swanson, Jeff Tayman, and Lucky Tedrow
-New research evaluating quality of Hamilton-Perry by Tom Wilson, and both quality and innovations by Baker, Swanson, Tayman… a lot of interest in it, which I don't recall was in place in the 2000s

One resource I don't have an image of here, but especially neat: a PhD dissertation by Webb Sprague, 'Wood's Method -- a Method for Fitting Leslie Matrices from Age-Sex Population Data, with some Practical Applications' which included inference for components of change where two of three are specified…


RESEARCH ARTICLE

Indirect Estimates of Total Fertility Rate Using Child Woman/Ratio: A Comparison with the Bogue-Palmore Method

Matt Hauer, Jack Baker, Warren Brown

Published: June 24, 2013 • https://doi.org/10.1371/journal.pone.0067226

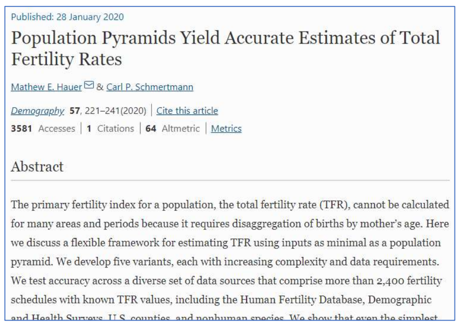
Article | Authors | Metrics | Comments | Media Coverage

Abstract
Introduction
Materials and Methods
Results
Discussion
Author Contributions
References

Abstract

Indirect estimation methodologies of the total fertility rate (TFR) have a long history within demography and have provided important techniques applied demographers can use when data is sparse or lacking. However new methodologies for approximating the total fertility rate have not been proposed in nearly 30 years. This study presents a novel method for indirectly approximating the total fertility rate using an algebraic rearrangement of the general fertility rate (GFR) through the known relationship between GFR and TFR. It then compares the proposed method to the well-known Bogue-Palmore method. These methods are compared in 196 countries and include overall errors as well as characteristics of the countries that contribute to fertility behavior. Additionally, these methods were compared geographically to find any

journals.plos.org


Published: 28 January 2020

Population Pyramids Yield Accurate Estimates of Total Fertility Rates

Mathew E. Hauer & Carl P. Schmertmann

Demography 57, 221–241(2020) | Cite this article

3581 Accesses | 1 Citations | 64 Altmetric | Metrics

Abstract

The primary fertility index for a population, the total fertility rate (TFR), cannot be calculated for many areas and periods because it requires disaggregation of births by mother's age. Here we discuss a flexible framework for estimating TFR using inputs as minimal as a population pyramid. We develop five variants, each with increasing complexity and data requirements. We test accuracy across a diverse set of data sources that comprise more than 2,400 fertility schedules with known TFR values, including the Human Fertility Database, Demographic

springer.com

$$iTFR = \frac{\beta - \alpha}{n} \cdot \frac{C}{W}$$

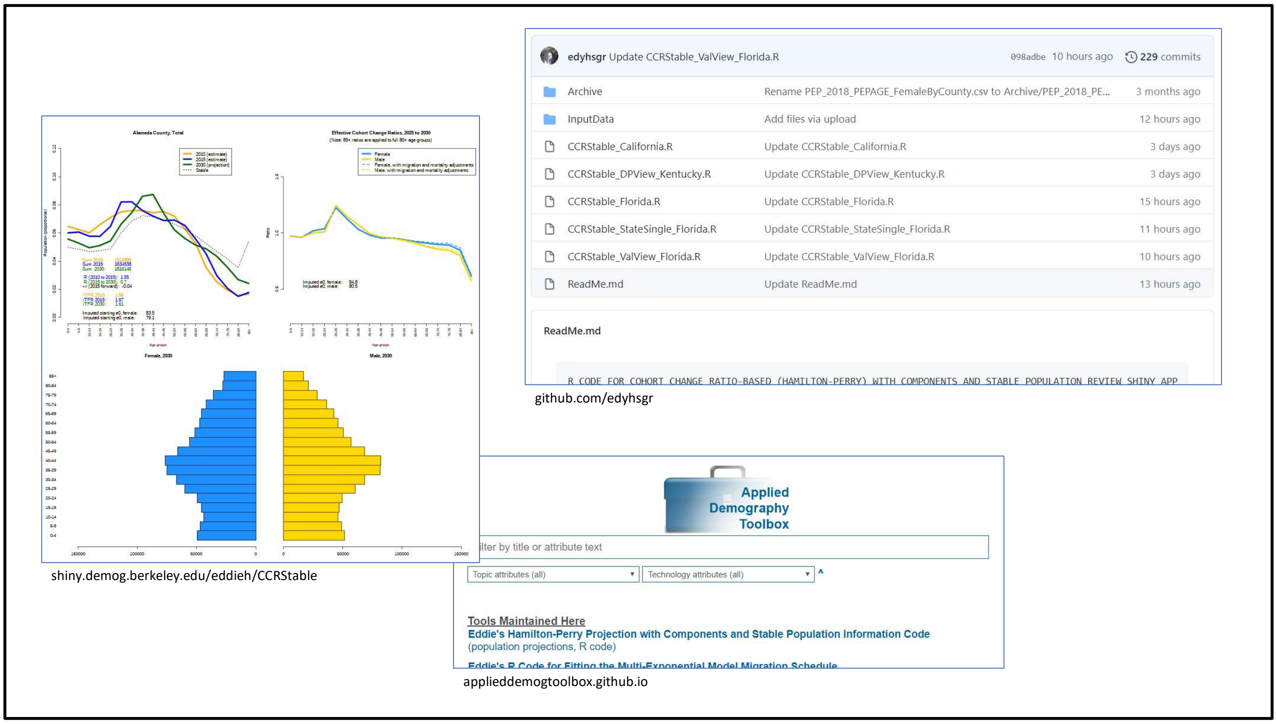
$\beta - \alpha$ is childbearing ages
n is age group sizes
C is the youngest age group
W is women in childbearing ages

$$C = iTFR \cdot \frac{W}{\beta - \alpha} \cdot n$$

And at the same time, work by Hauer, Baker, and Warren Brown, and by Matt Hauer and Carl Schmertmann, provided new indirect estimates for the Total Fertility Rate, which is an index that's readily interpreted and used so often in demographic research and population projections in particular

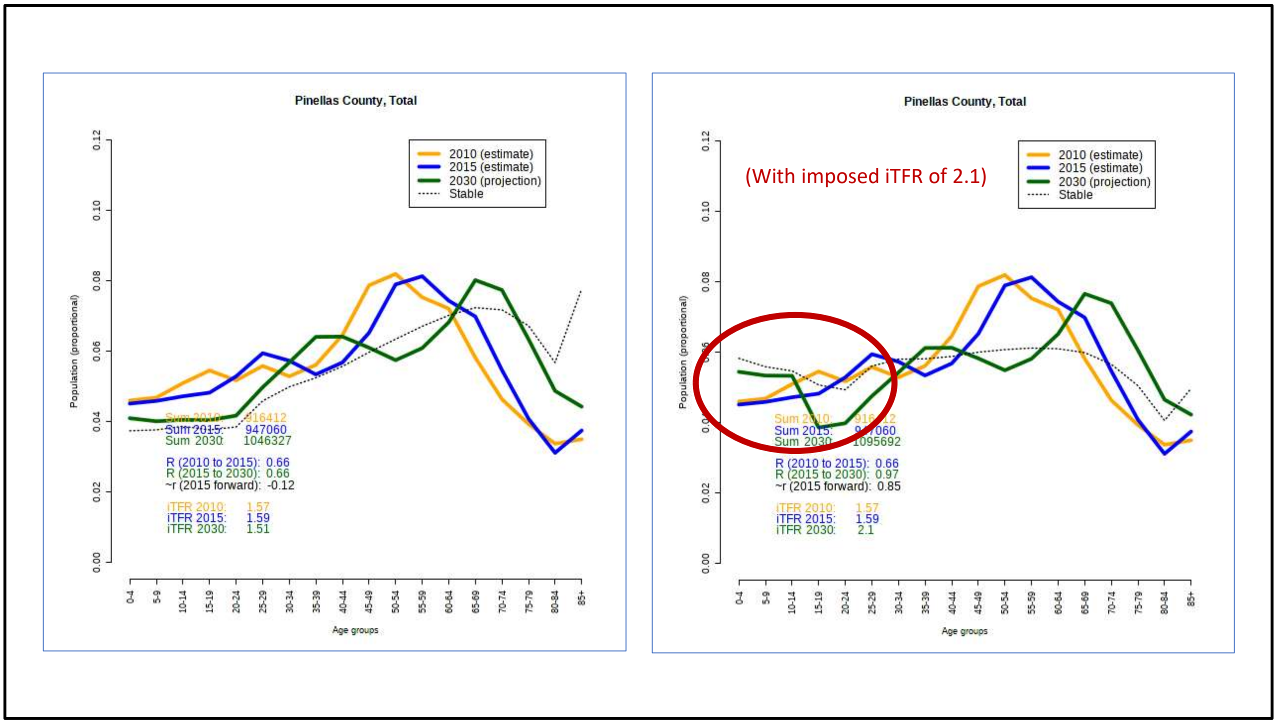
iTFR (it's actually one of several measures proposed in Hauer and Schmertmann) is such a neat age measure – I think not just because it can give a good estimate of fertility, but because it can predict a fertility index that's so often used (including in time series) and readily understandable. I think it must be generating interest and new possibilities to better understand human fertility in so many contexts.

And we can apply it very directly in population projections (as suggested in the Hauer et al publications), following the above formula.



Last summer and fall, related to all of this that I've talked about, and work with the State of California to develop a simple comparison projections approach, I put together some new population projection R code that simply uses cohort change ratios in projection matrices and includes the new iTFR index as an input option, and also projects out very far to estimate that stable population age structure I was interested in too…

With the basic framework, I've expanded on it over the last year and used it for different things (population projection, estimates evaluation, Differential Privacy evaluation…), and I or anyone can pretty readily bring in other states, so for today of course brought in Florida… [Others can run the shiny application of course – if you'd like to run a FL app (just copy/paste the code in R) see (three applications):
https://github.com/edyhsgr/CCRStable/blob/master/CCRStable_Florida.R
https://github.com/edyhsgr/CCRStable/blob/master/CCRStable_ValView_Florida.R
https://github.com/edyhsgr/CCRStable/blob/master/CCRStable_StateSingle_Florida.R
(each via https://github.com/edyhsgr/CCRStable)]

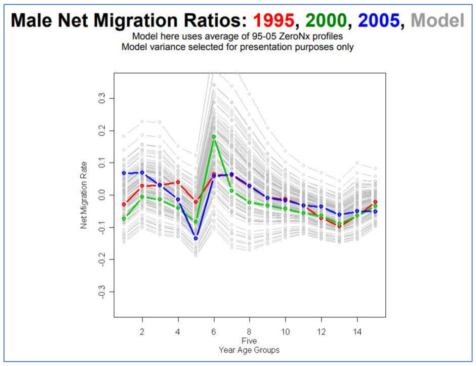


A look at a population projection for Florida counties, first with a look at applying the iTFR measure.

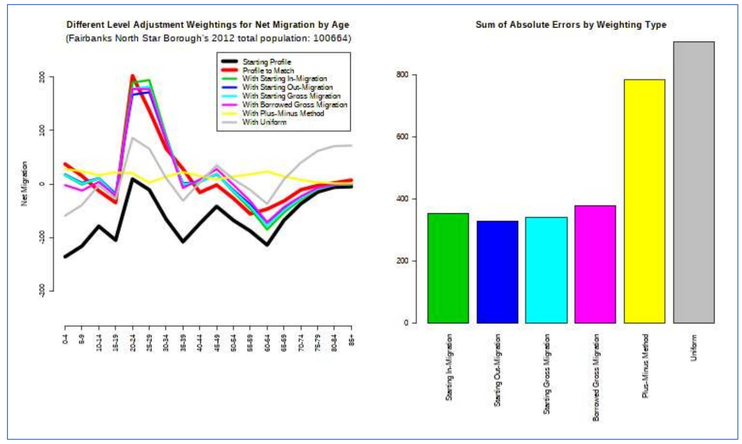
This ability to reasonably adjust a ~TFR measure without fertility data is a big deal to me, because of its practical use – again that it's a well-known and often used and powerful index for the component, and I can readily apply it with age structures alone…and it's one of three components down/managed in cohort change ratio projection.

When I apply and look at the iTFR though, I also know that it's not real fertility data – I think of it as an age measure as much as a fertility measure – and that it has with it not just births, but also the mix of mortality and migration, and so if I apply it, I don't make further adjustments to that youngest age group specifically (though it may change in size based on adjustment to women in childbearing ages).

I also include a simple sex ratio at birth input in the interface – largely because of a neat paper (Zhang and Li, 2020) specifying the effect of sex ratio at birth on deaths, births, and aging.



edyhsgr.github.io



shiny.demog.berkeley.edu/eddieh/NMAdjustCompare

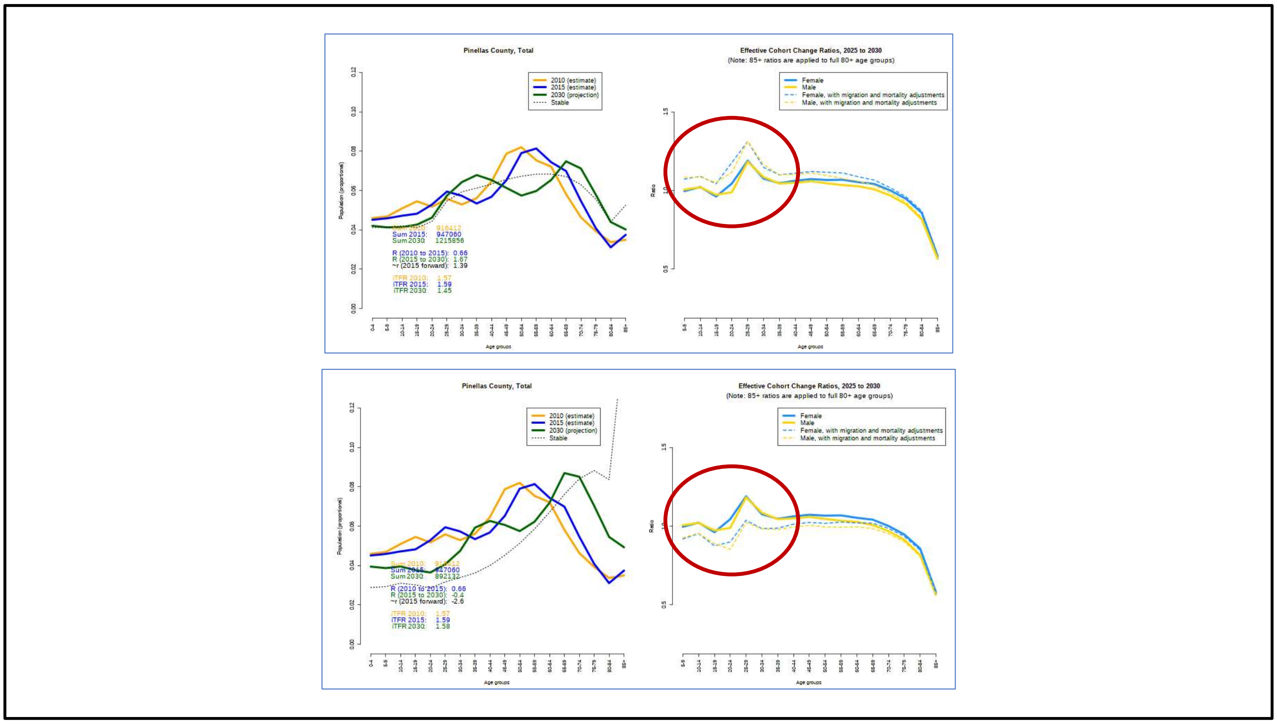
So of course in doing it, I started to think about the other two components, and especially about how cohort change ratios and age-specific net migration ratios have overlap in how they're understood: As "correction" ratios…

(I should point out here that net migration by age is most commonly managed as a vector that's added to a projected population – independent of the projection matrix or other components…)

So, correction ratios: That is, that net migration ratios aren't really migration – real rates of migration can only be calculated and implemented through multi-regional data and models – net migration ratios residuals from what we assume to be movers… and while real migration has a pretty ubiquitous age characteristics (I think of Rogers-Castro, and you can see it in the Juneau data), net migration ratios will vary from place to place (having those "signature patterns" (Johnson, Voss, et al 2005))… Very interestingly to me though – so clearly impressed in my mind from working with Alaska population and migration data – that for a given place they are often quite stable over time.

(On the chart to the left are net migration profiles for Alaska, adjusted to sum to zero-level, with many random levels in grey; the chart to the right shows effects of some different net migration adjustment methods for Fairbanks North Star Borough, Alaska (an extreme case, but I think helpful).)

So in making cohort component model-based population projections over the years, I explored the effects of adjusting migration by different approaches, and learned a few things (perhaps obvious things, but they've stuck with me and come up multiple times):
-Net migration by age is perhaps best-adjusted with a gross migration profile. (That is, when we want to change a level of net migration and distribute that level change across ages, it's most-reasonable to weight that distribution to ages in which people move the most.)
-Whether we use in-migration data, out-migration data, gross-migration data…matters little, so long as it's scaled correctly. What matters is that we use a basic gross migration profile.
-Even real migration data from another area seems to work well enough.
-Methods that don't weight movers based on real migration (such as "plus-minus" adjustments or uniform adjustments) can create unreasonable results.
-Weighted level adjustment still leave the potential effect of change in general turnover (that is, intensity, rather than level) out though… I think changes in level happen much more swiftly and are what we typically grapple with, but changes in intensity, and work to predict migration intensity especially for long-term projections, are very important too…



To manage migration adjustments in the cohort change ratio-based projection method I've been working on, I include both a net migration (that is, level) and a gross migration (that is, intensity) adjustment option.

This graph gives two example net migration adjustments (+1 and -1 annual percent of total population), and their effects.

For net migration, I simply add a net migration vector, which is level adjustment weighted by a generic gross migration profile.*

I don't actually change the projection matrices for the net migration adjustment (though, you'll see, I do for gross migration and for mortality)….

I then project the population and re-calculate "effective cohort change ratios" (the dotted lines in the graphs to the right) that include all of the migration (and mortality) adjustments I make. This approach allows me to use a *familiar* index to adjust net migration with: the ratio of total net migration to total population.

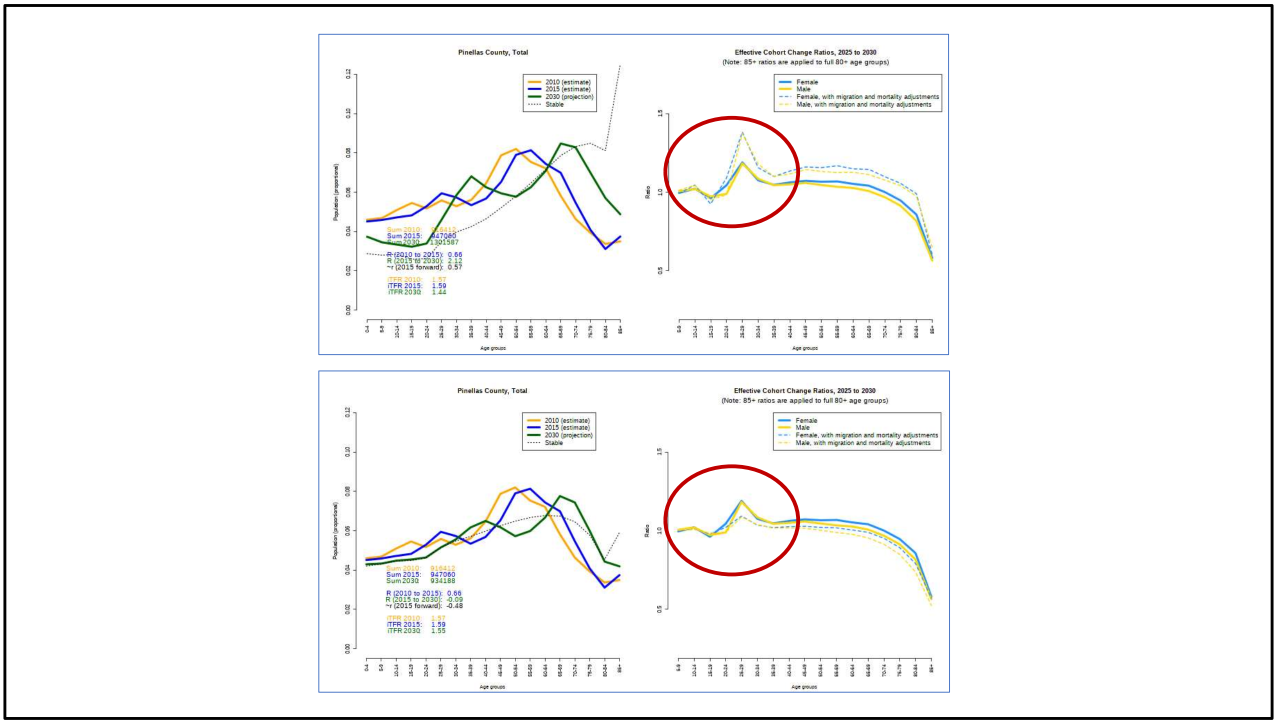
When we need to make an adjustment to the total population of a projection – perhaps to control to totals (value in this noted by Wilson, as well as Baker, Swanson, and Tayman), –

net migration, weighted by gross migration – this is I think a good and reasonable candidate for where to make that adjustment.

Thinking about the generic gross migration profile that I use to adjust net migration with, to scrutinize that a bit, the interface includes several options (FL, Sarasota County, CA, KY) – I think selecting a good profile can be helpful, but using a real (not net or uniform) migration profile is the most important part.

--

*Thanks – additional to the US Census Bureau, of course – thanks to IPUMS USA, University of Minnesota



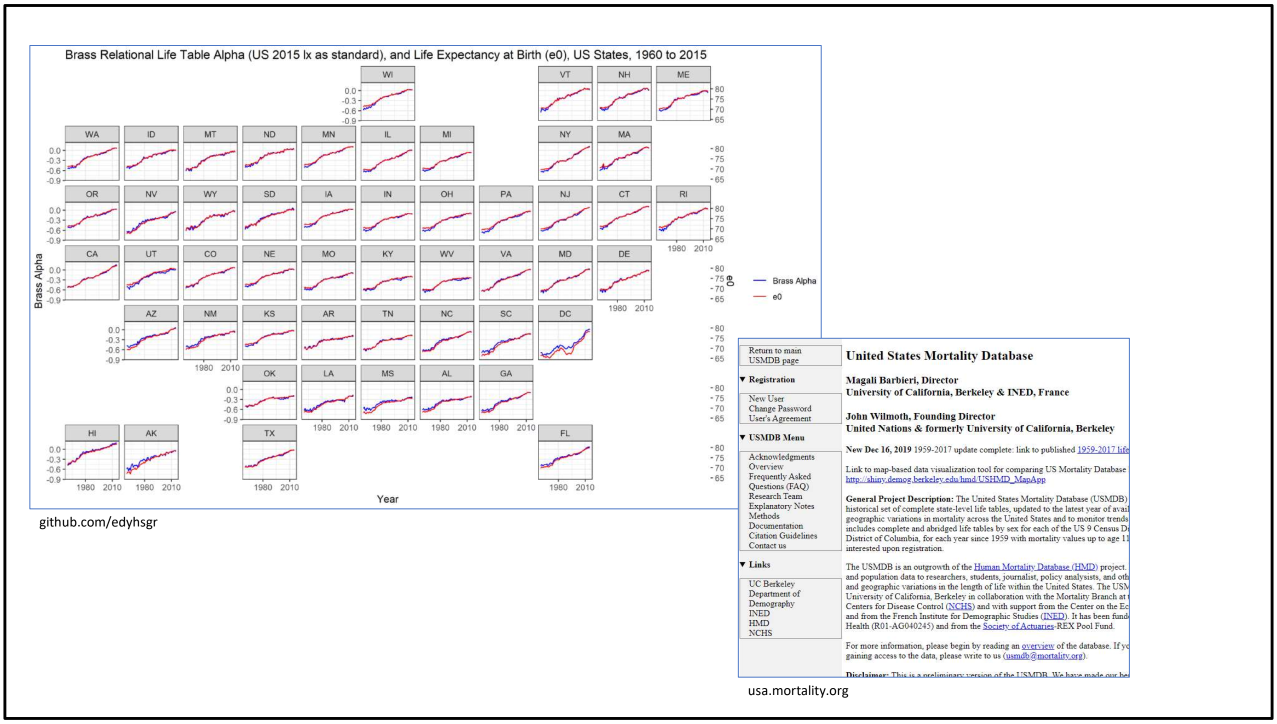
Thinking about gross migration as an input to adjust, and the importance of migration intensity (maybe it can be thought of as turnover intensity): Though net migration index levels seem to change more abruptly and dramatically, gross migration (much larger numbers than net migration, which is just a residual) can change too, and even with a fixed net migration *level* that can affect future age structure (fewer people moving in or out can mean more rapid aging for some areas, for instance).

These graphs show effect of increasing and decreasing gross migration (200 percent and 50 percent, respectively) – migration intensity – which I simply assume to proportionally increase the cohort change ratios (adjusted for mortality, so I think of it as a layer on top of mortality, and independent of mortality).

Though this gross migration adjustment method makes the most sense to me so far, it also seems very simplistic and so it's something I've been questioning for sure – where for net migration and iTFR I'm not really changing the projection matrices, I am here, and so it's a little bit deeper water.

Here, again for Pinellas County, sort of interestingly, an increase in migration intensity leads to more rapid population aging (compared to an increase in net migration, which drags on aging).

You can see the decrease for gross migration still follows a survivorship curve: Survival is managed separately, and I’ll cover that next.



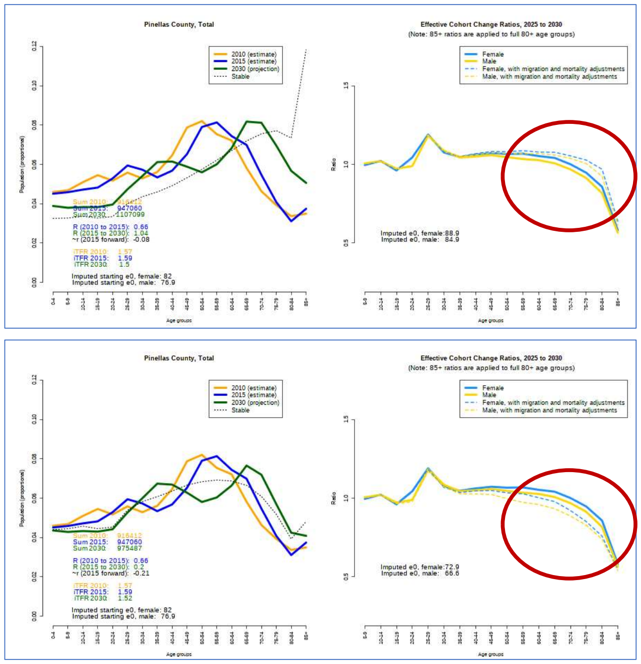
The net migration option that I included is just an adjustment option – allowing us to add or subtract movers – and it still leaves mortality unknown. So to manage mortality I simply assume it, with inputs for both a beginning and an ending point.

I'm not sure if this approach has been used by others – and additional to the migration intensity adjustment – it's something that I want to be especially cautious with (as it affects the those cohort change ratios and matrices directly), and I'm staying ready to go back to the drawing board, but it seems to work well/helpfully.

I start with a life table from usa.mortality.org and apply a Brass Relational Logit Model alpha adjustment as desired for the starting point (for instance, if I want adjust it to another year, or for an area that's known to have higher or lower mortality). I then input an alpha level for the ending point…

The map here shows Brass' alpha (in blue) with life expectancy at birth, and how the two track well over time.

Similar to iTFR and the net migration ratio, this gives me a single index that's readily understood and useful.



Importantly, the life tables only affect the cohort change ratios insofar as the beginning and ending life tables (beginning or ending Brass' alpha indexes) are *different from one another* (and as a level for the gross migration input to work with). They don't change the cohort change ratios initially, but allow the ratios to be adjusted over time (short or long) in the projection.

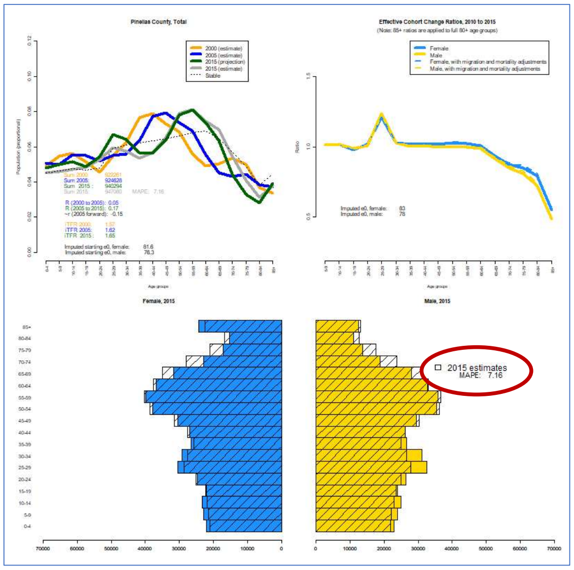
Indeed, the net migration option as well holds and builds on the "correction ratios" which I think is really valuable, to keep estimates and projections in step with on another.

So survivorship is treated as a modifiable layer that migration can change on top of, but its rates aren't affected by migration – it makes me think of Rogers-Castro fitting, where you can select a level, then fit childhood on top of that, then fit labor on top of that, then fit retirement and so on.

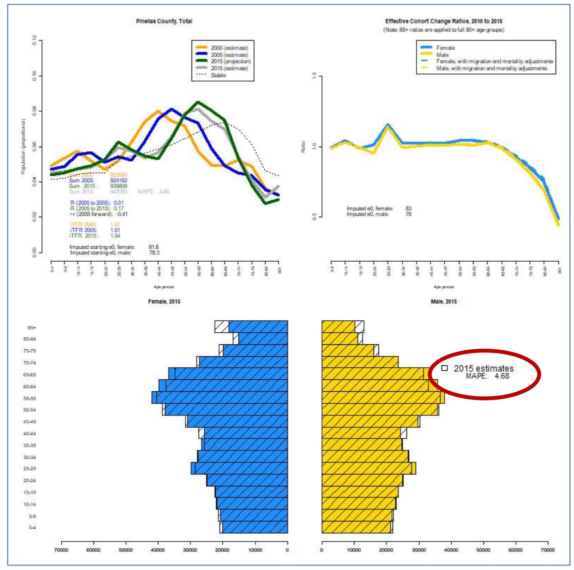
One drawback of this simple model is that life tables can vary in more ways than Brass' alpha of course… But it gives a level to work with, and some nice things about it are that it seems to have a fairly linear trend (that can be helpful for interpolation). Also, we can adjust both the beginning and the end to account for lower- and higher- mortality areas (and the effects will be different).

In the Florida interfaces I include life tables from a few states as input options to explore.

From Postcensal Estimates (2000 to 2005 Ratios)



From Intercensal Estimates (2000 to 2005 Ratios)



Here is an interface [go to interface] to review at errors in the projections from 2000s data, and a couple things stand out from looking at the different counties: (1) that (perhaps obviously) the intercensal estimates reduce error quite a bit. This may be due to improve estimates, and of course simple alignment of the intercensals with the next decade's estimates.

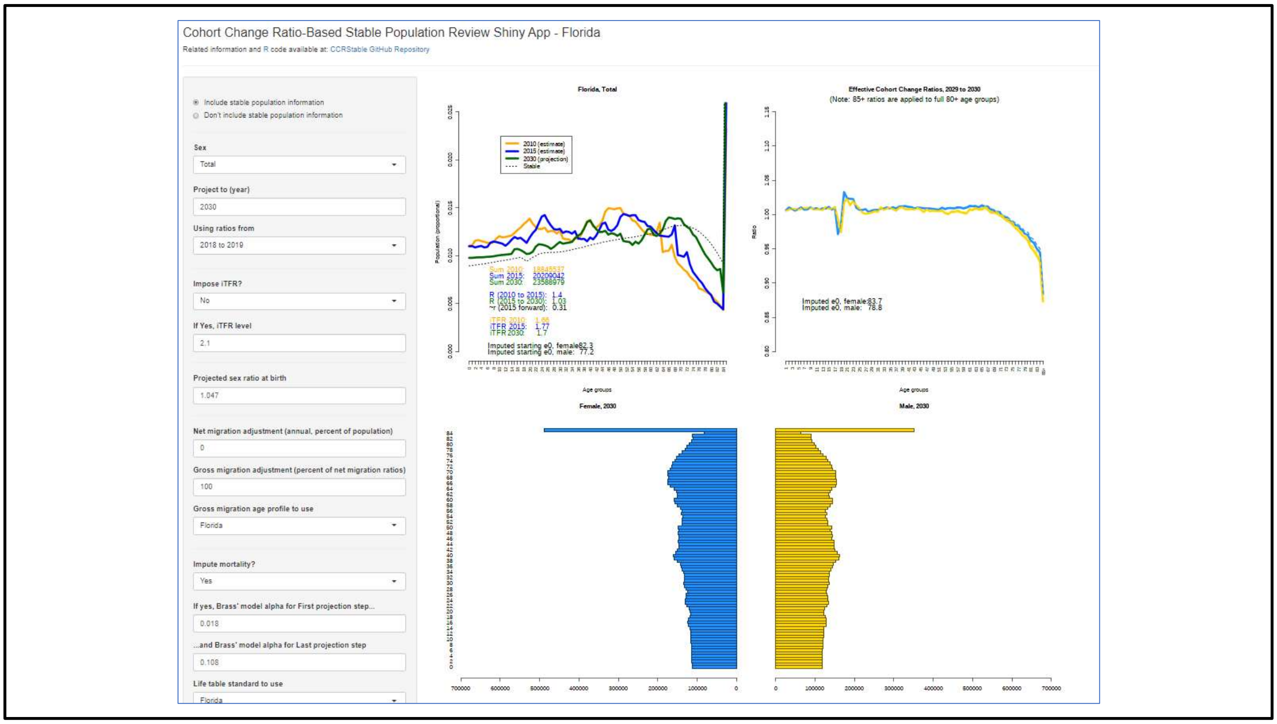
(2) That it appears meaningful improvements (meaningful reduction in MAPE) are possible with the indices I've covered (and net migration in particular).

Indeed, I had some thought about optimizing fit some parameters, which might be interesting if the combination and results are reasonable (for example, that we don't have mortality filling gaps that should be from migration…).

I also think it would be neat to connect these with employment-based (could be very simply done) and stochastic projections work.

I think it's been such a neat method/approach to learn, and I'm not sure how many (if any) people do similar, but I think that working this way, rather than by traditional cohort component, could put demographers in a stronger spot: To have a system with especially strong consistency in its age structures and between estimates and projections, easy-to make projection adjustments, and the ability to bring more helpful info in as well.

Lastly but for sure not leastly I'll be very sure to note: These projections, and the data we see, are a form of the US Census Bureau's estimates that they're very simply extrapolated from, and what's really stuck in my mind in recent months is the very apparent quality – whether by model, data, or both – of the Census Bureau's estimates, in that they appear so compatible with concepts for age structure and components of change.



A last one to share, and to include an image of the interface inputs too: Florida (just statewide) single year of age…

Thank You

More information: github.com/edyhsgr/CCRStable

Email: eddie.hunsinger@transamerica.com

Thanks very much for your time and look forward to any questions or follow-up!